13TH INTERNATIONAL SYMPOSIUM ON PROCESS SYSTEMS ENGINEERING (PSE 2018)

PART A

Edited by
MARIO R. EDEN
MARIANTHI G. IERAPETRITOU
GAVIN P. TOWLER

COMPUTER-AIDED CHEMICAL ENGINEERING, 44

13TH INTERNATIONAL SYMPOSIUM ON PROCESS SYSTEMS ENGINEERING (PSE 2018)

PART A

This page intentionally left blank

13TH INTERNATIONAL SYMPOSIUM ON PROCESS SYSTEMS ENGINEERING (PSE 2018)

PART A

Edited by

Mario R. Eden
Department of Chemical Engineering
Auburn University
Auburn, Alabama, USA

Marianthi G. Ierapetritou
Department of Chemical and Biochemical Engineering
Rutgers University
Piscataway, New Jersey, USA

Gavin P. Towler
Honeywell Performance Materials and Technologies
Des Plaines, Illinois, USA

ELSEVIER

Amsterdam – Boston – Heidelberg – London – New York – Oxford
Paris – San Diego – San Francisco – Singapore – Sydney – Tokyo

Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, USA

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN (Part A): 978-0-444-64243-1
ISBN (Set): 978-0-444-64241-7
ISSN: 1570-7946

Publisher: Joe Hayton
Acquisition Editor: Kostas Marinakis
Editorial Project Manager: Emily Thomson
Production Project Manager: Paul Prasad Chandramohan
Designer: Greg Harris

Typeset by SPi Global, India

Contents

Contributed Papers: Process Dynamics and Control

This page intentionally left blank

Preface

This book contains papers presented at the 13th International Symposium on Process Systems Engineering (PSE 2018), held at the Manchester Grand Hyatt in San Diego, California, USA, July 1-5, 2018. The PSE series is a triennial conference which has been held since 1982, organized on behalf of the international PSE Executive Committee with representation from countries in Asia-Pacific, Europe, and the Americas. The goal is to create an academic and industrial dialogue, a critical assessment of existing enabling technologies, a discussion on research, education, and industrial needs, and an international forum for new directions, challenges and opportunities in process systems engineering.

The PSE symposia bring together researchers, educators, and practitioners to discuss the latest developments in the field of Process Systems Engineering (PSE), including applications of methods, algorithms, and tools to solve a wide range of problems as well as provide the venue for discussion of new scientific challenges in our field. This symposium will feature more than 425 presentations including invited plenary and keynote lectures, as well contributed papers (both oral and poster) encompassing a large number of core and cross-cutting PSE themes.

The core PSE themes include:

- Process and Product Design/Synthesis
- Process Dynamics and Control
- Scheduling and Planning
- Supply Chain Management and Logistics
- Decision Making Under Uncertainty
- Integration of Process Operations and Design/Synthesis
- Modeling, Analysis and Simulation
- Optimization Methods and Computational Tools
- Education

The cross-cutting themes embrace new and exciting PSE applications in the areas of:

- Energy, Water, Food, Waste Nexus
- Multi-scale Systems Engineering
- Sustainable and Renewable Systems Engineering
- Modular Process Intensification
- Healthcare Systems Engineering
- Industrial Biotechnology
- Molecular and Materials Systems Engineering
- Smart Manufacturing and Plant of the Future
- Smart Grid Systems
- Urban/Building Systems Engineering
- Big Data Research and Development
- Man-machine Interface Systems Engineering
- Safety Driven Systems Engineering

This book includes 11 invited papers and extended abstracts as well as 407 contributed papers. All papers have been reviewed by at least two members of the International Programming Committee. We are very grateful to the members of the International Programming Committee for their assistance and constructive feedback during the review process. We would also like to thank the Elsevier editorial team, particularly Mr. Kostas Marinakis and Ms. Emily Thomson, for their support on this project, which provides an archival and fully indexed record of the conference.

We are dedicating PSE 2018 to the memory of the late Professor Christodoulou A. Floudas, a giant in the field of process systems engineering. When Chris passed away, our community lost a visionary leader and many of us lost a close friend.

Christodoulos A. Floudas was the Director of the Texas A&M Energy Institute, and the Erle Nye ’59 Chair Professor for Engineering Excellence at the Artie McFerrin Department of Chemical Engineering at Texas A&M University. Before moving to Texas A&M he was the Stephen C. Macaleer '63 Professor in Engineering and Applied Science, and Professor of Chemical and Biological Engineering at Princeton University for 29 years. Chris was a world-renowned authority in mathematical modeling and optimization of complex systems, with research interests at the interface of chemical engineering, applied mathematics, and operations research. His research accomplishments were recognized by numerous awards and honors, including election to the U.S. National Academy of Engineering in 2011, to the Academy of Athens in 2015, and to the U.S. National Academy of Inventors in 2015. He was the author of over 300 refereed publications and had delivered over 330 invited lectures, seminars, and named lectureships.

We will honor Chris’ life and legacy by establishing the Christodoulou A. Floudas Distinguished Lectureship sponsored by the Elsevier journal Computers & Chemical Engineering. The inaugural C.A. Floudas Lecture will be given by Professor Art Westerberg, a true pioneer in our field and a close friend of Chris’.

We are indebted to our conference manager, Mrs. Robin Craven, for all of her help and hard work. Similarly, we greatly appreciate the support from the CACHE Corporation and the AIChE Computing and Systems Technology (CAST) Division.

We hope PSE 2018 will foster constructive interaction among thought leaders from academia, industry, and government and that this book will serve as a useful reference for the latest research in all areas of process systems engineering

Mario R. Eden, Marianthi G. Ierapetritou, and Gavin P. Towler
PSE 2018 Chairs

International Programming Committee

Conference Chairs

Mario R. Eden	Auburn University
Marianthi G. Ierapetritou	Rutgers University
Gavin P. Towler	Honeywell/UOP

Members

Luke Achenie	Virginia Tech
Thomas A. Adams	McMaster University
Claire Adjiman	Imperial College London
Raymond A. Adomaitis	University of Maryland
Rakesh Agrawal	Purdue University
Ioannis Androulakis	Rutgers University
Suttichai Assabumrungrat	Chulalongkorn University
Thomas Badgwell	ExxonMobil
Parisa Bahri	Murdoch University
Bhavik R. Bakshi	The Ohio State University
Subash Balakrishna	Optience
Michael Baldea	The University of Texas at Austin
Ana Barbosa-Povoa	Technical University of Lisbon
Paul I. Barton	Massachusetts Institute of Technology
Matthew Bassett	Dow AgroSciences
Mona Bavarian	University of Nebraska
Lorenz Biegler	Carnegie Mellon University
Heriberto Cabezas	US Environmental Protection Agency
Kyle Camarda	University of Kansas
Ian Cameron	University of Queensland
Pedro Castro	University of Lisbon
Chuei-Tin Chang	National Cheng Kung University
Nishanth Chemmangattuvalappil	University of Nottingham – Malaysia
Xi Chen	Zhejiang University
Leo Chiang	The Dow Chemical Company
Min-Sen Chiu	National University of Singapore
Donald Chmielewski	llinois Institute of Technology
Luis A. Cisternas	University of Antofagasta
Lloyd Colegrove	Dow Chemical
Selen Cremaschi	Auburn University
Prodromos Daoutidis	University of Minnesota
Jim Davis	University of California Los Angeles
Maria Soledad Diaz	Plapiqui
Michael Doherty	University of California, Santa Barbara
Thomas Edgar	The University of Texas at Austin
Nimir Elbashir	Texas A&M University - Qatar
Fadwa Eljack	Qatar University

Sebastian Engell	University of Dortmund
Antonio Espuna	Universitat Politecnica de Catalunya
Dominic Foo	University of Nottingham – Malaysia
Kevin Furman	ExxonMobil
Ferenc Friedler	Pazmany Peter Catholic University
Rafiqul Gani	PSE for SPEED
Salvador Garcia Munoz	Eli Lilly
Michael Georgiadis	Aristotle University of Thessaloniki
Nevin Gerek	Schneider-Electric
Chrysanthos E. Gounaris	Carnegie Mellon University
Ignacio Grossmann	Carnegie Mellon University
Martin Guay	Queens University
Ravindra D. Gudi	Indian Institute of Technology Bombay
Iiro Harjunkoski	ABB Corporate Research
Faruque Hasan	Texas A&M University
Shinji Hasebe	Kyoto University
Q. Peter He	Auburn University
Gabriela P. Henning	INTEC
Diane Hildebrandt	University of South Africa
Yinlun Huang	Wayne State University
Jun Huang	Pfizer
Robert Huss	Eastman Chemical Company
Mohd Azlan Hussain	University of Malaysia
Arturo Jimenez	Technological Institute of Celaya
Conchita Jimenez-Gonzalez	GlaxoSmithKline
Xavier Joulia	INP-ENSIACET
Iftekhar Karimi	National University of Singapore
Antoon ten Kate	AkzoNobel
Worapon Kiatkittipong	Silpakorn University
Paisan Kittisupakorn	Chulalongkorn University
Antonis Kokossis	National Technical University of Athens
Zdravko Kravanja	University of Maribor
Costas Kravaris	Texas A&M University
Carl Laird	Sandia National Laboratories
Ajay Lakshmanan	AspenTech
Jean-Marc Le Lann	INP-ENSIACET
Jay H. Lee	KAIST
Zheng Li	Tsinghua University
Zukui Li	University of Alberta
Jinghai Li	Chinese Academy of Sciences
Patrick Linke	Texas A&M University – Qatar
Andreas Linninger	University of Illinois at Chicago
Vladimir Mahalec	McMaster University
Thokozani Majozi	University of the Witwatersrand
Zainuddin Abd Manan	Universiti Teknologi Malaysia
Sam Mannan	Texas A&M University
Christos T. Maravelias	University of Wisconsin – Madison
Francois Marechal	Ecole Polytechnique Federale de Lausanne
Mariano Martin	University of Salamanca
Henrique Matos	University of Lisbon

David Miller	Department of Energy
Ruth Misener	Imperial College London
Alexander Mitsos	RTWH Aachen
Il Moon	Yonsei University
Iqbal M. Mujtaba	University of Bradford
Zoltan Nagy	Purdue University
Costantinos C. Pantelides	Imperial College London
Lazaros Papageorgiou	University College London
Eduardo Perez-Cisneros	University of Autonoma Metropolitana
Fernando Pessoa	Federal University of Rio de Janeiro
Patrick Piccione	Syngenta
Sauro Pierucci	Politecnico de Milano
Efstratios Pistikopoulos	Texas A&M University
Luis Puigjaner	Universitat Politecnica de Catalunya
James B. Rawlings	University of Wisconsin – Madison
Matthew Realff	Georgia Institute of Technology
Gintaras Reklaitis	Purdue University
Nikolaos V. Sahinidis	Carnegie Mellon University
Yudi Samyudia	Universitas Prasetiya Mulya
Bent Sarup	Alfa Laval Copenhagen
Jeffrey Seay	University of Kentucky
Warren Seider	University of Pennsylvania
Jeffrey J. Siirola	Purdue University
John D. Siirola	Sandia National Laboratories
Sigurd Skogestad	Norwegian University of Science and Technology
Rajagopalan Srinivasan	Indian Institute of Technology Gandhinagar
Hongye Su	Zhejiang University
Christopher L.E. Swartz	McMaster University
Ana Torres	Universidad de la Republica Oriental del Uruguay
Metin Turkay	Koc University
Dimitrios Varvarezos	Aspen Technology, Inc.
Ashlee Ford Versypt	Oklahoma State University
Dionisios Vlachos	University of Delaware
Jin Wang	Auburn University
John Wassick	The Dow Chemical Company
Tapio Westerlund	Abo Akademi University
Yoshiyuki Yamashita	Tokyo University of Agriculture and Technology
Erik B.Ydstie	Carnegie Mellon University
Fengqi You	Cornell University
Suojiang Zhang	Chinese Academy of Sciences
Jinsong Zhao	Tsinghua University
Danielle Zyngier	Hatch

This page intentionally left blank

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50001-X

Future directions in process and product synthesis and design

Mariano Martín[a*], Thomas A. Adams II[b]

[a]*Departamento de Ingeniería Química. Universidad de Salamanca, Salamanca, 37008, Spain*
[b]*Department of Chemical Engineering. McMaster University, 1280 Main Street West, Hamilton, Ontario, L8S 4L7, Canada*

Abstract

The incorporation of renewable energy sources to the market has brought a new golden era for process synthesis along with new challenges and opportunities. In this work, we give a brief review of the state of the art and then present challenges and future directions in this exciting area. The biggest driver is the rapid improvement in computer technology which greatly increases the number of factors that can be considered during the design process. Thus, some of the key future directions lie in integrating the design process with other aspects of process systems engineering, such as scheduling, planning, control, and supply chain management. In addition, sustainability is now a major consideration in which the design process must consider not only economic sustainability, but social sustainability and environmental sustainability when making design decisions. The tools available to address these challenges are limited but we are in a position to develop them based on strong chemical engineering principles following a multidisciplinary approach with contributions from other disciplines such as biology and biochemistry, computer science, materials, and chemistry.

Keywords: Modelling, process synthesis, raw materials, heuristics, optimization, process design, product design

1. Introduction

The first contribution in the field of process synthesis is dates back to 1973 (Rudd et al., 1973). Traditionally, this is methodological creation of a process to produce chemical products using experience, heuristics, and algorithms. Over time, the boundaries of the analysis expanded as more and more complex concepts and techniques developed, allowing for more layers of rigor to be considered. This included fields such as optimization, modeling, economics, and control, which ultimately resulted in better process designs. However, product design was traditionally considered to be more related to consumer marketing than chemical engineering. Only recently has product design become a part of the scope of process engineers (i.e. Ng et al. 2013). For example, the popular textbook Product and Process Design Principles (Seider et al., 2017) did not get the "Product and" part of the title until the second edition. However, it is only natural that the boundaries of process design have now expanded to include product design since the next level up from the question "What is the best process to make this product?" is "What product should we make to meet market needs?" These products can range from custom chemicals such as catalysts for reactors, solvents for CO_2 capture, or thermal fluids for Rankine cycles to complex products such as cosmetics, pharmaceuti-

cals, and processed food. Furthermore, chemical processes must be controlled and operated carefully as they are the core of the product supply chain. So multiscale analysis in both time and size scales is becoming increasingly important. In addition, a new focus on using renewable resources such as biomass, solar, or wind energy adds additional challenges with regard to consistency and irregular availability. This paper briefly reviews developments and predicts future directions in these areas.

2. A bit of history

2.1. Heuristic-based design

Over the years graphical and heuristic based methods have been proposed and used with high acceptance in industry. The most used examples lie in the area of heat exchanger network (HEN) design from rules proposed by Masso and Rudd (1969) or Ponton and Donaldson (1974), to the pinch method (Linnhoff and Hindmarsh, 1983). Energy integration has evolved into total site integration aiming at exploiting synergies between processes in large complexes to save utilities and resources (Klemes et al., 1997) with an extension to the use of renewables as they entered the energy mix (Varbanov and Klemes, 2010). Water consumption is an interesting case: an effort driven by industry (Carnes et al., 1973), was originally presented as a mathematical problem (Takama et al., 1980). However, the difficulty in solving it lead to the use of heuristic-based conceptual design for many years. El-Halwagi and Manousiouthakis (1989) proposed a targeting graphical method for mass exchange networks. A few years later, Wang and Smith (1994) proposed the so called “water pinch” for minimization of water consumption. Heuristic rules have also been used in the design of sequences of distillation columns (Seader and Westerberg, 1977) while geometric rules were proposed for reactor selection (Glasser et al., 1987). The main procedures for complete process design are the hierarchical decomposition (Douglas, 1985) and the onion model (Smith, 2005). This heuristic based approach has alsobeen applied to process integration and intensification aiming at the efficient use of resources (Klemes et al., 2014).

2.2. Mathematical-based design

The development of computers and solvers has extended the use of mathematical based approaches. For example, early HEN’s were designed using sequential methods (Floudas et al., 1986) and later simultaneous design based on superstructure optimization (Yee and Grossmann 1990). However, large scale problems are still quite challenging; sequential framework approaches are able to handle larger problems, but not prove global optimality (Anantharaman et al., 2010). In the area of mass exchange networks, mathematical optimization for water management was addressed in the early 90’s (Papalexandri et al., 1994). Galán and Grossmann (1998) revisited Takama’s et al. (1980) work and several reformulations have been presented afterwards (Karuppiah and Grossmann, 2006; Ahmetovic and Grossmann, 2011). Simultaneous water and energy integration was presented in a two stage procedure by Grossmann and Martín (2010). Later, simultaneous methods such as targeting procedures (Yang and Grossmann, 2013) or superstructure optimization have been presented (Baliban et al., 2012). Power and steam production systems have also been addressed (Papoulias and Grossmann, 1983). Design procedures that consider flexibility in production have only started to appear recently due to the complex models required (Martín and Martín, 2013), but have been extended to polygeneration systems (Chen et al., 2012) and total site optimization (Nemet et al. 2015). Reactor networks aim at optimizing the conversion and require

considering the dynamics of the kinetics in the formulation (Achienie and Biegler, 1986). The link between reaction and separation led to considering both simultaneously in the formulation (Balakrishna and Biegler, 1993). In terms of separation, with distillation columns the most energy intense units in industry, a number of early works deal with the optimal sequence of columns to separate mixtures (Andrecovich and Westerberg, 1985), including heat integration among them (Floudas and Paules, 1988) as well as the rigorous design of columns (Viswanathan and Grossmann, 1993). More recent advances focus on optimal multicomponent distillation sequencing with the possibility of non-traditional, thermally coupled sequence configurations, which explores a very large combinatorial space of possible designs (Shenvi et al., 2012). PROSYM (now MYPSYN) was developed as an optimization based software for process synthesis (Kravanja and Grossmann, 1990; Kravanja, 2010). The first attempts into general representation for process synthesis date back to Papalexandri and Pistikopoulos (1996). Recently, Super–O has been based on a stage-wise process structure (Constantinou and Gani 2004; Bertran et al., 2017) and automatic process flowsheet generation using blocks and a grid structures have also been presented (Demirel et al., 2017).

For years, design has been process focused while product design was addressed as a separate problem. Typically product design focused on molecular design (Gani and Brignole, 1983), catalyst design (Gounaris et al., 2006), and protein folding (Floudas et al., 2005), for which global optimization became the driver via software like BARON (Sahinidis, 1996) or sbb (Androulakis et al., 1995). General packages for product design have been developed such as the VPPD Lab (Sawitree et al., 2015). However, sometimes process and product design must be solved simultaneously. Examples include the production of specialized ingredient based products (Martín and Martínez, 2013; Gani and Ng et al., 2015), the production of fuels within environmental regulations (Zhao and Wang, 2009), product/process design related to separations (Eden et al., 2004), the design of algae for fully renewable biodiesel production (Martín and Grossmann, 2013), the design of specific molecules with a desired properties such as reactant media (Struebing et al., 2017) or solvent selection (Gopinath et al., 2016), and addressing process design considering product characteristics (Moggridge and Cussler, 2000; Gani, 2004).

2.3. Hybrid approaches

In most cases the actual number of technologies is so high that it makes sense to follow a two stage evaluation procedure to discard those less promising due to their early stage of development or the extreme costs (Martín and Grossmann, 2011). Alternatively, if the number of alternatives is small, a scenario based approach can be performed using rigorous process simulators (Zhang et al., 2013). Finally, stochastic optimization using process simulators is interesting to take advantage of detailed models of distillation columns, liquid-liquid extraction columns, and other models that are implemented in rigorous process simulators (Adams and Seider, 2008; Ramirez et al., 2017).

3. Challenges in process and product design

Figure 1 summarizes the grand challenges of chemical process and product design, from the use of new materials to the development of novel technologies and processes involving the choice of pretreatments or separation schemes. Optimality is usually characterized by economics or business objectives, minimizing wastes, minimizing environmental impact, or combinations of these are becoming more important.

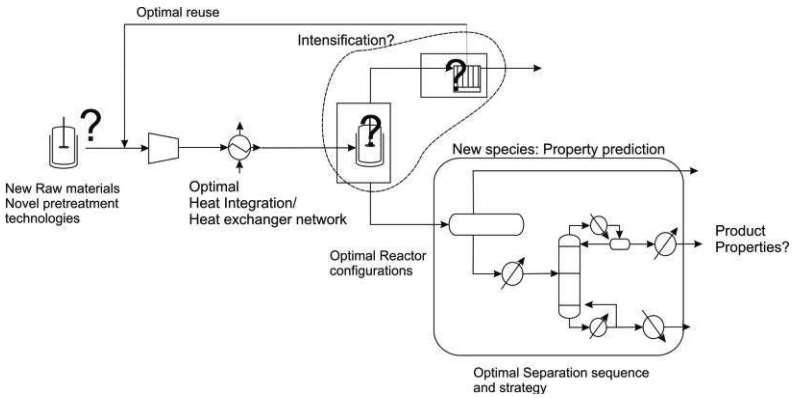


Figure 1. Challenges in process and product design.

3.1. New sources and products

The recent focus on environmental concerns for the chemical industry has brought several new potential feedstocks to the forefront. These especially include biomass, waste, nuclear, solar, and wind as source of energy. However, the variability of renewable sources greatly impacts process design (Martín, 2016) and operation (Zhang and Grossmann, 2017). Uncertainty, either exogenous (e.g. wind velocity) or endogenous (e.g. novel technologies), plays an important role in process design and operation (Pistikopoulos and Ierapetritou, 1995) and the size of the problems require novel solution procedures (Lara et al., 2017; Heuberger et al., 2017).

Similarly, new environmentally-motivated products are now of interest too, such as biodiesel from different alcohols, glycerol ethers (de la Cruz, 2014), biobutanol (Dalle Ave and Adams, 2018), bio-butyl butyrate (Birgen et al., 2016), and dimethyl ether (Ballinger and Adams, 2017). Furthermore, advanced and high added value materials are the key for economically promising bio-refineries. Polymers (Bueno et al., 2015), drugs, proteins, and carotenoids (Psycha et al., 2014) can be produced from biomass and waste. These new materials show the link between process and product design by targeting the properties that meet society needs, which is made possible by good physical property estimation models (Marrero and Gani, 2001). Furthermore, the link between fuels (chemicals) and power must be further analyzed due to the possibility of using power to produce chemicals that allow energy storage and shipping (Martín and Grossmann, 2017), going a step beyond batteries, where power is also stored chemically but whose capacity is still limited (Dunn et al., 2011).

3.2. Electricity and Energy storage

Chemical process engineers have rapidly moved into the field electricity generation processes, traditionally in the domain of electrical and mechanical engineering. However, the interface between electricity and chemical production has become increasingly important. Commonly, this link occurs as a waste handling step in chemical production, in which high-energy off-gases from a process used as fuel for a power plant in various ways, leading to polygeneration processes (Adams and Ghouse, 2015). However, because electric power production is driven by demand, and because electric grids are increasingly dependent on renewables in which power generation cannot be controlled, there is a clear mismatch between production and consumption. Grids with limited amounts of renewables traditionally handled this mismatch by raising or lowering natu-

ral gas combustion rates in response to demand. However, more modern grids with large amounts of integrated renewables face a new problem in which more electricity is occasionally produced than consumed. This leads to a modern problems associated with electricity disposal, which is sometimes resolved by selling it at negative prices.

A number of chemical engineering solutions have been discussed to address this issue. To mitigate the temporary absence of a resource, the integration of several of them provides backup (Yuan and Chen, 2012). Furthermore, thermal or chemical storage of excess energy generated by renewables is increasingly of interest. For example, concentrated solar power plant designs include thermal storage options such as molten salts, hot water, or waxes (Martín and Martín, 2013). Alternatively chemical storage options use reversible exothermic/endothermic reactions such that energy is stored as chemical potential (Lai and Adams, 2017). While these address day-to-day variation, seasonal variation might be better dealt with by integrating solar energy with biomass combustion during less sunny months (Vidal and Martín, 2015), since biomass is how nature already stores solar energy. Instead of thermal pathways, electrolytic pathways can be used to create H_2 from water via electrolysis when extra electricity is available. When combined with CO_2, high energy products such as methane, methanol, DME, (Davis and Martín, 2014, Martín, 2016) can be produced, thus storing the energy in fuel form. Physical pathways are also possible, in which excess energy (or electricity or mechanical work) is stored by compressing a fluid such as air (Nease and Adams, 2013). However, cost remains a major challenge with these storage systems, which is a major focus of current research.

3.3. New targets and metrics

Economic objectives have been the obvious choice for characterizing plant quality, and most heuristics and rules of thumb in the chemical industry have been developed with this in mind. However, they are no longer socially acceptable. A second objective is environmental impact minimization. There are a number of metrics (e.g. GREET, TRACI, ReCiPe) that measure the impact generated by a process, with life cycle analyses the most common technique for obtaining these metrics to be used in process optimization (Kniel et al., 1996; Azapagic, 1999). More recently, social issues are also being included. For example, the Jobs and Economic Development Impact model (JEDI) developed by the NREL can estimate the economic impacts of constructing and operating chemical and power facilities, which can be included in process synthesis (You et al. 2011). Safety, originally presented by (Klenz, 1984) has also been included as an objective for process synthesis (Ruiz-Femenía et al., 2017) and for product and process design (Martínez, 2017). To address them all as multiobjective, several strategies have been used to address its solution, either by producing a single-weighted objective function such as ecoprofit (Cucek et al., 2012) or RePSIM (Martín, 2016), or by using a constraints-based approach. Alternatively, methods on reducing the number of objectives without eliminating solutions of the problem (i.e. Pozo, et al. 2012) or to evaluate the target yields of a particular technology for it to become competitive (Limleamthong et al., 2016) have been developed.

4. Near-Term Future Directions

4.1. Big data approaches to process and product modelling

Modelling is now the root of rigorous process and product design, often striking a balance between model rigour and model speed or complexity. Machine learning to create

data-based models is growing in popularity. Examples include estimating product properties and performance (e.g. wine taste, appropriate smell), which may not even exist yet. Given the more recent inclusion of social objectives, data and theory from outside engineering (including social science) will need to be included in big data approaches for product design. Artificial intelligence (AI) approaches will also become more in use, such as training artificial neural networks (ANN) to recognize heuristics and trends in human-designed processes and then use that to create new processes from the ANN models. For example, IBM's Watson platform is being trained with data from engineering operations documentation in order to make models of oil production operations. Engineers can ask plain language questions to get information about day to day operations or get suggestions on how to deal with problems as they appear (Lewis, 2017). Although not used for design currently, the same approach could be applied with an appropriate data set.

4.2. Process operation: Integrated scheduling and operation

Renewable-based processes rely on resources that are uncertain and seasonal. Thus, dynamics and scheduling will play an important role. For example, demand-side management is starting to address this issue with integration of processes and their operation (Zhang et al 2017). However, key challenges include the problem size, the various time and spatial scales involved, and understanding dynamic process characteristics and operation such as start-up and shut down. This is especially prevalent with systems that include energy storage technologies such as batteries, thermal fluids, and chemicals, in which technology degradation over time becomes an important factor.

4.3. Process control

Integrated process design and control has been growing in popularity to create a more robust or profitable steady state design and control system by considering the inherent controllability of the design in the face of uncertain disturbances (Ricardez-Sandoval, 2012; Washington and Swartz, 2014). In a renewables-driven process, uncertain disturbances such as minute-to-minute wind velocity fluctuation can be incorporated under this framework. However, more predictable disturbances (such as seasonal solar and biomass cycles) are better considered in an integrated design, control, and scheduling framework, which is at its infancy (Pistikopoulos and Diangelakis, 2016). Again, the dynamics of the systems plays a major role in these areas.

4.4. Supply chain management.

Although process and supply chain are at different scales, the design of the supply chain becomes directly linked to the design of the process when environmental considerations are made. In addition, the operation of flexible multiproduct processes that use unreliable feeds (biomass in particular) relies on the information of their supply chains. Either way, the link between process design and supply chain design will become increasingly important in an environmentally conscious world.

4.5. Computer science

Nearly all of the future directions put forward in this paper, from process modelling advances to integrated supply chain management, create large mathematical problems which will require novel algorithms to solve. Efficient decomposition algorithms, modelling approaches, as and better software (or a combination of each) will allow us to solve larger and larger MILP or MINLP problems. Global optimization, despite its many advances, is still limited to small problems. Many of the future directions put

forward in this paper are rooted in the solution of ever-larger problems with more complex models in order to capture and exploit more nuances and characteristics of the system. This means including more detail, non-linearity, and non-convexity, and so global solvers will need to be extended and improved. However, as engineers, a suboptimal solution is better than none, therefore solvers that handle large problems (but not guarantee global optimality) will be a necessary tool for practical process synthesis and operation. These include stochastic, evolutionary, or non-deterministic optimization algorithms which may not be elegant or theoretically attractive, but are undeniably useful.

5. The PSE Technology Tree

To help us prepare for this article, we created a technology tree for process systems engineering. A technology tree maps the progression of research from one technology to another, and how they combine to form new technologies. For example, in order to create a sequential modular process simulator, you must first create an algorithm to solve individual unit operation models, and also create an algorithm to solve sequences of unit operations with recycle using tear streams. Such trees are useful for getting a bird's eye view of both the history of the field and its future directions. However, every researcher might have a different idea about what the tree should look like and how detailed it should be, and the field (and tree) is always changing. That is why, in the spirit of this conference, we created a Technology Tree Wiki on the website psecommunity.org/pse-technology-tree/. We encourage readers to visit the website and contribute by adding to or changing the structure of the tree. Readers can also create, discuss, or edit articles about each individual node, such as providing descriptions, examples, images, videos, or literature citations in each area. In this way, readers can participate in the open creation of an ever-evolving encyclopaedia for our field. An example screen shot is shown in Figure 2.

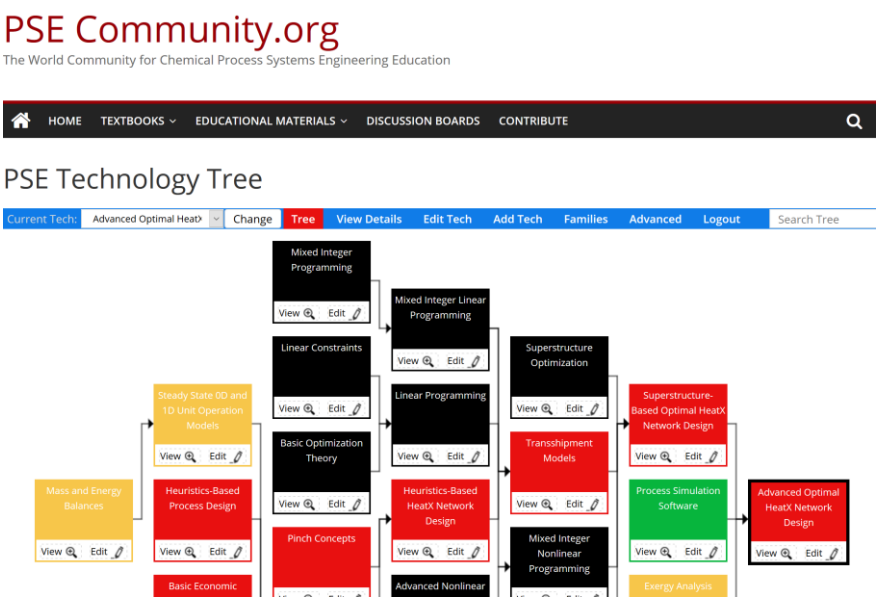


Figure 2. A screen capture from the PSE Technology Tree Wiki at psecommunity.org

6. Conclusions

From a big picture perspective, the historical progression of the field of process design is perhaps best characterized by an ever-increasing problem scope. Our fundamental understanding behind major chemical unit operations and their thermophysical phenomena has not changed much over the decades. Instead, the *methodology* of process design has expanded to include more and more rigorous mathematical, dynamics, control, planning, scheduling, product design, uncertainty, economics, and environmental factors, although not yet all at once. All of this is made possible by continually improving numerical methods (particularly optimization and modelling) and software tools which make larger and larger problems more tractable.

Based on the extrapolation of this trend, we envision that the future of process design will be largely driven by software that essentially includes all of these areas integrated together in one large global optimization problem. The designer would specify key objectives (perhaps in plain language) that relates to either particular products or merely product properties. The software would then formulate and solve an optimization problem which automatically synthesizes the design of the dynamic process and its products, together with its dynamic supply chain, plans and schedules, target markets, control systems, and considering all operational details including product flexibility and transitions, startup, shutdown, safety, reliability, uncertainty, and disturbance response. Candidate processes would be ranked on metrics that include a mix of environmental, business, and social factors. Data for environmental impacts, markets, supply chains, unit operations, and chemicals would be automatically drawn from massive databases or database-driven models. The process synthesis would be achieved through AI algorithms which draw upon massive models or databases of both human-derived and AI-created processes, heuristics, and best-practices. All of these aspects already exist in one form or another, it is merely the integration of everything together into one mega-problem that remains.

7. References

L.E.K, Achenie, L.T. Biegler, 1986, Ind. Eng. Chem. Fund., 25, 621
T.A. Adams II, W.D. Seider, 2008, Comput. Chem. Eng., 32, 9, 2099-2012
T.A. Adams II, P.I. Barton, 2009, Int. J. Hydrogen Energy, 34, 21, 8877-8891
T.A. Adams II, J. Nease, D. Tucker, P.I. Barton, 2012, Ind. Eng. Chem. Res., 52, 9, 3089-3111
T.A. Adams, II, J.H. Ghouse, 2015, Current Opinion Chem. Eng, 10, 87-93
R. Anantharaman, I. Nastad, B. Nygreen, T. Gundersen, 2010, Comput. Chem. Eng., 34, 1822-1830
M.J. Andrecovich, A.W. Westerberg, 1985, AIChE J., 31, 9, 1461-1474
I.P. Androulakis, C.D. Maranas, C.A. Floudas, 1995, J. Global Opt., 7, 4, 337-363
A. Azapagic, 1999, Chem. Eng. J., 73, 1-21
S. Balakrishna, L.T. Bielger, 1993, Ind. Eng. Chem. Res. 32, 1372
M. Baldea. Comput. Chem. Eng. 81, 104-114
R.C. Baliban, J.A. Elia, C.A. Floudas, 2012, Comput. Chem. Eng., 37, 297-327
S. Ballinger, T.A. Adams II, 2017, Comput. Chem. Eng., 105, 197-211
M.-O. Bertran, R. Frauzem, A.-S. Sanchez-Arcilla, L. Zhang, J.M. Woodley, R. Gani, 2017, Comput. Chem. Eng., 106, 892-910
A. Bhosekar, M. Ierapetritou, 2018, Comput. Chem. Eng, 108, 250-267
C. Birgen, H.A. Preisig, A. Wentzel, S. Markussen, B. Wittgens, U. Sakar, A. Ganguly, S. Saha, S. Baksi, 2016, Comput. Aided Chem. Eng., 38, 577-582
F. Boukouvala, F.J. Muzzio, M.G. Ierapetritou, 2010, J. Pharma, Innov., 5, 3, 119-137
L. Bueno, C.A. Toro, M. Martín, 2015, Chem. Eng. Res. Des., 93, 432-440

J. Caballero, I.E. Grossmann, 2013, AIChE J., 59, 4, 1139–1159
B.A. Carnes, D.L. Ford, S.O. Brady, 1973, Proc. Nat. Conf. Complete Water Reuse, Washington, DC
Y. Chen, X. Li, T.A. Adams II, P.I. Barton, 2012, AIChE J., 58, 19, 3080-3095
A. Cozad, N.V. Sahinidis, D.C. Miller, 2014, AIChE J., 60, 6, 2211-2227
L. Čuček, R. Drobež, B. Pahor, Z. Kravanja, 2012, Comput. Chem. Eng., 42, 87-100
G. Dalle Ave, T.A. Adams II, 2018, Energy Conversion. Manage., 156, 288-300
W. Davis, M. Martín, 2014, J. Cleaner Prod., 80, 252-261
V. de la Cruz, S. Hernández, M. Martín, I.E. Grossmann, 2014, Ind. Eng. Chem. Res., 53, 37, 14397–14407.
I. Dejanović, L.J. Matijašević, Z. Olujić, 2010, Process Intens., 49, 6, 559-580
S. E. Demirel, J. Li, M. M. F. Hasan, 2017, Comput. Chem. Eng, 150, 2-38.
J.M. Douglas, 1988, Conceptual Design of Chemical Processes, McGraw Hill, New York
B. Dunn, H. Kamth, J.M. Tasacon, 2011, Science, 334, 6068, 928-935
M.R. Eden, S.B. Jørgensen, R. Gani, M.M. El-Hawagi, 2004,Chem. Eng. Processing: Process Intens. 43, 5, 595-608
M.M. El-Halwagi, V. Manousiouthakis, 1989, AIChE J., 35, 1233–1244
C.A. Floudas, A.R. Ciric, I.E. Grossmann, 1986, AIChE J., 32, 276-290
C.A. Floudas, H.K. Fung, S.R. McAllister, M. Monnigmann, R. Rajgaria, 2005, Chem. Eng. Sci., 61, 966-988
C.A. Floudas, A.M. Niziolek, O. Onel, L.R. Mathews, 2016, AIChE J., 62, 3, 602-523
B. Galan B, I.E. Grossmann, 1998, Ind. Eng. Chem. Res., 37, 4036–4048.
R. Gani, E.A. Brignole, 1983, Fluid Phase Equil., 13, 331-340
R. Gani, 2004, Comput. Chem. Eng., 28, 2441–2457
R. Gani, K.M. Ng, 2015, Comput. Chem. Eng., 81, 70-79
D. Glasser D, C. Crove, D. Hilderbrandt, 1987, Ind. Eng. Chem. Res., 26, 9, 1803.
S. Gopinath, G. Jackson, A. Galindo, C.S. Adjiman, 2016, AICHE J., 62, 3484-3504
C.E. Gounaris, C.A. Floudas, J. Wei, 2006, Chem Eng Sci, 61, 24, 7933–7948.
I.E. Grossmann, J. Caballero, H. Yeomans, 1999, Korean J. Chem. Eng., 16, 4, 407–426
I.E. Grossmann, M. Martín, 2010, Chinese J. Chem. Eng, 18, 6, 914-922
C.F. Heuberger, I. Staffell, N. Shah, N.M. Dowell, 2017, Comput. Chem. Eng., 107, 247-256
M. Hostrup, P. Harper, R. Gani, 1999, Comput. Chem. Eng., 23, 10, 1395–1414
J. Marrero, R. Gani, 2001, Fluid Phase Equilib., 183, 183-208
R. Karuppiah, I.E. Grossmann, 2006, Comput. Chem. Eng., 30, 4, 650–673.
J.J. Klemeš, V.R. Dhole, K. Raissi, S.J. Perry, L. Puigjaner, 1997, App. Thermal Eng., 17, 8-10, 993-1003
J.J. Klemeš, P.S. Varbanov, S.R.W.W. Alxi, Z.A. Mana, 2014, De Gruyter. Berlin. Germany
T. Kletz, 1984, Cheaper, safer plants, or wealth and safety at work: notes on inherently safer and simpler plants, IChemE, ISBN 0852951671
G.E. Kniel, K. Delmarco, J.G. Petrie, 1996, Environ. Progr., 15, 4, 221–228
Z. Kravanja, I.E. Grossmann, 1990, Comput. Chem. Eng., 14, 1363-1378.
C. Lara, D. Mallapragada, D. Papageorgiou, A. Venkatesh, I.E. Grossmann, 2017, Euro. J. Operat. Res., in press, preprint available at http://www.optimization-online.org/DB_FILE/2017/08/6162.pdf
K. Lewis, 2017, Cognitive energy with Woodside, IBM Internet of Things Blog, accessible https://www.ibm.com/blogs/internet-of-things/cognitive-energy-woodside/
P. Limleamthong, M. Gonzalez-Miquel, S. Papadokonstantakis, A.I. Papadopoulos, P. Seferlis, G. Guillen-Gosalbez, 2016, Green Chem., 18, 6468-6481
M.L. Liu, N.V. Sahinidis, 1995, Ind. Eng. Chem. Res., 35, 4154–4165
S. Malmierca, R. Díez-Antoínez, A.I. Paniagua, M. Martín, 2017, Ind. Eng. Chem. Res., 56, 6, 1525–1533
L. Martín, M. Martín, 2013, Appl. Therm. Energy, 59, 627-633
M. Martin, I.E. Grossmann, 2011, AIChE J., 57, 12, 3408-3428
M. Martin, I.E. Grossmann, 2012, Comput. Aid. Chem. Eng., 30, 16-20
M. Martín, I.E. Grossmann, 2013, AIChE J., 59, 8, 2872-2883

M. Martín, 2016, Comp. Chem. Eng., 92, 43-54
M. Martín, 2016, Energy, 114, 833-845
M. Martin, I.E. Grossmann, 2017, App. Energy, DOI: 10.1016/j.apenergy.2017.10.121
J. Martínez, J. Peña, J.M. Ponche – Ortega, M. Martín, 2017, J. Environ. Manag., DOI: 10.1016/j.jenvman.2017.07.001
A.H. Masso, D.F. Rudd, 1969, AIChE. J., 15, 10-17.
G.D. Moggridge, E.L. Cussler, 2000, Chem. Eng. Res. Des., 78, 1, 5-11
J. Nease, T.A. Adams II. 2013, J. Power Sources, 228, 281-293
A. Nemet, J.J. Klemeš, Z. Kravanja, 2015, Comput. Chem. Eng., 72, 159-182
K.M. Ng, R. Gani, K. Dam-Johansen (eds), 2007, Comp. Aided Chem. Eng. 23. Elsevier. Oxford
S. Papoulias, I.E. Grossmann, 1983, Comput. Chem. Eng., 7, 6, 707–721
K.P. Papalexandri, E.N. Pistikopoulos, 1996, AIChE J., 42, 4, 1010-1032
K.P. Papalexandri, E.N. Pistikopoulos, C.A. Floudas, 1994, Chem. Eng. Res. Des., 72, 3, 279-294
E.N. Pistikopoulos, N.A. Diangelakis, 2016, Comput. Chem. Eng., 91, 85-92
P.M. Portillo, F.J. Muzzio, M.G. Ierapetritou, 2007, AIChE J., 53, 1, 119-128
E.N. Pistikopoulos, M.G. Ierapetritou, 1995, Comput. Chem. Eng., 19, 10, 1089-1110
J.W. Ponton, R.A.B. Donaldson, 1974, Chem. Eng. Sci., 29, 2375-2377
C. Pozo, R. Ruíz-Femenia, J. Caballero, G. Guillén-Gosálbez, L. Jiménez, 2012, Chem. Eng. Sci., 69, 1, 146-158
M. Psycha, K. Pyrgakis, P.J. Harvey, A. Ben-Amotz, A.K. Cowan, A.C. Kokossis, 2014, Proc. 8th Int. Conf. Found. Comput. Aid. Process Des. Mario R. Eden, John D. Siirola and Gavin P. Towler (Editors), 591-596
C. Ramírez-Márquez, M. Vidal, J.A. Vázquez-Castillo, M. Martín, J.G. Segovia-Hernández, 2017, J. Cleaner Prod., in press, DOI: 10.1016/j.jclepro.2017.09.126
L. A. Ricardez-Sandoval, 2012, Comput. Chem. Eng., 43, 91-107
D. Rudd, G. Powers, J. Siirola, 1973, Process synthesis. Englewood Cliffs, NJ: Prentice-Hall
R. Ruiz-Femenia, M.J. Fernández-Torres, R. Salcedo-Díaz, M.F. Gómez-Rico, J.A. Caballero, 2017, Ind. Eng. Chem. Res., 56, 25, 7301–7313
N.V. Sahinidis, 1996, J. Global Opt., 8, 2, 201-205
K. Sawitree, H. Rehan, E. Nimir, R. Gani, 2015, Proceed. 25th Euro. Symp. Comput. Aid. Proc. Eng., 1415–1420
J.D Seader, A.W. Westerberg, 1977, AIChE J., 23, 951
J.G. Segovia-Hernández, S. Hernández, A.B. Petriciolet, 2015, Chem. Eng. Process., 97, 134-43
W.D. Seider, D.R. Lewin, J.D. Seader, S. Widago, R. Gani, Ng K. M, 2017, Product and Process Design Principles: Synthesis, Analysis, and Evaluation. Wiley.
A.A. Shenvi, V.H. Shah, J.A. Zeller, R. Agrawal, 2012, AIChE J., 58, 8, 2479-2494.
E.M.B. Smith, C.C. Pantelides, 1995, Comput. Chem. Eng, 19(S), s83–s88.
R. Smith, 2005, Chemical Process Design and Integration. John Wiley and Sons. McGraw Hill.
M. Steffens, E. Fraga, I. Bogle, 1999, Comput. Chem. Eng, 23, 10, 1455–1467.
H. Struebing, S. Obermeier, E. Siougkrou, C.S. Adjiman, A. Galindo, 2017, Chem. Eng. Sci., 159, 69-83
N. Takama, T. Kuriyama, K. Shiroko, T. Umeda, 1980, Comput. Chem. Eng., 4, 251
P.S. Varbanov, J.J. Klemeš, 2011, Comput. Chem. Eng., 35, 9, 1815-1826
M. Vidal, M. Martín, 2015, Comput. Chem. Eng, 72, 273-283
J. Viswanathan, I.E. Grossmann, 1993, Ind. Eng. Chem. Res., 32, 2942
Y.P. Wang, R. Smith, 1994, Chem. Eng. Sci., 49, 981–1006
I.D. Washington, C.L.E. Swartz, 2014, AIChE J., 60, 9, 3151-3168
T.F. Yee, I.E. Grossmann, 1990, Comput. Chem. Eng., 14, 1165-l184
F. You, L. Tao, D.J. Graziano, S.W. Snyder, 2012, AIChE J., 58, 4, 1157-1180
Z. Yuan, B. Chen, 2012, AIChE J., 58, 11, 3370-89
Q. Zhang, M. Martín, I.E. Grossmann, 2017, Comp Aided Chem Eng, Vol 40
1879-1884
Y. Zhang, M.A. Dube, D.D. McLean, M. Kates, 2003, Bioresour. Tech., 89, 1
X. Zhao, Y. Wang, 2009, Proc. 2009 Int. Conf. Computational Intell. Natural Comput., CINC 2009, 2, 5231039, 84-87

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50002-1

Multilevel and multiscale PSE: Challenges and opportunities at mesoscales

Jinghai Li, Wenlai Huang, Wei Ge

State Key Laboratory of Multiphase Complex Systems, Institute of Process Engineering (IPE), Chinese Academy of Sciences (CAS), Beijing 100190, People's Republic of China
jhli@ipe.ac.cn

Abstract

This presentation discusses the challenges and opportunities at different levels of process engineering. It is indicated that all processes involve in multilevel structures, each multiscaled, and the complexity of each level exists always at the corresponding mesoscale at the specified level. On the other hand, these mesoscale structures at different levels are all subject to operating regimes, showing the most complex spatiotemporal behavior at mesoregimes between two relatively simple limiting regimes. There is no thereby available approach for mesoscale problems in mesoregimes, leading to the proposition of mesoscience based on the recognition of a common principle for all these mesoscale problems. The presentation will introduce what is the physical principle and mathematical formulation of mesoscience, why it is critical to the paradigm shifting of process engineering, particularly to the realization of virtual process engineering (VPE), which is the dream for process systems engineering (PSE). The presentation will be concluded with perspectives in mesoscience.

Keywords: Mesoscale, Mesoregime, Mesoscience, Multilevel, Multiscale.

1. Introduction: Levels, scales, and mesoscales

To study a system, it is important, and should be the first step to define its domain and its relationship with surroundings. In doing so, it is essential to clarify which level we are studying, which scales should be involved in the study, and what is the boundary and interaction between the system and its surroundings. Subsequently, with these definitions for the specified system, we should further know in which regime the system is operated. Unfortunately, it is quite often that these aspects were not specified well during study, making confusions and debates. Especially, the division of levels, the definition of mesoscale, and the regime transition in a specified level have not received sufficient attention, hindering the development of PSE and the realization of VPE.

As shown in Figure 1, the whole procedure for PSE concerns three levels, i.e., the **material** level (from the scale of an atom or a molecule to the scale of a macroscopic particle), the **reactor** level (from the scale of a macroscopic particle to that of a reactor), and the **factory** level (from the scale of a reactor to that of a factory). These levels belong to different disciplines, i.e., chemistry and materials science for the material level, chemical engineering for the reactor level, and PSE for the factory level, as reviewed by Grossmann and Westerberg (2000). Each level exhibits multiscale characteristics, involving the element scale, the system scale, and the in-between mesoscale where complexity always emerges (Li and Huang, 2014). Currently, lots of investigations have been undertaken at the element scales and the system scales, but

understandings are quite limited on the mesoscale problems at these three levels, e.g., self-assembling at the material level, particle clustering at the reactor level, and process synthesis superstructure (Floudas et al., 1986; Yee et al., 1990) at the factory level. There is no satisfactory physical theory or mathematical approach for describing these mesoscale problems, showing knowledge gaps. In PSE, although the interactions among unit operations are taken into account, and formulated as multi-objective variational or optimization problems, however, the mesoscale problems at the material level and the reactor level are usually handled with averaging (e.g., coarse-graining) approaches, and therefore, the fidelity of the physical models is frequently neglected.

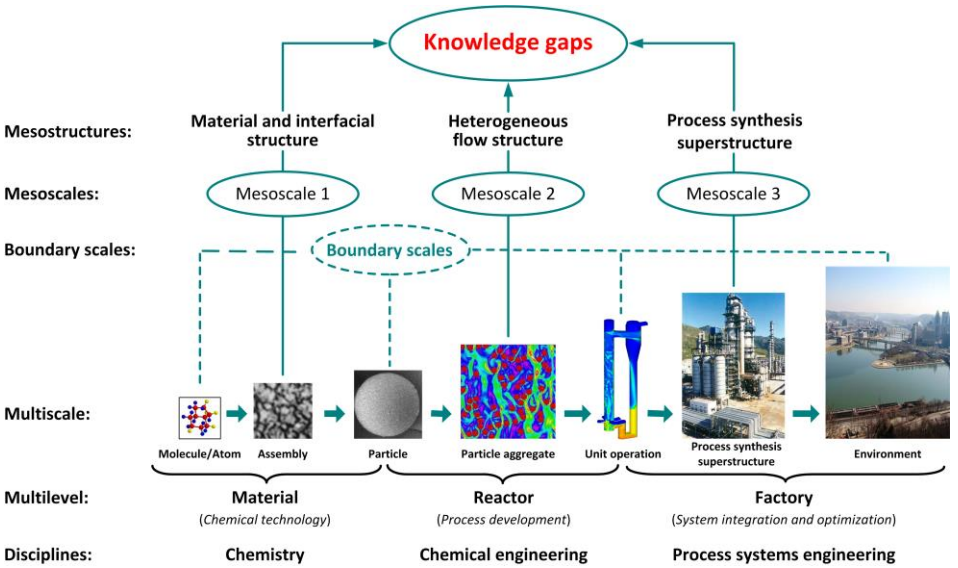


Figure 1 Three levels and corresponding mesoscales of chemical processes. Modified from (Li, 2015a), Copyright 2018, with permission from Elsevier

The complexity of mesoscale problems led to the proposition of nonequilibrium thermodynamics (Onsager, 1931a; 1931b), the theory of dissipative structures (Nicolis and Prigogine, 1977), complexity science (Wikipedia, 2018), etc. Recent investigations on the mesoscale structures at various levels revealed that traditional theories fail to offer satisfactory descriptions, and some common principle seems to exist, along with multi-objective formulations mathematically. Therefore, the concept of mesoscience was put forward (Li et al., 2013; 2014), which will be adopted in this work. In the language of mesoscience, three regimes (Li, 2016; Li et al., 2016) of the mesoscale structures may appear with the change in the operating conditions, i.e. (taking the systems with two dominant mechanisms as an example), two limiting regimes (mechanism *A*- or *B*-dominated) with simple structures, and the in-between mesoregime (*A*-*B* compromising) with complex structures. The complexity in this *A*-*B* compromising regime lies in the joint dominance of different mechanisms. Mesoscience concerns not only mesoscales, but also mesoregimes (Huang et al., 2017; Li, 2017). Therefore, the key (challenging) issues in PSE include bridging different scales in each

level, correlating different levels in a whole process, and distinguishing different regimes at each level.

2. Commonality and speciality of the three levels

All the complex systems at the above-mentioned three levels have common features: (1) They are composed of many elements; (2) There exist complex spatiotemporal structures at the mesoscales; (3) The mesoscale structures can be determined using stability conditions based on the principle of compromise in competition (Li, 2015a; 2015b; Li et al., 2016; Li et al., 2018; Li and Huang, 2018); (4) The stability conditions can be expressed as multi-objective variational problems in mathematics. However, level-specific features exist as well:

1) Dominant mechanisms are level-specific. At the material level, the mechanisms dominating mesoscale structures might be related to the processes of reaction, diffusion, etc. (Wang et al., 2013; Sun et al., 2016). At the reactor level, the dominant mechanisms might be subject to interaction between different phases, and also dependent on the reactor types. At the factory level, in addition to physical and chemical constraints, dominant mechanisms are also subject to the indices of economy, environments, etc. which are set to meet the production requirements.

2) For the first two levels, the processes and corresponding complex structures depend on the intrinsic nature at each level itself, and the interplay between the two levels. However, for the third level, the processes and structures are dependent not only on the constraints from the first two levels, but also on artificial factors (e.g., economic and ecological indices) (Grossmann, 2005; Bakshi, 2002; Triebl et al., 2013), bringing more complexity to PSE.

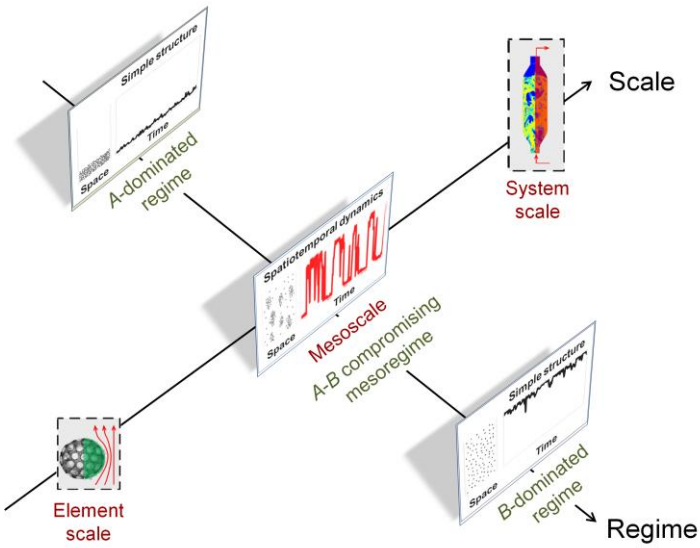


Figure 2 Regime-specific feature of mesoscales, exemplified with the reactor level

3) For the first two levels, the structures are determined by physical mechanisms under specified conditions. However, for the third level, the optimization is conducted firstly according to the specified conditions, and then the optimized

results should be returned to the first two levels to check the possibility of realization, so iterations and correlations between levels are necessary.

The above features already lead to lots of complexity, but moreover, the mesoscale structures at each level may change with the changes in specified conditions, usually exhibiting three distinct regimes, as shown in Figure 2 for the reactor level. Once such regime transitions occur during optimization, the situation should be more complicated. This is beyond our current capability in reactor designing and optimizing, it should be more difficult to account for.

Additionally, there exist problems in the literature, due to the ambiguity or even mistakes in defining levels, scales and mesoscales, which enhance the complexity and difficulty, as follows:

1) Identification of level: Naturally, one level corresponds to one kind of systems, and neighbouring levels might be taken as constraints of such a system, that is, the environment. However, sometimes one blurs different levels into a single problem, and sometimes one defines a "system" (e.g., a reactor) as a "mesoscale". Such treatments usually lead to mistakes or difficulties due to the confusions of different mechanisms at different levels.

2) Definition of mesoscale parameters: Mesoscale, as reviewed in (Li and Huang, 2018), is a relative concept to describe the scale with complex phenomena, between the element scale and the system scale at each level. Therefore, mesoscale parameters should reflect the intrinsic nature behind these phenomena, that is, the related dominant mechanisms should be taken into account, and interaction between the phases corresponding to such mechanisms should be formulated. For instance, in gas-solid fluidization, particle clustering and interaction between the dilute and the dense phases must be represented in its mesoscale model (Li and Kwauk, 1994).

3) Inter-level transfer of parameters: The bottom-up transfer of the parameters at the material level to the reactor level is relatively easy. However, the top-down transfer of the parameters optimized at the factory level to the reactor and material levels is very difficult (even more difficult when considering the regime-specific feature). It is a great challenge in practical optimization, and more transdisciplinary approaches and knowledge are necessary.

Therefore, under such circumstances, the future PSE should address the following issues:

1) Division of levels: Partitioning the overall system into correlating sub-systems, covering different levels, and identifying the dominant mechanisms in each sub-system.

2) Definition of scales at each level: Appropriate parameters should be given for the element scale, the system scale, and the structural unit at the mesoscale. Defining the mesoscale parameters is especially critical.

3) Correlation between levels: Determining the inter-level correlations in both top-down and bottom-up ways.

4) Resolution of mesoscales: Solving mesoscale problems depends on the understanding and description of mesoscale structures and the identification of the regime-specific features. Mesoscience is a promising way to this end.

Addressing these issues properly will pave the way for PSE to a new era, and lay cornerstones for VPE, where mesoscience or something equivalent will be the key.

3. Mesoscience bridges element behaviour and system performance

Over the past three decades, investigations on various mesoscale problems in process engineering have been conducted at IPE, CAS. Different from the methodology in traditional thermodynamics, which aims to find a single-objective variational function, and the complexity science that aims to directly correlate elements with the system, we focus on the complex structures at the mesoscales, and realize that there must exist knowledge gaps. Through analysing the compromise between competing mechanisms, we find the stability conditions for mesoscale structures, and express them as multi-objective variational problems (Li et al., 2004). As shown in Figure 3, such a EMMS principle originated from our EMMS model (Li and Kwauk, 1994) for gas-solid fluidization, and then was confirmed (Li et al., 1999; Ge et al., 2007; Huang and Li, 2016; Han et al., 2016) in turbulence, granular flow, gas-liquid flow, micro-fluidics, foam drainage, emulsion, heterogeneous catalysis, materials, micro-reactors, protein, impinging streams, and other systems. The commonality recognized from these different systems stimulated exploration of its universality, leading to the proposition of mesoscience.

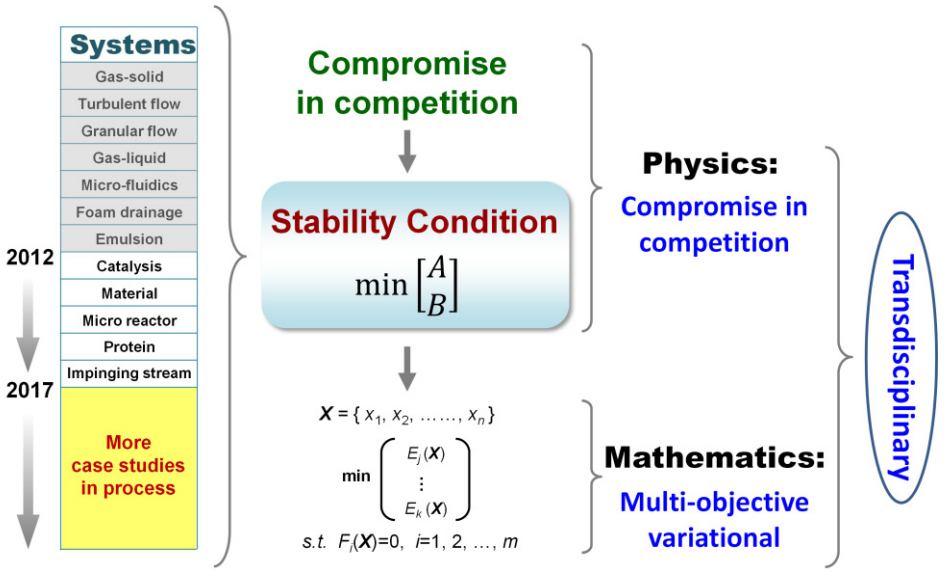


Figure 3 From case studies to common principle

Mesoscale structures are actually the focus of the theory of dissipative structures (Nicolis and Prigogine, 1977) and the complexity science (Wikipedia, 2018). After three decades of exploration, we reckon that the theory of dissipative structures is really revolutionary, but it is a pity that it does not succeed completely in describing complex systems, partly due to the limitation in directly searching for a single-objective

variational function. Therefore, the key is to find multi-objective variational functions, instead of single-objective variational functions. This is possibly the common principle we have revealed through the concept of mesoscience.

As we stated recently in an editorial (Li, 2017), the concept of mesoscience is based on the core principle of compromise in competition. Instead of looking for a single-objective variational function directly, different dominant mechanisms at the mesoscales in mesoregimes are first analyzed. The resulting variational functions are then defined and their compromise in competition is analyzed, mathematically, leading to multi-objective variational formulation. Physically, the origin of complexity is related to the alternate appearance of states dominated by different mechanisms. This shift in research strategy from single-objective to multi-objective optimization might offer a glimmer of hope for understanding the complex world in which nature exits. The solution of such a multi-objective variational problem represents a continuing grand challenge. However, in some cases the possibility to transform it into a single-objective problem, particularly based on physical analysis, should certainly not be excluded, as shown in our EMMS model (Li and Kwauk, 1994).

4. VPE perspective: Towards a new paradigm

The dream of PSE is to realize virtual reality (VR)-like VPE. The key lies in reasonable physical models, sufficient computational capability with high speed and efficiency, and the capability of handling big data. Mesoscience will contribute to all of these aspects, as summarized in Figure 4. Firstly, the establishment of mesoscience will clarify the relationship among the element, the system, and mesoscale structrues, i.e., identifying the intrinsic logic in complex systems. The complex mechanisms will be revealed for the mesoscale structures at each level in the mesoregime, and the physical principle (compromise in competition) and mathematical formulation (multi-objective variational) will be established. Secondly, based on the intrinsic logic, reasonable physical models can be achieved, guaranteeing the reliability of computation. Then, according to the common logic and structural features of mesoscale structures, reasonable computers can be designed, which will provide reasonable logic and structure for the computational science aiming at solving complex problems, establish the logic consistency among the problem, the model, software, and hardware, and hence improve the computational efficiency, achieving high speed and efficiency. This will supply powerful tools for mesoscience. Meanwhile, the analytic processing of the big experimental data will also be performed under the logic framework determined via mesoscience. This will greatly expedite the data processing and improve the capability of revealing underlying mechanisms, providing more evidence for the development of mesoscience and online data for the realization of virtual reality. At present, we have established a VPE center (Ge et al., 2011; Liu et al., 2012), the computational speed supports two-dimensional realtime virtual experiment (Lu et al., 2016). In other words, if one changes the operating conditions, the dynamic evolution of the system can be observed immediately on the screen. If such a scene is realized for three-dimensional industrial apparatuses, realtime VR simulations will be achieved, and VPE will not be far away.

Mesoscience originated from chemical engineering where there are lots of mesoscale phenomena. In developing chemical engineering, we can collect more evidence to verify the universality of mesoscience, and improve the intelligence in chemical engineering (Quantrille and Liu, 1991).

In summary, the complexity of mesoscale structures at different levels exhibits the regime- and level-specific nature. This is the common bottleneck at all the levels in chemical engineering, and also the missing link in current knowledge. The proposed mesoscience aims at solving such challenging problems. Certainly, more evidence is still needed to support our current understandings on the physical principle and the mathematical expression.

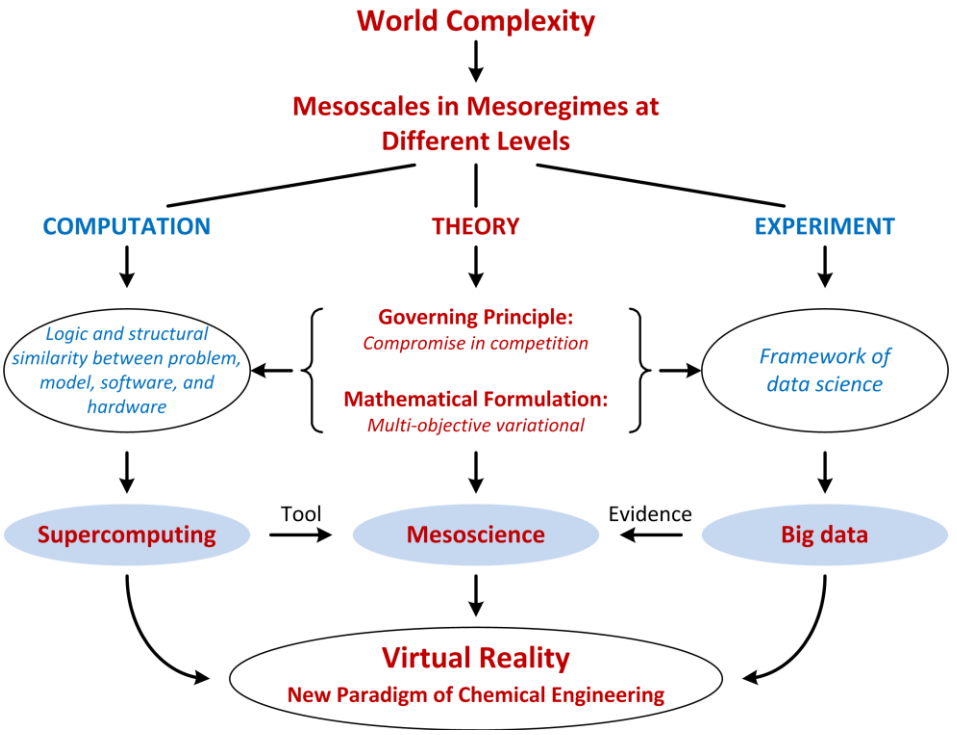


Figure 4 Relationship between big data, supercomputing, mesoscience, and virtual reality, showing the path to a new paradigm of chemical engineering. Modified from (Li, 2015b), Copyright 2018, with permission from Elsevier

Since the complexity in nature, mostly prevailing in mesoregimes, originates from the coexistence and compromise in competition between different dominant mechanisms, it seems difficult for traditional single-objective variational functions to describe such complexity. Different mechanisms follow different rules, and additionally the world is full of diversity, so directly establishing a single-objective description seems more difficult than finding a multi-objective description. Therefore, mesoscience deserves exploring in various fields. Once we collect sufficient evidence to verify the universality of the principle of compromise in competition, and develop the data and computational science as shown in Figure 4, we will greatly promote the realization of VPE—a new paradigm of chemical engineering!

References

B.R. Bakshi, 2002, A thermodynamic framework for ecologically conscious process systems engineering, Computers and Chemical Engineering, 26, 269-282

C.A. Floudas, A.R. Ciric, I.E. Grossmann, 1986, Automatic synthesis of optimum heat exchanger network synthesis, AIChE Journal, 32, 276-290

C. Triebl, V. Nikolakis, M. Ierapetritou, 2013, Simulation and economic analysis of 5-hydroxymethylfurfural conversion to 2,5-furandicarboxylic acid, Computers and Chemical Engineering, 52, 26-34

F. Sun, W.L. Huang, J. Li, 2016, Structural characteristics of adlayer in heterogeneous catalysis, Chemical Engineering Science, 153, 87-92

G. Nicolis, I. Prigogine, 1977, Self-organization in nonequilibrium systems: From dissipative structures to order through fluctuations, Wiley, New York

H. Wang, W. Huang, Y. Han, 2013, Diffusion-reaction compromise the polymorphs of precipitated calcium carbonate, Particuology, 11, 3, 301-308

I.E. Grossmann, A.W. Westerberg, 2000, Research challenges in process systems engineering, AIChE Journal, 46, 1700-1703

I. Grossmann, 2005, Enterprise-wide optimization: A new frontier in process systems engineering, AIChE Journal, 51, 1846-1857

J. Li, 2015a, Approaching virtual process engineering with exploring mesoscience, Chemical Engineering Journal, 278, 541-555

J. Li, 2015b, Mesoscales: The path to transdisciplinarity, Chemical Engineering Journal, 277, 112-115

J. Li, 2016, Exploring the logic and landscape of the knowledge system: Multilevel structures, each multiscaled with complexity at the mesoscale, Engineering, 2, 276-285

J. Li, 2017, Mesoscale spatiotemporal structures: Opportunities from challenges, National Science Review, in press, DOI: 10.1093/nsr/nwx124

J. Li, J. Zhang, W. Ge, X. Liu, 2004, Multi-scale methodology for complex systems, Chemical Engineering Science, 59, 1687-1700

J. Li, M. Kwauk, 1994, Particle-fluid two-phase flow: The energy-minimization multiscale method, Metallurgical Industry Press, Beijing

J. Li, W. Ge, W. Wang, N. Yang, W. Huang, 2016, Focusing on mesoscales: From the energy-minimization multiscale model to mesoscience, Current Opinion in Chemical Engineering, 13, 10-23

J. Li, W. Ge, W. Wang, N. Yang, X. Liu, L. Wang, X. He, X. Wang, J. Wang, M. Kwauk, 2013, From multiscale modeling to meso-science: A chemical engineering perspective, Springer, Berlin

J. Li, W. Huang, 2014, Towards mesoscience: The principle of compromise in competition, Springer, Berlin

T.E. Quantrille, Y.A. Liu, 1991, Artificial Intelligence in Chemical Engineering, Academic Press, Inc., San Diego

J. Li, W. Huang, 2018, From multiscale to mesoscience: Addressing mesoscales in mesoregimes of different levels, Annual Review of Chemical and Biomolecular Engineering, in press, DOI: 10.1146/annurev-chembioeng-060817-084249

J. Li, W. Huang, J. Chen, W. Ge, C. Hou, 2018, Mesoscience based on the EMMS principle of compromise in competition, Chemical Engineering Journal, 333, 327-335

J. Li, Z. Zhang, W. Ge, Q. Sun, J. Yuan, 1999, A simple variational criterion for turbulent flow in pipe, Chemical Engineering Science, 54, 1151-1154

L. Lu, J. Xu, W. Ge, G. Gao, Y. Jiang, M. Zhao, X. Liu, J. Li, 2016, Computer virtual experiment on fluidized beds using a coarse-grained discrete particle method—EMMS-DPM, Chemical Engineering Science, 155, 314-37

L. Onsager, 1931a, Reciprocal relations in irreversible processes. I, Physical Review, 37, 405-426

L. Onsager, 1931b, Reciprocal relations in irreversible processes. II, Physical Review, 38, 2265-2279

M. Han, J. Xu, Y. Ren, J. Li, 2016, Simulations of flow induced structural transition of the β-switch region of glycoprotein Ibα, Biophysical Chemistry, 209, 9-20

T.F. Yee, I.E. Grossmann, Z. Kravanja, 1990, Simultaneous optimization models for heat integration—I. Area and energy targeting and modeling of multi-stream exchangers, Computers and Chemical Engineering, 14, 1151-1164

W. Ge, F. Chen, J. Gao, S. Gao, J. Huang, X. Liu, Y. Ren, Q. Sun, L. Wang, W. Wang, N. Yang, J. Zhang, H. Zhao, G. Zhou, J. Li, 2007, Analytical multi-scale method for multi-phase complex systems in process engineering—bridging reductionism and holism, Chemical Engineering Science, 62, 3346-3377

W. Ge, W. Wang, N. Yang, J. Li, M. Kwauk, F. Chen, J. Chen, X. Fang, L. Guo, X. He, X. Liu, Y. Liu, B. Lu, J. Wang, J. Wang, L. Wang, X. Wang, Q. Xiong, M. Xu, L. Deng, Y. Han, C. Hou, L. Hua, W. Huang, B. Li, C. Li, F. Li, Y. Ren, J. Xu, N. Zhang, Y. Zhang, G. Zhou, G. Zhou, 2011, Meso-scale oriented simulation towards virtual process engineering (VPE)—the EMMS paradigm, Chemical Engineering Science, 66, 4426-4458

W. Huang, J. Li, P.P. Edwards, 2017, Mesoscience: Exploring the common principle at mesoscales, National Science Review, in press, DOI: 10.1093/nsr/nwx083

W.L. Huang, J. Li, 2016, Mesoscale model for heterogeneous catalysis based on the principle of compromise in competition, Chemical Engineering Science, 147, 83-90

Wikipedia, 2018, Complex system, https://en.wikipedia.org/wiki/Complex_system

X. Liu, L. Guo, Z. Xia, B. Lu, M. Zhao, F. Meng, Z. Li, J. Li, 2012, Harnessing the power of virtual reality. Chemical Engineering Progress 108, 28-33

This page intentionally left blank

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50003-3

Chemical Product Design: Advances in Research and Teaching

Ka M. Ng[a], Rafiqul Gani[b]

[a]*Department of Chemical and Biological Engineering, The Hong Kong University of Science & Technology, Hong Kong (kekmng@ust.hk)*
[b]*PSE for SPEED, Skyttemosen 6, DK-3450 Allerod, Denmark (rgani2018@gmail.com)*

Abstract

After its launch at the turn of the millennium, integrated product and process design has gradually taken root in research and teaching within chemical and biochemical engineering. The transition from primarily process design in which the product is relatively well defined to include product design is primarily driven by economics. With globalization, any product that can be made by multiple producers would eventually exert enormous financial pressure to lower product price, resulting in squeezed profit margin. The only way for a company to survive and prosper is to invent innovative products. Also, with the rapid advances in computer technology, emergence of new business models, and requirements in social responsibility and sustainability, a chemical engineer should and is well positioned to contribute to the entire product life cycle. This article identifies the technical and non-technical issues/problems in integrated product and process design, the relationships among which are captured in a Grand Product Design Model. The methods/techniques and computer aided tools for the design, analysis, and development of molecular products, formulated products, functional products, and devices are discussed. Many of these recent developments have been included in teaching product design to prepare the new generation of chemical engineers.

Keywords: Product design, Grand Product Design Model, Formulated products, Functional products, Devices

1. Introduction

Chemical product design is a diverse subject. In addition to the basic science and engineering underlying the product under consideration, consumer preference, budget, competing products, pricing, supply chain analysis, government policy, corporate social responsibility, sustainability, and so on also need to be taken into account in designing and developing a product. This is captured in the Grand Product Design and Development Model (see Figure 1), which shows the relationships among the different tasks / problems in the product design and development process (Fung et al., 2016). Normally, the development process begins with consumer preferences, which define the desired product quality. The ingredient and production process are then appropriately selected (designed) to yield product properties and product structure, which together provide product attributes that meet product quality requirements. Note that supply chain analysis is used to optimize the selection of product ingredients. After accounting for product cost and non-manufacturing expenditure, a pricing model is needed to maximize profit while meeting corporate social responsibility. Other issues may also be considered such as sustainability, company strategy, aesthetics, and so on. This model is part of a hierarchical, multidisciplinary framework for chemical product design (Seider et al., 2017).

There are 5 elements (highlighted through different colors in Figure 1) of the design framework that are used for carrying out product design. Model-based methods, rule-based methods, and databases do not involve hardware while computational and experimental tools do. Note that not all models from Figure 1 are considered in all product design problems.

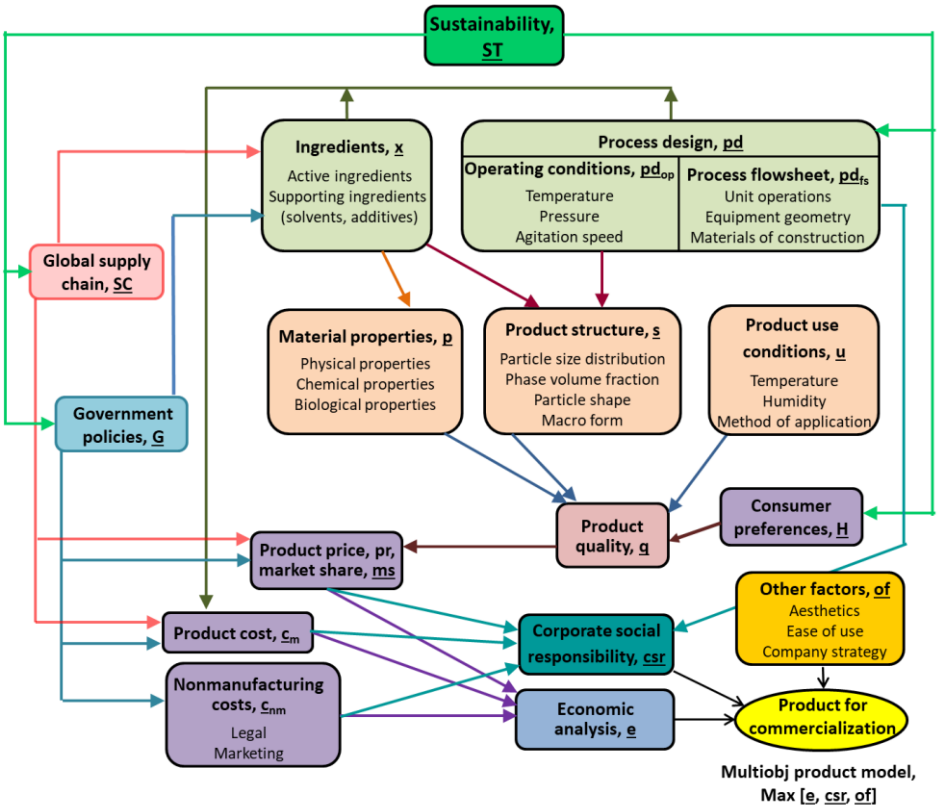


Figure 1: Grand Product Design and Development Model (Fung et al., 2016).

Many chemical products have been considered in product design. These include fuel additives (Sundaram et al., 2001; Hada et al., 2014), refrigerants (Sahinidis et al., 2003), perfumes (Mata et al., 2005), slow release deodorizer (Street et al., 2008), medical diagnostic products (Heflin et al., 2009), disinfectant (Omidbakhsh et al., 2012), biofuel (Dahmen and Marquardt, 2016), solvents for reaction synthesis (Struebing et al. 2017) and so on. A classification of the chemical product types is given by Seider et al. (2017). A review of these publications shows that one or more of the five elements of the design framework is used for design of these products. Model-based design methods often play a prominent role. For example, Bernardo and Saraiva (2015) treated a product design problem as the inversion of three design functions: quality, property, and process functions. Xiao and Huang (2009) developed a model for paint design. Often, two or more of the elements are used synergistically. For example, Wibowo and Ng (2001), Cheng et al. (2009), Smith and Ierapepritou (2010), Conte et al. (2011) and Mitrofanov et al. (2012) used a combination of rule-based methods, model-based meth-

ods, and data-bases in the design of different types of product. Experimental tools are needed in product design either for obtaining data such as material properties that are not available in the database or for fabricating a product prototype for testing its performance (Conte et al., 2012). Similarly, computer-aided tools for product design are highly desirable (Kalakul et al., 2017).

Most of the publications in chemical product design focus on Ingredients, Process design, Material properties, and Product quality of the Grand Product Design Model (Figure 1) and relatively little has been done on product pricing, corporate social responsibility and government policies, sustainability, and so on. This is natural because materials and processing are Chemical Engineers' core competencies. However, the other problems need to be considered in order to answer the seemingly simple question of "what to make?" Making a product that has the best quality on the market does not guarantee commercial success. The product can be priced beyond the consumers' budget even if it has all the product attributes to satisfy the consumer preferences. In product design, nonmanufacturing costs such as legal and advertising costs can be substantial. This is in stark contrast to process design where the raw material cost easily constitutes over half of the product cost. Government policy has a huge impact on what to make. Without government subsidy, electric vehicles (and all the chemical products associated with them) and solar panels would take much longer to open up the market to drive down the product cost. A product proposed by the marketing and engineering teams may not be approved by management because of social responsibilities.

In this article, the aforementioned problems are reviewed in more detail focusing on the design and analysis of the major types of chemical products – molecular, formulated, functional, and devices. Special emphasis is placed on the PSE tools and methods for their design. In view of the fact that product design is being incorporated into the chemical engineering curriculum around the world, the recent developments on the educational front are also discussed. The article ends with a discussion of the challenges, gaps, and opportunities in product design. Unless otherwise mentioned, "product design" in this article refers to "chemical product design"

2. Design, analysis and development

Figure 2 shows the relationships among the four major types of chemical products in product development. Single species products can be further classified as small molecules (refrigerants, solvents) or large molecules (active ingredients, surfactants). Usually, these chemical products have process applications (separation, reaction) as well as product applications (part of formulated or functional products). The single species small molecular products are usually produced in bulk quantities where the process costs are a key to their success, in addition to their product attributes. Design methods for these products are well established (Gani, 2004, Mitrofanov et al. 2012, Samudra and Sahinidis, 2013). Because of the size of the molecules and the product attributes (not all can be modelled), design methods for large or complex molecular products are not well established, although there is increased interest to develop them. For example, Karunanithi and Mehrkesh (2013) reported a method for ionic liquid design, Zhang et al. (2015) for surfactant design, and Gerbaud et al. (2017) for bio-based molecules. Formulated products are obtained by mixing selected components together to get the desired product attributes (Zhang et al., 2017). These include creams and pastes, paints,

shampoos, insect repellents, inks, and so on. They often contain different ingredients to fulfill the desired product attributes. For example, a sunscreen lotion can easily contain around 20 ingredients. In the design of such a lotion, the selection of more complex ingredients including UV absorbing agent, emollient, emulsifier, stabilizer, neutralizer, humectant, film former, thickener, and preservative depends a great deal on experience in the form of rule-based methods and databases. The rules and heuristics for ingredient selection are often based on an understanding of the basic science. In contrast, the solvents used for formulated products can be reliably designed through computational tools because of the extensive database and available models for thermodynamic properties. Suitability of an ingredient is decided not only by its physiochemical properties but also by its supply and cost. Prototyping is performed to check whether the formulation meets the product requirements. Iteration is often required to come up with the final product.

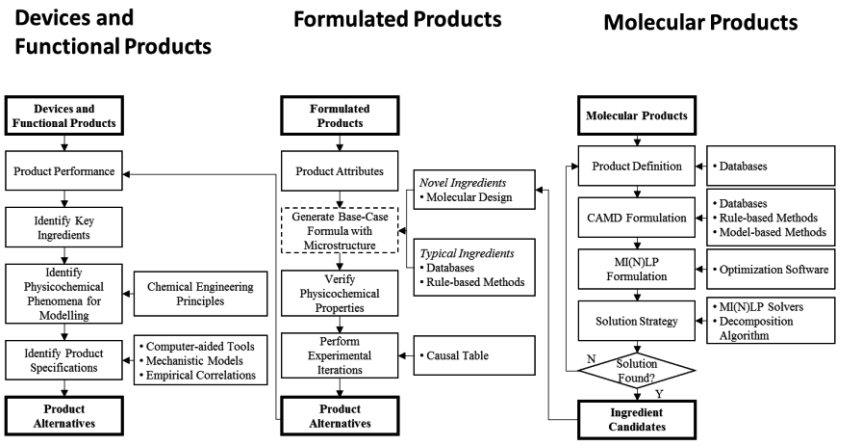


Figure 2: Relationships among the four major types of chemical products.

Functional products are those chemical products made up of materials that perform a desired function. Generally, these products do not have feed and outlet streams, and do not involve mechanical and electrical parts. They serve as barriers (food packaging), delivery vehicles (controlled release granule, transdermal patch), absorbers (silica gel), etc. The performance of functional products depends on the material properties of the ingredients, and the structure, form, shape or configuration of the product. Figure 3 shows a reservoir type transdermal patch with five major parts (Cheng et al., 2010). A drug reservoir that holds the pharmaceutical active ingredient (API) in a carrier. A rate-controlling membrane that controls the rate at which the API is released from the reservoir. A liner that prevents unwanted release. A backing layer that prevents unwanted release and protects the API from the environment. The adhesive layer simply bonds the liner to the rate-controlling membrane. In designing a functional product, each part of the functional product has to play its role to deliver the overall product performance, which can be modelled using chemical engineering principles. Clearly, diffusion, permeation, and adhesion are the key mechanisms in this functional product, and experimental tools such as a Franz diffusion cell system are needed in product design.

Chemical devices are those chemical products that perform a particular purpose, especially those with mechanical and electrical parts. Chemical engineers can view a chemi-

cal device as a single miniaturized chemical processing equipment item, a group of interconnected equipment items, or even a chemical plant. Often, a feed stream to a chemical device is transformed into an outlet stream with characteristics specified in the product attributes by performing reactions, fluid flow, heating/cooling, separations, addition/formation, and/or removal/destruction. The first four items are conventional chemical engineering principles which can be used to model the chemical device. For example, an air purifier with UV-based catalytic decomposition of VOCs can be modelled based on Langmuir-Hinshelwood kinetics (Seider et al., 2017). As will be explained in more detail in the section on Education below, the last two items include unconventional processing techniques such as solution coating, granulation, etching, breakage, and so on. As the boundary of the chemical engineering discipline expands and overlaps with bioengineering, biomedical engineering, material science, data analytics, and so on, it is expected that the integration of systems engineering principles with the newly developed domain knowledge can spur the development of innovative chemical devices, and expand the capability and reach of our profession.

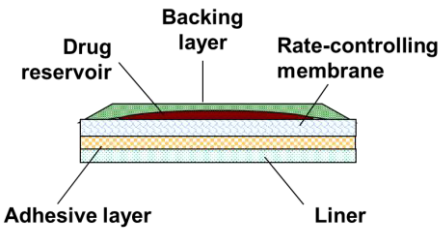


Figure 3: A reservoir type transdermal patch.

3. Computer Aided Techniques & Tools

Like process synthesis and design, chemical product synthesis and design can also be formulated mathematically and solved with many of the available numerical tools. Since the early 1980s, when the first CAMD (computer aided molecular design) technique related to solvent design was proposed, many developments in various directions have been reported. In all cases, the synthesis-design problem is formulated as an optimization (MILP or MINLP) problem, which is then solved in a variety of ways, rule-based methods (Gani 2004), graphical-visual methods (Solvason et al. 2009), decomposition-based methods (Karunanithi et al., 2005) as well as direct solution of the optimization problem (Samudra and Sahinidis, 2013, Zhang et al., 2015; Jonuzaj et al., 2016) with an appropriate numerical solver. The objective of these computer-aided techniques is not to determine the final design but to quickly identify a number of promising candidates, which can be verified by focused experimental means. Therefore, an integration of the computer-aided techniques with experiments is recommended. The text here is divided into two main sub-sections: computer aided techniques and computer-aided tools.

3.1 Computer-aided Techniques

3.1.1 CAMD Techniques: In addition to the earlier generate and test paradigm, computer-aided options have been reported for refrigerants (Sahinidis et al. 2003), ionic liquid design (Karunanithi and Mehrkesh, 2013), enzyme design (Frusicheva et al., 2014), surfactant design (Zhang et al., 2015), for solvents for reaction synthesis (Zhou et al., 2017), lipids or bio-based molecular design (Gerbaud et al. 2017), chemical substitution (Jhamb et al., 2018) and many more, using a variety of methods mentioned above. In all

these problems, single species molecules are synthesized, screened and/or designed by combining building blocks (to represent the molecule) and property prediction tools to estimate the desired (target) molecular properties. The CAMD techniques are well developed because of the availability of large amounts of data and therefore, property models. Integration with process issues is an important factor for these single species products (especially small molecules) and integrated product-process design techniques have also been reported (Hostrup et al., 1999; Li et al., 2007; Lampe et al., 2014).

3.1.2 CAM^{b}D techniques: Unlike the CAMD technique, in this class of problems, a basic set of molecules and their properties are known but which molecules will be present in a single-phase liquid mixture together with their composition in the mixture are unknown. Like CAMD problems, these mixture or blend design problems have also been formulated as MILP or MINLP problems and solved by a variety of ways. Tailor-made fuel blends (Yunus et al., 2014; Ordruei and Elkamel, 2015; Dahman and Marquardt, 2017; Choudhury et al., 2017) and solvent mixtures for gas separation (Jonuzaj et al. 2016) are examples from this area. Although the design steps (work-flow) are the same for different blended products, the molecules, their properties and product attributes are different, thereby increasing the dependence on data and models for a wider application range. Also, a computer-aided framework that is able to provide users with the necessary data and models needed for each step of the design process, becomes a very useful option to have.

3.1.3 Formulation design techniques: Unlike the CAM^{b}D techniques, where molecules from the same set are identified for the blend, here the set of molecules to define the formulation are divided into separate classes: ingredients (performs the main function of the formulated product); additives of different types (improves the quality or performance of the formulated product). Note, however, the formulated product must be a single liquid phase. Although, the potential application range of this kind of technique is very large, their success depends to a large extent on the available data and property models. Conte et al. (2011) employed a decomposition-based technique using data, property models and product performance calculation models to design a variety of formulated products. Here also, the design steps are similar for different formulated products, but the molecules, their properties and product attributes are different, thereby increasing the dependence on data and models for a wider application range. Again, availability of a computer-aided framework becomes a very useful option to have.

3.1.4 Functional products and devices: Here, most of the design techniques are based on heuristics, database, experiments and very limited modelling. For example, Mattei et al. (2015) proposed a knowledge-based system for designing emulsified products. Morales-Rodríguez and Gani (2007) provided models for evaluation of the performance of functional products and devices. Most of developed computer-aided techniques, however, are product specific. Consequently, generic computer-aided frameworks are not yet developed for these types of products.

3.2 Computer Aided Product Design Tool

Although a number of systematic computer-aided techniques for design of various types of products have been reported, only one so-called chemical product simulator, ProCAPD, with similar functions as a process simulator has been reported (Kalakul et al., 2017) and available for general use (see Figure 4). The ProCAPD software tool contains

a large collection of data organized in terms of databases (organic chemicals, solvents, active ingredients, lipids, ionic liquids, and many more); property models organized in terms of pure component properties, functional properties, bulk mixture properties and phase-equilibrium related properties (a total of 56 properties and their corresponding models are currently available); product performance calculation routines (phase stability, solvent evaporation rate, controlled release, etc.); product design templates (solvents, fuel-blends, formulations, functional products and devices); product analysis (model based verification of product attributes); and, utility tools (design template generation, property model creation, etc.). Many of the computer-aided techniques mentioned above (section 3.1) are available in ProCAPD. It is suitable for teaching (See examples in Seider et al., 2017), getting quick estimates for promising new products, evaluating product attributes and many more.

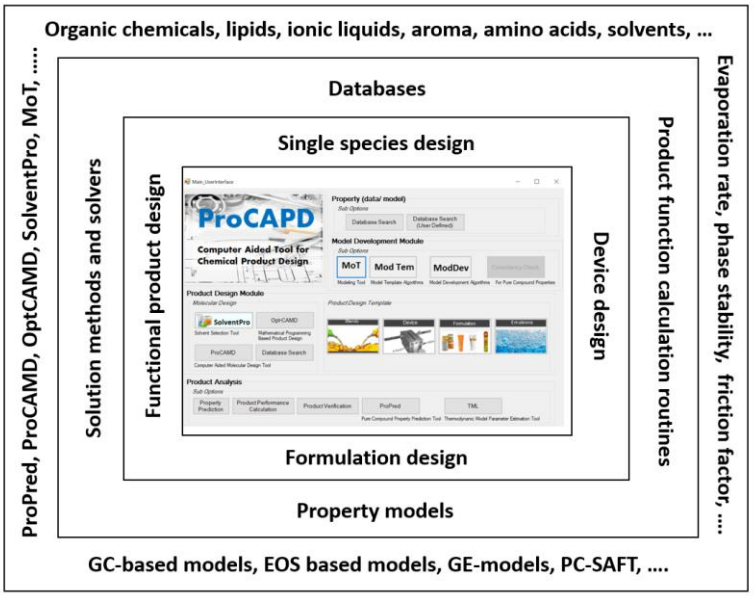


Figure 4: Structure of ProCAPD (design options, models-calculation routines, tools)

4. Education

Generations of chemical engineering students have been taught to design a process to produce a specified product, usually a bulk chemical. There has not been any discussion as to why the particular product is selected and who decided to make it? This is a serious omission for two reasons. From a wider perspective, it takes away the opportunity to challenge the students to innovate to come up with chemical products to meet human needs and to expand the company's product line. Another reason is that an optimal process for an inferior product does not really provide the best return on investment. In other words, the optimal product should be the best product-process combination. In chemical engineering education, it is fine to initially concentrate on the production of single species small molecules in bulk quantities where the process issues play a key role. Then, product design is introduced for single species large molecules or multi-species formulated and functional products, where the molecular structure and the ingredient identity and composition play also a major role. These concepts are included in

product design courses taught at the Hong Kong University of Science and Technology. Table 1 shows the topics covered in a course for chemical engineering seniors. Most of the topics are included in a textbook (Seider et al., 2017). There are still a number of challenges that require a lot more work. Two gaps are highlighted below.

Table 1: Contents of a chemical product design course

Topic	**Sub-topics**
Introduction	*The Diversity of Chemical Products; Companies Engaging in Production of Chemical Products*
Product Design and Development	Multidisciplinary Hierarchical Product Design (Elements of the Product Design Framework for Task Execution); Project management (Objective-time Chart, RAT^{2}IO); Market study (Consumer preference, Conjoint analysis, QFD, House of Quality, Business model canvas); Research & development (Innovation map, Patent search, Technology platform)
Design of Devices, Functional Products, and Formulated Products	*The Use of Models in Designing Devices and Functional Products; Design of Formulated Products*
Design of Processes for Manufacturing B2C Products	Unconventional Processes
Business Decision-Making in Product Development	Cash Flow Diagram; Financial Analysis for Product Development; Make-or-buy Analysis
Microeconomics of Product Design	Supply and Demand; Consumer Behavior; Individual and Market Demand; Pricing
Molecular and Mixture Design	*Framework for Computer-Aided Molecular-Mixture Design Case Studies; ProCAPD (chemical product design software)*
Optimization in Product Design	*Optimization Basics; Design Problem Formulation & Solution with OptCAMD (with GAMS)*
The Grand Product Design Model	*Government and Social Impact; Supply Chain Analysis; Sustainability Analysis*
Case Studies	*Refrigerants; solvents; surfactants; tailor-made fuel-blends; mosquito repellent; creams and pastes; inkjet inks; die attach adhesive; solar control interlayer; desiccant dehumidifier; biosensor; and so on (presented throughout the course duration)*
Presentations	Presentation by students of their design projects

Note: The topics marked in *italics* are parts of a graduate level course on computer-aided chemical product design given by the co-authors at different institutions.

4.1 Unconventional processing techniques

The way unit operations such as distillation, crystallization, absorption, adsorption, filtration, and so on is taught, with support from transport phenomena, thermodynamics and kinetics, has not fundamentally changed for decades. While this is sufficient for the production of commodity chemicals in the form of single species molecular products, it

is grossly inadequate for the manufacture of formulated products, functional products, and devices. As shown in Figure 5, unconventional processing techniques has two major categories. Addition/formation includes solution coating, physical vapor deposition, granulation, tableting, lamination, sintering, and so on while removal/destruction includes etching, homogenization, nanomization, breakage, and so on. These topics require a significant amount of effort to organize the materials for teaching.

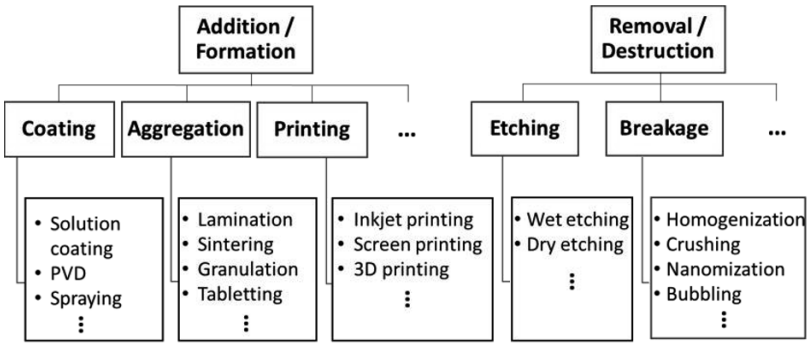


Figure 5: Unconventional processing techniques

4.2 Interface between the underlying science and product design

A thorough understanding of why the product works is necessary to properly design a chemical product. The constituent atoms, configuration, electronic structure, and intramolecular bonds of a molecular product determine its physicochemical properties. The physicochemical properties of the constituents, and microstructure/macrostructure of a formulated product, functional product, and device determine its performance and product attributes. Often, the underlying physics and chemistry is not covered in a typical chemical engineering curriculum. For example, to properly design an inkjet ink, phenomena including suspension dispersion, adhesion, wetting, droplet formation, and so on need to be modelled (Tam et al., 2016). Similarly, understanding of the interaction between biomolecules and fluorescent agents is essential in the design of an L-lactate biosensor (Zhang et al., 2017). Because of the diversity of these chemical products, the instructor may want to emphasize those scientific platforms that are suitable for the students. Case studies can be included in the course to reinforce all the concepts learned (Ng and Seider, 2018).

5. Perspective and Conclusions

Much has been achieved since the launching of product design research and teaching in the chemical engineering community at the turn of the new millennium (Cussler, 1999; Stephanopoulos, 2003; Cussler and Wei, 2003; Hill, 2004; Gani, 2004; Zhang et al., 2016). Despite the ups and downs in its development, it is firmly believed that product design is going to play a much more important role in our profession for a singular reason: the information age has led to changing business models in many market sectors. The chemical processing industry is no exception. Fluctuating revenues and stagnant profit margins in commodity chemical companies are quite common these days. The chemical sectors with high growth potential include healthcare products, high performance materials, consumer products, etc. All of these sectors require the capability to

come up with the product concept and the process to make the product. It is necessary to collaborate and work with other disciplines and faculty from bioengineering, biomedical engineering, materials, and chemistry, to name a few.

There are, however, many gaps to fill and challenges to meet. From a research and development point of view, the methods and computational tools need to be generic, for which sufficient knowledge is still not available. Development of many products is still based on expert knowledge and/or iterative trial and error because the integration of the underlying science and engineering is still not clear for these products. There is a large amount of data available in different forms and at different sources – they need to be collected, checked for consistency and then employed through smart search, retrieve and data-manipulation tools. Efforts should be made to systematically identify gaps in data so that experiments can be designed and planned to fill them, thereby leading to the developments of generic and predictive property models. The integration of product-process needs to be further developed together with issues related to supply chain, sustainability and economics. The business model with appropriate costing and pricing models also need to be developed.

Product design as described in this article is expected to bring together the people in the new areas and the people in process systems engineering (PSE). This integration suggests that PSE should take on a different meaning - Product/Process Systems Engineering. The methods and tools from PSE, when properly refined and applied, can significantly reduce the design and development times for chemical products.

6. References

Bernardo F P, Saraiva P M, A conceptual model for chemical product design. AIChE Journal, 2015; 6:802-815.

Cheng Y S, Lam K W, Ng K M, Ko R K M, Wibowo C, An integrative approach to product development—A skin-care cream. Computers & Chemical Engineering, 2009; 33:1097-1113.

Cheng, Y S, Ng, K M, and Wibowo, C, Product Design: a Transdermal Patch Containing a TCM Tincture, Ind. Eng. Chem. Res., 2010; 49:4904-4913.

Choudhury, H A, Intikhab, S, Kalakul, S, Khan, M, Tafreshi, R, Gani, R, Elbashir, N, Designing a Surrogate Fuel for Gas-to-Liquid (GTL) Derived Diesel. Energy & Fuels, 2017; 31:11266-11279.

Conte, E, Gani, R, and Ng, K M, Design of Formulated Products: a Systematic Methodology, AIChE Journal, 2011; 57:2431-2449.

Conte, E, Gani, R, Cheng, Y S, and Ng, K M, Design of Formulated Products: Experimental Component, AIChE J., 2012; 58:173-189.

Cussler, E L, Do changes in the chemical industry imply change sin curriculum? Chem. Engr. Edu., 1999; 33:12-17.

Cussler, E L, Wei, J, 2003, Chemical product engineering, AIChE Journal, 2003; 49:1072-1075.

Dahmen M, Marquardt W, Model-Based Design of Tailor-Made Biofuels. Energy & Fuels, 2016; 30:1109-1134.

Dahmen M, Marquardt W, 2017, Model-Based Formulation of Biofuel Blends by Simultaneous Product and Pathway Design. Energy & Fuels, 2017; 31:4096-4121.

Frushicheva M P, Mills M J L, Schopf P, Singh M K, Prasad R B, Warshel A, Computer aided enzyme design and catalytic concepts. Curr. Opin. Chem. Biol., 2014; 21:56–62.

Fung K Y, Ng K M, Zhang L, and Gani R, A grand model for chemical product design. Computers & Chemical Engineering, 2016; 91:15-27.

Gani, R, Chemical product design: challenges and opportunities, Computers and Chemical Engineering, 2004; 28: 2441-2457.

Gerbaud, Dos Santos, M T, Pandya, N, Aubry, J M, 2017, Computer aided framework for designing bio-based commodity molecules with enhanced properties, Chemical Engineering Science, 2017; 159:177-193.

Hada S, Solvason C C, Eden M R, Characterization-based molecular design of bio-fuel additives using chemometric and property clustering techniques. Front. Energy Res., 2014; 2: 1-12.

Heflin L, Walsh S, and Bagajewicz M. Design of medical diagnostics products: A case-study of a saliva diagnostics kit. Computers & Chemical Engineering, 2009; 33:1067-1076.

Hill, H. 2004, "Product and process design for structured products," AIChE Journal, 2004; 50:1656-1661.

Hostrup M, Harper P M, Gani R. 1999. Design of environmentally benign processes: integration of solvent design and separation process synthesis. Computers & Chemical Engineering, 1999; 23: 1395-1414.

Jhamb, S, Liang, X, Gani, R, Hukkerikar, A S, Estimation of physical properties of amino acids by group contribution method, Chemical Engineering Science, 2018; 175:148-161

Jonuzaj, S, Akula, P T, Kleniati, P M, Adjiman, C S, The formulation of optimal mixtures with generalized disjunctive programming: A solvent design case study

AIChE Journal, 2016; 62:1616–1633.

Kalakul, S, Eden, M R, Gani, R, The chemical product simulator – ProCAPD, Computer Aided Chemical Engineering, 2017; 40:979-984.

Karunanithi, A T, Mehrkesh, A, Computer-aided design of tailor-made ionic liquids, AIChE Journal, 2013; 59: 4627-4640.

Lampe M, Stavrou M, Bucker H M, Gross J, Bardow A, A simultaneous optimization of working fluid and processfor organic Rankine cyclesusing PC-SAFT. Ind. Eng. Chem. Res., 2014; 53:8821-8830.

Li J, Xiao J, Huang Y, and Lou H H, Integrated process and product analysis: A multiscale approach to paint spray. AIChE Journal, 2007;53(11): 2841-2857.

Mata V G, Gomes P B, Rodrigues A E, Engineering perfumes. AIChE Journal, 2005; 51:2834-2852.

Mattei, M, Kontogeorgis, G M, Gani, R, A Comprehensive Framework for Surfactant Selection and Design for Emulsion Based Chemical Product Design, Fluid Phase Equilibria, 2014; 362:288-299.

Mitrofanov I, Sansonetti S, Abildskov J, Sin G, Gani R, 2012. The Solvent Selection framework: solvents for organic synthesis, separation processes and ionic-liquids solvents. Computer Aided Chemical Engineering, 2012;30: 762-766.

Ng, K M, Seider, W D, Transforming Instruction to Chemical Product Design, in Proceedings

of PSE 2018; Editors Eden, M R, Ierapetritou, M, Towler, G.P (2018).

Omidbakhsh N, Duever T A, Elkamel A, Reilly P M, A Systematic Computer-Aided Product Design and Development Procedure: Case of Disinfectant Formulations. Industrial & Engineering Chemistry Research. 2012; 51:14925-14934.

Morales-Rodríguez, R, Gani, R, Computer-Aided Multiscale Modelling for Chemical Process Engineering, Computer Aided Chemical Engineering, 2007; 24:207-212

Sahinidis N V, Tawarmalani M, Yu M, Design of alternative refrigerants via global optimization. AIChE Journal, 2003; 49:1761-1775.

Seider, W D, Lewin, D R, Seader, J D, Widagdo, S, Gani, R, Ng, K M, Product and Process Design Principles, Synthesis, Analysis and Evaluation, 4th Ed., Wiley, 2017.

Smith B V, Ierapepritou M G, Integrative chemical product design strategies: Reflecting industry trends and challenges. Computers & Chemical Engineering, 2010; 34:857-865.

Solvason, C C, Chemmangattuvalappil, N G, Eden, M R, A systematic method for integrating product attributes within molecular synthesis, Computers and Chemical Engineering, 2009; 33:977-991

Stephanopoulos, G, 2003, Invention and Innovation in a Product-Centered Chemical Industry: General Trends and a Case Study, AIChE Institute Lecture.

Street C, Woody J, Ardila J, and Bagajewicz M, Product Design: A Case Study of Slow-Release Carpet Deodorizers/Disinfectants. Industrial & Engineering Chemistry Research, 2008; 47:1192-1200.

Struebing, H, Obermeier, S, Siougkrou, E, Adjiman, C S, Galindo, A, A QM-CAMD approach to solvent design for optimal reaction rates, Chemical Engineering Science, 2017; 159: 69-83.

Sundaram A, Ghosh P, Caruthers JM, and Venkatasubramanian V. Design of fuel additives using neural networks and evolutionary algorithms. AIChE Journal, 2001; 47:1387-1406.

Tam S K, Fung K Y, Poon G S H, Ng K M. Product design: Metal nanoparticle-based conductive inkjet inks. AIChE Journal, 2016; 62:2740-2753.

Wibowo, C, and Ng, K M, "Product-Oriented Process Synthesis and Development: Creams and Pastes," AIChE Journal, 2001; 47:2746-2767.

Xiao J and Huang Y, Microstructure–property–quality-correlated paint design: An LMC-based approach. AIChE Journal, 2009; 55:132-149.

Zhang L, Cignitti S, Gani R. 2015. Generic mathematical programming formulation and solution for computer-aided molecular design. Computers & Chemical Engineering, 2015; 78:79-84.

Zhang L, Babi D K, and Gani R, 2016, New Vistas in Chemical Product and Process Design. Annual Review of Chemical and Biomolecular Engineering, 2016; 7, 557-582.

Zhang, L, Fung, K Y, Zhang X, Fung, H K, and Ng, K M, An Integrated Framework for Designing Formulated Products, Computers & Chemical Engineering, 2017;107:61-76.

Zhang, Z, Kwok, R T K, Yu, Y, Tang, B Z, and Ng, K M, Sensitive and specific detection of L-lactate using an AIE-active fluorophore, ACS Applied Materials & Interfaces, 2017; 9:38153-38158.

Zhou, T, Qi, Z, Sundmacher K, Model-based method for the screening of solvents for chemical reactions, Chemical Engineering Science, 2014; 115:177-185.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50004-5

Real-Time Mixed-Integer Optimization for Improved Economic Performance in HVAC Systems

Michael J. Risbeck[a], Christos T. Maravelias[a*], James B. Rawlings[a]

[a]*Department of Chemical and Biological Engineering, University of Wisconsin–Madison*
christos.maravelias@wisc.edu

Abstract

Optimal operation of large-scale heating, ventilation, and air conditioning (HVAC) systems requires many discrete-valued decisions. For example, in large heating and cooling plants, operators must choose which combination of equipment to activate to meet a given load. Such discrete decisions are typically made by operators using heuristics, which can lead to suboptimal performance. In this paper, we discuss how relevant HVAC decision problems can be transcribed into mixed-integer linear programming (MILP) formulations. We first present the general modeling framework we adopt, which is very similar to the resource task network framework adopted in chemical production scheduling. Second, we discuss a series of reformulations, including linearizations of complex unit models. Third, we present solution methods, including decomposition approaches and the employment of surrogate models to approximate the performance of the system over a long planning horizon. Finally, we demonstrate how the resulting optimization problems can be solved online in closed loop to improve system performance.

Keywords: Scheduling, economic model predictive control, reformulations, solution methods, energy building optimization

1. Introduction

In large buildings or campuses, the heating, ventilation, and air conditioning (HVAC) systems are significant and inherently time-varying consumers of electricity (Powell et al., 2013). In such systems, heating and cooling is performed most efficiently by producing hot or chilled water in high-capacity equipment at a central facility (referred to as the "waterside") and then pumping that water to the buildings where it is used to raise or lower the air temperature of occupied spaces (called the "airside"). Due to the significant use of electricity by waterside and airside equipment, utility companies often impose time-varying prices and demand charges (assessed at regular intervals based on the maximum instantaneous rate of electricity purchase within the previous period). These pricing structures can be leveraged by employing thermal energy storage (TES) to temporally decouple utility purchase from heating/cooling demand (Touretzky and Baldea, 2016), leading to lower cost. Broadly, TES can be categorized as "active," storing energy in fluids in insulated storage tanks, or "passive," storing energy in the thermal mass of buildings (Henze, 2005). Thus, to achieve the lowest possible cost, the tradeoff between active and passive TES must be optimized, all while choosing the most efficient combination of central plant equipment wo meet a given load.

To address this problem, various strategies have been proposed that typically focus more heavily on either the airside or the waterside. For airside passive storage, model predictive control strategies have been developed (Oldewurtel et al., 2012; Ma et al.,

2015) to optimize the future temperature trajectory of the buildings using a dynamic model for temperature evolution; passive storage is then utilized by pre-heating or pre-cooling the building. Similar approaches also exist for active TES using models of storage tank dynamics (Ma et al., 2012; Touretzky and Baldea, 2016). In general, these approaches do not consider waterside equipment selection, which is commonly addressed as a separate static optimization problem (Henze et al., 2008; Powell et al., 2013). However, to achieve the greatest cost reduction while accounting for varying ambient or economic conditions, it is necessary to consider all of these layers together. Thus, the goal of this work is to propose a unified optimization problem that can be solved in real time to determine airside temperature trajectories, TES utilization, and equipment selection for large-scale HVAC systems.

2. Optimization Model

In the following subsections, we present a mixed-integer linear programming (MILP) formulation to optimally schedule equipment use and plan temperature trajectory in large-scale HVAC systems. We use lower-case Roman letters to denote subscripts, upper-case Roman letters for variables, Greek letters for fixed parameters, and bold upper-case Roman letters for sets. Additional details can be found in Risbeck et al., (2017).

2.1. Problem Overview

Given a finite discrete-time horizon indexed by times $t \in \mathbf{T}$, the optimization problem seeks to minimize operating costs of maintaining comfortable temperature in airside zones $i \in \mathbf{I}$ by running waterside equipment $j \in \mathbf{J}$ to consume and produce resources $k \in \mathbf{K}$. This abstraction to units and resources is similar in spirit to the resource-task network commonly used for chemical production scheduling (Pantelides, 1994), and it avoids the need for developing separate equipment-specific constraints. Waterside equipment is modeled by continuous variables Q_{jkt} which give the resource production (> 0) and consumption (< 0) rates of the equipment throughout the optimization horizon. Specific resources are consumed in the airside zones via variables G_{ikt} to either increase (hot water) or decrease (chilled water) zone temperatures T_{it}. Other resources are purchased from the utility market at time-varying prices ρ_{kt} via variables P_{kt} and then consumed in central plant equipment such as chillers (electricity) and boilers (gas) to produce airside resources. The remaining resources are produced by some equipment and consumed by other equipment within the waterside central plant (e.g., cooling towers produce cooling water, which is then consumed by chillers).

To accurately estimate system resource requirements, models of the airside system and waterside equipment must be included in the optimization problem. For the airside system, a discrete-time linear model predicts evolution of zone temperatures T_{it} as a linear combination of the current temperature and the current resource consumption G_{ikt}, with a time-varying term to account for occupancy and ambient conditions. For waterside equipment, it is assumed that transient dynamics can be neglected, and only static production/consumption relationships need be modeled. Although such relationships are typically nonlinear, they are approximated as piecewise-linear so that the resulting formulation can remain an MILP. The decision variables Q_{jkt} are then required to lie on the piecewise-linear operating surface. To allow all equipment to be completely shut off, each model includes the (possibly isolated) operating point $Q_{jkt} = 0$.

Based on this problem description, the optimization model can be formulated as an MILP. Waterside, airside, and linking constraints are defined in the following three subsections.

2.2. Waterside Constraints

To avoid symmetry for multiple pieces of equipment with the same model, identical pieces of equipment are treated as indistinguishable by aggregating them into a single index j. The on/off state and operating point for each individual unit can be chosen independently, with the model making no distinction between permutations of states for the identical equipment. For unique types of equipment, the piecewise-linear operating surface for Q_{jkt} is modeled using the following three equations:

$$Q_{jkt} = \sum_{m \in \mathbf{M}_j} \sum_{n \in \mathbf{N}_{jm}} \zeta_{jknt} Z_{jmnt},\ j \in \mathbf{J}, k \in \mathbf{K}, t \in \mathbf{T}, \tag{1}$$

$$\sum_{m \in \mathbf{M}_j} V_{jmt} = U_{jt},\ j \in \mathbf{J}, t \in \mathbf{T}, \tag{2}$$

$$\sum_{n \in \mathbf{N}_{jm}} Z_{jmnt} = V_{jmt},\ j \in \mathbf{J}, m \in \mathbf{M}_j, t \in \mathbf{T}. \tag{3}$$

Integer variables $U_{jt} \in \{0,1, \dots, \mu_j\}$ chooses how many of type j units are active, while $V_{jmt} \in \{0, \dots, \mu_j\}$ and $Z_{jmnt} \in [0, \mu_j]$ are auxiliary variables used to model the piecewise linear approximate equipment models. Sets $\mathbf{M}_j$ and $\mathbf{N}_{jm}$ index the linear subdomains and nodes for the operating surface, while the parameters ζ_{jknt} define the (possibly time-varying) values of the equipment model at the nodes.

To avoid rapid switching of waterside equipment, dwell times are enforced via

$$U_{jt} - U_{j(t-1)} = U_{jt}^{+} - U_{jt}^{-},\ j \in \mathbf{J}, t \in \mathbf{T}, \tag{4}$$

$$U_{jt} \geq \sum_{t\prime=0}^{\delta_j^{+}-1} U_{j(t-t\prime)}^{+},\ j \in \mathbf{J}, t \in \mathbf{T} \tag{5}$$

$$\mu_j - U_{jt} \geq \sum_{t\prime=0}^{\delta_j^{-}-1} U_{j(t-t\prime)}^{-},\ j \in \mathbf{J}, t \in \mathbf{T}, \tag{6}$$

in which the integer variables U_{jt}^{+} and U_{jt}^{-} indicate when units are switched on and off, which then restricts additional switching within the next δ^{+} and δ^{-} time units.

Storage tanks in the central plant are modeled using variables $S_{kt} \in [0, \Sigma_k]$ and the linear relation

$$S_{kt} = \sigma_k S_{k(t-1)} - Y_{kt},\ k \in \mathbf{K}, t \in \mathbf{T}, \tag{7}$$

in which $Y_{kt} \in [0, Y_k]$ gives the amount charged into (< 0) or discharged from (> 0) the storage tank in the given period. Storage inefficiency (e.g., due to imperfect insulation) is modeled by the decay rate σ_k. Initial conditions S_{k0} are fixed parameters. For resources that cannot be stored, the bounds Y_k and Σ_k are set to zero.

The waterside portion of the objective function consists of time-of-use charges for utilities (amounts $P_{kt} \in [0, \Pi_{kt}]$ with prices ρ_{kt}) in addition to peak demand charges

(amounts $P_k^{\max}$ and costs $\rho_k^{\max}$). There is also a penalty on unmet auxiliary demand (via the slack variable B_{kt}; see Section 2.4) These totals are determined via

$$F^{\mathrm{WS}} = \sum_{k\in\mathbf{K}} \left(\rho_k^{\max} P_k^{\max} + \sum_{t\in\mathbf{T}} (\rho_{kt} P_{kt} + \beta_{kt} B_{kt}) \right), \tag{8}$$

with peak demand $P_k^{\max}$ is calculated by

$$P_{kt} + \psi_{kt} \le P_k^{\max},\ k \in \mathbf{K}, t \in \mathbf{T}. \tag{9}$$

in which ψ_{kt} is a forecast of exogenous resource use (utility purchase outside of the HVAC system that is included in the system-wide demand charge calculation).

2.3. Airside Constraints

Airside zone temperature evolution is modeled using the discrete-time linear model

$$T_{it} = \sum_{i\prime\in\mathbf{I}} \left(\alpha_{ii\prime} T_{i\prime(t-1)} + \sum_{k\in\mathbf{K}} \omega_{ii\prime k}\, G_{i\prime kt} \right) + \theta_{it},\ i \in \mathbf{K}, t \in \mathbf{T}, \tag{10}$$

in which $\alpha_{ii\prime}$ and $\omega_{ii\prime k}$ give time-invariant system dynamics, and θ_{it} represents the time-varying disturbance (including the effects of occupancy, as well as ambient temperature and radiation). Resource consumption $G_{ikt} \in [0, \Gamma_{ikt}]$ is permitted for only a subset of the resources (e.g., chilled and hot water), which is enforced by setting the corresponding bounds Γ_{ikt} to zero. Note that the initial zone temperatures T_{i0} are fixed parameters in the optimization problem.

Comfort bounds on zone temperatures are enforced as soft constraints via

$$T_{it} \le \Theta_{it}^{\max} + T_{it}^{+},\ i \in \mathbf{I}, t \in \mathbf{T}, \tag{11}$$

$$T_{it} \ge \Theta_{it}^{\min} - T_{it}^{-},\ i \in \mathbf{I}, t \in \mathbf{T}, \tag{12}$$

in which $\Theta_{it}^{\max}$ and $\Theta_{it}^{\min}$ give the nominal temperature bounds, which are relaxed by nonnegative slacks T_{it}^{+} and T_{it}^{-}. The objective function for the airside consists solely of these penalty terms:

$$F^{\mathrm{AS}} = \sum_{t\in\mathbf{T}} \sum_{i\in\mathbf{I}} (\chi_{it}^{+} T_{it}^{+} + \chi_{it}^{-} T_{it}^{-}). \tag{13}$$

2.4. Coupling Constraints

The coupling of the airside and waterside systems is enforced via the demand balance

$$\sum_{j\in\mathbf{J}} Q_{jkt} + Y_{kt} + P_{kt} + B_{kt} \ge \phi_{kt} + \sum_{i\in\mathbf{I}} G_{ikt},\ k \in \mathbf{K}, t \in \mathbf{T}, \tag{14}$$

in which the sum of airside and auxiliary demand (G_{ikt} and ϕ_{kt} respectively) must be met by waterside supply, which is the net sum direct production Q_{jkt}, storage charge/discharge Y_{kt}, and purchase $P_{kt} \in [0, \Pi_k]$. The slack variable $B_{kt} \in [0, \phi_{kt}]$ is added to soften the auxiliary demand constraint, while the primary demand G_{ikt} is relaxed via the temperature slacks T_{it}^{+} and T_{it}^{-} From Eqs. (11) and (12).

The overall objective function is then

$$\min F^{\mathrm{WS}} + F^{\mathrm{AS}} \tag{15}$$

to minimize the sum of the waterside and airside terms.

3. Reformulations and Decompositions

The chief difficulty of the waterside formulation is the presence of the discrete variables to model equipment on/off states and enforce the piecewise-linear operating surface. In contrast, there are no discrete variables in the airside formulation, but optimization is nevertheless challenging due to the potentially large number of temperature zones. Thus, the problem can take a very long time to even find a feasible solution when solved as a single monolithic instance. To ease this burden, we present a simplified surrogate model and a decomposition strategy that can be employed to find near-optimal solutions to the combined problem in real time.

3.1 Surrogate Model

When there is a large amount of equipment in the central plant, solution progress can be slow due to the large number of integer variables. While the addition of more identical units is partially mitigated by the symmetry-free formulation, the inclusion of auxiliary equipment for both heating and cooling loops leads to a large number of discrete decisions that must be made at each timestep. To ease this burden, we propose a simplified surrogate model that seeks to aggregate multiple pieces of equipment into a single continuous curve.

To perform this computation, we start by defining subsets $\mathbf{K}^+$ of resources exported from the central plant and $\mathbf{K}^-$ for resources purchased by the central plant:

$$\mathbf{K}^+ \coloneqq \{k \in \mathbf{K}: \phi_{kt} > 0 \text{ or } \Gamma_{ikt} > 0 \text{ for some } t \in \mathbf{T} \text{ or } i \in \mathbf{I}\},$$

$$\mathbf{K}^- \coloneqq \{k \in \mathbf{K}: \Pi_k > 0\}.$$

For each major type of equipment (e.g., chillers, heat-recovery chillers, and boilers), we identify the supporting auxiliary equipment to define a small number of "operating groups." For example, the chiller operating group consists of conventional chillers, cooling towers, and chilled water pumps. For each operating group, we choose a representative resource $k^* \in \mathbf{K}^+$ and then pick a series of production rates ϕ_{k^*} that span the minimum and maximum total capacities of the current group. For each value of ϕ_{k^*}, we solve a one-period equipment selection period of the form

$$\begin{aligned}
\min \quad & \sum_{k \in \mathbf{K}} \rho_k P_k \\
\text{s.t.} \quad & \sum_{j \in \mathbf{J}} Q_{jk} + P_k \geq \phi_k, \; k \in \mathbf{K} \\
& Q_{jk} = \sum_{m \in \mathbf{M}_j} \sum_{n \in \mathbf{N}_{jm}} \zeta_{jkn} Z_{jmn}, \; j \in \mathbf{J}, k \in \mathbf{K} \\
& \sum_{m \in \mathbf{M}_j} V_{jm} = U_j, \; j \in \mathbf{J} \\
& \sum_{n \in \mathbf{N}_{jm}} Z_{jmn} = V_{jm}, \; j \in \mathbf{J}, m \in \mathbf{M}_j
\end{aligned}$$

which assumes that time-invariant representative values can be chosen for resource costs ρ_k and equipment models ζ_{jkn}. Note that each problem includes only the equipment in the current operating group. The goal of these subproblems is to determine the required purchase rate for resources in $\mathbf{K}^-$ to produce various values of resources in $\mathbf{K}^+$.

Once all the subproblems have been solved, the Q_{jk} variables are summed to define points on an aggregate operating curve for the current operating group. A straight line is then fit through these points to give an approximate linear operating surface that can be used instead of the individual piecewise-linear operating surfaces. In particular, this simplification means that the set **J** consists of the equipment operating groups, rather than individual pieces of equipment, and the set of resources **K** need only include $\mathbf{K}^+ \cup \mathbf{K}^-$, which eliminates auxiliary resources like cooling water. Furthermore, the equipment model constraints in Eqs. (1), (2), and (3) to be replaced by the much simpler

$$Q_{jkt} = \tilde{\zeta}_{jk} Z_{jt}$$

with $Z_{jk} \in [0,1]$ a single continuous variable, and $\tilde{\zeta}_{jk}$ the full-capacity resource production/consumption rates from the linear aggregate curves. With the removal of the discrete variables U_{jt}, the switching constraints in Eqs. (4), (5), and (6) can be omitted as well, leading to a fully continuous model with significantly fewer variables. While the aggregate operating curves no longer precisely calculate resource consumption and possibly ignore minimum equipment capacities and small total loads, accuracy is typically good enough to produce near-optimal utilization of active and passive storage. This surrogate model can then be employed as described in the next two sections.

3.2. Hybrid Model

When very large well-insulated storage tanks are present, it may be necessary to optimize over a long horizon (i.e., in excess of a week) to make optimal use of the storage tank. This longer horizon leads to many more discrete variables and constraints. However, forecasts of utility prices and auxiliary demand may not be extremely accurate in the later periods of the horizon. Furthermore, because optimization is performed in closed-loop, decisions made at the end of the horizon are never actually implemented. Thus, it is not necessary to fully optimize over the full decision space throughout the entire prediction horizon. This realization suggests the use of a hybrid approach, whereby the full model (with discrete variables and individual equipment models) is used over a modest horizon, with the surrogate model applied over a longer horizon to consider long-term behavior. By linking the appropriate variables (storage capacity S_{kt}, peak demand $P_k^{\max}$, and zone temperatures T_{it} if present) between the two models, the optimal near-term decisions can be determined while approximately accounting for potential storage utilization over a much longer period.

The hybrid optimization strategy can be implemented in one of two ways. In the “two-stage” approach, the surrogate model is first solved over a longer horizon. The values of the linking variables midway through the prediction horizon are then used as terminal constraints for a subsequent optimization of the full model over a shorter horizon. The advantage of the two-stage approach is that the full optimization is smaller in size. As an alternative, the “one-stage” approach solves a single optimization problem in which the first portion of the horizon uses the full model, while the remaining periods use the surrogate model. The advantage here is that the greater accuracy of the full model provides a more representative starting point for the surrogate model. These two approaches are diagrammed in Figure 1.

3.3. Hierarchical Decomposition

To improve solution times for the combined problem, the large model size from the airside can be separated from the discrete variables on the waterside. One method of separation is a hierarchical decomposition, whereby the surrogate model (as defined in

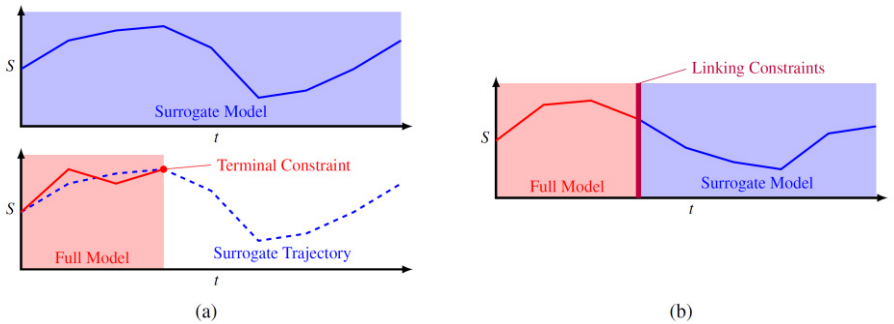


Figure 1: Diagram of two-stage (a) and one-stage (b) hybrid optimization strategies. Axes show different optimization problems, and shaded regions show prediction horizon and model type.

Section 3.1) is first optimized to determine airside demand, which is then sent down as a fixed parameter to a fully detailed waterside-only optimization. More precisely, after solving the surrogate model (which includes the full airside dynamic models but only the approximate operating group models for the waterside), total demand is calculated as

$$\widetilde{\phi}_{kt} \coloneqq \phi_{kt} + \sum_{i \in \mathbf{I}} G_{ikt}\,,\ k \in \mathbf{K}, t \in \mathbf{T}.$$

The total demand $\widetilde{\phi}_{kt}$ is then used in place of ϕ_{kt} in Eq. (14) in the full model, but with the airside variables T_{it}, G_{ikt}, etc. and constraints (10) through (13) now removed from the problem. This decomposition allows passive storage to be determined in the upper-level optimization (which can be solved as a large linear program), while equipment allocation and active storage are optimized in the lower level (a modestly sized MILP). A feasible overall solution is then constructed by combining the airside variables from the upper problem with the solution to the lower problem. To obtain a lower bound, the demand constraint (14) can be dualized (in the Lagrangian sense), which allows dual airside and waterside subproblems to be solved. A guess for the multipliers can be taken from the top-level surrogate model.

4. Example Simulations

4.1. Hybrid Model

To start, we simulate the closed-loop performance of the hybrid model on a waterside-only system that must meet forecasted cooling demand. Using 22 weeks of hourly chilled water demand and electricity prices provided by Johnson Controls, Inc., closed-loop simulations were performed using various prediction horizons for the full and surrogate models. Each simulation covers 1 week of data, re-optimizing every hour. Because this system is small enough, the optimal week-long schedule can also be determined by solving the full model in a single week-long optimization, which gives a baseline for closed-loop cost. Figure 2 shows a sample optimal trajectory for the system, and Table 1 shows cost performance for the simulations.

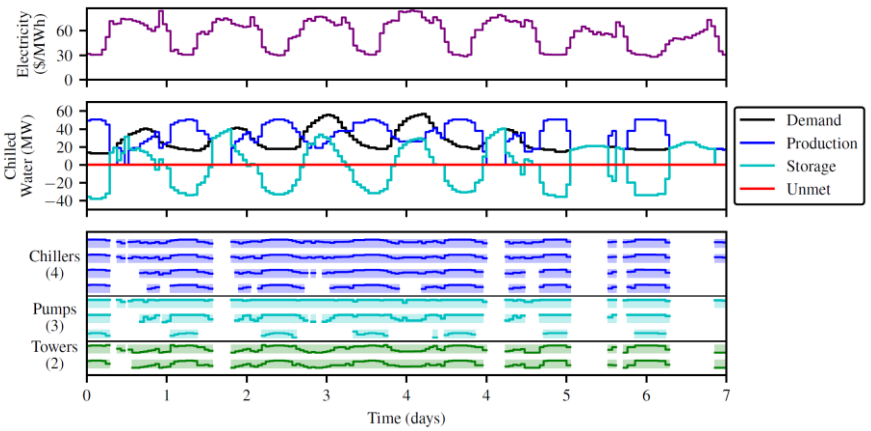


Figure 2: Sample optimal trajectory for the cooling-only system. Central plant equipment is shown in the Gantt chart with light bars showing equipment on/off and solid lines showing loading between minimum and maximum capacity.

From the results in Table 1, we see that use of the one-stage hybrid optimization strategy leads to slightly better performance than simply solving the corresponding standard model, while the two-stage strategy is only helpful for short horizons. In terms of closed-loop cost, the one-stage optimization with a surrogate horizon of 48 performs almost as well as the full model by itself with the same horizon. Thus, including the surrogate model can significantly reduce computational requirements (by reducing the number of discrete variables) without sacrificing closed-loop performance.

Of note in Table 1, the biggest difference between the closed-loop solutions and the optimal solutions is in tradeoff between demand and use charges. Because the closed-loop problems are looking over shorter horizons, they slightly underestimate the effects of having a large peak electricity usage and instead more aggressively exploit low use charges. This change in behavior leads to higher demand charges and slightly lower use charges than in the optimal solution. However, since demand charges are typically assessed on a month-long basis, even these "optimal" solutions may not represent the true month-long minimum. Thus, additional strategies may be needed to address the timescale differences between hourly use charges and monthly demand charges, which could lead to even better performance for the hybrid optimization strategies.

4.2 Decomposition Strategy

To test the decomposition strategy on a realistically sized system, a very large instance is created. The central plant consists of 5 chillers, 3 heat-recovery chillers (HRCs), 2 boilers, 10 pumps, and 5 cooling towers. There is auxiliary demand ϕ_{kt} for both hot and chilled water. The airside system consists of 20 separate buildings, each with 50 interacting temperature zones. Half of the zones represent occupied air zones in which comfortable temperature must be maintained using chilled water, while the other half of the zones are slowly evolving mass zones with no temperature constraints. For this system with 7 day horizon (1 h timestep), the MILP solver Gurobi 7.0 is unable to find a feasible solution to the full problem within 2 h of computation time. Instead, the hierarchical decomposition strategy is applied to the problem. Sizes for the various subproblems are shown in Table 2.

Table 1: Total cost breakdown across 22 simulations (each 1 week long) for the cooling-only system. Demand charges account for most of the performance difference. Costs (%) are relative to optimal 1-week solutions.

Strategy	Full Horizon	Surrogate Horizon	Use Charges	Demand Charges	Total Charges
Standard	12	–	+0.86	+12.42	+1.95
One-Stage	12	48	+0.18	+6.60	+0.79
Two-Stage	12	48	−0.16	+14.49	+1.23
Standard	24	–	−0.11	+10.65	+0.91
One-Stage	24	120	+0.02	+6.97	+0.67
Two-Stage	24	120	−0.23	+12.43	+0.97
Standard	48	–	−0.22	+8.67	+0.62
One-Stage	48	120	+0.16	+2.46	+0.37
Two-Stage	48	120	−0.39	+11.40	+0.72

Table 2: Instance sizes for subproblems in hierarchical decomposition.

Subproblem	Variables ($\times 10^6$)	Integers ($\times 10^3$)	Constraints ($\times 10^3$)	Nonzeros ($\times 10^6$)
Centralized	1.19	0	341.21	3.29
Airside Dual	0.84	0	336.34	2.60
Waterside Dual	1.20	6.38	18.15	1.25
Waterside Primal	0.03	6.38	18.15	0.07

After solving the upper-level subproblem (1.5 min) and the lower-level waterside feasibility problem (10 min, 1.6% optimality gap), a feasible solution was obtained as shown in Figure 3. Using the dual multipliers from the upper-level subproblem the dual airside (0.4 min) and waterside (10 min, 1.2% gap) subproblems, a lower bound was obtained, resulting in a duality gap of 4.3%. Thus, while the full combined problem was completely intractable for this large system, both decomposition strategies produced a satisfactory solution quickly enough for real-time application.

5. Conclusions

In this work, we have presented an MILP model for economic optimization of large-scale HVAC systems. Using a symmetry-free abstract formulation, the model can be applied to a wide variety of systems. With additional reformulations and decompositions, near-optimal solutions can be obtained for large-scale problems within 15 minutes, which enables online application. By employing these optimization methods, HVAC systems can react optimally in real time to changing ambient or economic conditions, leading to improved economic performance over heuristic or other less-frequent scheduling strategies.

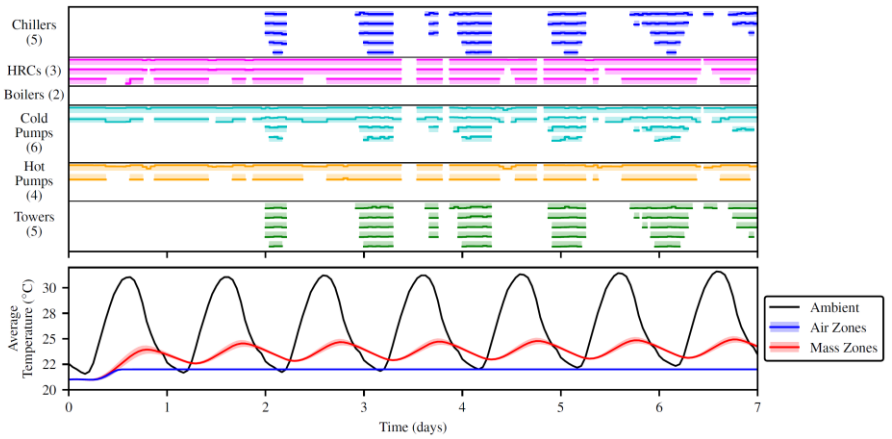


Figure 3: Hierarchical decomposition solution for big instance. Temperature plot shows average $\pm$ standard deviation for 500 air and mass zones.

References

G.P. Henze. Energy and cost minimal control of active and passive building thermal storage inventory. *J. Solar Ener. Eng.*, 127(3):343–351, 2005.

G.P. Henze, B. Biffar, D. Kohn, and M.P. Becker. Optimal design and operation of a thermal storage system for a chilled water plant serving pharmaceutical buildings. *Energ. Buildings*, 40(6):1004–1019, 2008.

Y. Ma, J. Matusko, and F. Borrelli. Stochastic model predictive control for building HVAC systems: Complexity and conservatism. *IEEE Trans. Cont. Sys. Tech.*, 23(1):101–116, 2015.

Y. Ma, F. Borrelli, B. Hencey, B. Coffey, S. Bengea, and P. Haves. Model predictive control for the operation of building cooling systems. *IEEE Ctl. Sys. Tech.*, 20(3):796–803, 2012.

F. Oldewurtel, A. Parisio, C.N. Jones, D. Gyalistras, M. Gwerder, V. Stauch, B. Lehmann, and M. Morari. Use of model predictive control and weather forecasts for energy efficient building climate control. *Energ. Buildings*, 45:15–27, 2012.

C.C. Pantelides. Unified frameworks for optimal process planning and scheduling. In *Proceedings on the second conference on foundations of computer aided operations*, pages 253–274. Cache Publications New York, 1994.

K.M. Powell, W.J. Cole, U.F. Ekarika, and T.F. Edgar. Optimal chiller loading in a district cooling system with thermal energy storage. *Energy*, 50(0):445–453, 2013.

M.J. Risbeck, C.T. Maravelias, J.B. Rawlings, and R.D. Turney. A mixed-integer linear programming model for real-time cost optimization of building heating, ventilation, and air conditioning equipment. *Energ. Buildings*, 142:220–235, 2017.

C.R. Touretzky and M.Baldea. A hierarchical scheduling and control strategy for thermal energy storage systems. *Energ. Buildings*, 110:94–107, 2016.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50005-7

Design for dynamic operation – A review and new perspectives for a dynamic manufacturing environment

Christopher L.E. Swartz[a*] and Yoshiaki Kawajiri[b]

[a]*Department of Chemical Engineering, McMaster University, 1280 Main Street West, Hamilton, L8S 4L7, Canada*
[b]*Department of Materials Process Engineering, Nagoya University, Furo-cho 1, Chikusa, Nagoya 464-8603, Japan*
swartzc@mcmaster.ca

Abstract

In this paper, we review frameworks for design, control, and operation of dynamic processes and their applications. Recognition that design characteristics of a process limit its achievable control performance dates back to at least the 1940s. Isolated studies appeared in the literature since then, with a watershed in the 1980s, when quantitative analysis of this phenomenon was advanced in a series of studies using the internal model control (IMC) framework. Dynamic optimization was subsequently utilized as a platform to analyze the impact of a plant design on its dynamic performance, and to perform simultaneous plant and control system design. This paper presents a review of approaches toward the analysis of the interaction between design and dynamic performance, and offers some new perspectives on this topic. In particular, the trends of increased globalization, market volatility, and variation in electricity prices have created dynamic conditions that require responsive operation in order to remain competitive. In addition to such processes where transient dynamics between steady states can be considered and exploited, there are intrinsically dynamic processes which cannot be designed based on steady state mass and energy balances. In intrinsically dynamic processes, the intra-cycle and inter-cycle dynamics should be taken into account when optimizing the design, control, and operation. We discuss adsorption processes as an example. Our study includes a description and discussion of advances in these areas.

Keywords: integrated design and control, dynamic optimization, dynamic operability, dynamic resilience, simulated moving bed, intentionally dynamic operation.

1. Introduction

Process plants do not operate in a static environment, and need to respond adequately to disturbances in order to meet safety, operational, environmental and product quality constraints. A plant's achievable dynamic performance is strongly influenced by its design – a phenomenon recognized at least as far back as the 1940s (Ziegler and Nichols, 1943). Isolated studies on this topic appeared over a span of several years (e.g. Anderson, 1966), but it was only in the 1980s and 1990s that this area began to resonate within the process systems engineering community. The traditional plant and control system design process is sequential, with the control system considered after completion of the plant design phase. However, recognition that poor control performance could potentially be avoided through alternative design choices led to advocacy for integrated plant and

control system design approaches (Downs and Doss, 1991) and to research efforts toward the development of systematic approaches for incorporation of dynamic considerations into plant design calculations.

Two key paradigms in the analysis of the interaction between design and control are the notion of a performance limit regardless of controller type, and dynamic performance under the action of a particular controller type and tuning (which could be included in the optimization decision space). Morari (1983) refers to the former as dynamic resilience, and through use of the internal model control (IMC) framework, identified control performance-limiting factors as time delays, right-half-plane transmission zeros, input constraints and model uncertainty. Analyses of their impact on achievable performance are given in Morari (1983) and Holt and Morari (1985a, 1985b). This framework has given rise to open-loop controllability indicators, several of which are described in Skogestad and Postlethwaite (1996). Many subsequent studies have followed an optimization-based approach, some of which incorporate explicit characterizations of dynamic resilience.

Integrated plant and control design paradigms are typically based on the assumption of operation around a nominal steady-state point (which can be included as an optimization decision in the design procedure). However, an increasingly global marketplace with increased variation in product demand, supply, and utility prices has created an increasingly dynamic operating environment for which Backx et al. (1998) advocate "intentionally dynamic operation" in order to be responsive and remain competitive. Applications of this type are relatively sparse in integrated design and control studies. Many processes, on the other hand, are intrinsically dynamic in their operation, examples of which are pressure swing adsorption (Jiang et al., 2004), semicontinuous distillation (Adams and Pascall, 2012), periodically operated reactors (Zahn et al. 2009) and simulated moving beds (Kawajiri and Biegler, 2006). These processes do not have a steady state, and thus process dynamics must always be in taken into account.

In this paper, we present an overview of key approaches for inclusion of dynamic considerations in plant design, and thereafter focus on applications to intentionally dynamic and intrinsically dynamic operation. Section 2 presents a general dynamic optimization formulation which underpins a large class of integrated design approaches, and reviews some key approaches for design with dynamic performance considerations. Optimization-based design of intentionally dynamic and intrinsically dynamic operation is discussed in Sections 3 and 4, with conclusions and future research directions presented in Section 5.

2. General dynamic optimization formulation

Design problems discussed in the sequel can be considered within the context of a general dynamic optimization framework. An optimization-based formulation for design subject to dynamic performance constraints may be stated as follows:

$$
\begin{aligned}
&\min_{\boldsymbol{d},\boldsymbol{u}(t)} \mathcal{J} := E_{\boldsymbol{\theta}\in\Gamma}\left\{\phi\left(\boldsymbol{x}(t_f),\boldsymbol{z}(t_f),\boldsymbol{u}(t_f),\boldsymbol{d},\boldsymbol{\theta},t_f\right)\right\}\\
\text{st: } &\dot{\boldsymbol{x}}(t) - \boldsymbol{f_d}(\boldsymbol{x}(t),\boldsymbol{z}(t),\boldsymbol{u}(t),\boldsymbol{d},\boldsymbol{\theta},t) = \boldsymbol{0}\\
&\boldsymbol{f}_a(\boldsymbol{x}(t),\boldsymbol{z}(t),\boldsymbol{u}(t),\boldsymbol{d},\boldsymbol{\theta},t) = \boldsymbol{0}\\
&\boldsymbol{g}(\boldsymbol{x}(t),\boldsymbol{z}(t),\boldsymbol{u}(t),\boldsymbol{d},\boldsymbol{\theta},t) \le \boldsymbol{0}\\
&\boldsymbol{d} \in [\boldsymbol{d}^L,\boldsymbol{d}^U],\ \ \boldsymbol{u}(t) \in [\boldsymbol{u}^L,\boldsymbol{u}^U]\\
&\Gamma = \{\boldsymbol{\theta} \mid \boldsymbol{\theta} \in [\boldsymbol{\theta}^L,\boldsymbol{\theta}^U]\}
\end{aligned}
\tag{1}
$$

where $\boldsymbol{x}$, $\boldsymbol{z}$, $\boldsymbol{u}$ and $\boldsymbol{d}$ represent differential, algebraic, input and design variables respectively, $\boldsymbol{\theta}$ is a vector of uncertain parameters, $\boldsymbol{f}_d$ and $\boldsymbol{f}_a$ represent differential and algebraic functions, respectively, of the DAE model in semi-explicit form, and E is the expectation operator. An economic objective function is typically optimized, with dynamic performance implicitly accounted for through the path constraints, $\boldsymbol{g}$. The above formulation accommodates open- or closed-loop dynamics through exclusion or inclusion of controller equations. The design decisions can include equipment sizing and plant configuration, a steady-state operating point, controller configuration and tuning, or combinations thereof. The uncertain parameters may be conveniently handled through discretization of the uncertainty set, resulting in a multiperiod dynamic optimization problem. The formulations discussed next utilize various forms of Eq. (1), from a steady-state version of the problem that includes constraints associated with dynamic performance, to a comprehensive dynamic optimization formulation that includes equipment design, control configuration selection and uncertainty.

Luyben and Floudas (1994) utilize open-loop indicators of controllability in the evaluation of design alternatives in a multiobjective optimization framework. While the open-loop controllability metrics are relatively easy to compute, a key drawback is their restriction to individual performance-limiting characteristics. Swartz (1996, 2004) proposes a computational framework for resiliency assessment based on Q- (or Youla) parametrization of all linear stabilizing feedback controllers. It permits performance-limiting design characteristics to be simultaneously accounted for, and also provides a linear feedback controller capable of satisfying the performance criteria, if achievable by linear feedback control. The strategy is applied in Ross and Swartz (1995) to the dynamic operability analysis of flotation circuits, and extended in Ross and Swartz (1997) to account for model uncertainty.

It is widely recognized that the economically optimal steady-state operating point of a plant typically lies at the intersection of constraints (Arkun and Stephanopoulos, 1980). This characteristic is utilized in Narraway et al. (1991) to quantify the economics of control performance through the extent to which the steady-state operating point needs to "back off" from the constraints in order to maintain feasible operation in the presence of disturbances. The back-off magnitude is estimated using frequency response analysis techniques, and related to an economic loss using Lagrange multipliers of the steady-state economic optimization problem. Figueroa et al. (1996) compute the back-off through a dynamic optimization problem in which the economically optimal operating point is determined, subject to path constraints on the dynamic response, and utilize this in the performance assessment of alternative controllers. Soliman et al. (2004) consider back-off under constrained predictive control. The presence of MPC optimization subproblems along the back-off calculation horizon results in a multi-level optimization problem that is transformed into single-level mathematical program with complementarity constraints (MPCC) by replacing the MPC quadratic programming subproblems by algebraic constraints corresponding to their first-order optimality conditions, with the complementarity constraints in turn reformulated as mixed-integer linear constraints.

Lenhoff and Morari (1982) propose a dynamic performance index based on an optimal control formulation which they use in conjunction with steady-state plant economics for plant performance evaluation within a multi-objective setting. Mohideen et al. (1996) present a more general dynamic optimization framework for integrated plant and control system design. Plant economics are optimized with dynamic performance implicitly accounted for through the imposition of path constraints. A mixed-integer formulation

permits discrete design decisions such as number of trays in a distillation column and control structure selection. Baker and Swartz (2004) propose mixed-integer and complementarity constraint formulations for rigorously incorporating actuator saturation effects within a simultaneous design and control formulation. Inclusion of constrained MPC within the design formulation was subsequently considered using an MPCC formulation such as that described earlier (Baker, 2006), but with plant design parameters included as optimization decision variables and the single-level optimization problem solved using an interior point approach. Sakizlis et al. (2003), on the other hand, include MPC within an integrated design framework through the use of a multi-parametric MPC formulation.

Several other studies and applications on the interaction between design and dynamic performance have appeared, utilizing to various extents the concepts and approaches described above. Comprehensive reviews on this topic may be found, inter alia, in van Schijndel and Pistikopoulos (1999), Sakizlis et al. (2004), and Ricardez-Sandoval et al. (2009). We consider in the next two sections particular applications to intentionally and intrinsically dynamic operation.

3. Intentionally dynamic operation

Chemical plants operate in an environment of increased competition in a global marketplace, with increased variation in product demand and raw material supply. The deregulation of electricity prices in many jurisdictions has resulted in large fluctuations in electricity price. It is therefore becoming increasingly important for plants to be able to transition rapidly in order to respond to such variation in order to maximize profits and increase competitiveness. Cryogenic air separation units (ASUs) are a prime example of such plants due to their high electricity consumption, demand variation and complexity of operation through tight thermal integration.

White et al. (1996) present an optimization formulation for switchability analysis that is posed as a dynamic optimization problem that seeks to minimize a measure of transition time subject to dynamic model and path constraints. The formulation is applied to integrated distillation columns in an air separation plant and a binary distillation system, with the columns in both applications modelled using reduced-order compartmental models. They conclude that there are no process constraints limiting the speed of reduction in gas oxygen product, but that a transition between product qualities in the binary system is improved by a reduction in tray holdup. Schenk et al. (2002) present an integrated design and control formulation for an air separation plant in which both disturbances and ramping between operating points are considered. A mixed-integer dynamic optimization problem is posed, with design decisions comprising distillation column diameters and numbers of trays, and the control structure.

Cao et al. (2015) propose a two-tiered optimization strategy to explore design limitations to transition agility in a nitrogen plant, illustrated in Figure 1(a), in response to demand and electricity price changes that is formulated as follows:

- *Tier 1*: Solve a steady-state economic optimization problem to determine the new operating point.
- *Tier 2*: Solve a dynamic optimization problem to minimize a measure of transition time from the current operating point to that determined in Tier 1.

Various constraints are imposed, such as flooding, compressor surge, liquid fraction in turbine inlet, and product purity. Figure 1(b) shows that the flooding constraint is violated

if the transition in response to a 20 % increase in nitrogen gas demand is applied via single step changes in the inputs, motivating the need for determining the optimal input trajectories through dynamic optimization. The flooding and turbine inlet constraints were found to be active for certain of the case study scenarios, indicating design bottlenecks in the column and primary heat exchanger design. The potential benefit of using stored liquid product as additional reflux during transition was also explored, and found to reduce transition times when the product impurity limit was tightened. A current extension involves formulation of the ASU design problem under transition as a two-stage stochastic optimization problem, with preliminary results presented in Jaydeep et al. (2017).

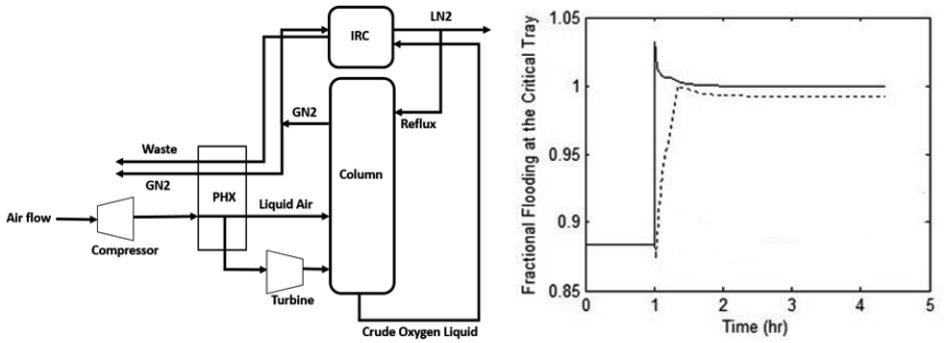


Figure 1 (a) Schematic of nitrogen plant, (b) Ratio of vapor velocity to flooding velocity for step changes in manipulated variables (solid line) and optimized inputs (dot-dashed line)

4. Intrinsically dynamic operation

Here we discuss intrinsically dynamic processes where the conventional approaches for steady state design cannot be applied. Such processes can be characterized by cyclic and repetitive operations. These processes never reach a steady state, and we must deal with process dynamics even when we analyse mass and energy balances over a long time period. To design and control such processes, we consider formulating a dynamic optimization problem in the form of Eq. (1).

In this paper, we distinguish batch processes from intrinsically dynamic processes. In batch processes, a sequence of operations is repeated in each batch, and all variables are re-initialized at the beginning of each batch. On the other hand, intrinsically dynamic processes we discuss in this paper are semi-continuous periodic processes where accumulations occur in some or all variables. Due to the accumulations, there exists inter-cycle dynamics in addition to the intra-cycle dynamics. While these processes do not have a steady state, they typically reach a cyclic steady state (CSS) after repeating a certain number of cycles where variables have periodic profiles. The inter-cycle profiles in the transition between CSSs should also be optimized and controlled (Bentley, et al. 2014, Toumi and Engell 2004).

We consider periodic adsorption processes as an example. In adsorption processes, a single or series of columns connected to each other is packed with adsorbent particles. Using these columns, a mixture can be separated into purified products utilizing a difference in the affinity towards the adsorbent. In these processes, profiles of state and differential variables, such as concentration, pressure and temperature, keep propagating

through the columns. These processes are operated in a cyclic manner, where the same sequence of operations (steps) are repeated. As a result, the state variables have periodic dynamics, characterized as the CSS. The dynamics are described as a system of partial differential algebraic equations (PDAEs), which can be discretized in space to formulate an optimization problem shown as Eq. (1).

4.1. Chromatographic processes

Chromatographic separation is a type of adsorption process where multiple components in a mixture travel at different speeds in a packed column, and these components are fractionated at the outlet of the column when they elute out. The simplest design of chromatography is the single-column process, as shown in Figure 2(a). In this conventional batch process, a mixture is injected periodically from the inlet of the column. The mixture is subsequently pushed towards the outlet by the desorbent. As the components travel through the column, the component with the stronger affinity for the adsorbent moves more slowly, while that with the weaker affinity moves faster. These components are fractionated at the outlet of the column.

To improve the performance of chromatography, quantified by product purity, recovery, productivity (throughput), and desorbent consumption, a number of multi-column semi-continuous chromatographic processes have been proposed. A well-known example is simulated moving bed (SMB) chromatography. This process is widely used in industry, where applications include xylene isomer separation, sugar purification, and chiral separation for pharmaceuticals. The principle of the SMB process is shown in Figure 2(b). In this process, a series of columns is connected to form a circulation loop, and the two products, raffinate and extract, are withdrawn continuously while the feed and desorbent are supplied also continuously. The supply and withdrawal ports are switched periodically in the direction of the liquid flow. This switching operation realizes the counter-current flow of the solid phase in a "simulated" manner.

Design of multi-column chromatographic processes can be challenging due to the complex operations. In addition to the standard design and operation shown in Figure 2(b), there have been a number of modifications which have been validated to improve the performance (Sreedhar and Kawajiri 2014). In addition, the SMB process is often integrated with other unit operations in a process flowsheet. In Figure 2(c), the SMB process for enantiomer separation is integrated with two crystallizers, which partly carry out the separation. It has been shown that the capital cost of this hybrid separation process can be substantially lower (Kaspereit, et al. 2005). Figure 2(d) shows a flowsheet that consists of an enzymatic rector and membrane for sugar production. In an example of D-psicose production from D-fructose, the unreacted reactant is separated by SMB and recycled back to the enzymatic reactor (Bechtold, et al. 2006). To design such flowsheets, we can formulate an optimization problem in the form of Eq. (1), but at a significantly larger scale.

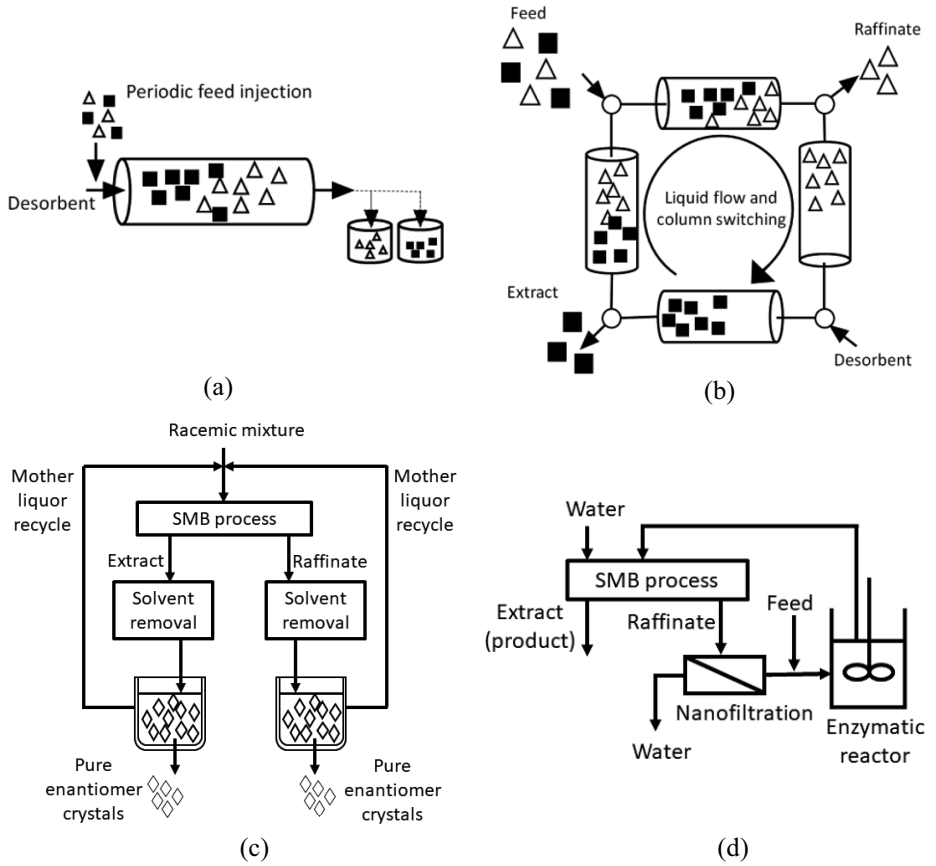


Figure 2 Chromatography processes; (a) single-column batch chromatography, (b) simulated moving bed (SMB) process, (c) integration with crystallizers for enantiomer purification (Kaspereit, et al. 2005), (d) integration with membrane and enzymatic reactor (Bechtold, et al. 2006)

4.2. Challenges for intrinsically dynamic processes

While there are many past studies for optimization of design and operation of periodic processes, there remain many challenges. First, due to the complex dynamics governed by PDAEs and CSS conditions, the problem size can be large and a solution may not be found easily. In multi-column adsorption processes, the number of columns must be treated as integer variables, and process developers must consider a large number of design alternatives. Furthermore, finding a robust design, operation, and control under model uncertainty would be very challenging in the framework of the dynamic optimization. In chromatography, for example, a substantial degree of model uncertainty must be taken into account, since the mass transfer kinetics and adsorption equilibria are not fully understood in many novel applications such as biopharmaceuticals. These difficulties are even more pronounced when designing a flowsheet that embeds an intrinsically dynamic process, as well as in optimizing and controlling inter-cycle transitions between CSSs.

5. Conclusions

The impact of the design of a process on its dynamic performance has led to analysis techniques for assessment of a dynamic resilience, and design formulations that account for dynamic response behavior. While many applications focus on operation around a steady-state point, the present manufacturing of environment is highly dynamic in nature, requiring transient operation in response to often large variations in parameters such as demand and electricity price. In addition, certain processes are inherently dynamic in their operation, necessitating inclusion of plant dynamics in the design procedure. In this paper, we have presented an overview of this topic, highlighting key paradigms that have been followed. A key remaining challenge is solution of the large-scale DAE and PDAE systems that arise when considering larger and more complex configurations, compounded by consideration of uncertainty within the formulation. These problems are moreover often mixed-integer in nature, adding a significant level of complexity. A further interesting research direction is consideration of the interplay between dynamic operation, scheduling and design, a concept that has recently been considered in Patil et al. (2015) and Pistikopoulos and Diangelakis (2016).

References

Adams II, T.A., Pascall, A., 2012. Semicontinuous thermal separation systems. Chem. Eng. Technol. 35(7), 1153-1170.

Arkun, Y., Stephanopoulos, G., 1980. Studies in the synthesis of control structures for chemical processes. Part IV. Design of steady-state optimizing control structures for chemical process units. AIChE J. 26(6), 975-991.

Anderson, J.S., 1966. A practical problem in dynamic heat transfer. Chem. Engr. CE97-CE103.

Backx, T., Bosgra, O., Marquardt, W., 1998. Towards intentional dynamics in supply chain conscious process operations. AIChE Symp. Ser. 320, 5-20.

Baker, R., Swartz, C.L.E., 2004. Simultaneous solution strategies for inclusion of input saturation in the optimal design of dynamically operable plants. Optimization and Engineering. 5, 5-24

Baker, R., 2006. Interior point solution of complementarity constrained problems arising in integrated design and control. Ph.D. Thesis, McMaster University.

Bechtold, M., Makart, S. Heinemann, M. Panke, S., 2006**,** Integrated operation of continuous chromatography and biotransformations for the generic high yield production of fine chemicals. J Biotechnol 124 1, 146-62.

Bentley, J., Li, S., Kawajiri, Y., 2014. Experimental validation of optimized model-based startup acceleration strategies for simulated moving bed chromatography, Ind. Eng. Chem. Res. 53(30), 12063-12076

Cao, Y., Swartz, C.L.E., Baldea, M., Blouin, S., 2015. Optimization-based assessment of design limitations to air separation plant agility in demand response scenarios. J. Proc. Contr. 33, 37-48.

Downs, J.J, Doss, J.E., 1991. Present status and future needs - A view from North American Industry. In Chemical Process Control - CPC IV, Y. Arkun, W.H. Ray, eds., CACHE and AIChE.

Figueroa, J.L., Bahri, P.A., Bandoni, J.A., Romagnoli, J.A., 1996. Economic impact of disturbances and uncertain parameters in chemical processes – A dynamic back-off analysis. Comp. Chem. Eng. 20(4), 453-461.

Holt, B.R., Morari, M., 1985a. Design of resilient processing plants – V. The effect of deadtime on dynamic resilience. Chem. Eng. Sci. 40(7), 1229-1237.

Holt, B.R., Morari, M., 1985b. Design of resilient processing plants – VI. The effect of right-half-plane zeros on dynamic resilience. Chem. Eng. Sci. 40(1), 59-74.

Jaydeep, R., Swartz, C.L.E., Cao, Y. and Flores-Cerrillo, J., 2017. Design of cryogenic air separation units for transition responsiveness, 10th World Congress of Chemical Engineering, Barcelona.

Jiang, L., Fox, V.G., Biegler, L.T., 2004. Simulation and optimal design of multiple-bed pressure swing adsorption systems. AIChE J. 50(11), 2904-2917.

Kaspereit, M., Sainio, T., 2011**,** Simplified design of steady-state recycling chromatography under ideal and nonideal conditions. Chem. Eng. Sci. 66 21, 5428-5438.

Kaspereit, M., Gedicke, T., Zahn, V., Mahoney, A. W., Seidel-Morgenstern, A., 2005**,** Shortcut method for evaluation and design of a hybrid processes for enantioseparations. Journal of Chromatography A 1092, 43-54.

Kawajiri, Y., Biegler, L.T., 2006. Large scale nonlinear optimization for asymetric operation and design of Simulated Moving Beds. Journal of Chromatography A. 1133, 226-240.

Lenhoff, A.M., Morari, M. 1982. Design of resilient processing plants – I. Process design under consideration of dynamic aspects. Chem. Eng. Sci. 37(2), 245-258.

Luyben, M.L., Floudas, C.A., 1994. Analyzing the interaction of design and control – 1. A multiobjective framework and application to binary distillation synthesis. Comp. Chem. Eng. 18(10), 933-969.

Mohideen, M.J, Perkins, J.D. and Pistikopoulos, E.N. 1996. Optimal design of dynamic systems under uncertainty, AIChE J. 42(8), 2251-2272.

Morari, M., 1983. Design of resilient processing plants – III. A general framework for the assessment of dynamic resilience, Chem. Eng. Sci. 38(11), 1881-1891.

Narraway, L.T., Perkins, J.D., Barton, G.W., 1991. Interaction between process design and process control: economic analysis of process dynamics. J. Process Control. 1, 243-250.

Patil, B.P., Maia, E., Ricardez-Sandoval, L.A., 2015. Integration of scheduling, design, and control of multiproduct chemical processes under uncertainty. AIChE J. 61(8), 2456-2470.

Pistikopoulos, E.N., Diangelakis, N.A., 2016. Towards the integration of process design, control and scheduling: Are we getting closer? Comp. Chem. Eng. 91, 85-92.

Narraway, L.T., Perkins, J.D., Barton, G.W., 1991. Interaction between process design and process control: economic analysis of process dynamics. J. Process Control. 1, 243-250.

Ricardez-Sandoval, L.A., Budman, H.M., Doulas, P.L., 2009. Integration of design and control for chemical processes: A review of the literature and some recent results, Annual Reviews in Control 33, 158-171.

Ross, R., Swartz, C.L.E., 1995. The application of dynamic operability assesment to flotation circuit design. Proc. 8th International Symposium on Automation in Mining, Mineral and Metal Processing, IFAC Proceeedings Volumes, 28(17), 83-89.

Ross, R., Swartz, C.L.E., 1997. Inclusion of model uncertainty in a computational framework for dynamic operability assessment. Comp. Chem. Eng., 21, Suppl, S415-S420.

Sakizlis, V., Perkins, J.D., Pistikopoulos, E.N., 2003. Parametric controllers in simultaneous process and control design optimization. Ind. Eng. Chem. Res. 42, 4545-4563.

Sakizlis, V., Perkins, J.D., Pistikopoulos, E.N., 2004. Recent advances in optimization-based simultaneous process and control design, Comp. Chem. Eng. 28, 2069-2086.

Schenk, M., Sakizlis, V., Perkins, J.D., Pistikopoulos, E.N., 2002. Optimization-based methodologies for integrating design and control in cryogenic plants. In European Symposium on Computer-Aided Process Engineering-12, J. Grievink and J. van Schijndel, eds., 331-336.

Skogestad, S., Postlethwaite, I., 1996. Multivariable Feedback Control. Analysis and Design. Wiley, New York.

Soliman, M., Swartz, C.L.E., Baker, R., 2008. A mixed-integer formulation for back-off under constrained predictive control. Comp. Chem. Eng. 32, 2409-2419.

Sreedhar, B, Kawajiri, Y., 2014**,** Multi-column chromatographic process development using simulated moving bed superstructure and simultaneous optimization - model correction framework. Chem. Eng. Sci. 116, 428-441.

Swartz, C.L.E.,1996. A computational framework for dynamic operability assessment, Comp. Chem. Eng., 20(4), 365-371

Swartz, C.L.E., 2004. The use of controller parametrization in the integration of design and control. Chapter B3 in The Integration of Process Design and Control, P. Seferlis & M.C. Georgiadis, eds., Elsevier, 239-263.

Toumi, A., Engell, S., 2004. Optimization-based control of a reactive simulated moving bed process for glucose isomerization, Chem. Eng. Sci. 59(18), 3777-3792

van Schijndel J., Pistikopoulos, E.N. 1999. Towards the integration of process design, process control & process operability – Current status & future trends. Proc. FOCAPD 1999, Snowmass, Colorado.

White, V., Perkins, J.D., Espie, D.M., 1996. Switchability analysis. Comp. Chem. Eng. 20(4), 469-474.

Ziegler, J.G., Nichols, N.B., 1943. Process lags in automatic-control circuits. Transactions of the ASME, 65, 433-444.

Zahn, V. M., Mangold, M., Krasnyk, M., Seidel-Morgenstern, A., 2009. Theoretical Analysis of Heat Integration in a Periodically Operated Cascade of Catalytic Fixed-Bed Reactors. Chem. Eng. Technol. 32 (9), 1326-1338.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50006-9

Including Nature in Engineering for Innovation and Sustainability: Promise, Progress and Peril

Bhavik R. Bakshi[a]

[a]*Lowrie Department of Chemical and Biomolecular Engineering, The Ohio State University, Columbus, OH 43210, USA*
bakshi.2@osu.edu

Abstract

Modern engineering keeps ecological systems outside its decision boundary, even though goods and services from nature are essential for sustaining all its activities. This has been the system boundary at least since the industrial revolution when the human footprint was quite small, and nature seemed infinite. With this situation having reversed in the modern world, there is an urgent need to change the engineering paradigm from one that takes nature for granted to one that accounts for the role of nature and works with it while respecting its limits. In this article we argue that such a paradigm shift holds the promises of enabling sustainability and innovation by developing engineering that benefits from nature's ability to treat emissions and provide resources in an economically viable and environmentally sustainable manner. We provide an overview of recent progress in including nature in engineering, for tasks such as design of processes and supply chains, and for life cycle assessment. Despite the promise of including nature in engineering, there are also risks if this inclusion is not done properly. Current methods tend to commodify nature due to which holistic characteristics of ecosystems such as biodiversity and resilience could be lost. The tendency of wanting to over-engineer ecosystems may also need to be restrained.

Keywords: Ecosystem services; Sustainability; Innovation; Design; Life Cycle Assessment; Paradigm shift.

1. Promise

Goods and services from nature are essential for sustaining all human activities. This statement is widely accepted and understood yet most disciplines and even methods meant to enable sustainability do not account for the role and capacity of ecosystems in supporting human activities. Contributions from nature toward human well-being are substantial: worth trillions of dollars. It includes goods such as minerals, biomass, wood and water, and services such as photosynthesis, biogeochemical cycles, climate regulation, and air and water quality regulation.

Conventional engineering has focused on a relatively narrow boundary of a selected product or process. In the last few decades, this boundary has expanded to include processes in supply chains and the corporate enterprise (Grossmann and Westerberg, 2000). By adopting sustainability assessment methods such as waste minimization and life cycle assessment, the boundary has further expanded to account for the environmental impact of industrial activities. However, most methods do not account for the role of nature in supporting and even enhancing industrial activities. Existing methods also routinely ignore the limits of nature's ability to supply resources and

absorb wastes and implicitly assume nature to be an infinite source and sink. Ignoring nature is not limited just to the discipline of engineering. Even conventional economics ignores or undervalues goods and services from nature. For example, if we consider the role of trees, ecosystem goods such as fruits and wood have monetary value due to their being traded in markets, but services such as carbon sequestration by storing atmospheric carbon dioxide as biomass, air quality regulation by mitigating NOx, SOx and particulate matter, and water provisioning by reducing storm water runoff and recharging aquifers have no value since are outside the market. This results in the implicit assumption that these services do not matter and are unlimited. Such ignorance has at least two negative side effects.

- *Unsustainable Engineering.* Keeping nature outside the system boundary can result in decisions that contribute to exceeding nature's capacity, causing ecological degradation, and resulting in an unsustainable engineering.
- *Lost Opportunities for Innovation.* Ecosystems are capable of satisfying many human and industrial needs while relying on only renewable resources, being resilient to disturbances, and sustaining themselves over long periods of time. By ignoring nature, engineered systems loose opportunities of benefitting from innovative designs that could work in harmony with nature.

Recent efforts for quantifying ecosystem goods and services is encouraging the development of frameworks such as techno-ecological synergy (TES) (Urban et al., 2010; Bakshi et al., 2015), which aims to include the role and limits of ecosystems explicitly in engineering to establish mutually beneficial synergies between human and natural systems. The novelty of this approach stems from the fact that it considers nature not just as a resource or a system to be protected, but as a holistic entity with which to build mutually beneficial synergies for sustaining human and other activities. TES accounts for goods and services from ecosystems that directly benefit industry and society. As depicted in Figure 1, in TES technology relies on ecosystems to absorb wastes and produce resources, while ecosystems rely on technological systems to provide essential nutrients, land, and other resources. In addition to meeting industrial needs, ecosystems also provide other cobenefits that could benefit society. Thus, TES goes beyond and combines the best features of existing approaches like circular economy, industrial symbiosis, and cradle-to-cradle design. Increasing synergy can result in a system where there is no net flow of raw materials or pollutants in Figure 1. Such a system could be self-sustaining.

This paper provides an overview of the recent progress in methods for including nature in engineering and its applications, with emphasis on chemical engineering problems. This is the topic of the next section with emphasis on process design, supply chain design, and life

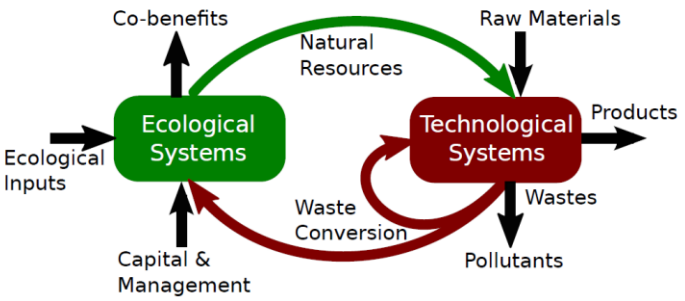


Figure 1. Framework of Techno-Ecological Synergy (Bakshi et al., 2015).

cycle assessment. At the smallest scale of process design, ecosystems are considered analogous to unit operations. At the supply chain scale, availability of ecosystem services influences decisions about choosing suppliers. At the life cycle scale, ecosystems are included just like process modules and the resulting TES-LCA method provides insight about absolute sustainability based on determining the extent to which the limits of ecosystem services are exceeded. Various applications are described to convey the potential benefits of the convergence of engineering with ecology are demonstrated with the help of several examples relevant to chemical engineering. The subsequent section discusses the challenges of including nature in engineering and whether this approach will lead to sustainable development. It identifies the potential pitfalls of this approach and the issues that require close attention in future work.

2. Progress

2.1. Process Analysis

Increasing understanding about the goods and services provided by nature and their importance led to work on determining whether these goods and services are relevant to industrial activity and whether nature could satisfy industrial needs. The most successful commercial use of ecosystems in industry is that of wetlands for treating industrial waste (Vymazal, 2011), as demonstrated by many successful implementations. Most such installations have been used for the last finishing step of water treatment, and are found to be not only economically and environmentally attractive, but also provide additional cobenefits to society through preserving biodiversity, maintaining the water table, and providing recreational and educational opportunities. One such implementation is estimated to have $283 million in net present value savings over the project's life time as compared to a conventional technological alternative (DiMuro et al., 2014).

The Nature Conservancy and Dow Chemical evaluated the ability of vegetation to mitigate air pollutants such as NOx, SOx and particulate matter (TNC-Dow, 2013). This study found that this nature-based solution can be less expensive than conventional technological solutions such as selective catalytic reduction of NOx and scrubbers for removing SOx. Another study (Gopalakrishnan et al., 2016) evaluated a biodiesel manufacturing site with a combined heat and power process for supporting the biodiesel facility. If available land in the vicinity of this site had native trees that were only 15 years old, they were shown to be capable of mitigating a significant fraction of criteria air pollutants. Older trees could drive the site toward zero net emissions. Here, vegetation was considered to go beyond environmental regulations, which is the stated sustainability goal of many corporations. Older trees were also found to have excess capacity than what was needed for the manufacturing process, which can be a cobenefit that the trees provide to the region by mitigating emissions from other local sources. Wetlands on this biodiesel site could also provide fresh water to the local river or aquifer, or back to the process after further treatment. Comparing the net present value of the process with technological versus ecological solutions for removing identical amounts of pollution showed that ecosystems had a higher NPV. Over time, ecosystems become even more attractive than technological systems since ecosystems appreciate over time while technology depreciates.

Comparing air emissions with the capacity of vegetation to take them up in each county across the United States also demonstrates the benefits of relying on nature to go

beyond regulations. Vegetation has substantial capacity to take up SO_2 emissions in 49% of US counties, 13% for PM_{10}, 4% for $PM_{2.5}$ and 5% for NO_x. Restoration of available land to the average current vegetation in each county is found to be more cost effective than using technology for the same task in 75% of US counties. Nature-based solutions are found to be most effective for mitigating emissions from activities such as farming; mining, oil and gas; electricity generation; and residential systems. Such studies convey the benefits of including ecosystems in industrial processes, and have encouraged further work. (Gopalakrishnan, 2017)

2.2. Process design

Reducing the environmental impact of chemical processes has been an area of active research for many decades. These efforts evolved from reducing resource use in individual equipment and process to reducing resource use and impact of emissions over the entire life cycle (Grossmann, 2004). Recent work on including nature in process design is including relevant ecosystems as if they were another unit operation, like distillation columns or heat exchangers. The optimization formulation of this integrated design of technological and ecological systems is as follows,

$$\max Z_1(x, x_e, y); \quad \max Z_2(x, x_e, y)$$

$$\text{s.t. } f(x, y) \leq 0;\ f_e(x_e, y) \leq 0$$

Here, the two objectives represent economic and environmental goals, and subscript e denotes ecological models and variables. Variables x and x_e are continuous while y are integers. The key difference with conventional design is the simultaneous presence of technological and ecological models and variables to enable integrated and synergistic design.

As described in Section 2.1, wetlands are already being used for treating industrial waste water. However, most such uses do not integrate wetlands into the flowsheet like other pieces of equipment. That is, wetlands are usually included as end-of-pipe solutions, meaning that the process is designed first in a techno-centric manner, after which the wetland is added to treat the waste water. Recent work on including ecosystems as unit operations in process design has shown the benefits of integrating wetlands in the process design (Gopalakrishnan and Bakshi, 2018). Integrated design with wetlands results in a design where the main process operates in a way that accommodates the use of wetlands for treating waste water. This results in decisions that avoid chemicals that cannot be treated by the wetland, and larger concentration of those chemicals that can be treated. The wetland is also designed to have the capacity and residence time for treating the industrial waste and producing freshwater of desired quality to satisfy needs of the technological system. Such an integrated design with objectives of maximizing net present value and minimizing net water use is found to be superior to an end-of-pipe design and a design with a technological alternative (anaerobic baffle reactor).

Application of the TES design approach to several problems demonstrates that inclusion of ecosystems can expand the design space as compared to the design space available from conventional, techno-centric designs. The nature of this expansion is depicted in Figure 2. As shown, the Pareto curve for TES design is in a region that is infeasible for conventional, techno-centric design. The additional design space made available by

TES design is indicated by the shaded region. Thus, TES design could provide innovative "win-win" designs that may be economically and environmentally better than conventional designs. Research to date has been demonstrated this for design of a biodiesel process, management of biosolids from urban water treatment, design of agricultural landscape for biofuels, and design of residential buildings. Integration of a local biomass farming system, biofuel manufacture and waste water treatment is also shown to be more self-sustaining than conventional disintegrated approaches (Martinez-Hernandez et al., 2017) and have benefits in terms of the food-energy-water nexus (Hang et al., 2016).

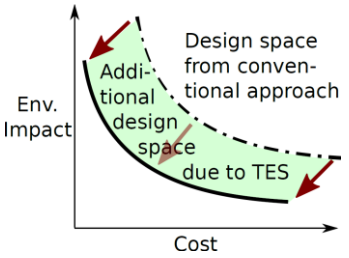


Figure 2. Expansion of design space due to including nature in engineering.

A unique challenge from including nature in engineering is due to the vastly different dynamic behavior of technological and ecological systems. The former exhibit relatively little variation with time, while the latter have high temporal variability with time of day, seasons, etc. For example, deciduous trees shed their leaves in autumn resulting in a drastic reduction in their capacity to mitigate air pollutants, and the productivity of wetlands in temperate regions may be much lower in the winter. Therefore, viable synergies between technological and ecological systems need to address such differences in their dynamics. Efforts toward addressing such issues have considered a local bioenergy-farming-waste water treatment system (Martinez-Hernandez et al., 2015) and a biodiesel design problem (Gopalakrishnan, 2017). The latter problem has been solved over four seasons for a location near Cincinnati, Ohio as a multi-period optimization problem. For mitigating air emissions, it is found that ecosystems can still be viable for capturing a large fraction of emissions by selecting a combination of native deciduous and evergreen tree species, adjusting the use of technologies for capturing emissions, and varying the amount of biodiesel produced in each season. Seasonal variation does deteriorate the overall performance of trees as compared to having the process in a region without seasonal variation, but using trees for going beyond regulations still remains economically more attractive than using technology. Also, evergreen trees continue to mitigate pollutants in the winter and are included in the design.

2.3. Supply chain design

If ecosystems are included in the design of supply chains, the overall goal of minimizing violation of ecological limits encourages inclusion of those suppliers that are closest to satisfying this requirement. It also encourages protection and restoration of ecosystems at and around supplier sites. An illustrative example on designing the supply chain of a biofuel finds a solution that reduces the impact of emissions and enhances ecosystem services, as compared to a conventional solution that reduces only the impact of emissions (Ghosh and Bakshi, 2018).

Ecosystem services may be represented in monetary units to allow aggregation with conventional cost analysis. By using such aggregation and calculating a Green GDP, a biofuel supply chain design problem has been solved recently with ecosystem services (Garcia and You, 2017). This work demonstrates how accounting for ecosystem services in the selected supply chain steps and its geographical region affects the resulting design and optimum value of the objective function.

2.4. Life Cycle Assessment

It is widely accepted that the life cycle of products and processes must be considered to prevent shifting of impacts along the supply or demand network. The method of life cycle assessment (LCA) has been formalized and adopted in process systems engineering for developing designs that aim to reduce life cycle impact along with conventional monetary goals. However, as discussed in Section 1, a significant shortcoming of conventional LCA is its ignorance of ecosystem capacity and goods and services from nature. The idea of developing synergies between human and natural systems has been extended to the approach of techno-ecological synergy in life cycle assessment (TES-LCA). This approach explicitly accounts for the demand and supply for ecosystem services in each step of the life cycle (Liu and Bakshi, 2018).

Accounting for nature requires some modifications and additions to the four steps of conventional LCA, which are goal and scope definition, inventory analysis, impact assessment and improvement analysis. The main goal of conventional LCA is to compare alternatives to choose one that has the smallest life cycle impact. This comparative approach results in relative sustainability metrics. TES-LCA also has this goal, but in addition, it also aims to protect and restore ecosystems that provide needed goods and services and encourage efforts toward staying within nature's capacity. Such an approach can provide absolute sustainability metrics by comparing the demands imposed on nature by industrial activities with the capacity of nature to meet the demand. Absolute environmental sustainability requires the demand to be less than or equal to the supply at the scale of the serviceshed of the selected ecosystem service. The serviceshed is the geographical region that supplies ecosystem services to a beneficiary. For CO2, the serviceshed is the planet since an emitted molecule can be absorbed anywhere on earth. For water the serviceshed is the watershed, and for pollination by insects the serviceshed is determined by the distance travelled by the insects. With this insight, the boundary of TES-LCA includes the local and serviceshed scales. The inventory analysis step of TES-LCA requires information about ecosystem services, which may be obtained from ecological models and remote sensing data. Since such data can have significant spatial variation, it is best to develop spatial models for TES-LCA.

Impact assessment in TES-LCA calculates the sustainability metric, $V_k = (S_k - D_k) / D_k$, for the k-th ecosystem service at selected spatial scales, including the serviceshed. Environmental sustainability requires $V_k \geq 0$ at the scale of the serviceshed. If this criterion is satisfied then the emissions could have zero net impact under the assumption of steady state. However, if $V_k < 0$ then the emissions not taken up by nature will have an environmental impact. We use conventional life cycle impact assessment to determine this impact. The consequences of this impact on degradation or enhancement of ecosystem services may also be quantified.

The unique characteristics of TES-LCA are illustrated in Figure 3 for a corn ethanol life cycle. Here, the x-axis is V_k at a local scale, while the y-axis is V_k at the scale of the serviceshed. If both metrics are positive (Quadrant I), it implies local and absolute sustainability, which is true for the water provisioning ecosystem service in this example. This is due to location of the farming in the water-rich region of Ohio. If an activity is in Quadrant II, the activity demands more from nature than available within the local boundary. However, demand at the serviceshed scale is within nature's limits. Thus, we have absolute sustainability, in this case, but not local sustainability. V_k in

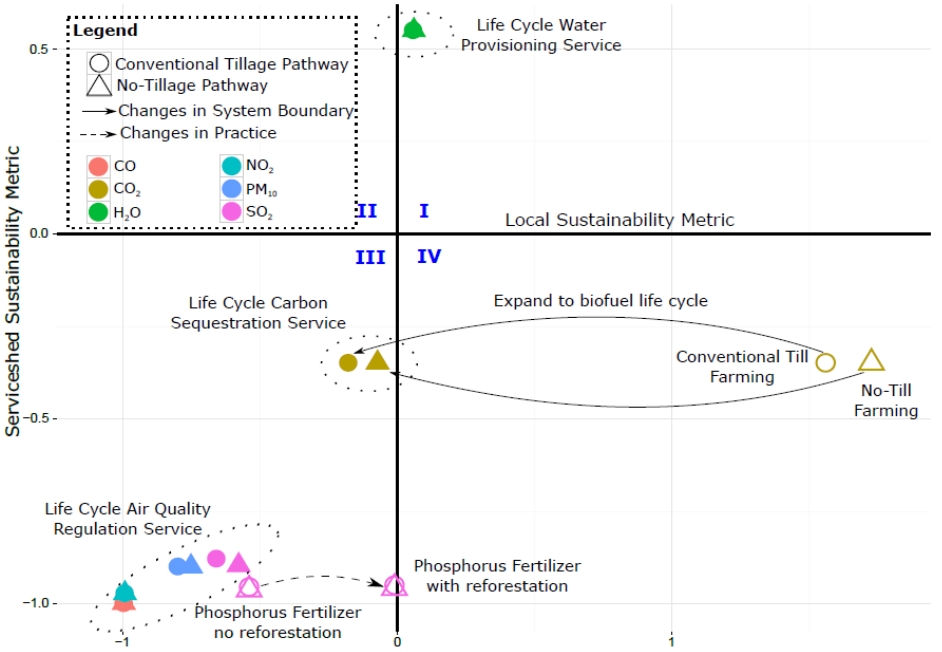


Figure 3. TES-LCA of carbon sequestration, water provisioning, and air quality regulation ecosystem services in the corn ethanol life cycle. (Liu and Bakshi, 2018)

Quadrant III implies local and absolute unsustainability, while Quadrant IV indicates an island of sustainability since the activity is within nature's limits at a local scale, but not in the serviceshed. As we can see from this figure, farming activities are an island of sustainability (Quadrant IV), but including the life cycle causes them to shift to Quadrant III. The Phosphorus fertilizer manufacturing process is in Quadrant III for air quality regulation, but if there was a forest around the facility, the activity moves toward Quadrant IV due to the capacity of vegetation to take up air emissions. Such insight is unique to TES-LCA since it cannot be obtained from conventional LCA.

3. Peril

Progress in the inclusion of nature in engineering decisions described in Section 2 conveys the potential benefits of greater attention to this effort. Many research challenges need to be addressed such as the following.

Data and models. The complex nature of ecosystems means that they are often not as well understood as human-designed systems. However, models of many ecosystems are available that are validated, reliable and widely used. Due to advances in remote sensing and geographical information systems, data about land cover and ecosystem services across the world are increasingly available. Advances in process systems engineering are needed to use these data and models along with technological information for developing innovative synergies.

Spatio-temporal variation. Nature is "self-designed" while technology follows the approach of "imposed design." Self-design of nature means that it requires little direct intervention and develops organically into a system that is resilient and self-sustaining,

but intermittent across space and time. This is because at scales larger than an organism nature follows homeorhesis. In contrast, technological systems are designed for homeostasis and to do few tasks with input of high quality resources that are often non-renewable and unsustainable. High efficiency of technological systems makes them predictable but fragile. Many ecosystems function at a reduced level at night and in the cold. To work with such systems, technological systems need to adapt their functioning or an appropriate combination of technological and ecological systems needs to be used. Integrating nature into engineering systems requires methods to deal with these differences between human and natural systems.

Sustainability and resilience. Technological systems are designed and operated with objectives such as maximizing economic value and minimizing environmental impact. Such strategies may not be appropriate for ecosystems since the objective function of nature is not known. Designing ecosystems with anthropocentric objectives such as maximizing efficiency could compromise their resilience and ability to provide diverse ecosystem services. Thus, conventional optimization-based approaches that have been very successful in engineering design may have to be modified for TES design.

Commodification of Ecosystem Services. Current research in quantifying and accounting for ecosystem services considers each service separately or emphasizes some services more than others. For example, methods exist for quantifying services such as climate regulation and water provisioning in physical or monetary units. However, quantification of cultural services such as spiritual, religious, and aesthetic aspects is difficult to quantify. Therefore, current approaches may result in less emphasis on those services that are more difficult to quantify. In addition, greater emphasis on a few ecosystem services could translate into efforts toward protecting them at the cost of others. For example, interest in being carbon neutral may result in efforts for maximizing the carbon sequestration of vegetation. This could encourage the development and use of monocultures of genetically modified high carbon sequestering trees. Such ecosystems would lack the holistic and self-sustaining properties of nature, and could cause unintended harm to other ecosystem services. Thus, thinking of ecosystem services as separate commodities may result in solutions that are not likely to be truly sustainable.

Ecological literacy. Successful efforts for including nature in engineering requires engineers to have knowledge about ecological systems. Surveys have shown that engineers are among the least ecologically literate. This is an educational challenge that requires a change in the engineer's attitude from considering nature to be controlled and dominated to one of wishing to work with and respecting nature.

There has been a successful convergence between biology and engineering at the micro or reductionist level resulting in many advances and new disciplines such as biomedical and biological engineering and systems biology. Such convergence between ecology and engineering at the macro or systems level may be initiated by efforts toward techno-ecological synergies, as discussed in this paper.

4. Conclusions

The current engineering paradigm keeps nature outside its boundary, which can unknowingly encourage ecological degradation and miss opportunities for innovation by working with nature to benefit from its ability to satisfy industrial and human needs.

This work summarizes the potential benefits, recent progress and outstanding challenges in including nature in engineering decisions. Sustainability and innovation are the two key promises of including nature in engineering. This promise is demonstrated by recent work on integrating nature-based solutions such as the ability of vegetation to treat air emissions and of wetlands to regulate water quality, with conventional technologies. Methods have also been developed for including ecosystem services in life cycle assessment. Full realization of the promise requires advances to address the fundamentally different behavior of technological and ecological systems such as the self-design of nature versus the imposed design of technology. Conventional optimization-based engineering design may also need to be modified to retain the resilience of nature. Engineers also need to learn about ecology to enable a successful and synergistic partnership between technological and ecological systems

Acknowledgements. Partial financial support was provided by the National Science Foundation and the United States Department of Agriculture.

References

B. R. Bakshi, G. Ziv and M. D. Lepech, 2015, Techno-Ecological Synergy: A Framework for Sustainable Engineering, Environmental Science & Technology, 49, 3, 1752-1760.

J. L. DiMuro, F. M. Guertin, R. K. Helling, J. L. Perkins, S. Romer, 2014, A Financial and Environmental Analysis of Constructed Wetlands for Industrial Wastewater Treatment *Journal of Industrial Ecology*, 18, 631-640

D. J. Garcia, F. You, 2017, Introducing Green GDP as an Objective to Account for Changes in Global Ecosystem Services Due to Biofuel Production, in Espuna, A.; Graells, M. & Puigjaner, L. (Eds.), *27th European Symposium on Computer Aided Process Engineering*, Elsevier, 40, 505 – 510

T. Ghosh, B. R. Bakshi, 2018, Including ecosystem services in sustainable process design across multiple spatial scales, PSE 2018, accepted

V. Gopalakrishnan and B. R. Bakshi, 2018, Ecosystems as Unit Operations for Local Techno-Ecological Synergy: Integrated Process Design with Treatment Wetlands, *AIChE Journal*, accepted

V. Gopalakrishnan, B. R. Bakshi, and G. Ziv, 2016, Assessing the Capacity of Local Ecosystems to Meet Industrial Demand for Ecosystem Services, AIChE Journal, 62, 9, 3319-3333.

V. Gopalakrishnan, 2017, Nature in Engineering: Modeling Ecosystems as Unit Operations for Sustainability Assessment and Design, Ph. D. Thesis, The Ohio State University

I. E. Grossmann, A. W. Westerberg, 2000, Research Challenges in Process Systems Engineering, *AIChE Journal*, 46, 1700

I. E. Grossmann, 2004, Challenges in the new millennium: product discovery and design, enterprise and supply chain optimization, global life cycle assessment, *Computers & Chemical Engineering*, 29, 29-39

M. Y. L. P. Hang, E. Martinez-Hernandez, M. Leach, A. Yang, 2016, Designing integrated local production systems: A study on the food-energy-water nexus, *J. Cleaner Production*, 135, 1065-1084

X. Liu and B. R. Bakshi, 2018, Ecosystem Services in Life Cycle Assessment while Encouraging Techno-Ecological Synergies, *Journal of Industrial Ecology*, accepted

E. Martinez-Hernandez, M. Y. L. P. Hang, M. Leach, A. Yang, 2017, A Framework for Modeling Local Production Systems with Techno-Ecological Interactions, *Journal of Industrial Ecology*, 21, 815-828

E. Martinez-Hernandez, M. Leach, A. Yang, 2015, Impact of Bioenergy Production on Ecosystem Dynamics and Services—A Case Study on U.K. Heathlands, *Environ. Sci. Technol.*, 49, 9, 5805–5812

P. M. Tikka, M. T. Kuitunen, S. M. Tynys, 2000, Effects of Educational Background on Students' Attitudes, Activity Levels, and Knowledge Concerning the Environment, *The Journal of Environmental Education*, 31, 12-19

TNC-Dow, The Nature Conservancy - Dow Collaboration: 2012 Progress Report, 2013, https://www.dow.com/en-us/science-and-sustainability/collaborations/nature-conservancy, accessed February 11, 2018

R. A. Urban, A. Baral, G. F. Grubb, B. R. Bakshi, and W. J. Mitsch, 2010, Towards the sustainability of engineered processes: Designing self-reliant networks of technological-ecological systems, Computers and Chemical Engineering, 34, 9, 1413-1420.

J. Vymazal, 2011, Constructed Wetlands for Wastewater Treatment: Five Decades of Experience, *Environmental Science & Technology*, 45, 61-69

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50007-0

Innovations in Process Control Education: A Flipped Classroom/Studio Approach

B. Wayne Bequette

Rensselaer Polytechnic Institute, 110 Eighth St., Troy, NY 12180-3590, USA
bequette@rpi.edu

Abstract

A flipped classroom where students view screencasts and read textbook material and take an on-line quiz before class has been implemented in a process dynamics and control course. The class periods involve brief lectures summarizing what they have learned, and include discussions and advanced problem solving using MATLAB and Simulink. The weekly assignments include a mix of analytical and numerical solutions. Motivating examples, such as automated insulin delivery are used throughout the course. Although a physical laboratory experiment is not used, the class benefits from a tour of the campus boilerhouse system.

Keywords: Learning modules, flipped classroom, simulations, interactive learning

1. Motivation and Background

Process control is a core course in the chemical engineering undergraduate curriculum in most departments, yet it sometimes suffers from an over-emphasis on analytical mathematics and without proper motivation from real process challenges. This paper presents an overview of a flipped classroom approach that best makes use of classroom time to solve realistic, challenging process control problems.

1.1 Process Control Textbooks and Curricula

Current process control textbooks cover much of the same analysis content as Coughanowr and Koppel (1965), which was the textbook used in the course that I took in 1979. Most books have added topics and examples and make use of computer-aided software, such as MATLAB. Many texts cover far too many topics for a one-semester course, so it is important that an instructor be selective about the material covered.

At roughly ten-year intervals the course structure is revisited at a control or education conference; see, for example Edgar (1990), Edgar et al. (2006), and Silverstein et al. (2016). The survey by Silverstein et al. (2016) indicates that 70% of courses use project or problem-based learning and 45% of courses use computers in the classroom.

1.2 Evolution of Teaching by the Author (30 years and counting)

My initial experience teaching process control was at UC-Davis in 1987-88. I used the Stephanopoulos (1984) textbook and analytical solutions. I arrived at RPI in 1988 and was impressed with the computing facilities and software, so I scheduled a weekly computer laboratory for the process control class. Initially I used the IBM simulation package, CSMP, but switched to MATLAB after a year or so. The regular classroom sessions remained largely lecture-based with handwritten derivations, but with as much engagement and active discussions with students as possible. To further motivate

students we developed a set of case studies that the students, in groups of two, would work on during the final 4 weeks of the course (Bequette et al., 1998). They would select from a set 4 topics, perform a literature review, develop models based on step-testing, design SISO controllers, then use the relative gain array (RGA) to decide on controller pairings for the multivariable case. Requirements included short reports each week, a final written report and a final presentation. I would serve as the advisor for one case study, while the TA and two of my other graduate students would be the advisors for the other three case studies. This was very time-intensive but the students were clearly motivated by the opportunity to choose a case study in their area of interest.

1.3 Studio Classrooms

The studio classroom approach was pioneered at RPI, with initial courses in physics and calculus in 1993 (Wilson, 2002). The first classrooms had networked computer workstations, with 2-4 students per computer. I started using a studio classroom for process dynamics and control in 1999 (Bequette et al., 1999). That classroom could handle up to 40 students that could face forward during lectures and discussions, then swivel 180º to work at computer workstations in groups of two. This was not an efficient use of space, and in 1999 RPI began requiring that entering freshmen (class of 2003) purchase laptop computers supplied with relevant computational software.

2. The Flipped/Inverted Classroom at RPI

2.1 Screencasts

A common characteristic of the flipped classroom is that students read and view important lecture material outside the class so that more time can be spent using engaging, interactive learning techniques in the classroom. I feel that it is better for the students to view a couple of short "screencasts" of 5-7 minutes each, rather than viewing a videotaped lecture of 50 minutes in length. The idea is to focus on a specific topic or technique and to have their full attention during the short screencast. The University of Colorado has developed a large set of screencasts for most of the core chemical engineering courses, available at learncheme.com, with 58 videos related to process control. Sometimes the notation is slightly different than that used in my course, so I warn students of this both pre- and post-viewing.

2.2 Brief On-line Quizzes

The flipped classroom works best if all of the students have read or viewed the lecture material in advance. I require the students to take a brief on-line quiz, automatically graded, before class. This typically consists of 5 multiple choice or true-false questions that are relatively easy if they have covered the material. Questions are sometimes related to the previous lecture, often involving impromptu discussions. For example, when Hurricane Harvey hit the Gulf Coast in August 2017, we discussed refining capacity and the effect of plant shutdowns on gasoline prices. I noted that I had worked at a refinery in Port Arthur, Texas and asked the students to name a famous former blues/rock singer from that area; since no one could name Janis Joplin the next quiz included questions based on a Wikipedia page on Janis Joplin.

2.3 In-class Simulation Exercises

In-class simulation exercises change in difficulty throughout the semester. During early exercises I have the students construct simple Simulink diagrams. As we get into more

difficult problems, e.g. cascade control, I supply them with .mdl and script files that they can modify. I feel that over the years some Simulink blocks have gotten too flexible and difficult for students to follow; I need to spend time explaining and discussing the different ways to implement a continuous PID controller, for example. Indeed, when I cover digital control I supply my own block that corresponds to my preferred derivation and implementation of digital PID.

2.4 Weekly Homework Assignments

Homework problems are assigned weekly and cover analytical and MATLAB/Simulink-based solutions. At the time of the assignment the particular topics have not been covered yet, so we quickly go over the assignment and the particular challenges to motivate them for the techniques being covered that week. The daily in-class exercises are used as a springboard for the solution of the weekly homework assignments.

2.5 Case Studies and Detailed Problems

Most years I have used case study projects for the final month of the semester, as discussed in section 1.2. This past year (2017) I decided to use a distillation control problem as a multivariable study example. Another detailed example used and discussed throughout the course is an automated insulin delivery system, as presented in section 4.

2.6 Mid-term and Final Exams

While the in-class exercises and weekly homework assignments involve the use of MATLAB and Simulink, the two mid-term exams and the final exam are based on analytical solution techniques and fundamental modeling and control understanding.

2.7 Additional Topics: Safety, Process Design, Boilerhouse Tour

Process safety is discussed, in one way or another, in almost every lecture. Most control strategies involve a discussion of whether a fail-open or fail-closed value should be used, for example. Early on most students do not have an appreciation of how a valve can be specified for a particular failure mode, so I present cross-sectional diagrams of typical valves and actuators for discussion. When discussing liquid surge drums we think about realistic values for the high and low level alarms. We also take groups of students on a boilerhouse tour with the operator pointing out the various vessels, sensors, actuators and the control room and control system. This is of particular benefit to students that have not had summer industrial experience.

3. Course Topic Overview

Chemical Process Dynamics and Control is a 4-credit course scheduled for 3 days/week (Tu/We/Fr) for one hour and 50 minutes each class period. As a practical matter, to give adequate time for homework problem solving and to accommodate the instructors travel schedule, during many weeks the third class period serves as a recitation, with the TA reinforcing material, and assisting with MATLAB, as examples. Since it is a 4-credit course, we tend to cover about 33% more topics than is covered in most 3-credit process control courses. A typical 15-week semester will include 32 lecture/regular classes, 7 recitations, 2 in-class exams, 1 campus boilerhouse tour, 1 no-class (Thanksgiving week), 1 final course review, and 1 final exam (during finals week).

Table 1. Course Topics	Table 2. Example Processes
Incentives/Motivation for Process Control Mathematical modeling Linearization, state space models Laplace transforms, transfer function analysis Connection between poles and eigenvalues Dynamic behavior, step and impulse inputs Introduction to feedback (FB) control, PID Closed-loop stability Real PID, filtering, derivative on output Internal Model Control and IMC-based PID Digital PID and digital filtering Cascade (CC) and feedforward (FF) Anti-reset-windup Nonlinear PID – surge vessel example Closed-loop control interaction, RGA Singular Value Decomposition (SVD) Plantwide control, startup/shutdown (brief) Model predictive control (brief) Statistical process control (brief) Optimization: Linear Programming (brief)	Gas and liquid surge vessels Insulin pharmacokinetics Biochemical reactor Jacketed reactors (batch & continuous) Steam drum, 3-mode (FF/FB/CC) Gasoline blending – ratio control Automated insulin delivery for type 1 diabetes: (i) response to meals w and w/o feedforward, (ii) effect of sensor bias, quantization & noise Drug infusion in anesthesia – multiple loops Distillation control – multiple loops Gasoline blending – LP problem

4. Motivating Example: Automated Insulin Delivery

In this learning module we first review the natural physiologic control loops involved when a healthy pancreas regulates blood glucose. We then describe the lifestyle of an individual with type 1 diabetes, who must monitor blood glucose levels and administer insulin, either through multiple daily injections or using a continuous insulin infusion pump. We describe typical ranges for basal (steady-state) insulin infusion, carb-to-insulin ratios for meals, and correction factors (amount of insulin required to reduce a glucose level by a certain amount). Someone with type 1 diabetes serves as a controller (feedforward and feedback) through constant diligence; this motivates the development of an automated insulin delivery system shown in Figure 1.

After reading this module and performing the studies a student is able to:

- Appreciate the challenges someone with type 1 diabetes faces in managing blood glucose. Meals increase and insulin decreases blood glucose. One challenge is providing the correct insulin bolus to compensate for a meal. Uncertainty in the carbohydrates in a meal makes it safer to under-bolus to avoid *hypoglycemia*.
- Appreciate the importance of maintaining a consistent healthy value of blood glucose. Blood glucose that is too low (*hypoglycemia*) can cause short-term dangers, such as drowsiness or a diabetic coma. Blood glucose that is too high (*hyperglycemia*) causes long-term risks, such as micro- and macro-vascular problems resulting in retina and other problems.
- Understand the metrics used in reporting clinical performance of glucose control strategies, such as mean glucose, time-in-range (70-180 mg/dL), time in *hypoglycemia* (less than 70 mg/dL), time in *hyperglycemia* (greater than 180 mg/dL), and total daily insulin dose (insulin used each day).
- Approximate the dynamic behavior between insulin and meal inputs and blood glucose outputs using low-order transfer function models.

- Understand the importance of sensors that report discrete quantized values (for example, a continuous glucose sensor reports to the nearest 1 mg/dL).
- Understand the effects of sensor noise and bias on control system performance. A mis-calibrated glucose sensor can be off by 10-20 mg/dL or more.
- Understand that providing an insulin bolus at mealtime (feedforward control) leads to improved glucose control performance compared to feedback-only control.

Figure 1. Example of an automated insulin delivery system currently under study. The components, using Bluetooth communication, include a continuous glucose monitor (CGM), a control algorithm on the smart phone and insulin infused by the insulin pump.

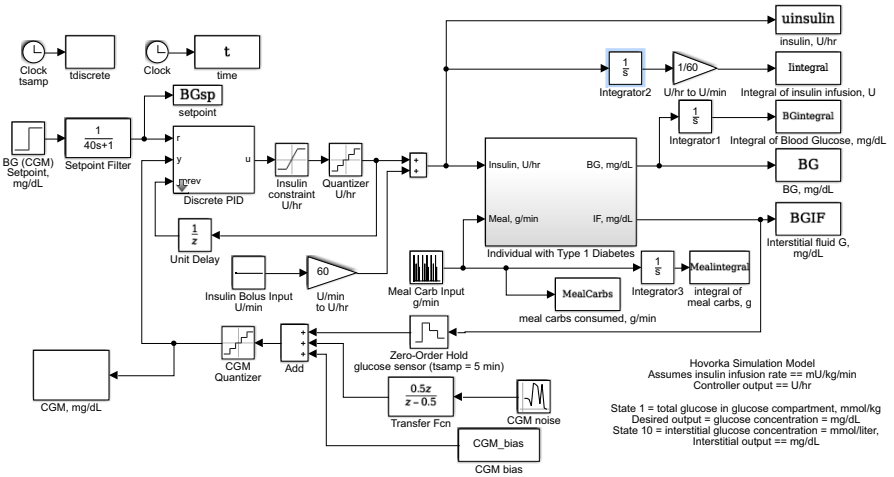


Figure 2. Simulink Diagram for Automated Insulin Delivery. Includes open-loop behaviour by setting the controller gain to 0.

Students are provided with a Simulink diagram (Figure 2) that includes the insulin-meal-blood glucose model of Hovorka et al. (2004) and Wilinska et al. (2005). We added additional lags for the dynamics between the blood and interstitial fluid and the sensor lag, with possible bias and measurement noise. The simulation model also includes a digital PID controller, and it is suggested that the students begin with the tuning parameters proposed by Palerm (2011). An outcome of the simulation exercise is to suggest controller tuning and an appropriate blood glucose setpoint if the sensor can be biased by up to 20 mg/dL, with the objective of reducing the risk of hypoglycaemia to less than 5% (that is, less than 72 minutes per day spent less than 70 mg/dL).

5. Discussion

In a process dynamics and control course the specific examples used may not be as important as the instructors enthusiasm in presenting them. I have been fortunate to be involved in a wide-range of automation and control related problems, ranging from classic chemical processes (petroleum refining, pharma batch reactors) to biomedical systems (automated insulin delivery, drug infusion in anesthesia). It certainly helps that I can bring specific examples that I have worked on into the classroom – such as responding to level sensor failures while serving as an operator in the refining industry, and conducting clinical trials of an automated insulin delivery system. The final course survey results, specifically for the instructor-added questions, shown in Table 3 indicates, that students were generally satisfied with the course; indeed it appears that the on-line quizzes actually provided encouragement for them to read the textbook.

Table 3. Course survey responses to instructor-added questions

Instructor-Added	Number Responding/(%age of Respondents)								
	1	**2**	**3**	**4**	**5**	**Omit**	**Avg.**	**Med.**	**s.d.**
The screencasts have been useful	1 (4%)	3 (11%)	3 (11%)	9 (32%)	12 (43%)	0	4.00	4.28	1.16
The on-line quizzes have been useful	2 (7%)	5 (18%)	6 (21%)	10 (36%)	5 (18%)	0	3.39	3.60	1.20
The studio approach of combining lectures and simulations has been useful	1 (4%)	3 (11%)	1 (4%)	10 (36%)	13 (46%)	0	4.11	4.40	1.13
I have benefited by reading the textbook	1 (4%)	1 (4%)	3 (11%)	14 (50%)	9 (32%)	0	4.04	4.14	0.96
I have adequate time to do the homework	0 (0%)	1 (4%)	3 (11%)	15 (56%)	8 (30%)	1	4.11	4.13	0.75
I appreciate the importance of process control in industry	0 (0%)	0 (0%)	2 (7%)	10 (36%)	16 (57%)	0	4.50	4.62	0.64
MATLAB has been easy to use	7 (25%)	5 (18%)	8 (29%)	5 (18%)	3 (11%)	0	2.71	2.75	1.33
The boiler house tour was useful	0 (0%)	0 (0%)	4 (14%)	8 (29%)	16 (57%)	0	4.43	4.62	0.74
Key **1 = Strongly Disagree**	**2 = Disagree**		**3 = Neutral**		**4 = Agree**		**5 = Strongly Agree**		

6. Approaches and Material from Other Sources

Huang (2017) discusses the results of integrating three MATLAB-based learning modules into a flipped classroom by using teaching videos to be viewed a week before class and quizzes taken at the beginning of class. Marlin (2017) discusses "blended learning" and develops a flipped classroom as one example. E-lessons consist of pre-prepared slides with audio that can be studied via the internet at any time, followed by an un-graded quiz. Class time then involves a mini-lecture reviewing the e-lesson, then workshops (hands-on, interactive learning) to reinforce learning material, followed by a mini-lecture on the next e-lesson). Seames (2017) takes a simulation-based approach, including a new textbook, and requires outside material be studied, with on-line quizzes; re-quizzes can be taken up to 3 weeks later. It should be noted that the text by Svrcek et al. (2014) also takes a simulation-based approach.

There are many learning resources available on the internet. For example, cache.org contains links to material from many sources. The learncheme.com site is particularly strong in the use of screencasts. The Hedengren process dynamics and control course at BYU has lecture material and a low-cost temperature control experiment.

7. Conclusions

A flipped classroom makes better use of a students and instructors time. Students come to class prepared, after taking an on-line quiz, and are more engaged in discussions and problem solutions. Because of the interactive nature of the class-time, it is important to have a teaching assistant willing to engage with the students.

8. Acknowledgments

I wish to acknowledge the outstanding efforts of Edmund Tang and Sambit Ghosh, who served as the TAs the first two times I used a flipped classroom approach.

References

B.W. Bequette, K.D. Schott, V. Prasad, V. Natarajan and R.R. Rao. Case Study Projects in an Undergraduate Process Control Course. *Chem. Eng. Educ.* **32**(3) 214-219 (1998).

B.W. Bequette, J.H. Chow, C.J. Li, E. Maby, J. Newell and G. Buckbee. An Interdisciplinary Control Education Studio. in *Proc. Conf. Decision & Cont.*, Phoenix, pp. 370-374 (1999).

B.W. Bequette and B.A. Ogunnaike. Chemical Process Control Education and Practice. *IEEE Control Systems*, **21**(2), 10-17 (April, 2001).

B.W. Bequette. *Process Control: Modeling, Design and Simulation*, Prentice Hall, Upper Saddle River, NJ (2003).

B.W. Bequette. A Laptop-based Studio Course for Process Control. *IEEE Control Systems*, **25**(1), 45-49 (February, 2005).

T.F. Edgar, "Process control education in the year 2000: A round table discussion," *Chem. Eng. Educ*., vol. 24, no. 2, pp. 72-77, 1990.

T.F. Edgar, B.A. Ogunnaike, J.J. Downs, K.R. Muske and B.W. Bequette. Renovating the Undergraduate Process Control Course. *Comp. Chem. Engng*. 30, 1749-1762 (2006).

J. Hedengren. Dynamics and Control. http://apmonitor.com/che436/index.php/Main/HomePage (accessed 7 February 2018).

R. Hovorka R, Canonico V, Chassin LJ, Haueter U, Massi-Benedetti M, Federici MO, Pieber TR, Schaller HC, Schaupp L, Vering T, Wilinska ME. Nonlinear model predictive control of glucose concentration in subjects with type 1 diabetes. *Physiol. Meas*. 2004; 25: 905-920.

Learncheme.com. Educational Resources for Chem. Eng. Univ of Colorado, Boulder.Process Control Screencasts. http://www.learncheme.com/screencasts/process-controls

X. Li, Z. (Jacky) Huang. An Inverted Classroom Approach to Educate MATLAB in Chemical Process Control. *Education for Chemical Engineers*. 2017;19:1-12.

T. Marlin. Flipping the Chemical Engineering Process Control Class with e-Lessons. 2017 ASEE Ann. Mtg., Columbus, OH, June 2017. http://cache.org/files/sum17-flipping-class-asee.pdf. http://pc-education.mcmaster.ca/Flipping%20class/Flipping%20the%20Classroom.pdf

C.C. Palerm. Physiologic insulin delivery with insulin feedback: A control systems perspective. *Computer Methods and Programs in Biomedicine*, 2011;102(2), 130-137.

W.S. Seames. Updating the Process Controls and Dynamics Course for the 21st Century. Presented at the 2017 AIChE Ann. Mtg., session 145d, Minneapolis, MN (Oct. 2018). Abtract: https://www.aiche.org/conferences/aiche-annual-meeting/2017/proceeding/paper/145d-updating-process-controls-and-dynamics-course-21st-century (accessed 7 February 2018).

D.L. Silverstein, M.A. Vigeant and M. Staehle. *How We Teach Process Control: 2015 Survey Results*. Presented at 2016 ASEE Ann. Conf. & Expo., New Orleans, LA. 10.18260/p.25495. https://peer.asee.org/how-we-teach-process-control-2015-survey-results (accessed 8 Feb 18).

M.W. Wilinska, L.J. Chassin, H.C. Schaller, L. Schaupp, T.R. Pieber and R. Hovorka. Insulin kinetics in type-1 diabetes: continuous and bolus delivery of rapid acting insulin. *IEEE Trans. Biomed. Eng*. 2005; 52(1):3-12.

J. Wilson. The Development of the Studio Classroom. In *Technology-Enhanced Learning: Opportunities for Change*. P.S. Goodan (ed.), Lawrence Erlbaum Assoc. Mahwah, NJ (2002). Preprint: http://www.jackmwilson.net/ArticlesTalks/Studio2000.pdf. (accessed 7 Feb 18).

References: Selected Process Control Textbooks

P.C. Chao. Process Control. A First Course with MATLAB. *Cambridge University Press* (2002).

D.R. Coughanowr and L.B. Koppel. *Process Systems Analysis and Control*. McGraw-Hill, New York (1965).

D.R. Coughanowr and S.E. LeBlanc, *Process Systems Analysis and Control*, 3rd ed., McGraw-Hill, New York (2009).

M.L. Luyben and W.L. Luyben, *Essentials of Process Control*, McGraw-Hill, New York (1997).

W.L. Luyben, *Process Modeling Simulation and Control for Chemical Engineers*, 2nd ed., McGraw-Hill, New York (1990).

T.E. Marlin. *Process Control: Designing Processes and Control Systems for Dynamic Performance*, 2nd ed., McGraw-Hill, New York (2000).

B.A. Ogunnaike and W.H. Ray, *Process Dynamics, Modeling and Control*, Oxford, NY (1994).

J.B. Riggs and M.N. Karim, *Chemical and Bio-Process Control*, 3rd ed. Ferret Publishing, Lubbock, Texas (2007).

J.A. Romagnoli and A. Palazoglu. *Introduction to Process Control*. 2nd ed. CRC Press (2012). ISBN 9781439854860

D.E. Seborg, T.F. Edgar, D.A. Mellichamp and F.J. Doyle III, *Process Dynamics and Control*, 4th ed., Wiley, New York (2016).

W.S. Seames. *Designing Controls for the Process Industries*. CRC Press, 2017.

C.A. Smith and A.B. Corripio, *Principles and Practice of Automatic Process Control*, 3rd ed. Wiley, New York (2005).

C.A. Smith. *Practical Process Control: Tuning and Troubleshooting*. Wiley (2009).

C.A. Smith. Advanced Process Control: Beyond Single Loop Control. Wiley (2009).

Stephanopoulos, G., *Chemical Process Control*, Prentice Hall, Englewood Cliffs, NJ (1984).

Svrcek, W.Y., D.P. Mahoney, and B.R. Young, *A Real-Time Approach to Process Control*, 3rd ed., Wiley, Chichester (2014).

P. Woolf (ed.). Open Textbook: *University of Michigan Chemical Engineering Process Control*. https://open.umich.edu/find/open-educational-resources/engineering/che-466-process-dynamics-controls. (accessed 7 February 2018).

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50008-2

Reinforcement Learning – Overview of Recent Progress and Implications for Process Control

Thomas A. Badgwell[a], Jay H. Lee[b], Kuang-Hung Liu[a]

[a]*ExxonMobil Research & Engineering Company, Clinton, NJ, USA, thomas.a.badgwell@exxonmobil.com*
[b]*Korea Advanced Institute of Science and Technology, Daejeon, Korea jayhlee@kaist.ac.kr*

Abstract

This paper provides a brief introduction to Reinforcement Learning (RL) technology, summarizes recent developments in this area, and discusses their potential implications for the field of process control. The paper begins with a brief introduction to RL, a machine learning technology that allows an agent to learn, through trial and error, the best way to accomplish a task. We then highlight two new developments in RL that have led to the recent wave of applications and media interest. A comparison of the key features of RL and Model Predictive Control (MPC) is then presented in order to clarify their similarities and differences. This is followed by an assessment of five ways that RL technology can potentially be used in process control applications. A final section summarizes our conclusions and lists directions for future RL research that may improve its relevance for process control applications.

Keywords: Reinforcement Learning, Model Predictive Control, Process Control.

1. Introduction

In March 2016, a computer program named AlphaGo defeated Lee Sedol, an 18 time world champion, four games to one, at the ancient Chinese game of Go, generating widespread media attention. In May 2017, a new and improved version called AlphaGo Master soundly defeated Ke Jie, the reigning world No. 1 ranked player, three games to none. It is said that Google DeepMind was already in possession of AlphaGo Zero, a version much stronger than the Master version, at the time (Silver et al., 2017). Go was invented more than 2,500 years ago and is believed to be the oldest board game widely played today. Despite its relatively simple rules, Go is significantly more complex than chess in terms of combinatorial possibilities and defeating the best human players had been a goal of Artificial Intelligence researchers for many years. AlphaGo owes its success to recent developments in Machine Learning (ML), specifically to a branch of ML known as Reinforcement Learning (RL) (Sutton and Barto, 2016). Other notable successes of RL include manoeuvring self-driving cars, soccer robots, and optimizing the operation of a data center (Knight, 2017). One can find many interesting demonstrations of Reinforcement Learning on YouTube, including a robot that learns to defend a soccer goal and one that learns how to flip pancakes.

Given the recent successes of RL technology, and given the similarities between RL tasks and process control problems, it is natural to ask if there are any implications of these developments for the field of process control. This paper attempts to address this question, building on a prior assessment of Machine Learning presented at the

FOCAPO/CPC 2016 meeting by Lee et al. (2016). The paper begins with introduction to RL, a machine learning technology that allows an agent to learn, through trial and error, the best way to accomplish a task. We then highlight two new developments in RL that have led to the recent wave of applications and media interest. A comparison of the key features of RL and Model Predictive Control (MPC) is then presented in order to clarify their similarities and differences. This is followed by an assessment of five ways that RL technology can potentially be used in process control applications. A final section summarizes our conclusions and lists directions for future RL research that may improve its relevance for process control applications.

2. Reinforcement Learning

Machine Learning is customarily divided into three classes of algorithms: Supervised Learning (SL), Unsupervised Learning (UL), and Reinforcement Learning (RL) (Bishop, 2012). In Supervised Learning, an agent is provided with labelled examples (X and Y) and learns a mapping that allows it to predict values of Y for new values of X. For example, an agent may be provided with a number of pictures of animals, each labelled properly with the types of animals in the picture. The agent then learns a mapping that allows it to determine, for example, which of the pictures contain a particular type of animal. In Unsupervised Learning the agent is provided with unlabelled data (X only) and learns about how the data is distributed ($p(X)$). For example, the agent may be able to group pictures according to the types of animals they contain. Reinforcement Learning differs significantly from both Supervised and Unsupervised Learning. A Reinforcement Learning agent has the goal of learning the best way to accomplish a task through repeated interactions with its environment (Sutton and Barto, 2016). In order to accomplish this the agent must evaluate the long-term value of the actions that it takes.

Of course the concept of teaching a computer to accomplish a task has been around since the earliest days of the computer age. Alan Turing, for example, wrote the following in his remarkable paper on computing and intelligence (Turing, 1950):

> *Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's? If this were then subjected to an appropriate course of education, one would obtain the adult brain…We normally associate punishments and rewards with the teaching process. Some simple child machines can be constructed or programmed on this sort of principle…I have done some experiments with one such child machine, and succeeded in teaching it a few things, but the teaching method was too unorthodox to be considered really successful.*

The development of RL from such early concepts is well documented in what is probably the best overall introduction to the field, the classic text of Sutton and Barto (2016). Two essential threads of research contributed to the modern theory of RL: Animal Psychology contributed key concepts related to learning a task by trial and error, and Optimal Control Theory provided guidance on how to improve task performance as well as how to know when the task was completed in the best possible way.

2.1 Reinforcement Learning Basics

The basic setting for RL, as abstracted from Animal Psychology, is illustrated in Figure 1. In this setting there are only two entities, the Agent and the Environment. At time t the Agent takes an action A_t that affects the environment, causing it to transition from

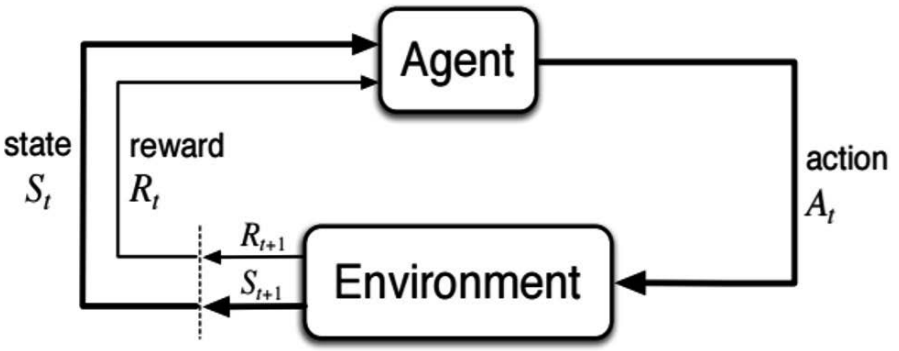


Figure 1 - Basic setting for Reinforcement Learning

state S_t to state S_{t+1}. As a result of the action A_t, and the subsequent transition from state S_t to state S_{t+1}, an immediate reward R_{t+1} is generated. The Agent then uses the state information S_{t+1} and immediate reward R_{t+1} to generate the next action A_{t+1}, continuing the cycle. The goal of the Agent is to learn a mapping from states to actions called a policy $\pi(A_t = a|S_t = a)$ that maximizes a *long-term* sum of future rewards called a value function $v_\pi(s)$:

$$v_\pi(s) = E\{R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} \cdots |S_t = s\} \quad (1)$$

Here $\gamma \in [0,1]$ is a discount factor that determines the importance of future rewards. When $\gamma = 0$ the Agent acts in a greedy manner, maximizing the reward at the next time step. When $\gamma = 1$ the Agent will place equal importance on all future rewards. Note that every value function $v_\pi(s)$ is associated with a particular policy π that determines the future path through the space of states.

RL algorithms can be thought of as iterative updates that improve the policy so that the associated value function increases for all states. If the Agent modifies the policy appropriately at each iteration, then the policy continues to improve, leading to larger and larger values for the value function at any given state. It is natural to ask if the value function ever attains its best possible, or optimal value, $v_*(s)$. Optimal Control Theory answers this question in the form of Bellman's optimality equation, a condition which must be true for all states if the policy is optimal (necessary condition):

$$v_*(s) = \max_a \sum_{s',r} p(s', r|s, a)\,[r + \gamma v_*(s')] \quad (2)$$

Here the term $p(s', r|s, a)$ is the transition probability for the environment, known as the transition model. It provides, given the current state s and action a, the probability that the environment will transition to state s' with reward r. For a discrete set of actions and states, and for a given process model, Bellman's optimality equation provides a set of nonlinear equations that can, in principle, be solved directly to obtain the optimal value function $v_*(s)$. Once the optimal value function is known, the associated optimal policy $\pi_*(s)$ can be found using the transition model. In practice, however, these equations usually cannot be solved directly because we either do not know the environment model, or the state dimension is so large (possibly infinite) that the solution cannot be found in a

reasonable time. Sutton and Barto consider all RL algorithms as approximate solutions to Bellman's optimality equation, dealing in various ways with these two limitations (Sutton & Barto, 2016).

For continuous states and actions, the most relevant case for process control applications, the state and action dimensions are infinite, requiring that function approximation methods be used to estimate the optimal value function $v_*(s)$ and policy $\pi_*(s)$. Note that while neural networks are often used for this, other types of function approximations can also be used. Let us assume that the policy function is parameterized by parameters θ:

$$\pi(a|s) \cong \pi_\theta(a|s) \tag{3}$$

Sergey Levine (2017) explains that for this case the goal of the RL algorithm is to solve the following optimization problem:

$$\theta_* = \underset{\theta}{argmax} E_{\tau \sim p_\theta(\tau)} \left[\sum_t r(s_t, a_t) \right] \tag{4}$$

This says that the parameters θ are to be chosen so that the resulting policy maximizes the expected value of future rewards. Here $r(s_t, a_t)$ is a reward function and the expectation is over the sequence τ of future actions and states, distributed according the probability $p_\theta(\tau)$:

$$\tau = (s_1, a_1, s_2, a_2, \ldots.) \tag{5}$$

From this perspective, Levine (2017) divides RL algorithms into four main classes:

- Model-Based (MB): estimates the transition model $p(s', r|s, a)$ and then uses this directly for control, to estimate the value function, or to estimate the policy.
- Value-Based (VB): estimates the so called Q function $Q_*(s, a)$ which combines the transition model and the optimal value function in (2), then determines the policy that selects the most beneficial (greedy) action from the current state.
- Policy-Gradient (PG): estimates the value function by summing observed rewards $r(s_{t,} a_{t,})$, then uses the gradient of the objective in (4) to improve the policy by gradient ascent
- Actor-Critic(AC): estimates optimal value function $v_*(s)$ in (2), uses this to predict future rewards $r(s_{t,} a_{t,})$, then uses gradient of objective in (4) to improve the policy by gradient ascent (combines best features of Policy-Gradient and Value-Based)

The last three of these are often referred to as model-free, since they do not require an estimate of the transition model. We will emphasize model-free RL methods here so as to provide a clear contrast with MPC techniques that are widely used in process control today.

The effectiveness of these algorithms on any particular problem may vary widely, due to a number of trade-offs (Levine, 2017):

- Sample efficiency: how many interactions with the environment are required to achieve good convergence?

- Stability and ease of use: how easy is it to choose the algorithm's meta-parameters so as to get reliable convergence?
- Stochastic versus deterministic: does the algorithm work better for stochastic or deterministic environments (and/or policies)?
- Continuous versus discrete: is the algorithm better suited for continuous or discrete states, actions, and/or rewards?
- Episodic or infinite horizon: is the algorithm better suited for episodic or infinite time environments?
- Parallelization: Can the algorithm be parallelized easily when a good simulation of the environment is available?

For example, PG algorithms typically require significantly more interactions with the environment than the other methods (low sample efficiency). However, the actual wall clock time to convergence may be lower if a good simulator is available, since PG algorithms are highly parallelizable. Most of the recent developments in RL theory and application are based on AC algorithms that use Deep Neural Networks (DNN) for function approximation. AC algorithms, first introduced by Witten (1977) and later popularized by Bart, Sutton, and Anderson (1983), combine some of the best features of VB and PG algorithms, resulting in faster, more stable learning.

2.2 Latest developments

We now highlight two relatively recent developments in RL technology that have enabled a significant expansion in the scope and capability of the technology: Deep Learning and Parallelization.

Deep Learning refers to a neural network architecture with a relatively large number of layers, in which the first several layers mostly extract significant features from the data (X), and the last layers are mostly engaged in correlating the extracted features to outputs (Y). This type of architecture, shown in Figure 2, is often referred to as a Deep Neural Network (DNN). Lee et al. (2017) provide more details on the development of Deep Learning technology. The use of DNNs to approximate the value and policy functions is a key factor in the many of the recent RL success stories (Silver, et al., 2016). The ability of a DNN to extract useful features significantly reduces the time and effort required for feature engineering. In the AlphaGo application, for example, the DNNs were able to use the raw 19x19 matrix of board positions to directly deduce the game state (Silver, et al., 2016). One implication for process control applications is that it may be possible to pass all available process measurements to a DNN and allow it to predict future process outputs directly, without having to design and implement a state estimator. This would greatly simplify the design of new control systems, and may provide a straightforward path to improving existing control systems.

A second recent development is the use of parallelization to speed up and stabilize the learning process, which might otherwise be a significant barrier for process control applications. A recent paper from Mnih, et al. at Google DeepMind (2016) discusses the advantages of this. Previous work had found that combining simple online RL algorithms with DNNs led to learning stability problems in that the DNN weights would not converge in a stable manner. This was thought to be due to the correlated nature of the updates that the RL algorithms received online. One popular solution to this problem is to store the

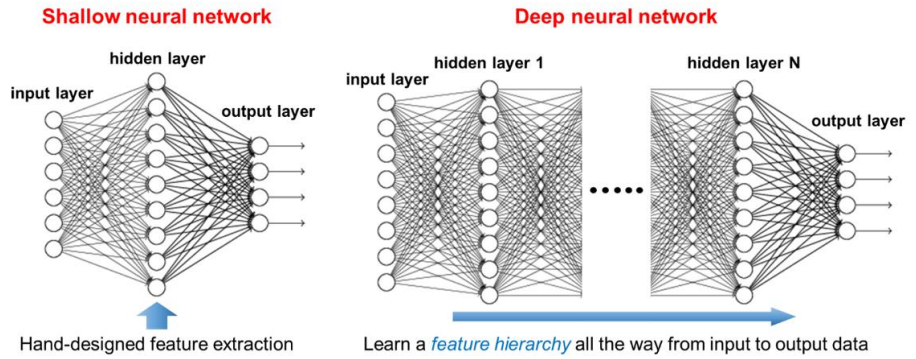


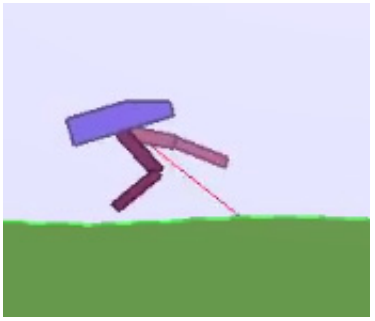
Figure 2 – Comparison of the structure of a Shallow neural network and a Deep Neural Network, from (Lee, at al., 2017)

online data in an experience replay memory, where they can be batched or randomly sampled to train the DNN. Mnih et al. (2016) present an alternative method in which a number of distributed agents execute in parallel on identical copies of the environment, each progressing on its own trajectory through the state space due to the use of different exploration policies. Gradients computed locally by the distributed agents are periodically transmitted to a master agent, where they are used to update a master policy. The master policy is, in turn, transmitted periodically to the distributed agents. There are two significant benefits to this approach. The first is that the learning rate increases dramatically, roughly in proportion to the number of distributed agents. The second is that learning is more stable because the gradient updates are uncorrelated. This means that there is no requirement for an experience replay memory. They demonstrate superior performance of their Asynchronous, Advantage, Actor-Critic (A3C) algorithm on a variety of games and control tasks that include both continuous and discrete actions. They conclude that their A3C algorithm is the most general and successful RL algorithm to date. Of course, all of this depends upon the existence of a quickly executing process simulator of sufficient accuracy to be useful.

An extension of this concept, which may prove useful in robust control applications, is to run a number of agents in parallel on slightly different environments, each corresponding to a possible realization of an uncertain process. The knowledge gained in each environment can be assembled in the master agent in the same way. The resulting master policy would be expected to produce more cautious, robust actions.

2.3 An example RL application – OpenAI Gym Bipedal Robot

We now present a simulated bipedal robot example. The bipedal robot environment comes from the OpenAI Gym, an open-source toolkit for comparing reinforcement learning algorithms. Figure 3 illustrates the bipedal robot, which consists of a head (also called a hull) and two legs. Each leg has two joints – at the hip and the knee. Twenty-four continuous states are measured, without error, to determine the dynamic state of the robot. These include position, speed, and angular velocity of the robot components, a sensor that determines if the legs are in contact with the ground, and ten Lidar distance measurements. The four continuous actions are torques applied to each joint. The reward



24 Continuous States	= [hull angle speed, angular velocity, horizontal speed, vertical speed, position of joints, joints angular speed, legs contact with ground, 10 LIDAR rangefinder measurements]
4 Continuous Actions	= [leg joint torques]
Reward	+x: Every time it moves forward. - 0.1: Apply leg motor torque. -100: Fall down.

Figure 3 - Bipedal robot example – states, actions, and rewards

is how many units of forward motion have been achieved with a small penalty for each action (torque) and a large penalty for falling down. The environment is episodic in that each episode begins with the robot standing at the edge of a field, and the episode ends when the robot either falls down or reaches the other edge of the field.

A Deep RL AC algorithm was implemented for this example. The results at various stages of learning are illustrated below in Figure 4. After 3k episodes the robot is able to make some forward progress but its rear leg is folded beneath it much of the time. After 40k episodes it is able to move forward much more rapidly. Further training leads to a very surprising policy, in which the robot hops on one leg, using the other leg as a tail to maintain its balance. This happens because the reward signal penalizes movement of the joints, so the robot has found a way to optimize its overall reward by decreasing its forward speed and moving its legs less vigorously. This points out an interesting feature of RL technology – RL agents can sometimes produce actions that a human control designer would never think of. The AlphaGo agent exhibited similar behaviour in that it sometimes made moves that no experienced Go player would consider. This can be viewed as either a problem or a learning opportunity depending on your point of view.

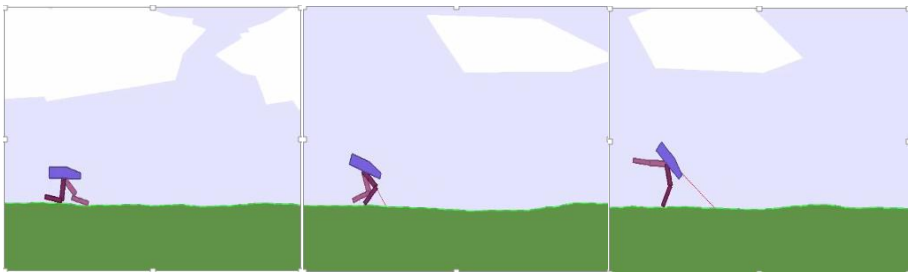


Figure 4 – Results for the bipedal robot using an AC algorithm. On the left after 3k episodes, in the middle after 40k episodes, and on the right after additional training.

Figure 5 presents results from a more sophisticated bipedal robot that learns to traverse a more challenging obstacle course consisting of steps, walls, and pits. This robot uses a parallel AC algorithm similar to that presented by Mnih et al. (2016) discussed above so

that it can learn from a number of environments simultaneously. Each of distributed environments contains the same types of features, but their number and location vary randomly. After sufficient training the robot is able to traverse the obstacles easily, occasionally raising its rear leg in a tail-like manner to restore its balance.

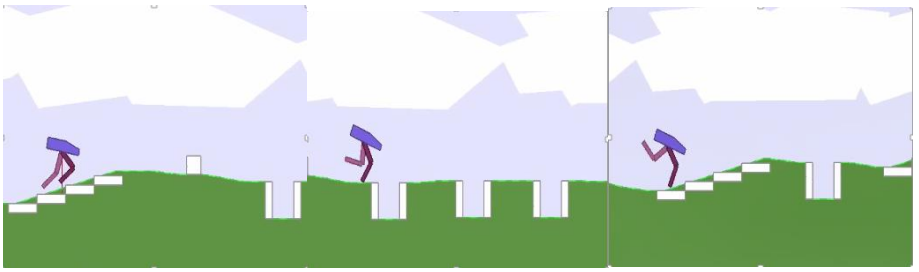


Figure 5 – Results for a bipedal robot using a parallel AC algorithm to traverse a more complex environment consisting of steps, walls, and pits.

3. Comparison of Reinforcement Learning and Model Predictive Control

In order to assess the strengths and weaknesses of RL for process control applications it is instructive to compare RL with MPC, the most widely-used technology developed by the process control community for complex, multivariable control applications (Qin and Badgwell, 2003). In doing so we consider a model-free, on-policy RL algorithm with continuous states and decisions applied to a game, since this is the type of application that has driven much of the recent effort in RL theory and applications. And we consider MPC with a fixed discrete time state-space model, since this is the form that has arguably driven most of the recent developments for this technology. With this in mind, Table 1 provides a comparison of the salient features of these technologies.

The two technologies make somewhat different assumptions about the underlying system. For MPC the underlying system is typically assumed to be a discrete-time state-space system of the form:

$$x_{k+1} = f(x_k, u_k, w_k)\ ;\ \ y_k = g(x_k, v_k) \tag{6}$$

Here x_k is the system state at time k, u_k is the input (action), and y_k is the measured output. The term w_k is a stochastic state disturbance, and the term v_k represents measurement noise. This model form describes how the state traverses through the state space, given a sequence of actions, state disturbances, and measurement noises. RL focuses instead on how the probability of being in a particular state changes from one time step to the next. This is described by a Markov Decision Process (MDP) with the following components (Levine, 2017):

- S - state space, states $s \in S$ (discrete or continuous)
- A - action space, actions $a \in A$ (discrete or continuous)
- T - transition operator, $p(s_{k+1}|s_k, a_k)$
- $r(s, a)$ - reward function, $r: SxA \rightarrow \mathbb{R}$

So basically, a MDP is an unstructured model in discrete state space, describing how the probability of being in a particular state changes over time. It is possible, in principle, to convert the discrete-time state-space model (6) into a MDP by sampling the state transitions. However model-free RL does not require a system model of the form in (6). So model-free RL has the advantage that no explicit system model is required. It also works naturally in a stochastic system environment. However, it has the disadvantage that the relevant system behaviour must then be learned by trial and error.

Table 1: A comparison of Reinforcement Learning and Model Predictive Control

	Reinforcement Learning (model-free)	**Model Predictive Control**
Underlying System Assumption	Markov Decision Process	Discrete-time state space system
Goal(s)	Simple, fixed goal: Just win (don't care about path through state-space)	Complex goal that changes with time: Estimate system state, determine best steady-state consistent with current constraint and setpoint values, follow smooth path consistent with constraints through state-space to best steady-state (smooth, stable path through state-space is critical)
Modeled component	Value function or Policy	Discrete-time state space system
Model learning paradigm	Adaptive model learned through trial and error either offline using simulation or online with real process	Fixed model developed through first principles and/or process identification experiments
Exploration/Exploitation	Simultaneous	Exploration first to get model/Exploitation thereafter
Feedback	Adapting value function or Policy	Disturbance estimation
Online execution time	Very fast - just a forward run of the policy network	Slow - potentially solving three optimization problems at each execution
Stability	Closed-loop stability is not considered	Closed-loop stability is essential.
State Constraints	State constraint enforcement is not considered	State constraint enforcement is essential
Failure tolerance	Failure is necessary for learning (simulation can be used)	Failure cannot be tolerated
Sweet-spot	Accomplishing a task (winning a game)	Keeping a stationary multivariable process within constraints and close to setpoints in the face of measurement noise and process disturbances

The typical goals of RL and MPC algorithms are also quite different. The goal of a RL application is often to complete a task, such as winning a game, where you do not care how the algorithm traverses the state space or how long it takes to win. In the AlphaGo application, for example, the agent does not care how long it takes to win, or how many of its stones get taken. All that matters is to win the game. And the goal of winning does not change during the game. In contrast, an MPC application cares very much about how it traverses the state-space. It is preferable to follow a short, smooth path from the current state to the desired steady-state, and state constraints must not be violated along the way. And the goal for the MPC algorithm can change frequently, in the worst case at every sample, as the operator changes setpoints or constraint limits, and as disturbances perturb the process. However, these distinctions should not be interpreted as fundamental limitations as each method can, in principle, be formulated to address problem type of the other. For example, an RL agent can carry a general stage-wise reward function, which can be used to shape or constrain the transient movement in the state space. And the RL agent can be trained using data that includes many setpoint changes, constraint limit changes, and disturbances. However this additional required training may be very challenging in practice.

For RL the "model-free" label is a bit misleading since these algorithms must still build up models of the value function and/or the policy. The value function for RL is similar to the objective function for MPC with the system model and control policy substituted into it (with a negative sign). So in this respect RL is like MPC where one builds a model of the "closed-loop" objective function rather than of the process itself.

RL learns the value and policy functions primarily through trial and error experiments. If a high-quality simulation is available, which is often the case for games, then one can initialize the value and policy function estimates by performing these experiments on the simulation. In the context of game playing where the environment follows a same set of rules as the agent, one can even train the agent by having it play copies of itself (Silver, et al., 2017). Once the agent is online, it typically spends most of its time exploiting the models to improve the average reward, with the remaining time dedicated to random actions that allow it to update the value and policy function models if necessary. This balance between exploitation and exploration allows the value and policy function models to adapt as the system changes, and is a fundamental aspect of RL technology. So for RL, model prediction error due to time varying nature of the problem is ultimately addressed (feedback) by adapting the value and policy function models.

With MPC the system model is typically developed through some combination of first principles theoretical arguments (e.g. mass and energy balances) and carefully chosen process identification experiments on the real system. It is assumed that the process will not change significantly in the near future so there is no need to update the system model, at least not on a continuous basis. This can be viewed as exploration first (through plant testing), followed by exploitation thereafter (through control with the identified model). Prediction error is integrated away by estimating disturbances that can enter the process through some combination of input, state, or output channels. This would seem to provide an advantage over RL technology for those cases where the core process dynamics do not change significantly and disturbance patterns remain the same.

The RL agent typically executes very quickly online since this involves only a forward run of the policy network. In contrast, MPC may take much longer to execute since it involves solving three optimization problems sequentially (state estimation, steady-state

target calculation, dynamic optimization). This is necessary in MPC because the control problem may change significantly from one time step to the next.

Closed-loop stability is a concept that does not receive much attention in the RL literature, probably because the emphasis is on batch tasks rather than continuous tasks. For MPC applications, closed-loop stability is essential and is now well-characterized following decades of study (Rawlings et al., 2017). Characterizing closed-loop stability for RL algorithms would seem to be a good research opportunity.

State constraint enforcement is something that is critical to MPC applications but does not seem be addressed by RL technology. When operating a fired heater, for example, it is critical to ensure that the tubeskin temperatures in the firebox remain below a critical value. With RL, by contrast, it does not matter where the agent goes on the board as long as it wins the game. As said before, this is not necessarily a fundamental limitation of RL as stage-wise reward or penalty function can be used to shape the transient performance. Working through the details of how to implement state constraints in RL algorithms would seem to be another good research opportunity.

The tolerance for failure in RL is also quite different that for MPC. Since much of RL learning is accomplished by trial and error using random actions, the system must be allowed to fail occasionally, at least during training, where failure is defined as the lowest possible reward. For MPC, failure in this sense cannot be tolerated. For RL this could be mitigated in a process control scenario in three ways. First a high quality simulation could be used initially to train the RL agent. Second, an imminent failure online could be detected through some form of feature engineering, generating an automatic action to restore the system to a better state. Third, the reward function can be designed to reflect transient constraints on the value function or policy learned, but a careful balance between exploration and exploitation should maintained, esp. during the initial learning period.

To summarize this comparison, RL technology appears to be best suited for batch tasks with a simple-fixed goal, carried out on a potentially non-stationary stochastic system, for which a high-quality simulation is available, and for which state and time constraints are not important. In contrast, MPC seems best applied to continuous tasks with complex, time-varying goals, carried out on stationary systems, for which high-quality simulations may not be available, and for which state and time constraints are very important. However, these distinctions are not necessarily fundamental and only reflect the types of applications each has targeted. Even though RL and MPC have been targeted at very different control problems thus far, they are in some ways complementary, and this opens the door for using RL, or at least some of its aspects, for process control applications.

4. Implications for Process Control

In general, a RL agent has the potential to perform any task that requires knowledge and experience gained by interacting with the process. However, as we have seen, this comes at the cost of requiring significant interaction (training) with the process, or at least with a high fidelity process simulation. Nevertheless there are several ways that RL technology may have an impact on the theory and practice of process control. Here we discuss five broad classes of opportunities.

4.1 Directly replace existing process control technology with RL

The most straightforward use of RL in process control involves using it to directly replace existing control technology. On the other hand, since RL technology was developed to

solve a very different class of problems, we view direct replacement to be the least likely path for RL to have an impact on process control. A much more likely way for RL to have an impact will be in combination with conventional control technology such as MPC, or when applied to different, but related problems where MPC would have some difficulty.

4.2 Integrate aspects of RL technology with MPC

Several authors, including some from the process control community, have discussed how RL can be used to complement existing MPC technology. These proposals make use of RL technology in some way to address weaknesses of MPC. Here we summarize three such proposals.

Lee and Wong (2010) point out that MPC has two inherent limitations. First is the high cost of online computation, which scales with the state dimension as well as the horizon length. The second inherent limitation is that the use of an open-loop dynamic optimization means that MPC cannot make use of information about future uncertainty as it is revealed. These limitations are particularly relevant in the case of nonlinear stochastic control problems for which MPC can be highly suboptimal. Lee and Wong propose to mitigate or remove these limitations through the use of Approximate Dynamic Programming (ADP), another name for the class of technology that we have labelled as RL. They propose several possibilities for integration of ADP with MPC. One is to use MPC in the initial stages of simulation sampling in order to determine likely state sequences and state space regions to focus on. A second method is to use the learned cost-to-go (negative of the value function) as the terminal penalty in a MPC formulation in order to reflect the effect of future uncertainties in closed loop and reduce the depth of prediction and optimization needed. A third way is to alternate between the MPC and ADP algorithms, switching to MPC when a new region of the state space is encountered in which the ADP algorithm cannot be trusted.

Morinelly and Ydistie (2016) present an interesting Dual MPC algorithm in which RL is used to introduce the effect of anticipated information. The algorithm is developed for the simple case of a discrete linear system with uncertain dynamics. They demonstrate that the algorithm converges to the optimal linear state feedback policy (Linear Quadratic Regulator) in this case. A simulation example shows good behaviour for the proposed algorithm, which they call RL dual MPC (RLdMPC). This proposal seems quite promising, provided that it can be generalized to full multivariable systems.

Kamthe and Deisenroth (2017) present an algorithm that combines data-efficient RL with probabilistic MPC. Their algorithm improves the data efficiency of RL while simultaneously ensuring the model uncertainty is properly accounted for in the MPC dynamic optimization. They accomplish this by learning a probabilistic transition model using Gaussian Processes (GPs) to incorporate model uncertainties into long-term predictions. This reduces the impact of model errors and results in a RL framework that makes maximum use of available data. Long-term predictions are obtained from the GP model by iterating forward using a deterministic moment-matching Gaussian approximation. Pontryagin's Maximum Principle is then used to solve the MPC dynamic optimization subject to constraints. They show remarkably efficient learning for two standard tasks – the under-actuated cart-pole swing-up and the fully-actuated double-pendulum swing-up - both of which are solved in less than 15 encounters.

We view these suggestions for integrating RL with MPC as very encouraging and worthy of further investigation.

4.3 Use RL to help manage process control systems – Proportional-Integral-Derivative (PID) controller tuning, MPC model gain adjustments, etc.

Process control systems require significant management by operators and control engineers in the form of adjusting limits and tuning parameters as process conditions change. When the feed to a unit is cut by 20%, for example, an experienced control engineer will know which PID and MPC controllers will need to be adjusted and will have a good idea for the types of adjustments that will be required. This type of application seems ideally suited for RL technology, if good process simulators or sufficient amounts of historical data are available.

4.4 Using RL to advise operators during unusual situations (process upsets/start-up/shutdown)

Operating a chemical process is similar in many ways to flying an airplane in that there are long periods of normal, quiet operation punctuated by brief periods of extreme tension when things go wrong. During a process upset, an operator will immediately attempt to bring in additional experienced operators to help diagnose and mitigate the problem. In cases requiring an immediate shutdown, the operator will typically consult a large manual of written instructions that carries with it an implicit assumption of the shutdown scenarios most likely to be encountered. In these situations, RL technology might be particularly helpful as a tool to advise the operator. The RL agent can be trained on the same process simulators that are employed to train the operators, but it can base its recommendations on many thousands of randomly generated scenarios, some of which may not have been anticipated before.

4.5 Use a hierarchical network of RL agents to simplify operation of a chemical plant.

In the not too distant future one can imagine a hierarchical network of RL agents that learn how to perform most of the tasks that human operators currently perform while operating a chemical plant. The hierarchies would reflect the operational priorities of the company operating the plant. These could be, for example:

1. Process Safety
2. Environmental Compliance
3. Reliability
4. Economics

Agents at the same level would cooperate in a distributed manner to optimize the overall goals at that level. Agents at different levels would act according to set priorities. For example, the actions of an economics agent on a particular piece of equipment could be interrupted at any time by a safety agent that has detected a problem and needs to shut down the equipment right away.

There has been significant work in the area of RL agent networks, by Foerster, et al. (2016), for example, but nothing that seems appropriate for off-the-shelf use on this problem.

One should be skeptical, however, of a proposal to replace *all* of the operators in a control room. The development of automation technology such as RL is more likely to lead to a

significant reduction in the need for human intervention, rather than the elimination of all operators in the control room. The march of automation for commercial aircraft has, for example, reduced the number of required cockpit crewmembers from four in the early days to two today. But it is unlikely that passengers will step onto an airliner with no cockpit anytime soon.

5. Future research directions

From our analysis there are several opportunities for process control researchers to have an impact on RL theory. These ideas are aimed augmenting RL technology to make it more appropriate for process control applications:

- Prove algorithmic convergence (value function and policy function approximations converge to optimal value and policy function) for a state-of-the-art RL algorithm such as A3C
- Prove closed-loop stability for the case of a state-of-the-art RL algorithm such as A3C acting on a nonlinear dynamic system
- Augment a state-of-art RL algorithm such as A3C so that it enforces state constraints
- Investigate additional ways to integrate RL and MPC, building on work already published
- Develop robust control technology based on an A3C algorithm in which each parallel environment represents a realization of the system uncertainty.
- Develop technology that allows a hierarchy of prioritized RL agents to accomplish a complex task similar to operation of a chemical plant.

6. Conclusions

In general, a RL agent has the potential to perform any task that requires knowledge and experience gained by interacting directly with the process. Recent technical developments, including deep learning and parallelization, greatly increase the potential scope and capability of RL. Five broad classes of potential process control applications for RL technology have been assessed, and several research directions aimed at improving the relevance of RL for process control applications have been suggested. In summary, we believe that RL technology has the potential to significantly impact both theory and applications in the field of process control.

Acknowledgements

The authors wish to thank Michael Kovalski of the ExxonMobil Technical Computing Company for his assistance in implementing the parallel AC algorithm.

References

A. Barto, R. Sutton, C. Anderson, 1983, Neuronlike elements that can solve difficult learning control problems, IEEE Transactions on Systems, Man, and Cybernetics, 13, 835-846.

C. Bishop, 2006, Pattern Recognition and Machine Learning, Springer.

J. Foerster, Y. Assael, N. Frietas, S. Whiteson, 2016, Learning to Communicate with Deep Multi-Agent Reinforcement Learning, 30th Conference on Neural Information Processing Systems (NIPS 2016)

W. Knight, 2017, 10 Breakthrough Technologies: Reinforcement Learning, MIT Technology Review.

J. Lee, W. Wong, 2010, Approximate dynamic programming approach for process control, Journal of Process Control, 20, 1038-1048.

J. Lee, J. Shin, and M. Reallf, 2017, Machine Learning: Overview of the Recent Progress and Implications for the Process Systems Engineering Field, FOCAPO/CPC 2017.

S. Levine, 2017, CS294 Deep Reinforcement Learning, Lecture 3, YouTube.

T. Lillicrap, J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, D. Wierstra, 2016, Continuous Control with Deep Reinforcment Learning, International Conference on Learning Representations (ICLR) 2016

S. Manthe, P. Deisenroth, 2017, Data-Efficient Reinforcement Learning with Probabilistic Model Predictive Control, arXiv:1706.06491v1

J. Morinelly, E. Ydstie, 2016, Dual MPC with Reinforcement Learning, IFAC Papers Online j.ifacol.2016.07.276

V. Mnih, A. Badia, M. Mirza, A. Graves, T. Harley, T. Lillicrap, D. Silver, K. Kavukcuoglu, 2016, Asynchronous Methods for Deep Reinforcement Learning, Proceedings of the 33rd International Conference on Machine Learning, New /York, NY.

S. Qin, T. Badgwell, 2003, A Survey of Industrial Model Predictive Control Technology, Control Engineering Practice, 11/7, 733-764.

J. Rawlings, D. Mayne, M. Diehl, 2017, Model Predictive Control: Theory, Computation, and Design, Knob Hill Publishing.

D. Silver, A, Huang, C. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lancot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, D. Hassabis, 2016, Mastering the game of Go with deep neural networks and tree search, Nature, 529, 484-489.

D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, D. Hassibis, 2017, Mastering the game of Go without human knowledge, Nature, 550, 354-359.

R. Sutton, A. Barto, 2016, Reinforcement Learning: An Introduction, Second Edition, in progress (online).

A. Turing, 1950, Computing machinery and intelligence, Mind, 59, 433-460.

I. Witten, 1977, An adaptive optimal controller for discrete-time Markov environments, Information and Control, 34, 286-295.

This page intentionally left blank

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50009-4

Challenges and Perspectives of Process Systems Engineering in Supply Chain Management

Ana Paula Barbosa-Póvoa[a] and José Mauricio Pinto[b]

[a]*Centre for Management Studies, Instituto Superior Técnico, University of Lisbon, Av. Rovisco Pais, Lisbon, Portugal*
[b]*Business & Supply Chain Optimization, Praxair Digital, Praxair Inc., 10 Riverview Dr., Danbury CT,06810, USA*
apovoa@tecnico.ulisboa.pt, jose_m_pinto@praxair.com

Abstract

Supply chain management is concerned with the efficient integration of interconnected entities so that products are sourced, manufactured and distributed in the right quantities, to the right locations and at the right time, thus satisfying customer requirements at minimal total system cost. Such systems face complexity due to the multiple material and information flows. In this context, the Process System Engineering (PSE) community has an important role to play in addressing such complexity. Departing from a real-world process supply chain, the industrial gas supply chain, while considering existing PSE contributions, this paper discusses challenges in terms or modeling and implementation. This work also presents perspectives in supply chain management from the industrial and academic standpoints that ought to be explored by the PSE community.

Keywords: Supply Chains, Industrial Gases, Uncertainty, Multiscale, Challenges, Perspectives.

1. Introduction

Supply chains (SC) are systems that comprise several entities placed in different geographical locations. These interconnected systems play a crucial role in organizations as they dictate product availability to the final consumer and consequently the financial success of the involved companies. The management of SC is complex due to the involvement of a large multiplicity of material and information flows, diversified characteristics of the associated entities and often-present conflicting objectives.

Within the process industry, this is not an exception and process supply chains must be flexible, resilient and efficient while guaranteeing customers' demands at minimum costs. Moreover, as the process industry is often associated with a high level of resource consumption and often deal with non-friendly environment processes and products, special attention to supply chain sustainability is required when managing such systems (Barbosa-Póvoa, 2014; Mota et al., 2018).

To deal with such complexity and underlying challenges, both industrial and academic communities identify the application of Process Systems Engineering (PSE) tools as a path to follow (Grossmann, 2012). Such tools support the integration of supply chain decisions under uncertainty while minimizing risk, maximizing service levels, sustainability levels and simultaneously guaranteeing the main SC goals – profitable and reliable operation.

The PSE community can contribute highly in the answer to such challenges (Barbosa-Póvoa, 2014) by studying real-world supply chain problems. These span from strategic (e.g. Network Design in Cardoso et al., 2013 and Mitra et al., 2014a) to tactical (e.g. Planning in You and Grossmann, 2013 and Malinowski et al., 2018) and operational decisions (e.g. scheduling, routing and inventory management in Amaro and Barbosa-Póvoa, 2008; Dong et al., 2014; and Zhang et al., 2015).

Several reviews provide a detailed analysis of such challenges. Shah (2005) identified the main requirements of process supply chains, focusing on the pharmaceutical case. Papageorgiou (2009) provided a comprehensive review of the design and planning of process supply chains and Stephanopoulos and Reklaitis (2012) highlighted the past and future contributions of the PSE community in several areas, from which the supply chain study was identified as essential. Grossmann (2012) presented the enterprise–wide optimization concept and highlighted the need of developing optimization tools that explore the integration of decisions across the different operations along the supply chain, considering different levels of detail. Lainez et al. (2012) reviewed the pharmaceutical supply chain and Barbosa-Póvoa (2014) analyzed the published papers on process supply chain considering the multiple decision levels. Recently, Barbosa-Póvoa et al. (2018) reviewed sustainable supply chains and conclude that particularly (bio) process supply chains have been actively addressing sustainability issues, which nevertheless present a large set of research opportunities.

All the above reviews tend to converge to a common set of critical problems, some already under study by the academic community, whereas many others call for further research. In the next section, we illustrate some of the problems in the context of an industrial gas supply chain. Additionally, while some challenges have been addressed, the implementation of PSE in industry is still far from effective. Significant progress must be achieved in critical issues that are discussed in this paper, both from an industrial and an academic perspective. We conclude that supply chain research has a major financial impact to the process industry and brings scientific challenges where the PSE community has an unquestionable role to play.

2. Process Supply Chain

Process Supply Chains deal with chemical and/or bio-based networks that involve a set of entities, materials and information that transform raw materials into final products using different types of transforming processes. Among these supply chains, several examples exist as is the case of the oil, pharmaceutical, biomass, and gas supply chains. Several challenges exist with respect to the management of such systems and consequently both industrial and the academic communities should be committed to address such problems in close collaboration.

We use an industrial gas supply chain as an example of process supply chains where its different characteristics are described and corresponding challenges identified. This is represented in Figure 1. The manufacturing of Oxygen, Nitrogen and Argon is considered, as well as the retail business (hardgoods) and the supply chain comprises upstream, production, distribution and consumer entities.

Electricity can be purchased from utility companies through various power contracts that may differ in price, availability, and penalty for under- or overconsumption (Zhang et al., 2015). Discount prices and penalties can also be defined with respect to the amount of electricity purchased over a certain period of time, which could be hours, days, or even

weeks. In practice, this means that the cumulative electricity purchase must be recorded, and there are pre-defined meter reading times at which the amount of electricity purchased since the last meter reading is computed. Based on this cumulative electricity purchase between consecutive readings, discounts and penalties are issued (Zhang et al., 2016a).

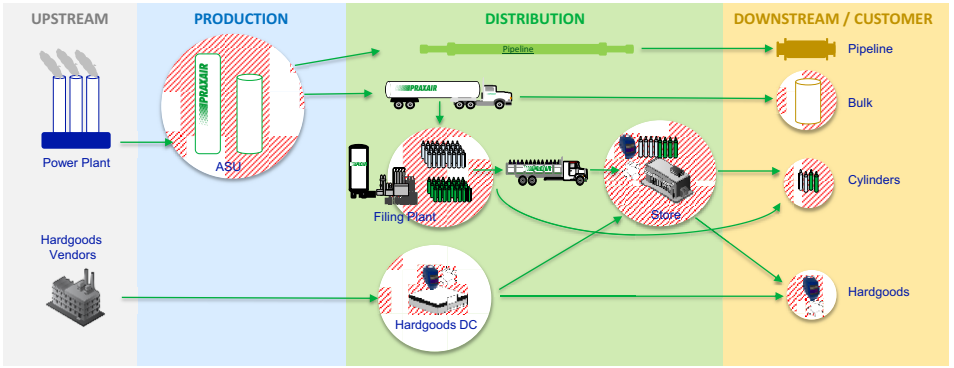


Figure 1- Industrial Gas Supply Chain (Praxair, Inc.)

Cryogenic air separation plants produce liquid oxygen (LO2), nitrogen (LN2) and argon (LAr) as well as gaseous oxygen (GO2) and nitrogen (GN2), all at high purities. Cryogenic process achieve separation through liquefaction followed by low-temperature distillation. All liquid products can be stored on-site in storage tanks. In contrast, gaseous products cannot be stored.

Regarding distribution, industrial gas companies serve customers through three primary distribution modes: large process plants, cryogenic liquid and packaged gases (Megan and Bruton, 2017).

For the largest customers, such as refineries and steel mills, industrial gas plants operate adjacent to their facilities and distribute products via pipeline. These plants act as a vital utility to those customers, similar to electricity and water, and provide an uninterrupted supply of industrial gases to support their operation.

Medium-volume customers, such as hospitals and universities, typically have liquid storage located at their facility, which is then used to provide product throughout their operations. Their inventory is monitored in real time with telemetry and the delivery of product is performed, using a tanker truck fleet, without the customer needing to place an order. This proactive, vendor-managed inventory model provides high reliability for customers while enabling the supplier to manage effectively their costs.

The third mode involves packaged gases. Here a wide variety of smaller packaged gases to laboratories, hospitals and other customers is at stake. These products can range from a cylinder of nitrogen to specialty gas mixtures needed for emissions testing, advanced manufacturing and semiconductor fabrication. As seen in Figure 1, cylinders are either distributed directly to customers or shipped to stores, from which additional deliveries are made or customers pickup their products. A variety of services is also offered, such as embedded regulators and telemetry. The cylinder assets, which helps ensure continuous supply to our customers are owned by the gas company.

In addition to cylinders, stores also retail hard goods that are purchased from vendors and shipped through multiple distribution centers. These in turn either ship products directly to customers or to the stores for pickup or further delivery.

3. Challenges

This section describes the main research and implementation challenges from an industrial perspective as well as these are addressed from the PSE community, including joint collaborations academia-industry.

3.1. Industry perspective

3.1.1. Supply Chain Modeling Challenges

Supply Chain Planning – Planning the merchant liquid supply chain is a very challenging problem. Liquid plants, while all making the same basic set of products, often vary in capacity and efficiency. As merchant liquid customers may receive shipped product from multiple locations, continuously optimizing this supply chain can be challenging. Uncertainties, such as varying customer demand and time-of-day electricity prices, make the system quite dynamic. There is a need for sophisticated forecasting tools and large mixed integer linear programming models in order to determine optimum production and distribution plans on a continuous basis. These planning tools, which should plan over a multi-week time horizon, then must guide the operational tools designed for minute-to-minute optimization of the plants and logistics.

Demand Side Management (DSM) – DSM refers to electric energy management on the consumers' side and encompasses energy efficiency and Demand Response (DR). DR presents challenges and opportunities, primarily the optimization of operational flexibility through the integration of production and energy management. On the strategic level, large industrial electricity consumers often enter into long-term contracts with favorable rates. However, such power contracts require the consumers to commit themselves to the amount that they are going to purchase years in advance when future demand is not yet known with certainty. Hence, there is the need to simultaneously optimize long-term electricity procurement and production planning while considering uncertainty in product demand (Zhang et al., 2018). Regarding mid-term (tactical) decisions, the main challenge is to integrate energy management, production, sourcing and customer-plant allocation (Zhang et al., 2017). This coordination problem gives rise to a multi-scale optimization problem because while a detailed production scheduling model must capture all critical operational constraints on a fine time grid, vehicle routing has to be considered in each time period of a coarser time grid.

Process Modeling - The traditional way of modeling a process involves heat and mass balances, which requires the detailed description of the system's performance (e.g. thermodynamics, kinetics). The disadvantage of this approach is that the model can become prohibitively hard to solve in the context of supply chain optimization due to its nonlinearities and its size. An alternative approach is to build surrogate models in reduced space, e.g. the product space. To determine the feasible region of the plant in the product space, production data can be obtained from extreme operating points or from a sequence of steady-state simulations. Moreover, production modes may represent state of equipment, e.g. "off", "production mode" or "ramp-up transition". Only one mode can be active, in other words the modes are disjoint. The data for each mode is represented as a collection of operating points (slates) that are the extreme points in terms of the products (Mitra et al., 2014a; Zhang et al., 2015).

Inventory Route Planning – Executing a VMI policy in an effective way is nontrivial, because it requires the integration of two components of SC management, inventory control and distribution routing (Dong et al., 2014). In inventory control, the goal is the

determination of orders (time and amount) of customers, while in distribution routing the goal is the generation of schedules to meet these deliveries. The integration of the two problems, which can have a dramatic impact on overall system performance, leads to the inventory routing problem (IRP) which is at the heart of all VMI policies. One of the main objectives in solving the IRP is to reduce the overall distance driven per volume delivered to customers, which has a sustainability impact by reducing fuel emissions.

3.1.2. Challenges in Modeling Approaches

Uncertainty – Uncertainty plays a crucial role in the management of the supply chain. Electricity prices may fluctuate on an hourly basis in certain markets; moreover, to ensure the stability of the power grid, backup capacities are called upon when electricity supply does not meet demand due to unexpected changes in the grid. As part of the demand response efforts in recent years, industrial gas companies are encouraged by financial incentives to provide such operating reserve in the form of load reduction capacities (interruptible load). However, a major challenge lies in the uncertainty that one does not know in advance when load reduction will be requested (Zhang et al., 2016b).

Although industrial gas sites operate at a very high level of availability, there are circumstances in which there are unplanned shutdowns that are caused by electricity supply and/or equipment failure. As mentioned in the previous section, typically plants are directly connected to customer by pipelines to supply gaseous products. Liquid storage tanks are usually installed onsite to be used as a backup when the plant is in outage. Assuming there are data for historical reliability data of individual assets, the main question is how to make design and operations decisions to maximize plant availability. Supply chain management plays an important role in supplying liquid products during plant downtime to guarantee uninterrupted supply to pipeline customers. Hence, the study of supply chain resiliency could bring benefits to the industry.

Besides electricity and plant availability, demand uncertainty plays a pivotal role in production planning & scheduling, inventory routing as well as in long-term capital investments (Mitra et al., 2014b). Figure 2 illustrates product demand uncertainty in short term and long-term decisions, which ultimately affect all operational (production, routing), tactical (safety stock) and strategic (tank sizing, plant capacity) decisions.

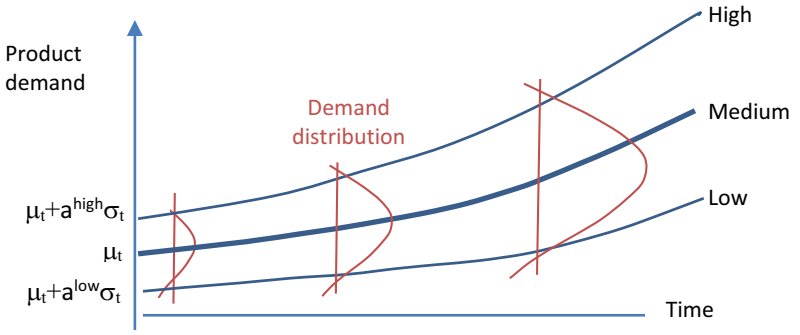


Figure 2. Product demand trajectories in short term and long term (Mitra et al., 2014a). Parameters μ_t and σ_t denote the expected demand and the standard deviation in time t.

Customer demand for packaged gases can also be received dynamically over the planning horizon. Routing plans should flexibly accommodate potential customers who have not

yet called in to request service. Subramanyan et al. (2017) model future potential customers as binary random variables, and seek to determine a visit schedule that remains feasible for all anticipated realizations of service requests.

Finally, in the context of inventory routing for the bulk business in addition to customer demand there are uncertainties with respect to travel times, service times at customers and delivery time windows, i.e., when customers allows access to bulk tanks. These may affect delivery times or even prevent deliveries at end of routes.

Multiscale Modeling – One of the biggest challenges on the design and long-term capacity planning of industrial gas plants is the incorporation of short-term operational decisions. Typically, capacity planning is performed over a 10-15 year horizon. Investigating the trade-off between capital or retrofits, and the operating costs, related to electricity prices, which can vary on an hourly basis, leads to a complex multiscale optimization problem (Mitra et al., 2014a; 2014b). On the mid-term level, there is also the need to coordinate multiple time scales for production (fine) and distribution (coarse), as in Zhang et al. (2017).

3.1.3. Data Challenges

Data Analytics – Industrial gas plants require a high degree of automation and data analytics to ensure that they continuously operate safely, reliably and efficiently (Megan and Bruton, 2017). Examples include condition and predictive monitoring of equipment such as motors and compressors to reduce unplanned downtime. As with large process plants, the vendor managed inventory model for bulk customers requires a variety of analytics, from optimizing assets to scheduling daily deliveries. Packaged gases are a very transaction-intensive business with many distinct products, which leads to many opportunities to use analytics to manage the supply chain, understand margins and better target the sales force.

Big Data Analytics – There has been a paradigm shift in leveraging data for supply chain, particularly external data, which has been experiencing tremendous growth in the past few years. The traditional approach has been to utilize internal data, from few core applications that track the activity of the company and automate various processes. The data is mostly structured and stored in data warehouse(s). Moreover, data is relatively clean and models activity related to customers and processes. With the arrival of external big data, the number of systems and applications may increase by one order of magnitude. Data provided may not be clean or structured; neither will be stored in data warehouses. Models will be a lot more complex as they will relate to both internal and external processes. In addition, a big data strategy must be defined and implemented.

3.1.4. Implementation Challenges

Computational Performance – The complexity of real-world industrial problems translates into large scale models, in terms of constraints and variables. Most SC problems result in mixed-integer linear programming models, which require the investigation of customized solution techniques to obtain good (not necessarily optimal) feasible solutions in reasonable time. Examples in strategic and tactical planning are: a hybrid bi-level decomposition scheme for a two-stage stochastic programming problem with mixed-integer recourse that results from a multiscale capacity planning problem with investment and operational decisions (Mitra et al., 2014b), and Helium SC planning in which a rolling horizon strategy is developed (Malinowski et al., 2018). Short term examples are as follows: IRP for which a dynamic preprocessing algorithm followed by a two-level

decomposition solution method is proposed (Dong et al., 2017), and for the production scheduling of air separation units under uncertainty, an integrated stochastic mixed-integer linear programming model is developed; Conditional value-at-risk is incorporated into the model as a measure of risk, and scenario reduction with multi-cut Benders decomposition are implemented to solve large-scale real-world instances (Zhang et al., 2016b).

Change management – One of the main challenges in the implementation of decision support tools is the management of change with decision makers as well as with the business organization. Unfortunately, the PSE community does not have the training and often oversees the impact of change management in the successful implementation of optimization tools. The change management area has been primarily addressed by the business management community (Cameron and Green, 2015) and ranges from mergers & acquisitions to projects. Although not core to the PSE community, change management should be part of any implementation project. Indeed, there are opportunities to bring design thinking into the area.

3.2 Academic perspective

Process supply chains have been studied by the PSE community with increased focus in recent years. Different challenges have been addressed in line with some of the industrial needs mentioned in the previous section. These span from operational to strategic decision levels and have been covering different types of problems (Barbosa-Póvoa, 2014), but further research is still required.

Multiscale Modeling - at the strategic level, optimal design and planning of supply chains is a well-known problem that, however, continuously faces new challenges. The integration of strategic and tactical decisions is still an area to explore where comprehensive models that account for different supply chain characteristics are required. Uncertainty; sustainability concerns; risk and resilience management should be targeted. Additionally, the availability of large amounts of data is nowadays a reality, which should be explored to lead to more accurate industrial representations and allow further decision-sharing (Ning and You, 2017). The integration of tactical-operational decisions is also an open issue and it has been seldom studied. Supply chain planning and operations appear as a research opportunity, where production planning, inventory management and logistics decisions should be considered simultaneously. Multi-scale supply chain models will help to answer to these challenges where an Enterprise Wide Optimization approach (Grossmann, 2012) should be explored. In this context, not only centralized supply chain decisions, as commonly treated by the academic community, should be considered but rather trade-offs between different supply chain entities need to be accounted for, where decentralized decisions may be at stake (Sahay and Ierapetritou, 2014).

Uncertainty, Risk, Resilience Modeling – uncertainty has been explored mainly using stochastic and robust approaches. The first one focuses on the establishment of representative scenarios to model uncertainty, as the use of a large number of scenarios leads to intractable models. One possible way to tackle this problem is by exploring the use analytics on uncertain parameters data, e.g. in demand and supply (see Figure 2). Statistical, data mining or machine learning techniques, amongst others should be explored to help the definition of more realistic possible scenarios. This has been studied by some authors (Yue and You, 2016; Lima et al., 2018) and constitutes an emerging research area. On the robust approach, research has been focused on how to minimize the conservativeness of the models, but a clear understanding of the problem is still needed.

The modeling of different exogenous and endogenous uncertainties is also a problem to be tackled. Linked to uncertainty arises risk modeling. Different measures have been explored and the authors identify the CVar as the most adequate one (Zhang et al., 2016a). But few works have explored risk when modeling supply chain decisions. With risk comes the need of guaranteeing supply chain resilience (Cardoso et al., 2015), a characteristics critical to supply chains due to the uncertain environment under which they operate. Nevertheless, resilience modeling is not yet well understood and there still exist a long way to trail on this topic.

Sustainability Modeling – increased concern for the management of sustainability in supply chain decisions has been observed in the last decade, often pressed by governmental regulations and societal pressures (Barbosa-Póvoa et al., 2018). This calls for decisions that consider trade-offs between the three sustainability pillars: economic; environmental and social. However, the academic community, when addressing supply chain problems, has been mainly focused on the economic pillar where typically the minimization of costs or the maximization of profit have been considered. But often important aspects have been left out as few models addressed international supply chain characteristics such as international taxes; transfer prices, duties as well as multi-modes and outsourcing options. Regarding the environmental and social pillars, a limited number of works have addressed such concerns (Mota et al, 2018), which are very important when addressing process supply chains. Often process supply chains deal with pollutant and/or hazardous products and/or processes, and a well-defined treatment of such concerns within the supply chain is required. For example, and as mentioned above, the optimization of energy consumption is a constant industrial need, which should be explored not only with an economic objective but also with an environmental concern. Moreover, researchers have been trying to understand the best method to quantify environmental impacts and how this could be used to establish a trade-off with economic concerns leading to a solution of compromise. The LCA method has been one of the main methods studied and along with it different approaches have been explored namely, the Eco-indicator 99, Recipe, and PEF. But comprehensive models that address in a sound form the environmental aspects when designing, planning and operating supply chains are still required. More recently, the social concern has been also studied, but in this case very few works studied this issue. This area is then a great challenge to the PSE community (Barbosa-Póvoa et al., 2018).

Efficient Solution Methods – the treatment of the above identified concerns evidently requires the development of models with increasing complexity. This need calls for new solution methods, where efficient decomposition methods, meta-heuristics, amongst others methods should be explored by the PSE community. Also the solution of non-linear models still needs further treatment (Su et al., 2018).

Finally, the so-called **"Supply Chain 4.0"**is nowadays a reality and the wide availability of data from supply chain process, customer demand, consumption of resources, as well as other associated internal and external activities opens a new challenging research area.

4. Conclusions and Perspectives

Supply Chain optimization presents multiple opportunities for implementation in the process industry and has been the focus of increasing research by the PSE community. However, there are many challenges that still prevent the successful implementation of SC solutions.

In this paper, we describe a real-world, process supply chain through an industrial gas SC whose goal is to illustrate the complexity of the decisions and information involved. Several challenges are discussed that relate to the modeling of entities (energy, production, distribution), the different approaches that result from integrated decision support systems, such as multiscale and mainly uncertainty and its treatment. Finally, we also discuss implementation challenges that result from large scale models that require custom solution methods, as well as the impact of decision support tools that requires the need for the management of change.

The supply chain work available in the literature indicates the need of comprehensive decision support tools required to coordinate cross-functional models. Such tools should allow the treatment of real supply chain characteristics in which different aspects must be considered: the presence of uncertainty - modeling resilience and risk; sustainability goals - accounting simultaneously for economic; environmental and social concerns; international taxes; transfer prices, duties as well as multi-modes and outsourcing options, amongst others. Such challenges lead to further complex, often multiscale models that demand the investment in efficient solution methods.

To conclude we want to raise the need to combine the traditionally PSE models and methods with big data analytics, machine learning, and advanced statistical methods, amongst others, so as to be able to inform the supply chain decision process supporting better decisions. This calls for a close integration between academia and industry aiming to reduce the gap between research and implementation.

References

A.C.S. Amaro, A.P. Barbosa-Povoa, 2008, Planning and scheduling of industrial supply chains with reverse flows: a real pharmaceutical case study, Comput. Chem. Eng. 32, 11, 2606-2625.

A.P. Barbosa-Póvoa, Process Supply Chains Management - Where are we ? Where to go next ? Frontiers in Energy Research - Process and Energy Systems Engineering, 2, 23.

A.P. Barbosa-Póvoa, C. da Silva, A. Carvalho, A. 2018, Opportunities and challenges in sustainable supply chain: an operations research perspective. Eur. J. Oper. Res., accepted.

E. Cameron, M. Green, 2015. Making sense of change management. Kogan Page Ltd., 4th Ed., Great Britain.

S. Cardoso, A.P. Barbosa-Póvoa, S. Relvas, 2015, Resilience metrics in the assessment of complex supply-chains performance under demand uncertainty, OMEGA 56, 53–73.

S. Cardoso, A.P. Barbosa-Póvoa, S. Relvas, 2013, Design and planning of supply chains with reverse logistics activities under demand uncertainty, Eur. J. Oper. Res. 226, 436-451

Y. Dong, C.T. Maravelias, J.M. Pinto, A. Sundaramoorthy, 2017, Solution methods for vehicle-based inventory routing problems. Comput. Chem. Eng. 101, 259-278.

Y. Dong, J.M. Pinto, A. Sundaramoorthy, C.T. Maravelias, 2014, MIP model for inventory routing in industrial gas supply chain. Ind. Eng. Chem. Res. 53, 17214-17225.

I. E. Grossmann, 2012, Advances in mathematical programming models for enterprise-wide optimization, Comput. Chem. Eng. 47, 2-18.

J.M. Lainez, E. Schaefer, G.V. Reklaitis, 2012, Challenges and opportunities in enterprise-wide optimization in the pharmaceutical industry. Comput. Chem. Eng. 47, 19-28.

C. Lima, S. Relvas, A.P. Barbosa-Póvoa, 2018, Stochastic programming approach for the optimal tactical planning of the downstream oil supply chain, Comp. Chem. Eng. 108, 314-336.

E. Malinowski, M. Karwan, J.M. Pinto, L. Sun, 2018, A mixed-integer programming strategy for liquid helium global supply chain planning. Transport Res E-Log, 110, 168-188.

L. Megan, K. Bruton, 2017, Corporate Profile: Praxair, Analytics, www.analitics-magazine.org, May/June, 56-62.

S. Mitra, J.M. Pinto, I.E. Grossmann, 2014a, Optimal multi-scale capacity planning for power-intensive continuous processes under time-sensitive electricity prices and demand uncertainty, Part I: Modeling, Comp. Chem .Eng. 84, 382-393.

S. Mitra, J.M. Pinto, I.E. Grossmann, 2014b, Optimal multi-scale capacity planning for power-intensive continuous processes under time-sensitive electricity prices and demand uncertainty. Part II: Enhanced hybrid bi-level decomposition, Comp. Chem. Eng. 65, 101-112.

B Mota, I. Gomes, A. Carvalho, A.P. Barbosa-Póvoa, 2018, Sustainable supply chains: an integrated modelling approach under uncertainty, Omega, in publication.

A. Nikolopoulou, M.G. Ierapetritou, 2012, Optimal design of sustainable chemical processes and supply chains: A review, Comp. Chem. Eng. 44, 94-103.

C. Ning, F. You, 2017, Data-driven adaptive nested robust optimization: General modeling framework and efficient computational algorithm for decision making under uncertainty, AIChE J., 63, 3790-3817.

L.G. Papageorgiou, 2009, Supply chain optimisation for the process industries: Advances and opportunities, Comp. Chem. Eng. 33, 1931–1938.

N. Sahay, M.G. Ierapetritou, 2014, Hybrid simulation based optimization framework for centralized and decentralized supply chains, Ind. Eng. Chem. Res. 53, 3996-4007.

N. Shah, 2005, Process industry supply chains: Advances and challenges. Comp. Chem. Eng. 29, 1225-1235.

G. Stephanopoulos, G.V. Reklaitis, 2012, Process systems engineering: From Solvay to modern bio- and nanotechnology.: A history of development, successes and prospects for the future, Chem. Eng. Sci. 66, 4272–4306.

L. Su, L. Tang, D.E. Bernal, I. E. Grossmann, 2018, Improved quadratic cuts for convex mixed-integer nonlinear programs, Comp. Chem. Eng. 109, 4, 77-95.

A. Subramanyam, F. Mufalli, J. Pinto, C.E. Gounaris, 2017, robust multi-period vehicle routing under customer order uncertainty. E-print: http://www.optimization-online.org

F. You, I.E. Grossmann, 2013, Multicut Benders decomposition algorithm for process supply chain planning under uncertainty, Ann. Oper. Res. 210, 191-211.

D. Yue, F. You, 2016, Optimal supply chain design and operations under multi-scale uncertainties: Nested stochastic robust optimization modeling framework and solution algorithm, AIChE J. 62, 3041-3055.

Q. Zhang, A. Sundaramoorthy, I.E. Grossmann, J.M. Pinto, 2015, A discrete-time scheduling model for continuous power-intensive process networks with various power contracts, Comp. Chem. Eng. 84, 382-393.

Q. Zhang, A. Sundaramoorthy, I.E. Grossmann, J.M. Pinto, 2017, Multiscale production routing in multicommodity supply chains with complex production facilities. Comp. Chem. Eng. 79, 207-222.

Q. Zhang, J.L. Cremer, I.E. Grossmann, A. Sundaramoorthy, J.M. Pinto, 2016a, Risk-based integrated production scheduling and electricity procurement for continuous power-intensive processes, Comp. Chem. Eng. 86, 90-105.

Q. Zhang, M.F. Morari, I.E. Grossmann, A. Sundaramoorthy, J.M. Pinto, 2016b, An adjustable robust optimization approach to scheduling of continuous industrial processes providing interruptible load, Comp. Chem. Eng. 86, 106-119.

Q. Zhang, A.M. Bremen, I.E. Grossmann, J.M. Pinto, 2018 Long-term electricity procurement for large industrial consumers under uncertainty, Ind. Eng. Chem. Res., Just Accepted Manuscript, DOI: 10.1021/acs.iecr.7b04589

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50010-0

Recent Developments towards Enhancing Process Safety: Inherent Safety and Cognitive Engineering

Rajagopalan Srinivasan[a], Andreja Nemet[b], Zdravko Kravanja[b*]

[a]*Department of Chemical Engineering, Indian Institute of Technology Madras, Chennai 600036, India*
[b]*Faculty of Chemistry and Chemical Engineering, University of Maribor, Smetanova ulica 17, Maribor 2000, Slovenia*
zdravko.kravanja@um.si

Extended Abstract

Safety is one of the critical operability issues and a prerequisite foundation for the prosperity of any process. It needs to be thoroughly respected, planned and maintained during all phases of product and process development, operation and even during the dismantling and salvaging of processes at the end of their life. Considering safety issues during the synthesis and design of processes is especially important, since their implementation at this early step is the most efficient and the least costly. With respect to different strategies, sequential and simultaneous ones can be applied, the former being prevalent in engineering practice. However, considering safety issues simultaneously during the synthesizing and designing processes could lead to significantly safer designs. Safety can be improved during the design phase by applying various direct and indirect measures. An example of direct measures would be performing synthesis of a process scheme simultaneously with a risk analysis in order to foster the *inherent safety* of the design. Simultaneous consideration of other operability issues like availability, reliability and controllability would also be appreciated, since this would indirectly increase the safety of the plant. The following mixed-integer program (S-MINLP) is therefore suggested for performing synthesis of safer and more operable process flowsheets:

$$\begin{aligned} \max \quad & z = c^{T}y + f(x) \\ \text{s.t} \quad & h(x) = 0 \\ & g(x) \le 0 \\ & By + Cx \le b \qquad \text{(S-MINLP)} \\ & Risk(x,y) \le 0 \\ & Oper(x,y)_i \le 0 \\ x \in X = & \{x \in R^n : x^{LO} \le x^{UP}\} \\ y \in Y = & \{0,1\}^m \end{aligned}$$

Any inherent safety increase directly contributes to an increase in process safety. Relying solely on a control system is quite problematic, since every control system has its gaps during any deviation event. Moreover, the significant rate (by some estimates, over 80%) of accidents is a consequence of human error, regardless of any cutting-edge control system. Therefore, minimizing inherent risk is one of the substantial aspects of the process planning. The main factors related to inherent safety are the amount and type of the various substances present in the equipment, and the quality (reliability) of the equipment used. Decisions about these factors are usually made at an early stage of planning;

therefore, inherent safety should be assessed simultaneously with the taking of those decisions, since later changes might be difficult or impossible to implement. In the case of a small-scale Heat Exchanger Network (Nemet et al., 2017), a significantly safer design was obtained while keeping economic performance at a similar level (the risk was halved, while the Net Present Value expense was negligible). This was made possible by the selection of compact heat exchangers that contain lower amounts of substances and, hence, less risk for the same amount of heat exchange. When the approach was applied to a larger scale integration of different plants via a central utility system (Total Site), improved safety was obtained only by making a significant compromise in the economic performance (obtaining a design that was twice as safe led to a 23% increase in TAC). A similar conclusion was reached after synthesis of a methanol production process (30% better risk performance was achieved at the cost of a 27% profit decrease). It is interesting to note that the safer design was obtained at the cost of lower overall process conversion, because the lower recycle rate leads to smaller process units containing lower quantities of the risky substances. There are certainly varied challenges to be considered in the future, e.g. those related to other trade-offs, besides that of economics vs. safety. Moreover, the challenge of selecting numerous pieces of equipment might be too high, since risk assessment of all of them may not be possible. Note also that inherent safety can be limited to each individual unit separately or to the process as a whole. Limiting it to individual units would lead to processes with a higher number of smaller sized units, while limiting it to processes would favor flowsheets with fewer numbers of units, each carrying higher risk. Furthermore, the economy of scale directs the economics-safety trade-off towards larger, hence more risky, processes, while on the other hand, the risk can be decreased for smaller processes with compact, high fidelity equipment. The inherent risk assessment could also serve for the selection of safer raw materials, technologies and products in order to upgrade production sustainability.

Chemical process industries routinely handle hazardous materials that are associated with inherent risk. Abnormalities may lead to incidents with varying consequences – from near-misses, to catastrophic accidents. Various layers of protection are usually deployed to reduce the likelihood of abnormal conditions, or to reduce the severity of their consequences. Even among the various layers of protection deployed in the plant, the role of humans through timely and correct action is critical to ensure proper functioning. Conversely, *human error* can lead to system failures. Moreover, such dependence on humans to 'actively' ensure safety is pervasive. Perhaps because of the belief that human error is inevitable and unpredictable, or that advances in automation will diminish the need for human intervention, the domain of human error has received limited attention. However, various studies across numerous domains, including the chemical industry, have established that human error accounts for over 70% of all accidents today.

Human errors can be grouped into various categories, one important category being cognitive errors, i.e., failures in problem solving. Cognitive errors, especially those made during abnormal situations, can be classified into errors in interpretation (incorrect problem identification), decision making (incorrect selection of task to compensate for the problem), or task planning (incorrect formulation of the sequence of actions based on problem constraints). All of these modes require the operator to utilize a mental model of the plant behavior (which is time-varying) and its state (also dynamic). Any mismatch between the plant's condition and the operator's conception is the precursor to cognitive

error. Hence, inference of such cognitive mismatches would engender a strategy to prevent human error.

Our research has sought to utilize cognitive sensors such as eye trackers to reveal the current focus of attention and the cognitive strategies among control room operators as they operate a (simulated) plant. Modern day, commercial, off-the-shelf eye trackers provide a non-invasive, moment-by-moment assessment of thought processes with millisecond scale temporal resolution, while the operator is seated comfortably in front of a DCS monitor. Our experimental studies with large cohorts of human subjects clearly demonstrate that both the operator's eye gaze and the extent of pupil dilation can offer rich insights into an operator's cognition while dealing with abnormal situations. Specifically, one study found that the gaze distribution of participants on different variables during simulated abnormal situations is related to their performance in handling the situation: participants whose gaze remained longer on uncorrelated variables predominantly failed in handling the situation. Pupillometry has been used to analyze the workload of participants during abnormal scenarios and has established that pupil diameter is a good indicator of cognitive workload during the execution of tasks. The workload of participants who failed in handling an abnormal scenario remained high towards the end, and so did the corresponding pupillary dilation. On the other hand, participants who successfully completed the tasks exhibited consistently decreasing pupil diameter after the first set of corrective actions. In another study, gaze entropy was used to quantify the level of situation awareness of a participant while handling abnormal situations and found that participants with adequate situation awareness looked at only the few variables responsible for the abnormal scenario, which resulted in lower gaze entropy. Lack of situational awareness resulted in the gaze lighting on a relatively large number of variables and thus in a larger value of gaze entropy.

Cognitive sensors such as eye trackers have the potential to be used for real-time performance assessment of plant operators; for assessing operator knowledge during training and for developing more user-friendly Human-Machine Interfaces (HMI), as well as for yielding direct insight into how information from various decision support systems, such as advanced control, alarm management, process monitoring and root-cause analysis, are utilized by the operator.

Keywords: inherent safety, simultaneous risk assessment, cognitive engineering

Reference

A. Nemet, J. S. Klemeš, I. Moon, Z. Kravanja, 2017, Safety analysis embedded in heat exchanger network synthesis, Computers & Chemical Engineering, 107, 357–380.

This page intentionally left blank

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50011-2

Solution Approaches to Stochastic Programming Problems under Endogenous and/or Exogenous Uncertainties

Selen Cremaschi[a], John D. Siirola[b]

[a]*Department of Chemical Engineering, Auburn University, Auburn, AL 36849, USA*
[b]*Center for Computing Research, Sandia National Laboratories, Albuquerque, NM 87185, USA*
selen-cremaschi@auburn.edu, jdsiiro@sandia.gov

Abstract

Optimization problems under uncertainty involve making decisions without the full knowledge of the impact the decisions will have and before all the facts relevant to those decisions are known. These problems are common, for example, in process synthesis and design, planning and scheduling, supply chain management, and generation and distribution of electric power.

The sources of uncertainty in optimization problems fall into two broad categories: endogenous and exogenous. Exogenous uncertain parameters are realized at a known stage (e.g., time period or decision point) in the problem irrespective of the values of the decision variables. For example, demand is generally considered to be independent of any capacity expansion decisions in process industries, and hence, is regarded as an exogenous uncertain parameter. In contrast, decisions impact endogenous uncertain parameters. The impact can either be in the resolution or in the distribution of the uncertain parameter. The realized values of a Type-I endogenous uncertain parameter are affected by the decisions. An example of this type of uncertainty would be facility protection problem where the likelihood of a facility failing to deliver goods or services after a disruptive event depends on the level of resources allocated as protection to that facility. On the other hand, only the realization times of Type-II endogenous uncertain parameters are affected by decisions. For example, in a clinical trial planning problem, whether a clinical trial is successful or not is only realized after the clinical trial has been completed, and whether the clinical trial is successful or not is not impacted by when the clinical trial is started.

There are numerous approaches to modelling and solving optimization problems with exogenous and/or endogenous uncertainty, including (adjustable) robust optimization, (approximate) dynamic programming, model predictive control, and stochastic programming. Stochastic programming is a particularly attractive approach, as there is a straightforward translation from the deterministic model to the stochastic equivalent. The challenge with stochastic programming arises through the rapid, sometimes exponential, growth in the program size as we sample the uncertainty space or increase the number of recourse stages. In this talk, we will give an overview of our research activities developing practical stochastic programming approaches to problems with exogeneous and/or endogenous uncertainty. We will highlight several examples from power systems planning and operations, process modelling, synthesis and design

optimization, artificial lift infrastructure planning for shale gas production, and clinical trial planning.

We will begin by discussing the straightforward case of exogenous uncertainty. In this situation, the stochastic program can be expressed completely by a deterministic model, a scenario tree, and the scenario-specific parameterizations of the deterministic model. Beginning with the deterministic model, modelers create instances of the deterministic model for each scenario using the scenario-specific data. Coupling the scenario models occurs through the addition of nonanticipativity constraints, equating the stage decision variables across all scenarios that pass through the same stage node in the scenario tree. Modelling tools like PySP (Watson, 2012) greatly simplify the process of composing large stochastic programs by beginning either with an *abstract* representation of the deterministic model written in Pyomo (Hart, et al., 2017) and scenario data, or a function that will return the deterministic Pyomo model for a specific scenario. PySP automatically can create the extensive form (deterministic equivalent) model from a general representation of the scenario tree. The challenge with large scale stochastic programs with exogenous uncertainty arises through managing the growth of the problem size. Fortunately, there are several well-known approaches to decomposing the problem, both stage-wise (e.g., Benders' decomposition) and scenario-based (e.g., Lagrangian relaxation or Progressive Hedging), enabling the direct solution of stochastic programs with hundreds or thousands of scenarios.

We will then discuss developments in modelling and solving stochastic programs with endogenous uncertainty. These problems are significantly more challenging to both pose and to solve, due to the exponential growth in scenarios required to cover the decision-dependent uncertainties relative to the number of stages in the problem. In this situation, standardized frameworks for expressing stochastic programs do not exist, requiring a modeler to explicitly generate the representations and nonanticipativity constraints. Further, the size of the resulting scenario space (frequently exceeding millions of scenarios) precludes the direct solution of the resulting program. In this case, numerous decomposition algorithms and heuristics have been developed (e.g., Lagrangean decomposition-based algorithms (Tarhan, et al. 2013) or Knapsack-based decomposition Algorithms (Christian and Cremaschi, 2015)).

Keywords: Multistage stochastic programming, endogenous uncertainty, exogenous uncertainty, planning and scheduling under uncertainty, decision-dependent uncertainty

References

B. Christian, and S. Cremaschi (2015). Heuristic solution approaches to the pharmaceutical R&D pipeline management problem. Computers & Chemical Engineering, 74, 34-47.

W.E. Hart, C.D. Laird, J.-P. Watson, D.L. Woodruff, G.A. Hackebeil, B.L. Nicholson, and J.D. Siirola (2017). Pyomo – Optimization Modeling in Python. Second Edition. Vol. 67. Springer.

B. Tarhan, I. E. Grossmann, and V. Goel (2013). Computational strategies for non-convex multistage MINLP models with decision-dependent uncertainty and gradual uncertainty resolution. Annals of Operations Research, 203(1), 141–166.

J.P. Watson, D. L. Woodruff, and W. Hart (2012), PySP: modeling and solving stochastic programs in Python. Mathematical Programming Computation 4, 109-149.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50012-4

Computer-Aided Design for Energy Saving in an Ammonia-based Post-combustion CO_2 Capture Process

Jialin Liu[a*], Ding-Sou Chen[b]

[a]*Tunghai University, Taichung, Taiwan, jialin@thu.edu.tw*
[b]*China Steel Corporation, Kaohsiung, Taiwan*

Abstract

This work investigated the aqueous ammonia-based post-combustion CO_2 capture (PCC) process using the rate-based model in Aspen Plus, by which the overall energy consumption was minimized under varying the important parameters, such as: the NH_3 concentration (2–5 M), the lean CO_2 loading (0.14–0.28 mol-CO_2/mol-NH_3), and the CO_2 stripper pressure (1–20 atm) on the target of 90 % CO_2 removal and 10 ppm of the slipped NH_3 from the emission. In general, the energy consumption of the CO_2 stripper is directly proportional to the lean solvent requirement, whereas the less lean solvent is required as the higher NH_3 concentration was applied. However, this study reveals that the high NH_3 concentration may result in the intensive energy consumption due to the solid precipitation being formed on the top of the stripper. On the other hand, the regeneration energy can be reduced by applying the pressurized CO_2 stripper; however, the electricity of pumping the rich-out solvent may offset the reduction of the energy consumption. In addition, the configuration of multiple inter-heaters along the CO_2 stripper is investigated. The result shows that the separation efficiency can be improved by adding inter-heaters on the stripper that leads to the energy consumption can be reduced around 18 % comparing with the case without implementing the inter-heaters.

Keywords: Post-combustion carbon capture; regeneration process; energy saving.

1. Introduction

Post-combustion CO_2 capture (PCC) using aqueous ammonia has received great attention as a promising technology because it provides many technical and economic advantages over conventional amine solvents; such as: low cost, low regeneration energy, no sorbent degradation, and an eco-friendly emission. However, the major drawbacks (e.g., slow absorption rate with CO_2 and high volatility of ammonia) necessitate a larger CO_2 absorber where an additional washing column (NH_3 absorber) was added on the top for controlling the NH_3 slipping. This study builds upon the previous work (Liu, 2017) to evaluate the energy consumption of a large-scale PCC process for a 500 MW coal-fired power plant under varying the NH_3 concentration, the lean CO_2 loading, and the CO_2 stripper pressure.

Karimi et al. (2012) investigated the effect of inter-heaters on the CO_2 stripper for the amine-based PCC process. They reported that the inter-heaters would not affect the amount of recovering sensible heat from the high-temperature lean solvent. In addition, the location of the inter-heaters attached onto the stripper might affect the amount of the energy reduction. In this study, the energy saving by implementing the inter-heaters on the CO_2 stripper is investigated for the ammonia-based PCC process. The remainder of

this paper is organized as follows. Section 2 gives a brief description of the simulation settings using the rate-based model in Aspen Plus. Additional details can be found in the previous work (Liu, 2017). In Section 3, the integrated PCC process using aqueous ammonia is described. Section 4 presents the results and discussions of the integrated process with inter-heaters. Finally, conclusions are presented in Section 5.

2. Simulation Model

The flus gas, after the SO_x scrubber, was cooled down to 50 °C prior to entering the PCC process. The flow rate was set to 2125 t/h for simulating the tailed gas from a 500 MW coal-fired power plant and the composition was approximated by 12.4 % CO_2, 11.8 % H_2O and 75.8 % N_2. A rate-based approach using an RADFRAC model was used to simulate the CO_2 and NH_3 absorbers. The design target of the PCC process was to capture three million tons of CO_2 per year, which is the 90 % CO_2 removal from the inlet. Due to the high volatility of NH_3, an ammonia scrubber (NH_3 absorber) was implemented onto the top of the CO_2 absorber in order to suppress the slipped NH_3 concentration at 10 ppm. The Redlich-Kwong equation of state and the electrolyte-NRTL method were used to compute the properties of the vapor and liquid phases, respectively. The parameters of the rate-based model were adopted according to the work of Zhang and Guo (2013), such as: (1) Billet and Schultes correlation for mass transfer, (2) Chilton-Colburn correlation for heat transfer, (3) reactor and mass transfer conditions set to default 0.5, and (4) the liquid holdup at 4% of the free volume. The model was validated again the pilot results of Yu et al. (2011) at Munmorah power station, and the details of the model settings can be found in the previous work (Liu et al., 2017).

3. Integrated PCC Process Design

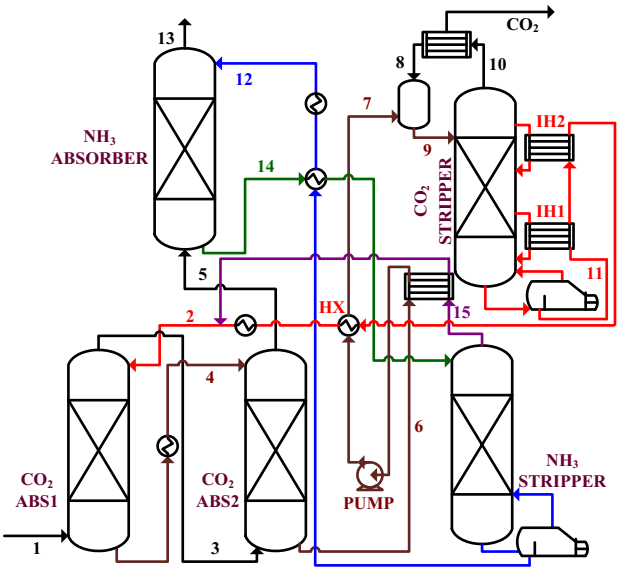


Figure 1. The integrated PCC process with the additional inter-heaters.

In this study, the PCC process flow diagram was modified from the previous work (Liu, 2017), as shown in Figure 1. In order to reduce the slipped NH_3 from the emission, a

two-stage absorber design was implemented (Jilvero et al., 2014). The flue gas (1) was fed into the bottom of the first absorber (CO_2 ABS1) where the lean solvent (2) was fed into the column from the top. Due to the exothermic reaction of CO_2 absorption, the semi-lean solvent (4) was cooled down to 15 °C before entering the top of the second absorber (CO_2 ABS2) where the gas flow from the top of the first absorber (3) was fed from the bottom for further removing CO_2. The CO_2-ridded flow (5) was cleaned by the washing tower (NH_3 ABSORBER) before the gas flow (13) was discharged into atmosphere.

Considering the formation of ammonium bicarbonate from the reflux of the CO_2 stripper, the reflux (8) was heated by the CO_2 rich solvent (7) in order to liquify the solid precipitation before entering into the CO_2 stripper. In this study, the inter-heater configuration was also investigated. As shown in Figure 1, two inter-heaters (IH1 and IH2) were evenly implemented along the CO_2 stripper where the lean solvent (11) was used as the heat source. After the inter-heaters, the remained sensible heat of the lean solvent was recovered by the cross-flow heat exchanger (HX). The separation efficiency of the CO_2 stripper could be enhanced because the top part of the column had a lower temperature, whereas the bottom part kept at the high temperature. In the following section, the result shows that the energy saving was dominated by the number of inter-heaters and the locations of the inter-heaters had a minor effect.

4. Results and Discussions

On the target of 90 % CO_2 removal, the lean solvent requirement was inversely proportional to the NH_3 concentration and directly proportional to the CO_2 loading, as shown in Figure 2. Figure 3 illustrates the regeneration energy of the lean solvent at atmospheric CO_2 stripper. Due to the limitation of the regenerable lean solvent (Liu et al., 2017), the lean solvent using 5 M NH_3 concentration only could be regenerated at 0.28 CO_2 loading where the regeneration energy was up to 7.65 GJ/t-CO_2. The high NH_3 concentration in the overhead vapor encouraged the formation of ammonium bicarbonate, which means the more amount of CO_2 was brought back into the stripper. In order to maintain the target of 90 % CO_2 removal, the more amount of the lean solvent needed to be vaporized as the higher solid fraction in the reflux.

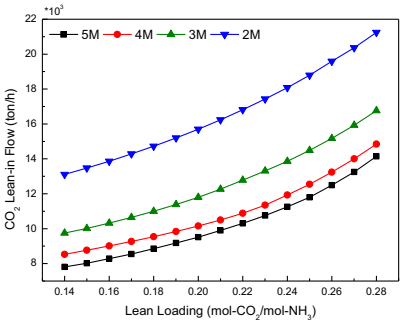


Figure 2. The required lean solvent for achieving 90 % CO_2 removal.

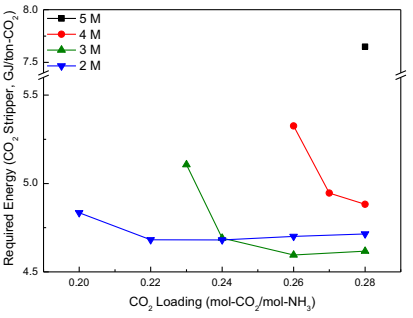


Figure 3. The required energy for the atmospheric CO_2 stripper.

Figure 4 shows almost 70 % solid fraction in the reflux when the lean solvent with 5 M NH_3 concentration was used where the vaporized ratio was 3.4 ton/t-CO_2. Comparing with the lean solvent using 3 M NH_3 concentration, the vaporized ratio was 1.9 ton/t-CO_2 that was 56 % of the lean solvent using 5 M NH_3 concentration. Therefore, the

regeneration energy may not be proportional to the required amount of the lean solvent. As shown in Figure 3, the minimum of the required energy for the CO_2 stripper was 4.59 GJ/t-CO_2 when the lean solvent using 3 M NH_3 concentration and 0.26 CO_2 loading. Meanwhile, the NH_3 stripper took 0.32 GJ/t-CO_2 for regenerating the washing water. The overall energy was 4.91 GJ/t-CO_2 for the atmospheric CO_2 stripper.

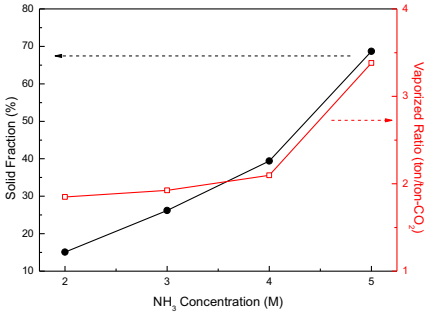
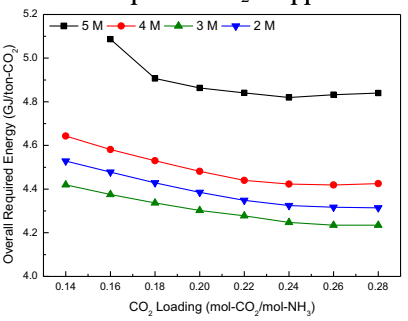
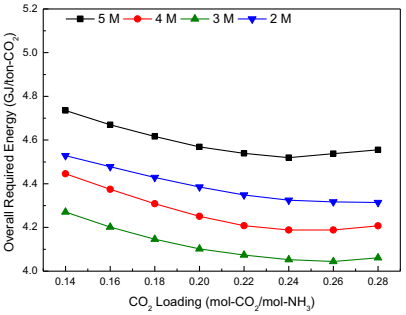


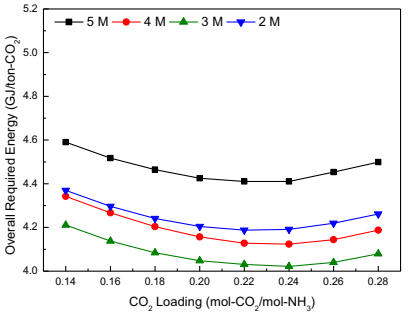
Figure 4. Solid fraction in the reflux and the vaporized lean solvent ratio.



(a)



(b)



(c)

Figure 5. Overall required energy for the pressurized CO_2 stripper, the operating pressure at (a) 5 atm, (b) 10 atm, and (c) 20 atm.

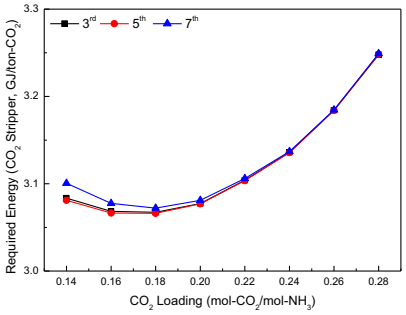


Figure 6. Energy requirement under varying the location of the single inter-heater.

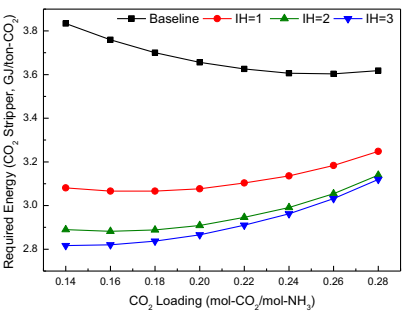


Figure 7. Comparing the reduction of required energy on varying the number of inter-heaters.

In this study, the overall required energy included the regeneration of CO_2 lean solvent and NH_3 washing water; in addition, the electricity of the pump was converted into thermal energy by 35.6 % conversion efficiency. Figure 5 compares the overall required energy under increasing the operating pressure of the CO_2 stripper. The figures show that the overall energy was decreasing as the stripper pressure was increasing; however,

due to the required electricity of the pump, the energy reduction was offset when the operating pressure was increased from 10 to 20 atm in Figures 5(b) and 5(c). Comparing Figures 5(b) and 5(c), the minimum of overall energy at 10 and 20 atm was comparable, which was 4.04 and 4.02 GJ/t-CO_2, respectively. However, considering the reboiler temperature of the stripper at the two operating pressure, which was 145 and 169 °C, respectively, the operating pressure at 10 atm was preferable.

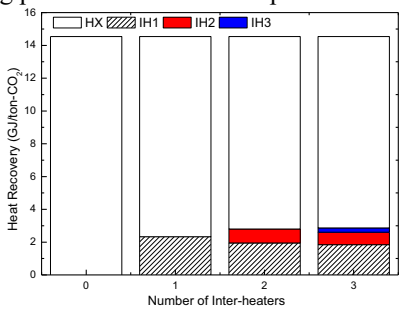
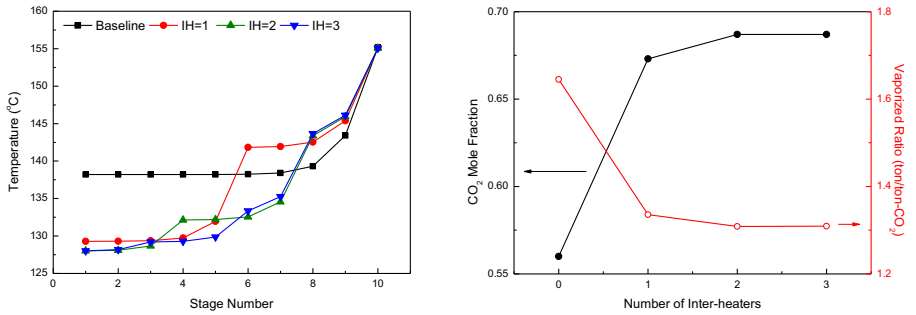


Figure 8. Overall heat recovery by the cross-flow heat exchanger and the inter-heaters.

According to the above discussions, the CO_2 stripper at 10 atm and the lean solvent with 3 M NH_3 concentration that was adopted to investigate the effect of inter-heaters for the energy saving. Figure 6 compares the required energy of the CO_2 stripper in which a single inter-heater was implemented on the different locations. The cold liquid flow was drawn from the third, fifth and seventh tray, respectively, and returned back to the next tray after heating by the inter-heater. The result shows that the required energy was not sensitive to the location of the single inter-heater. On the other hand, Figure 7 shows the effect of energy saving by varying the number of inter-heaters where the baseline indicates the results without inter-heaters by applying the same operating parameters. Figure 7 shows that the required energy can be reduced from 3.8 to 3.1 GJ/t-CO_2 at 0.14 CO_2 loading when a single inter-heater was applied, in which the cold liquid was drawn from the middle position of the column. However, the energy reduction was not significant when the number of inter-heaters was increasing from 2 to 3 where the inter-heaters were evenly attached along the column height.



(a) (b)

Figure 9. Investigation of the required energy reduction by the inter-heaters, (a) the temperature profile along the CO_2 stripper, (b) the CO_2 mole fraction at the top of the stripper and the vaporized lean solvent ratio.

Figure 8 investigates the heat recovery by the cross-flow heat exchanger (HX) and the inter-heaters (IHs) at the lean solvent with 0.14 CO_2 loading. The result shows that the

amount of overall recovery was not enhanced by the inter-heaters, which is consistent with Karimi et al. (2012) reported. The temperature profile along the column height was adjusted by implementing the inter-heaters, as shown in Figure 9(a), in which the top temperature of adding the 2nd and 3rd inter-heaters was close that means the CO_2 mole fraction in the vapor phase at the top of the stripper was also close, as shown in Figure 9(b). On the target of 90 % CO_2 removal, the vapor phase with the higher CO_2 mole fraction at the top can reduce the vaporized amount of the lean solvent, as also shown in Figure 9(b). The minimum of overall energy was found at 3.30 GJ/t-CO_2 where two inter-heaters were implemented and the lean solvent with 0.16 CO_2 loading was applied.

5. Conclusions

At a constant CO_2 absorption rate, the required amount of CO_2 lean solvent can be reduced by increasing the NH_3 concentration; however, the regeneration energy for the lean solvent may not be reduced when the less lean solvent was used. The precipitation of ammonium bicarbonate was easily formed at the low temperature of CO_2 condenser when the lean solvent had a high NH_3 concentration. Therefore, CO_2 was brought back into the stripper by the solid precipitation. In order to maintain the constant CO_2 removal rate, the more lean solvent needed to be vaporized, i.e., the more regeneration energy was needed. At atmospheric CO_2 stripper, the minimum of the overall energy was found at 4.91 GJ/t-CO_2. Since the high temperature at the bottom part of the CO_2 stripper favors the desorption reaction of CO_2 from the solvent, the pressurized stripper may relieve the energy burden, effectively. However, considering the electricity of the pump and the temperature of the low-pressure steam, the minimum of overall energy was found at 4.04 GJ/t-CO_2 where the lean solvent with 3 M NH_3 concentration and 0.24 CO_2 loading was applied and the operating pressure of the stripper was maintained at 10 atm. On the other hand, the low temperature at the top part of the CO_2 stripper may promote the CO_2 concentration in the vapor phase; therefore, the amount of vaporized solvent can be reduced on the target of 90 % CO_2 removal. At the conditions of 10 atm CO_2 stripper and the lean solvent with 3 M NH_3 concentration, two additional inter-heaters can further drive the overall requirement to 3.30 GJ/t-CO_2, at which the lean solvent with 0.16 CO_2 loading was applied. Comparing with the minimum of the atmospheric CO_2 stripper, the overall regeneration energy can be reduced 33% by implementing two inter-heaters along the pressurized CO_2 stripper at 10 atm.

References

H. Jilvero, F. Normann, K. Andersson, F. Johnsson, 2014, The Rate of CO_2 Absorption in Ammonia Implications on Absorber Design. Ind Eng Chem Res, 53, 6750-6758.

M. Karimi, M. Hillestad, H. Svendsen, 2012, Positive and Negative Effects on Energy Consumption by Interheating of Stripper in CO2 Capture Plant. Energy Procedia, 23, 15-22.

J. Liu, 2017, Process Design of Aqueous Ammonia-based Post-combustion CO_2 Capture, J Taiwan Inst Chem Eng, 78, 240-246.

J. Liu, D.S.H. Wong, S.S. Jang, Y.T. Shen, 2017, Energy-saving Design for Regeneration Process in Large-scale CO_2 Capture Using Aqueous Ammonia. , J Taiwan Inst Chem Eng, 73, 12-19.

K. Yu, S. Morgan, A. Allport, A. Cottrell, T. Do, J. Mcgregor, L. Wardhaugh, P. Feron, 2011, Results from Trialing Aqueous NH_3 Based Post-combustion Capture in a Pilot Plant at Munmorah Power Station: Absorption. Chemical Engineering Research and Design, 89, 1204-1215.

M. Zhang, Y. Guo, 2013, Process Simulations of Large-scale CO_2 Capture in Coal-fired Power Plants Using Aqueous Ammonia Solution. Int J Greenh Gas Control, 16, 61-71.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50013-6

Forest residues gasification integrated with electrolysis for production of SNG – modelling and assessment

Johan M. Ahlström[a*], Simon Harvey[a], Stavros Papadokonstantakis[a]

[a]*Chalmers University of Technology, Division of Energy Technology, Maskingränd 7b, 412 58, Gothenburg, Sweden*
johanah@chalmers.se

Abstract

This study investigates opportunities for integrating an electrolysis unit with a biomass gasifier for production of synthetic natural gas (SNG). Gasification is a key technology for production of biofuels and chemicals from lignocellulosic biomass, for which an increased demand is expected in the future. H_2 produced through an electrolyser can be used to increase the output of a gasifier by reaction with CO_2 to form CH_4. Four integrated flowsheet configurations are evaluated with respect to system energy efficiency and process operating revenue. The system energy efficiencies are in the range of 0.55 – 0.8, and the maximum value of operating revenues is 0.245 $/kWh$_{dry\ biomass}$. The results show that feeding the Sabatier reactor with the full product gas flow coming from the methanation unit, and separating the unreacted CO_2 afterwards, is the most attractive configuration with respect to operating revenue.

Keywords: Power-to-gas, Gasification, Biomass, SNG

1. Introduction

Large scale indirect gasification of biomass has been demonstrated at the GoBiGas plant in Gothenburg (Alamia, 2016) and in Güssing (Bolhàr-Nordenkampf et al., 2002). Direct fluidized bed gasification, which is the subject of this work, has been investigated by Hannula and Kurkela (2012) and Gassner and Maréchal (2012). However, no direct blown gasifier with a full scale downstream gas up-grading section have been constructed. When the raw gas from the gasifier is used for methane production, it needs to be upgraded to increase the methane content. The CO_2 remaining after the upgrading step has to be removed, which is done in expensive and energy-intensive sequences of separation steps. Hence, converting the remaining CO_2 into valuable products could significantly improve the economic performance of the process. For this purpose, a hydrogen source is required.

The share of intermittent electricity generation in the power mix is expected to continue to increase significantly. This will induce more volatility in power generation and thereby also in the electricity price (Woo et al., 2011), thereby creating opportunities for variable load applications with a short response time that can help balance the grid. Power-to-gas concepts which produce H_2 and O_2 when electricity prices are low, can contribute to this balancing. H_2 produced by electrolysis can react with CO_2 to produce methane through the Sabatier reaction, whereas O_2 can be used as a gasifying agent in the direct gasifier unit. In this way, the product output of the process is enhanced while the energy loads of

the air separation unit and the CO_2 separation sequence are decreased. To evaluate this concept, it is essential to estimate how the process performance is affected by the amount of H_2 fed to the process.

This paper investigates four different power-to-gas configurations integrated with a direct, oxygen blown, biomass gasifier. The gasifier is assumed to be fed with forestry logging residues. The configurations are compared with respect to system energy efficiency and operating revenues.

2. Process configuration

Forest biomass generally has a moisture content of 40-60%wt, and the fuel must therefore be dried before it is fed to the gasifier to increase efficiency. This study considers the belt dryer concept proposed by Alamia et al. (2015). The moisture remaining in the fuel is vaporized in the gasifier before the fuel volatile components are released. Thereafter, the char gasification starts. At the bottom of the gasifier, combustion of char occurs, which releases the energy that sustains the endothermic reactions in the gasifier.

The raw gas leaving the gasifier contains H_2, CO_2, CO, CH_4, H_2O, inorganic impurities (e.g. H_2S) and organic compounds such as tars. The ash and traces of char in the raw gas are removed in a cyclone, thereafter H_2S is removed. In a pre-methanation step the ratio of H_2/CO is adjusted through the water-gas-shift reaction. The methanation reaction occurs in a series of three adiabatic, fixed bed, reactors. Figure 2 provides an overview of the basic process flowsheet.

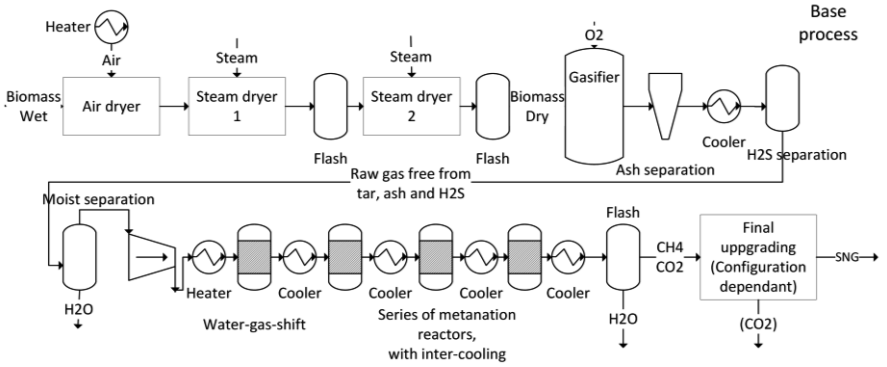


Figure 1. Process flowsheet

After the methanation section, the gas (containing only CH_4 and CO_2) is fed into the final CO_2 removing sequence. Two specification values are considered for the Wobbe index of the SNG product: 44.7-46.7 MJ/Nm^3 and 43.9-47.3 MJ/Nm^3, corresponding to the A, respectively B standards of the Swedish national gas grid. The Wobbe index essentially limits the concentrations of both CO_2 and H_2 in the gas product. A-type SNG is produced if possible, since it can be sold at a higher price, otherwise type B SNG is produced instead. Four possible configurations for the final CO_2 removal sequence are investigated (see Figure 2):

i. The gas from the methanation section is mixed with H_2 from the electrolyser and then compressed before being fed to the Sabatier reactor where H_2 reacts with CO_2 to increase the share of CH_4 in the gas. The gas stream is then dried before

removing the remaining CO_2, in a sequence of two amine-based CO_2 separators, each with a removal efficiency of 90%. The yield of the Sabatier reactor entails that there will be H_2 left in the gas after the reactor if all CO_2 is to be converted. Since configuration (i) does not include a H_2 separation sequence and the gas standards limits the concentration of H_2 in the gas product, there will always be CO_2 in the gas after the Sabatier reactor.

ii. H_2 is added to the gas mix in sufficient quantity to convert all remaining CO_2, thus removing the need for a CO_2 removal step. This configuration requires that the fraction of H_2 in the gas has to be decreased, which is achieved with a H_2 separation unit. As in configuration (i), the gas is dried before the final separation sequence. The separated H_2 is recirculated back to the mixing step before the Sabatier reactor.

iii. CO_2 is separated from the product gas and mixed with H_2 in the Sabatier reactor. The produced gas, containing mainly CH_4 but also some remaining CO_2, is dried and recirculated to the inlet gas stream before the CO_2 removal step.

iv. Similar to Configuration (iii), with the difference that a H_2 separation step is added to the process after the drying step. This provides a process in which all the CO_2 in the raw gas can potentially be reacted to methane, since the excess H_2 can be removed. This results in a more flexible process configuration.

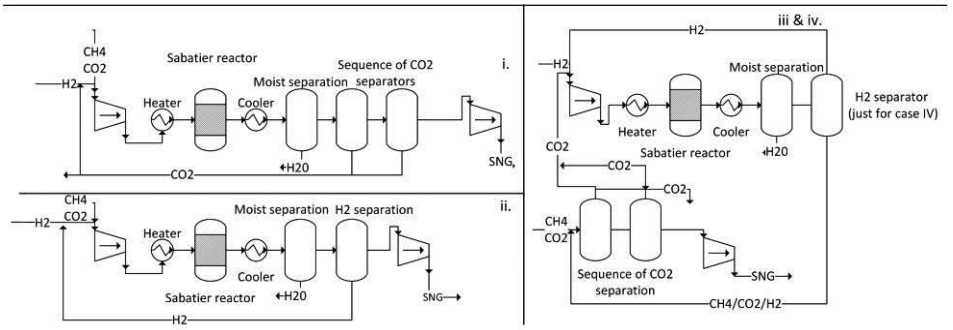


Figure 2. The four process configurations.

Candidate electrolyser technologies include alkaline and Polymer Electrolyte Membrane (PEM) electrolysers. PEM technology is characterized by a shorter start up time, but a lower efficiency. Alkaline electrolyser technology has reached a higher development level, and was therefore selected for this work.

The main difference between the process configurations investigated is the degree of operational flexibility. Configuration (i) is limited by the fraction of CO_2 that can be reacted, since there will be H_2 in the produced gas if all CO_2 is reacted to methane. Configuration (ii) is limited by the absence of CO_2 separation units, meaning that the Sabatier reactor must always be fed with enough H_2 to fully react the CO_2. Thus configuration (i) is more flexible than configuration (ii). In configurations (iii) and (iv), the CO_2 is separated before it is reacted with the H_2. Here configuration (iv) provides the more flexible option; enough H_2 to react all CO_2 can be fed to the process, since a H_2 separation sequence is included.

3. Methodology

All process modelling is performed in Aspen Plus Version 8.8. The biomass feedstock drying model assumes steady-state isothermal operation and negligible pressure drop. The assumed energy required for biomass drying is 0.51 MJ/kg (Alamia et al. (2015)). All simulation results are based on a feed of a 100 $kg_{dry\ biomass}$/h, which gives a flowrate of 2.55 kmol CO_2/h and 1.52 kmol CH_4/h from the methanation unit.

The gasifier is modelled to mimic the results published by Hannula and Kurkela (2012), within an error margin of 10% in terms of gas composition. It is assumed that the tar components are decomposed to CO and H_2. Pressure drop in the gasifier is assumed to be negligible. The methanation section is modelled assuming that the reaction reaches chemical equilibrium. The model accounts for the significant pressure drop in the methanation section (see Alamia (2016)). The Sabatier reactor is modelled as a plug flow reactor with Langmuir-Hinshelwood-Hougen-Watsonis kinetics, as described by Schlereth (2015). The CO_2 and H_2 separation sequences are not rigorously modelled; instead the major component recoveries are inferred by literature data (Heyne and Harvey (2014) and Mivechian and Pakizeh (2013)). Similarly, based on Brynolf et al. (2018), a H_2 yield of 23.7 kgH_2/MWh electricity is assumed for the Alkaline electrolyser.

The system energy efficiency suggested by Heyne and Harvey (2013) is used, including the process by-products and commodities, such as electricity and excess heat

$$\eta_{system} = \frac{\sum \dot{m}_p LHV_p + Q_- + E_-}{\sum \dot{m}_f LHV_f + Q_+ + E_+} \quad (1)$$

ṁ indicates the mass flow of either product, p, or of fuel, f. Q is heat and E electricity, - and +, refer to energy flows leaving or entering the process. Process stream heating and cooling requirements are used to perform heat targeting using pinch analysis and evaluate the possible heat recovery for the process.

The economic performance metric is the operating revenue, obtained by subtracting the cost of biomass feedstock and all utilities from the revenues of selling SNG, process excess heat and the oxygen produced in the electrolyser. Indicative prices for Swedish conditions are assumed; all input data and results are presented in US$. The reference electricity price is set to 0.024 $/kWh and since it is assumed that the electrolyser is run at periods when the electricity price is low, the electricity used for H_2 production is assumed to have a 50% lower cost. Revenue from sales of the SNG product is valued at 1.98 $/kg for A-grade quality and 1.5 $/kg for B-grade quality, the biomass price is 23 $/MWh. All other price data, e.g. catalysts, can be found in the report by Gambardella and Yahya (2017). Sensitivity analysis varying the feed of H_2, together with the recirculation of CO_2 is performed for all configurations except (ii), in which the H_2 flow is constant since the process is designed to react all CO_2 available in the gas.

4. Results

For configuration (ii) the H_2 flow is constant at 10 kmol/h, which is the flowrate required to convert all CO_2 in the gas mix. The system energy efficiency of configuration (ii) is 0.801 and the operating revenues are 0.245 $/$kWh_{dry\ biomass}$. Configuration (ii) results in

the highest revenues and a high system energy efficiency, which is due to the fact that all CO_2 in the raw gas is converted to CH_4. Sensitivity analysis, varying the feed of H_2, is performed for configurations (i), (iii) and (iv). The impact on the operating revenues and system energy efficiency is shown in Figure 3.

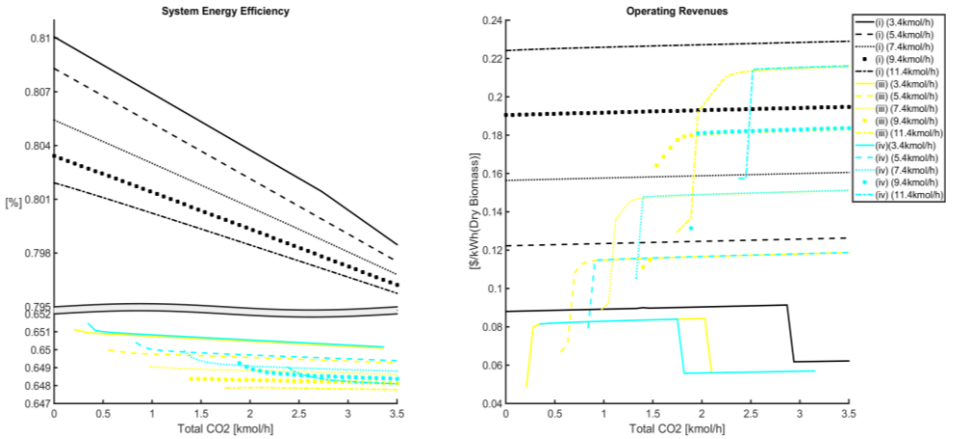


Figure 3. Total system energy efficiency and operating revenues as a function of CO_2 feed. The nuances correspondent to different configurations and the line types correspond to the amounts of H_2 feed.

η_{system}, decreases with increased H_2 feed and CO_2 recirculation rate. This is due to the conversion losses in the electrolyser. The ranges of each curve indicate the cases where the produced gas fulfills the standards of the Wobbe index. The process operating revenues increase for all configurations with H_2 feed and CO_2 recirculation. This indicates that the additional SNG produced by the increased addition of H_2 outweighs the cost of generating the H_2. The increase in revenues is essentially linear, except for some rapid increases and decreases in revenues. These rapid changes indicate the thresholds for the types of SNG produced, namely when the model has to change from type A to type B. Only configuration (i) has an increase in revenue that can be achieved without CO_2 recirculation, which is due to the fact that there is already CO_2 present in the incoming gas flow. For configurations (iii) and (iv), a certain amount of CO_2 has to be recirculated to provide the second reactant to the Sabatier reactor.

5. Conclusions

The results show that the operating revenue increases with increased addition of hydrogen. This indicates that there is an incentive for integration of an electrolyser unit with the process, provided that the investment costs of the processes does not outweigh the increase of operating revenue. This also suggests that the input feed of H_2 should be maximized if this type of power-to-gas concept is implemented. From an economic perspective, the best performing configuration is (ii), which reaches a maximum operating revenue of 0.245 $\$/kWh_{dry\ biomass}$ followed by (i). Even if configuration (ii) outperforms configuration (i) in terms of revenues, it is not necessarily the better alternative; configuration (i) is more flexible, since it allows for different flows of H_2, which could be a major advantage if electricity prices fluctuate a lot. Both configurations (iii) and (iv), in which CO_2 is separated before it is mixed with the H_2, generates lower revenues than the first two configurations. This is a result that highlights that it is more

beneficial to mix the H_2 with the raw gas, rather than to separate the CO_2 before the reactor.

6. Outlook and further work

An essential parameter for future evaluation is the impact of electricity price on the choice of process configuration and design. The model presented in this work constitutes a starting point for such evaluation. To investigate the impact of the electricity price on the process design, a rolling horizon, planning and scheduling optimization algorithm will be developed. By considering the possibility to store hydrogen, and co-running the optimization model with a model of the European electricity system, it becomes possible to determine the optimal process configuration design, accounting for the fluctuations in electricity price. To enable this sort of analysis, it is a necessity to first estimate the investment costs of the different process configurations

7. References

Alamia, A., 2016. Large-Scale Production and Use of Biomethane, Chalmers university of Technology.

Alamia, A., Ström, H., Thunman, H,. 2015. Design of an integrated dryer and conveyor belt for woody biofuels, *Biomass and Bioenergy,* 77, 92-109.

Bolhàr-Nordenkampf, M., Bosch, K., Rauch, R., Kaiser, S., Tremmel, H., Aichernig, C. & Hofbauer, H., 2002. Scale-up of a 100kWth pilot FICFB-gasifier to a 8 MWth FICFB-gasifier demonstration plant in Güssing (Austria), *Proc. 1st International Ukrainian Conference on Biomass For Energy, Kyiv, Ukraine*.

Brynolf, S., Taljegård, M., Grahn, M., Hansson, J., 2018. Electrofuels for the transport sector: A review of production costs, *Renewable and Sustainable Energy Reviews,* 81 (Part 2), 1887-905.

Gambardella, A. and Yasin, Y., 2017 Power-to-Gas concepts integrated with syngas production through gasification of forest residues, Chalmers university of Technology.

Gassner, M., and Maréchal, F., 2012. Thermo-economic optimisation of the polygeneration of synthetic natural gas (SNG), power and heat from lignocellulosic biomass by gasification and methanation, *Energy and Environmental Science,* 5 (2), 5768-89.

Hannula, I. and Kurkela, E., 2012. A parametric modelling study for pressurised steam/O2-blown fluidised-bed gasification of wood with catalytic reforming, *Biomass and Bioenergy,* 38 (Supplement C), 58-67.

Heyne, S. and Harvey, S., 2014. Impact of choice of CO2 separation technology on thermo-economic performance of Bio-SNG production processes, *International Journal of Energy Research,* 38 (3), 299-318.

Heyne, S. and Harvey, S., 2013. Assessment of the energy and economic performance of second generation biofuel production processes using energy market scenarios, *Applied Energy,* 101, 203-12.

Mivechian, A. and Pakizeh, M., 2013. Hydrogen recovery from Tehran refinery off-gas using pressure swing adsorption, gas absorption and membrane separation technologies: Simulation and economic evaluation, *Korean Journal of Chemical Engineering,* 30 (4), 937-48.

Schlereth, D., 2015. Kinetic and Reactor Modeling for the Methanation of Carbon Dioxide, (Dissertation, Technische Universität München, München. I).

Woo C.K., Horowitz, I., Moore, J., Pachecho, A,. 2011. The impact of wind generation on the electricity spot-market price level and variance: The Texas experience, *Energy Policy,* 39 (7), 3939-44.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50014-8

Plant-wide process design of producing dimethyl carbonate by indirect alcoholysis of urea

Li Shi[a], San-Jang Wang[b*], David Shan-Hill Wong[c*], Kejin Huang[a], En-Ko Lee[b], Shi-Shang Jang[c]

[a]*College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, P. R. China*
[b]*Center for Energy and Environmental Research, National Tsing Hua University, Hsinchu 30013, Taiwan*
[c]*Department of Chemical Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan*
wangsj@mx.nthu.edu.tw and dshwong@che.nthu.edu.tw

Abstract

Dimethyl carbonate (DMC) is a green compound with a broad variety of application. Recently, CO_2-based routes to produce DMC have attracted much attention due to the environment benefits from CO_2 utilization. In the study, we investigate the plant-wide process design of DMC production using CO_2 as a raw material by indirect alcoholysis of urea. The indirect alcoholysis route of urea shows many advantages because of cheap raw materials, mild and safe operation condition, and environmentally friendly chemicals. The DMC production contains the processes of urea synthesis, propylene carbonate (PC) synthesis, DMC synthesis, and DMC/methanol azeotrope separation. The methods of urea synthesis, DMC synthesis, and DMC/methanol azeotrope separation were well developed. In the study, the emphasis is put on the PC synthesis process. Some different methods of PC synthesis are proposed, designed, and optimized. Simulation results reveal that the intensified process containing a reactive distillation column and a conventional distillation column with internal vapor compression provides the most economical design by fully utilizing the special azeotrope characteristic of propylene carbonate and propylene glycol pair. A 21.8% reduction of total annual cost can be achieved by the intensified process in comparison with the traditional process. In addition, the complete plant-wide process for DMC production is designed. The net profit for this production process is also given.

Keywords: dimethyl carbonate, process design, urea.

1. Introduction

Dimethyl carbonate (DMC) is an important organic compound and chemical intermediate with the label of ‘‘green chemical’’. There are several reaction routes to manufacture DMC. Recently, DMC production utilizing CO_2 attracts much attention because it provides direct environment benefits when valuable products are manufactured from emitted and undesired CO_2. The CO_2-based methods to produce DMC include direct synthesis with methanol (MeOH), transesterification of propylene carbonate (PC) or ethylene carbonate with MeOH, direct and indirect alcoholysis of urea. In this work, DMC is produced by indirect alcoholysis of urea. This reaction route provides many benefits, such as cheap and easily available raw materials, mild reaction

conditions, safe operations, environmentally friendly chemicals, etc. Figure 1 shows the concept of the plant-wide process for the DMC production by this route. Urea, PC, and DMC are sequentially synthesized from the plant-wide process. The most important characteristic in this complete plant-wide process is that PG, a byproduct in the transesterification reaction for DMC synthesis, can be used as the reactant for PC synthesis and the released ammonia from PC synthesis can be recycled back to produce urea by reacting with CO_2. The overall process forms a green chemical cycle, which increases the utilization of raw materials for the DMC production. In this work, the emphasis is put on the PC synthesis because urea and DMC synthesis were studied by some literatures (Holtbruegge et al., 2013; Voskov and Voronin, 2016). Five different processes for PC synthesis by indirect alcoholysis of urea are proposed, designed, and optimized by minimizing total annual cost (TAC). These processes can be classified in terms of two operation types, near neat operation and excess reactant operation. Comparison of these processes by TAC is made and the complete plant-wide process for DMC production is also given.

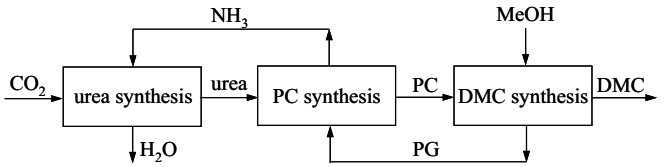


Figure 1. Concept of DMC production by indirect alcoholysis of urea.

2. Kinetics and thermodynamic models

In the reaction route for indirect alcoholysis of urea, three reaction steps are proceeded as follows.

$$CO_2 + 2\ NH_3 \rightleftharpoons Urea + H_2O \quad (1)$$

$$Urea + PG \rightarrow PC + 2\ NH_3 \quad (2)$$

$$PC + 2\ MeOH \rightleftharpoons DMC + PG \quad (3)$$

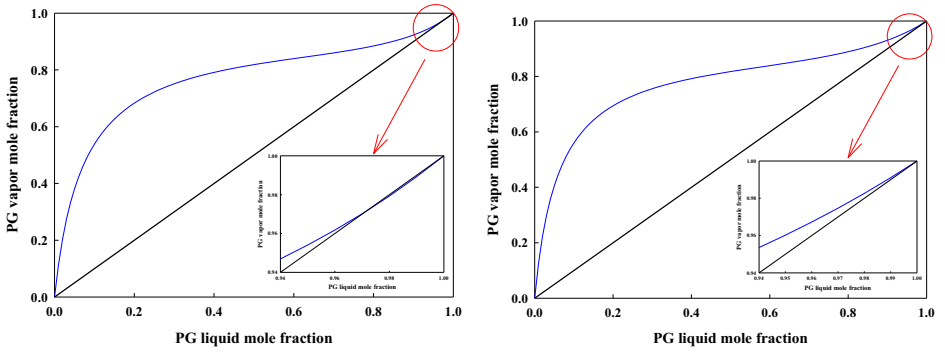
In the first step, urea is produced by reacting CO_2 and ammonia. Next, urea is reacted with PG to coproduce PC and ammonia. The final step is used to synthesize DMC by the transformation reaction of PC and MeOH. The byproducts, ammonia and PG produced from steps 2 and 3, respectively, are recycled back to steps 1 and 2, respectively, as reactants. There are very few studies about the process design for step 2 to synthesize PC. Therefore, in the present study, we primarily explore the design of different PC synthesis processes. Some possible configurations are proposed to synthesize PC.

Wang (2012) presented a kinetic model given below for PC synthesis by using MgO as the heterogeneous catalyst.

$$r_{PC} = 1.5888 \cdot \exp\left(-\frac{562.602}{T}\right) \cdot C_{urea} \cdot C_{PG} \quad (4)$$

where r_{PC} is the reaction rate of PC synthesis (mol liter^{-1} min^{-1}). T and C_i represent temperature (K) and concentration (mol liter^{-1}) of component i, respectively. The catalyst concentration is assumed to be 2 wt%.

In the study, the vapor-liquid equilibrium (VLE) relationship is described by UNIQUAC activity coefficient model in ChemCad software. In the PC synthesis reaction system, only the thermodynamic data of urea-ammonia and PG-PC pairs are found in the literatures (Mathuni et al., 2011; Voskov and Voronin, 2016). The phase-equilibrium relationships for the other pairs are determined in the study by the ideal model. Figure 2 shows the VLE relationship for PG-PC pair. It indicates that homogeneous minimum-boiling azeotrope can be found for this pair under low pressure (0.12 atm) (Mathuni et al., 2011). However, there is no azeotrope under high pressure (1 atm). This important characteristic is fully utilized to reduce energy consumption in the following process design for PC synthesis.



(a) (b)

Figure 2. VLE relationship for PG-PC pair under pressures (a) 0.12 atm and (b) 1 atm.

3. Optimal design of PC+DMC Synthesis

In the study, five configurations are proposed and classified in terms of two operation types, near neat and excess reactant operations. In the process design to synthesize PC, the objective is to minimize the TAC by adjusting design variables. TAC is defined as:

TAC = operating cost + capital cost/payback year (5)

The operating cost (OC) includes the costs of steam, cooling water, catalyst, and electricity. The capital cost (CC) comprises the costs of reactor, flash tank, column shell, stage, heat exchanger, and compressor. The formulas of these cost estimations are taken from Turton et al. (2012). Here, a payback of eight years is used. Figure 3 shows the optimal design performed by sensitivity analysis for two configurations with near neat operation. Figure 3a is a flowsheet for PC synthesis by using some CSTRs, flash tanks, and one conventional distillation (CD) column (named as CSTR+CD configuration). Two reactants are fed into the first CSTR and then the exit mixture from the CSTR is introduced into the first flash tank. Vapor product (mainly containing ammonia) is removed from the flash tank, and liquid product is fed into the second CSTR followed by the second flash tank. CSTRs are used for PC synthesis until the requirement of urea reaction conversion (99.98 mol%) is satisfied. Then the product of the last CSTR is introduced into the CD column for separation. The remaining ammonia, not removed

from the flash tanks, is withdrawn from the partial condenser at column top and high-purity PC is achieved from the column bottom. In the study, urea is fed with the flow rate of 40 kmol/h. The purity of ammonia removed from the flash tanks and the partial condenser is designed at 99 mol% while the PC purity of column bottom is set at 99.5 mol%. The CSTR and flash tank are operated at the same temperature (160 ºC as suggested by Wang (2012)) and pressure. Because PG, a byproduct in the transesterification reaction for DMC synthesis, is converted into the reactant for the PC synthesis by indirect alcoholysis of urea, the optimal reactive distillation (RD) column to synthesize DMC by reacting PC and MeOH is also designed. The kinetic model and the thermodynamic model from Holtbruegge et al. (2013) are adopted in the RD column design for DMC synthesis. DMC and MeOH form a minimum-boiling homogeneous azeotrope. The overhead product of the RD is a mixture of DMC and MeOH. The bottom PG product with 99.9 mol% is recycled back to the CSTR for further reaction. The optimal TAC for PC synthesis by this configuration is 1102.2×10^3 US$/year.

Figure 3b is a flowsheet for PC synthesis by using a RD. Two reactants are fed into the first RD column for PC synthesis. Ammonia is withdrawn from the partial condenser. PC is achieved from the column bottom and then fed into the second RD column for DMC synthesis. The optimal feed ratio of PG and urea is designed at 1.02, very close to the neat operation. Column pressure is 0.12 atm. High-purity (99.5 mol%) PC withdrawn from the bottom of the PC synthesis column is fed into the RD column for DMC synthesis. In the DMC synthesis column, DMC/MeOH mixture is also recovered from the column top while high-purity PG with 99.9 mol% is withdrawn from the column bottom and recycled back to the RD column for PC synthesis. The minimized TAC for the PC synthesis by this RD configuration is 973.4×10^3 US$/year. In comparison with the CSTR+CD configuration, this TAC can be reduced by 11.7 %.

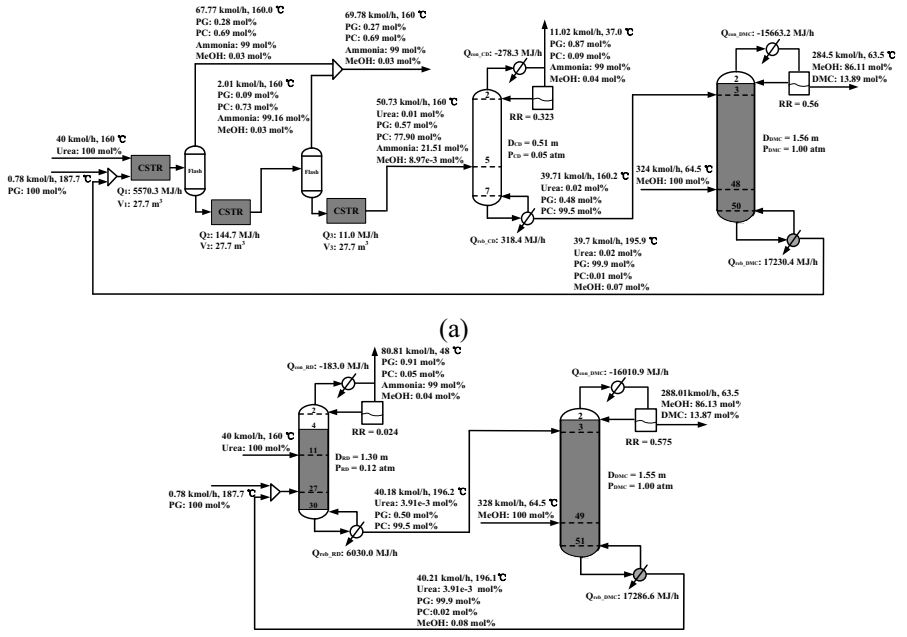


Figure 3. Optimal design for (a) CSTR+CD and (b) RD configurations.

Figure 4 shows the optimal design for three configurations with excess reactant operation. In this type of configuration, excess PG reactant is used in the RD column for PC synthesis. Bottom product, mainly containing PC and excess PG, of the RD column is introduced to a CD column for the separation of PC and PG. Heat integration between the RD and CD columns can then be implemented to reduce the energy consumption of the PC synthesis process by properly adjusting the pressures of RD and CD columns. Figure 4a is the first configuration of heat integrated RD+CD (named as RD+CD_HI1). The overhead temperature of CD column is designed to be greater than the bottom temperature of the RD column. Then the latent heat of the overhead vapor from the CD column can be released to the reboiler of the RD column. The optimal feed ratio of PG and urea is 1.25. The minimized TAC of this configuration is 1055.5×10^3 US$/year. In comparison with the CSTR+CD configuration, the TAC is reduced by 4.2 % for the RD+CD_HI1 configuration. However, the TAC is increased in comparison with the RD+CD configuration.

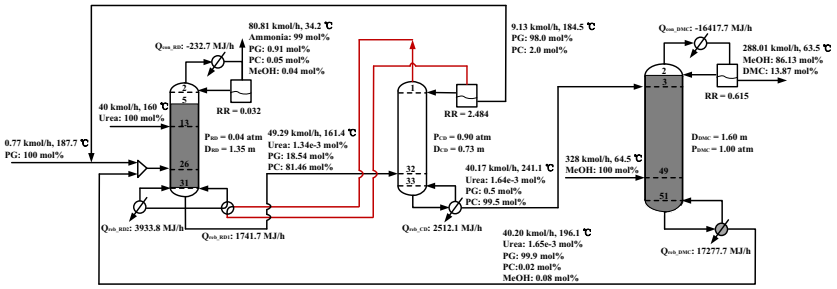


Figure 4. Optimal design for (a) RD+CD_HI1, (b) RD+CD_HI2, and (c) RD+CD_HI3 configurations.

Figures 4b and 4c are the two configurations with the CD column divided into top and bottom sections operated at high and low pressures, respectively. The overhead vapor of the bottom section is compressed by a compressor and used as a heating medium of the base of the top section. Much economical benefit can be achieved because PG-PC azeotrope vanishes under high pressure. Heat integration can be implemented by releasing the latent heat of overhead vapor to the reboiler in the CD column or in the RD column, respectively (named as RD+CD_HI2 and RD+CD_HI3). The TACs of these two configurations are 870.8×10^3 US$/year and 861.6×10^3 US$/year, respectively. RD+CD_HI3 configuration provides the most economical design. A 21.8% reduction of TAC can be achieved in comparison with CSTR+CD configuration.

In the addition to the PC synthesis, we also optimally design the processes for urea synthesis by Stamicarbon technology and DMC/MeOH azeotrope separation by extractive distillation. These two processes are intensified by heat integration to further reduce energy consumption. In this complete plant-wide process for DMC production, the TACs for urea synthesis, PC synthesis, DMC synthesis, and DMC/MeOH separation are 869.1×10^3 US$/year, 861.6×10^3 US$/year, 2775.4×10^3 US$/year, and 2913.1×10^3 US$/year, respectively. The annual costs of raw materials (CO_2 and MeOH) and product (DMC) are 1029.3×10^3 US$/year, 6625.2×10^3 US$/year, and 20667.1×10^3 US$/year, respectively. These indicate that the net profit is 5590×10^3 US$/year and 27% net profit margin can be obtained for this complete plant-wide DMC process.

4. Conclusions

Plant-wide process design of producing DMC by indirect alcoholysis of urea is investigated in this work. Five different configurations are proposed. Simulation results reveal that the RD+CD_HI3 configuration under excess reactant can provide the most economical design. In this configuration, the advantage of the azeotrope characteristic of PC-PG pair is fully taken by adding an internal compressor between top and bottom sections of the CD column. Heat integration is then implemented between column top and bottom to reduce energy consumption. In addition, the design of the complete plant-wide process is also given. Net profit margin with 27% can be achieved for this process.

References

J. Holtbruegge, M. Leimbrink, P. Lutze, A. Górak, 2013, Synthesis of dimethyl carbonate and propylene glycol by transesterification of propylene carbonate with methanol: Catalyst screening, chemical equilibrium and reaction kinetics, Chem. Eng. Sci., 104, 12, 347-360.

T. Mathuni, J. I. Kim, S. J. Park, 2011, Phase equilibrium and physical properties for the purification of propylene carbonate (PC) and γ-butyrolactone (GBL). J. Chem. Eng. Data, 56, 1, 89-96.

R. Turton, R. C. Bailie, W. B. Whiting, J. A. Shaeiwitz, 2012, Analysis, synthesis, and design of chemical processes, Pearson Education: Boston.

A. L. Voskov, G. F. Voronin, 2016, Thermodynamic model of the urea synthesis process. J. Chem. Eng. Data, 61, 12, 4110-4122.

Y. Wang, 2012, Green process development of dimethyl carbonate, Qingdao University of Science and Technology, Master Thesis.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50015-X

Design of a Pressure-Swing Distillation Process for the Separation of n-Hexane and Ethyl Acetate Using Simulated Annealing

Xiao-Ling Yang[a] and Jeffrey D. Ward[a*]

[a]*Department of Chemical Engineering, National Taiwan University, #1, Sec. 4 Roosevelt Rd. Taipei 10617 Taiwan*
jeffward@ntu.edu.tw

Abstract

A pressure-swing distillation process for the separation of n-hexane and ethyl acetate, which form a minimum boiling azeotrope, is proposed and optimized using simulated annealing and simulation-optimization. The design variables considered are the two column pressures, the feed stages and total number of stages of both columns. The optimal process design is somewhat more expensive than the optimal extractive distillation process (designed separately) but the pressure-swing process benefits from the fact that no additional component is introduced.

Keywords: pressure-swing distillation, simulated annealing, simulation optimization

1. Introduction

In some azeotropic systems, the azeotrope composition is a strong function of pressure. This property can be exploited to break the azeotrope without the introduction of a third component (an entrainer) by means of pressure-swing distillation. In general, the greater the change of the azeotrope composition with pressure, the more efficient pressure-swing distillation will be. General information about pressure-swing distillation can be found in textbooks including the book by Luyben and Chien (2010).

Mixtures of n-hexane and ethyl acetate are used as solvents for liquid chromatography in the pharmaceutical industry (Rodriguez-Donis et al., 2005). Rodriquez, et al. (1999) and Rodriguez-Donis et al. (2005) studied the design of batch distillation processes with entrainers for breaking the azeotrope. Yang (2017) studied the design of continuous extractive distillation processes for this mixture with several solvents. Vapor-liquid equilibrium data for this mixture have been measured by Acosta et al. (2002) and Gorbunova et al. (1965).

Simulated annealing is a stochastic optimization method that was developed by analogy to annealing in metalworking (Kirkpatrick et al., 1983). At the start of the optimization, the algorithm will accept a move to a higher energy state (larger value of the objective function which is to be minimized) with greater probability in order to avoid becoming trapped in a local minimum. As the optimization progresses, the likelihood of accepting a higher energy state decreases and finally the system becomes “frozen” in a low-energy state that is hopefully nearly optimal. This method of optimization has been applied successfully to many problems in chemical process engineering optimization, including heat-integrated distillation sequences (Wei-zhong and Xi-Gang, 2009) and pressure-swing distillation processes (Cardoso, et al., 2000).

In this work, simulated annealing is applied to the optimization of a process for separating mixtures of n-hexane and ethyl acetate by pressure-swing distillation. The distillation columns are modelled in Aspen Plus and the simulated annealing algorithm is coded in Matlab. These two programs communicate by means of the Aspen Plus automation server. (AspenTechnology, 2002) and Matlab automation client.

2. Process model

Figure 1 shows the Txy diagrams for mixtures of n-hexane and ethyl acetate at pressures of 0.5 and 10 atm. The data are calculated using the UNIQUAC activity coefficient model with parameters taken from the Aspen Plus databank. The model predicts that the composition of the azeotrope shifts from 70 mol% to 55 mol% n-hexane, which suggests that pressure-swing distillation may be a feasible method for separation. To facilitate comparison with an extractive distillation process designed previously (Yang 2017), the feed flowrate and composition are set the same as in the previous work: 100kmol/h of a mixture that is 66.87 mol% n-hexane and 33.13mol% ethyl acetate. Since the feed is rich in n-hexane, it is fed to the high-pressure column where n-hexane is collected as a product.

The steady-state design is shown in Figure 2. In the high-pressure column, nearly pure n-hexane is collected at the bottom and a mixture near the azeotropic composition at high pressure is collected at the top and fed to the low-pressure column. In the low-pressure column, nearly pure ethyl acetate is collected at the bottom of the column, and a stream with composition near to the azeotropic composition is collected at the top of the column and recycled to the high-pressure column. Required pumps are not shown on the diagram for simplicity.

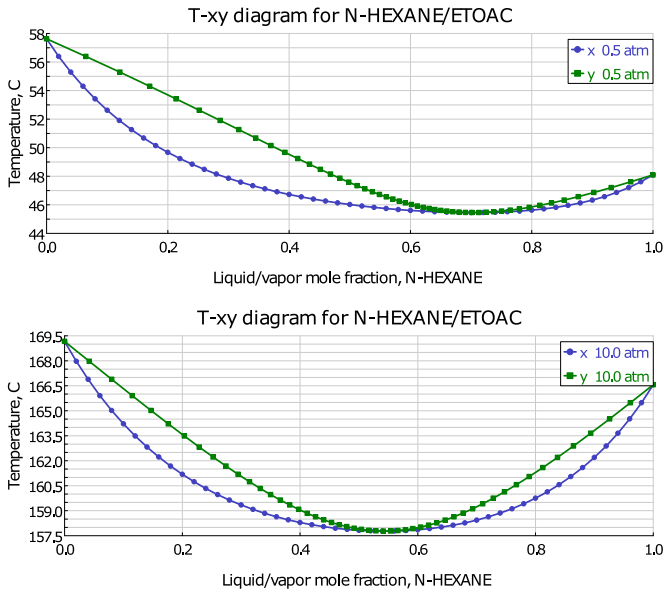


Figure 1. Txy diagrams for n-hexane-ethyl acetate at 0.5 and 10 atm

The variables to be determined by optimization are the total number of stages in the high pressure column (N1), the fresh feed and recycle stream feed tray locations (B1FS

and RS), the total number of stages of the low pressure column and feed tray location in the low pressure column (N2 and B2FS), and the two operating pressures (B1P and B2P). Design variables, their meanings, and the values of constrains are given in Table 1.

Both product purities are held at 99.5 mol% by manipulating the reboiler duties of the columns using the 'Design Spec/Vary' feature in Aspen Plus. In the high-pressure column, the distillate composition is specified to be 2.5 mol% greater than the azeotrope composition at the column pressure, and in the low pressure column, the distillate composition is specified to be 2.5 mol% less than the azeotropic composition at the column pressure. Distillate compositions are fixed by manipulating the reflux ratios in the columns using the Spec/Vary feature. To simplify the simulation, the recycle stream is torn in the Aspen Plus simulation and converged to within 0.1% by iteration in Matlab.

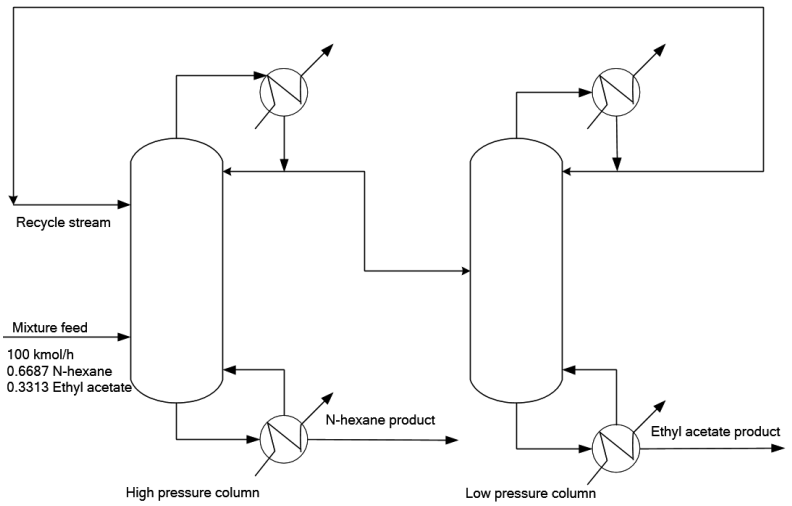


Figure 2. Proposed flowsheet of pressure-swing to separate n-hexane-ethyl acetate mixture

Table 1. Design variables for optimization

Design variable	Meaning	Constraints [min, max]	Nstep
B1P	Column 1 pressure	[8, 12]	0.1
B2P	Column 2 pressure	[0.1, 0.6]	0.01
N1	Number of stages in column 1	[5, 100]	1
B1FS	Fresh feed stage in column 1	[2, N1−1]	1
B1RS	Recycle feed state in column 1	[2, N1−1]	1
N2	Number of stages in column 2	[5, 100]	1
B2FS	Feed stage in column 2	[2, N2−1]	1

3. Optimization method

The SA algorithm was implemented in Matlab. A flowchart of the algorithm is shown in Figure 3. The procedure can briefly be described as follows:

[1] Start with given initial temperature, final temperature and solution space.

[2] Perturb the design variable vector d_i in the neighborhood to generate a new solution vector d_{i+1}. Run the simulation and check if the flowsheet is converged. Repeat this procedure if necessary until a solution vector d_{i+1} is found for which the simulation is converged.

[3] Iterate to converge the torn recycle stream. If the flowsheet fails to converge during the iteration, return to step [2].

[4] Evaluate the objection function $E(d_{i+1})$ and calculate $\Delta E = E(d_{i+1}) - E(d)$. If the $\Delta E \leq 0$, then accept the new value d_{i+1} by updating the current optimal solution. If $\exp(-\Delta E/T) \geq U(0,1)$, also accept d_{i+1}. $U(0,1)$ is a random number between 0 and 1.

[5] Return to step 2 until the system reaches thermodynamic equilibrium at temperature T.

[6] Decrease the temperature according to the cooling schedule. Then reset i=1 and repeat steps 2–5 until the final temperature is achieved.

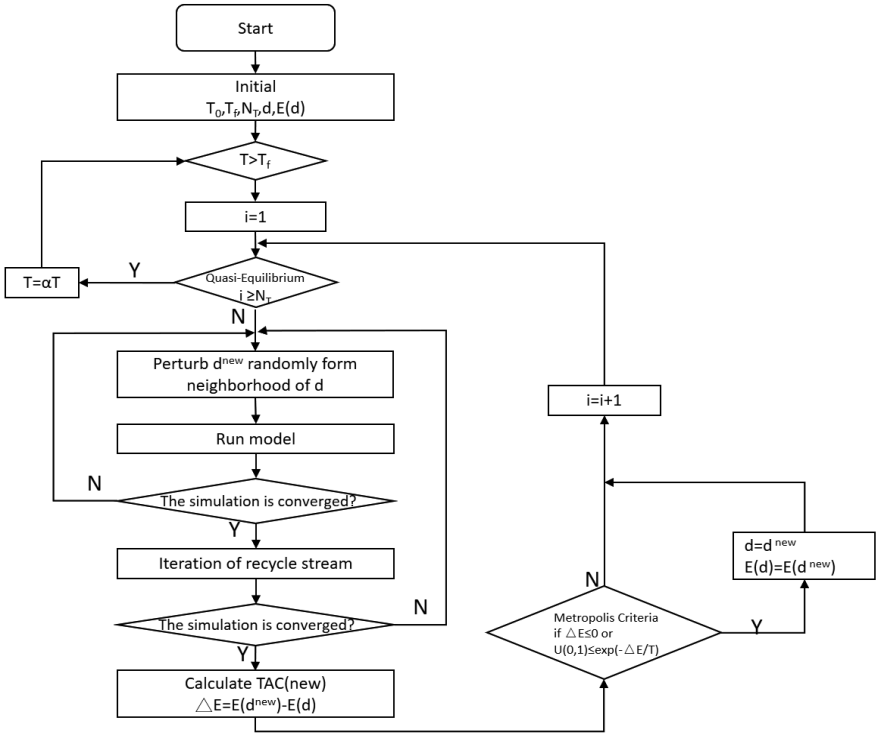


Figure 3. Flowchart of the simulated annealing algorithm

TAC calculations were performed using the methods and equations provided by Luyben (2006) except for the cost of the vacuum system which was determined using the methods and equations of Seider et al. (2009).

4. Results

The results of the annealing process are shown in Figure 4, and the detailed results are shown in Table 2. The lowest TAC obtained by the algorithm is 2,144,300$/yr with the two columns operated at 10atm and 0.19 atm. The SA algorithm was repeated starting at the optimal point determined by the first run, and the results of the second turn were similar to those of the first. The simulation was also repeated starting with different initial conditions and it was always found to converge to nearly the same point, suggesting that this point is near the global minimum.

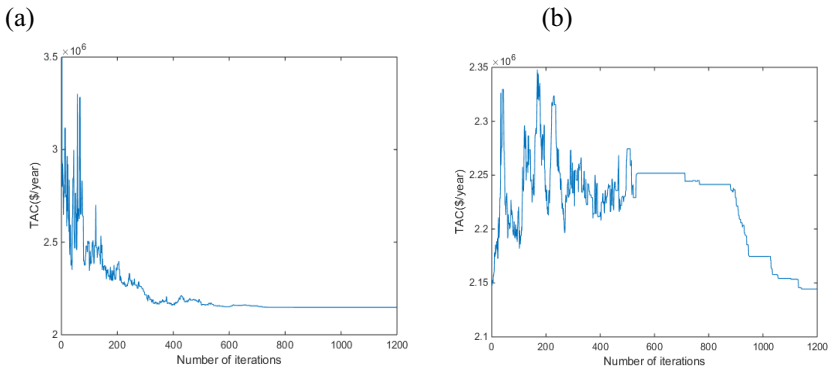


Figure 4. Results of the simulated annealing algorithm

Table 2. Optimal design result of SA algorithm in each turn

	SA(first turn)	SA(second turn)
B1P	10	10
B2P	0.19	0.19
N1	34	30
B1FS	12	10
RS	21	16
N2	23	24
B2FS	15	16
Vacuum cost($10^6$$/y)	0.2431	0.2428
TAC(10^6/y)	2.1478	2.1443

5. Conclusions

A pressure-swing distillation process for the separation of n-Hexane and Ethyl Acetate has been designed and optimized using simulated annealing. The process is modelled in Aspen Plus, and simulated annealing algorithm is implemented in Matlab. Connection between these two programs is facilitated by the Aspen Plus COM server.

The TAC of the pressure-swing distillation process is 2.14 million USD, compared with 0.92–1.76 million USD for entrainer-based separation alternatives depending on the choice of entrainer (Yang, 2017). Although the pressure-swing process is more expensive, it benefits from the fact that no additional component must be added to the system.

References

Acosta, J., Arce, A., Martinez-Ageitos, J., Rodil, E., Soto, A., 2002. Vapor-liquid equilibrium of the ternary system ethyl acetate plus hexane plus acetone at 101.32 kPa. Journal of Chemical and Engineering data, 47, 849-854.

An, W. Z., Yuan, X. G., 2009. A simulated annealing-based approach to the optimal synthesis of heat-integrated distillation sequences. Computers & Chemical Engineering, 33, 199-212.

Gorbunova, L. V.; Lutugina, N. V.; Malenko, Y. I. Zh. Prikl. Khim. (Leningrad), 1965, 38, 374-7 Relations between the boiling points and composition in binary systems formed by acetic acid, ethyl acetate, methyl ethyl ketone and hexane

Luyben, W. L. (2006). Distillation design and control using Aspen simulation. Hoboken, N.J.: Wiley-Interscience.

Luyben, W. L., Chien, I. L. (2010). Design and control of distillation systems for separating azeotropes. Hoboken, N.J.: John Wiley.

Rodriguez, I., Jauregui, U., Alvarez, J. C., Pardillo, E., 1999. Separation of the n hexane - Ethyl acetate mixture by azeotropic distillation. Latin American Applied Research, 29, 119-127.

Rodriguez-Donis, I., Acosta-Esquijarosa, J., Gerbaud, V., Pardillo-Fondevila, E., Joulia, X., 2005. Separation of n-hexane-ethyl acetate mixtures by azeotropic batch distillation with heterogeneous entrainers. Chemical Engineering and Processing, 44, 131-137.

Seider, W. D., Seider, W. D., Lewin, D. R., Widagdo, S. (2009). Product and process design principles : synthesis, analysis, and evaluation (3rd ed.). Hoboken, NJ: Wiley.

Wang, Y. L., Bu, G. L., Wang, Y. K., Zhao, T. R., Zhang, Z., Zhu, Z. Y., 2016. Application of a simulated annealing algorithm to design and optimize a pressure-swing distillation process. Computers & Chemical Engineering, 95, 97-107.

Yang X.-L. 2017. Extractive Distillation Optimization by Using Aspen Plus ActiveX Automation Server. Master's Thesis. National Taiwan University Taipei, Taiwan.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50016-1

Design and Economic Evaluation for Production of Ethyl Lactate via Reactive Distillation Combined with Various Separation Configurations

Shi-Bao Dai [a], Hao-Yeh Lee [b+], Cheng-Liang Chen [a*]

[a]*Department of Chemical Engineering, National Taiwan University, Taipei 10617, Taiwan*
[b]*Department of Chemical Engineering, National Taiwan University of Science and Technology, Taipei 10607, Taiwan*
haoyehlee@mail.ntust.edu.tw, CCL@ntu.edu.tw

Abstract

Ethyl Lactate (L_1E) has been holding as a promising replacement for the petroleum-based solvent in world markets because of its biodegradable and nontoxic identities. However, the process design for production and separation of L_1E is challenging due to its complex kinetics. This work provides a modified design by using a side draw to obtain high purity L_1E from the 1st reactive distillation (RD) column in reactive part. Besides, several separation configurations recycling reactants are implemented to investigate the most economic process design. The final result shows that the heat duty saved as much as 22.3 % in the reactive part by using a side draw. Also, the pervaporation (PV) module in the separation part is more economical than extractive distillation (ED) in terms of total annual cost (TAC).

Keywords: Ethyl Lactate; Process design; Reactive distillation; Pervaporation.

1. Introduction

Ethyl lactate is a common solvent in chemical industry. It is usually produced through the main esterification reaction of lactic acid (L_1) and ethanol (EtOH). However, a key issue is that when the reactant concentration of lactic acid is higher than 20 wt%, several oligomeric reactions, such as reactions (2) and (3) in Table 1 will appear. Consequently, process design definitely becomes more challenging.

Table 1. The overall reaction paths for L_1E system

Main reaction	$L_1 + EtOH \rightleftharpoons L_1E + H_2O$	(1)
Oligomeric side reaction (a)	$2L_1 \rightleftharpoons L_2 + H_2O$	(2)
Oligomeric side reaction (b)	$L_1 + L_2 \rightleftharpoons L_3 + H_2O$	(3)

RD is a technique which combines reaction and separation sections into a single unit thus can highly reduce capital cost. Meanwhile, high conversion and selectivity can be achieved by shifting the chemical equilibrium boundaries. Gao et al. (2007) reported a paper on the L_1E process by using single reactive distillation column. Whereas the process is on an experimental scale and the reaction kinetics only considered the main esterification of L_1 with EtOH. Rattanaphanee (2015) provided a process regarding commercial scale production of L_1E, which consists one RD column and three separation columns. Nevertheless, the kinetics described in the process are also too

simplified that neglected the oligomeric reactions. Miller et al. (2010) first disclosed a commercial scale process concept with completed reaction kinetics to produce L_1E. However, it required two RD columns and one product separation column in the reactive part, which costs a lot. Furthermore, they did not specify which configuration would be implemented in the downstream separation part.

This research centers on the steady-state design and economic comparison of different separation techniques of the commercial L_1E process. To be more specific, two separation techniques, ED and PV are implemented to find the lowest operating and fixed cost.

2. Model Building

2.1. Thermodynamic Property

Nonrandom two liquids (NRTL) activity coefficient model is used to account for the non-ideal vapor-liquid equilibrium (VLE) and possible vapor-liquid-liquid (VLLE) phase behaviors in this system. Also, to account for the dimerization and trimerization of lactic acid in the vapor phase, the second viral coefficients of Hayden -O'Connell (1975) is used. Neither VLLE nor LLE are found in the system. Temperature and composition of azeotropes are given in Table 2.

Table 2. Composition and temperature of the azeotropes for L_1E system (at 1 atm)

	Experimental		Computed	
	Composition	Temp. (°C)	Composition	Temp. (°C)
H_2O/ EtOH	(0.106, 0.894)	78.12	(0.112, 0.888)	78.12
H_2O/ L_1E	-	-	(0.969, 0.031)	99.85
L_1/ L_2	-	-	(0.404, 0.596)	215.38

2.2. Reaction Kinetics

The kinetic parameters from Su et al. (2013) are provided in Table 3. For this system, Amberlyst 15 is used as the catalyst. It should be noticed that this reaction kinetic model is catalyst-weight-based (m_{cat}). Thus, the convert of the tray volume between catalyst weight is necessary. One can solve by assuming that the solid catalyst occupies 50% of the liquid holdup in RD column trays and the density of the catalyst is 770 kg/m^3.

Table 3. Kinetic model for L_1E system

$$r_{L_1E} = m_{cat}(k_{f,L_1E}x_{L_1}x_{EtOH} - k_{r,L_1E}x_{L_1E}x_{H_2O})$$

$$k_{f,L_1E} = 6.52 \times 10^3 \exp\left(-\frac{48000}{RT}\right) \qquad k_{r,L_1E} = 2.72 \times 10^3 \exp\left(-\frac{48000}{RT}\right)$$

$$r_{L_2} = m_{cat}(k_{f,L_2}x_{L_1}x_{L_1} - k_{r,L_2}x_{L_2}x_{L_2})$$

$$k_{f,L_2} = 1.10 \times 10 \exp\left(-\frac{52000}{RT}\right) \qquad k_{f,L_2} = 5.54 \times 10 \exp\left(-\frac{52000}{RT}\right)$$

$$r_{L_3} = m_{cat}(k_{f,L_3}x_{L_2}x_{L_1} - k_{r,L_3}x_{L_3}x_{H_2O})$$

$$k_{f,L_3} = 4.56 \times 10^{-1} \exp\left(-\frac{50800}{RT}\right) \qquad k_{r,L_3} = 2.28 \times 10 \exp\left(-\frac{50800}{RT}\right)$$

r_i (kmol/s), m_{cat} (kg_{cat}), k_i (kmol/(kg_{cat}s)), R = 8.314 (kJ/(kmol/K)), T (K), x_i (mole fraction)

2.3. Pervaporation Model

In this work, a simpler lumped system model (Luyben, 2009) is used. It would be done by dividing the membrane unit into several equal size units which called lumps. Moreover, material and energy balances (Eq.s (4) to (6)) for each lump are applied.
Total mass balance at the steady state:

$$\frac{dM_R}{dt} = 0 = F_{R,n-1} - F_{R,n} - F_{P,n} \quad (4)$$

where M_R is the molar holdup in the membrane, F_n is molar flowrate from cell n, and the subscript R and P are represented as retentate or permeate side respectively. Component mass balance for species i:

$$M_R \frac{dZ_{R,n,i}}{dt} = F_{R,n-1}Z_{R,n-1,i} - F_{R,n}Z_{R,n,i} - F_{P,n}Z_{P,n,i} \quad (5)$$

where z_i is the molar fraction for species i. Energy difference of the inlet flow and outlet flow:

$$M_R \frac{dh_{R,n}}{dt} = F_{R,n-1}h_{R,n-1} - F_{R,n}h_{R,n} - F_{P,n}H_{P,n} \quad (6)$$

where h is the molar enthalpy of liquid in retentate side, H is the molar enthalpy of vapor in permeate side. The permeation rate of water (J_w) and other species i (J_i) are depicted below:

$$J_w = k_a \exp\left(\frac{-E_D}{RT}\right)\left[\exp\left(k_b w_{w,f}\right) - 1\right] \quad (7)$$

$$J_i = k_a \exp\left(\frac{-E_D}{RT}\right)\exp\left(k_b w_{w,f}\right) \quad (8)$$

Table 4. Pervaporation parameters for L_1E system

Component	k_a ($kgh^{-1}m^{-2}$)	E_D ($kJmol^{-1}$)	k_b ($kgh^{-1}m^{-2}$)
H_2O	1.19×10^7	49.96	2.17
EtOH	4.25×10^5	51.41	8.10
L_1E	1.93×10^3	40.93	9.58
L_1	9.72×10^8	76.89	6.34

The pervaporation parameters shown in Table 4 is taken from the paper published by Delgado et al. (2009). It is worth to notice that the membrane is really unfavorable for organic components. Also, the feed stream composed relatively small amount of L_2 and L_3 in the system. Therefore, the flux of L_2 and L_3 can be reasonably neglected.

3. Steady-state Design

The process designed here is to construct the commercial L_1E production process. The commercial simulator, Aspen Plus v9.0 is used for simulation. The membrane module for PV is developed and implemented into Aspen Plus interface via Aspen Custom Modeller. The annual L_1E production rate is set as 25 million pounds (roughly as 13.88 mole/min). The specification of L_1E is 0.990 (molar basis) to meet the product requirement. Besides, the purity of three feed streams, namely, EtOH, L_1, and H_2O are 0.900, 0.152 and 0.995 (molar basis), respectively.

3.1. Reactive Part

Figure 1 shows the base case simulation results of the reactive part from the process concept of Miller et al. (2010). It consists of two RD columns as denoted as RDC1 and

RDC2 and one conventional distillation column (C1) separating L_1E from the top. By analyzing the composition profile in RDC1, we found that the highest purity of the desired product, L_1E does not appear in the bottom. Instead, it arises on the 57th stage and shows the purity of 0.964 that is higher than 0.920 in the bottom stream. Therefore, it can be fairly considered to draw a side stream from the middle of RDC1 to replace C1.

Figure 2 shows the modified design of the reactive part. The product (L_1E) is taken from the 52th stage of RDC1 as a side draw. Additionally, by changing the operating condition, such as the reboiler duty, L_1E in the side draw can meet our specification of 0.990. By comparing this novel configuration with the original design, the energy has been saved for about 22.3 %. The Mix stream that composed mainly of EtOH and H2O will be sent to different configurations for further separation.

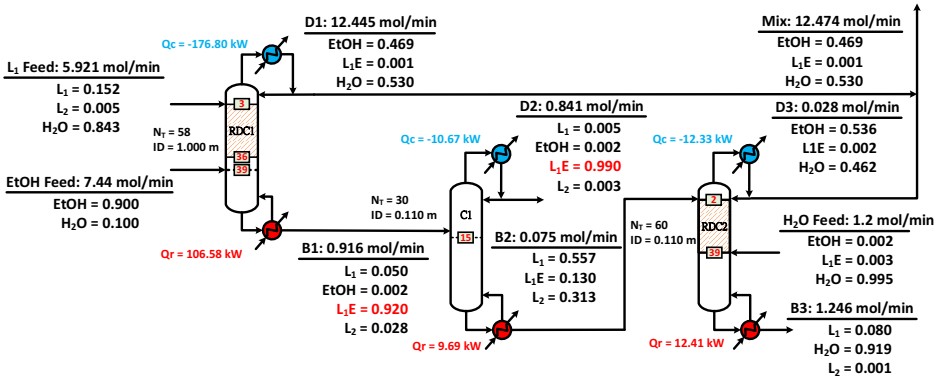


Figure 1. The simulation results from the process concept of Miller et al. (2010)

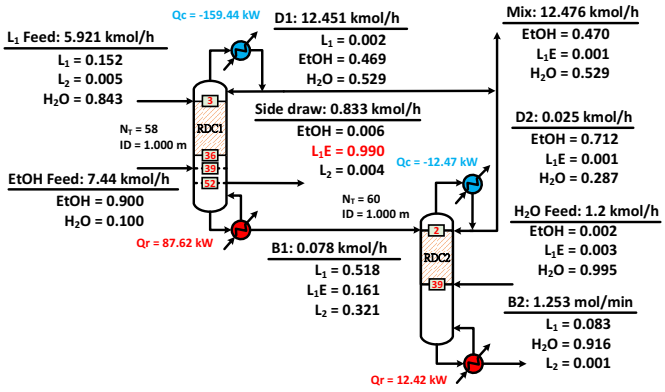


Figure 2. The modified process for reactive part

3.2. Reactive Distillation with Different Separation Configurations

In this section, we investigate the performance of different separation configurations. From the process described in Section 3.1, we will obtain a stream that composed mainly of EtOH and H_2O (stream Mix in Figure 3). Since EtOH will form a typical azeotrope with H_2O, it is impossible to obtain pure EtOH by using a single distillation

column. Accordingly, conventional separation technique called extractive distillation (ED) is being used. On the other hand, novel pervaporation module is also being utilized to find further advantages compared to ED. Figure 3 shows the RD + ED configuration. Since glycerol is non-toxic, it is being used as the entrainer in this system. C1 and C2 in this figure represent EtOH separation column and entrainer recycle column, respectively. Because the system will generate a large amount of H_2O, part of the purified H_2O from C2 will be sent as the reactant for RDC2. Figure 4 illustrates the RD + PV configuration. For physical limitation of the membrane, it is necessary to implement C1 to pre-concentrate EtOH in advance. Then the concentrated stream (D3 in Figure 5) will be sent to 4 parallel membranes. The purified EtOH from the retentate stream will be sent back to the system as the reactant. On the contrary, the permeate will be sent to C1 for further purification of H_2O.

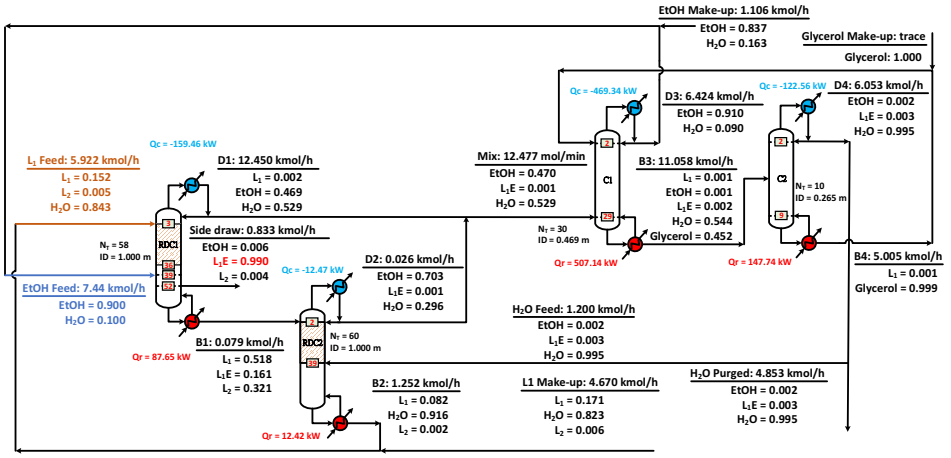


Figure 3. Reactive distillation with extractive distillation (RD + ED) configuration

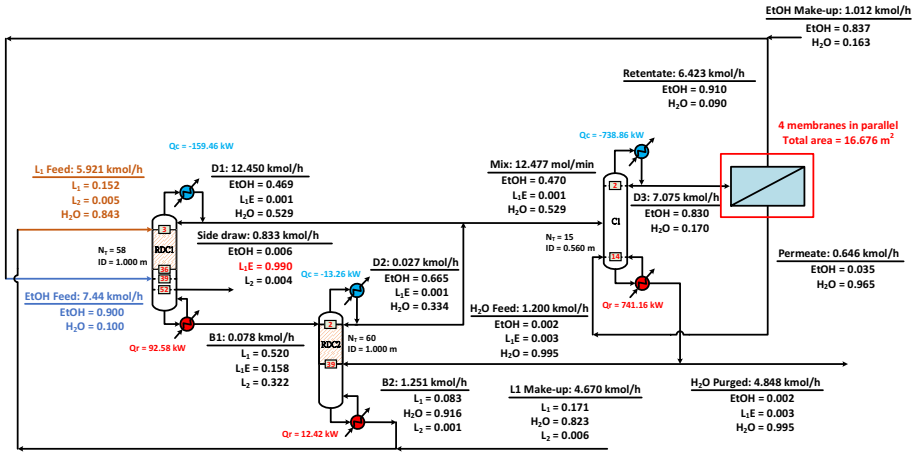


Figure 4. Reactive distillation with pervaporation (RD + PV) configuration

3.3. Economic Analysis

The calculation of TAC includes annual operating cost (AOC) and annual total capital cost (TCC) which is shown in Eq.(9). The payback period is set as 3 years. Table 5 lists the TAC of two configurations. It is clearly that the RD + PV is much more competitive than RD + ED since the TAC can be saved for about 56,000 USD.

$$\mathrm{TAC} = \mathrm{AOC} + \frac{\mathrm{TCC}}{\text{payback period}} \tag{9}$$

Table 5. Comparison of TAC between two configurations

Configuration	Fixed Cost (1000 USD)	Operating Cost (1000 USD)	TAC (1000 USD)
RD + ED	1173.84	32.79	1206.63
RD + PV	1121.47	29.12	1150.59

4. Conclusion

Two commercial L_1E production processes had been constructed. For the reactive part, taking a side draw to replace product column can save 22.3 % energy. For the separation part, PV shows more potential than traditional ED process because it would reduce TAC for around 56,000 USD.

References

P. Delgado; M. T. Sanz; S. Beltrán, 2009, Pervaporation of the Quaternary Mixture Present during the Esterification of Lactic Acid with Ethanol, J. Membr. Sci., 332, 113-120

J. Gao; X. M. Zhao; L. Y. Zhou; Z. H. Huang, 2007, Investigation of Ethyl Lactate Reactive Distillation Porcess, IChemE., 85, 525-529

B. H. Lunelli; E. R. de Morais; M. R. W. Maciel; R. M. Filho, 2011, 10th International Conference on Chemical and Process Engineering, 24

W. L. Luyben, 2009, Control of a Column/ Pervaporation Process for Separating the Ethanol/ Water Azeotrope, Ind. Eng. Chem. Res., 48, 3484–3495

D. J. Miller; N. Asthana; A. Kolah; D. T. Vu; C. T. Lira, 2010, Process for Production of Organic Acid Esters, U. S. Patent 7,652,167, January 26

C. S. M. Pereira; V. M. T. M. Silva; A. E. Rodrigues, 2011, Ethyl Lactate as A Solvent: Properties, Applications and Production Processes – A Review, Green Chem., 13, 2658-2671

C.-Y. Su; C.-C. Yu; I-L. Chien; J. D. Ward, 2013, Plant-Wide Economic Comparison of Lactic Acid Recovery Processes by Reactive Distillation with Different Alcohols, Ind. Eng. Chem. Res., 52 (32), 11070–11083

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50017-3

Potentials for CO_2 Utilization: Diethyl Carbonate Synthesis from Propylene Oxide

Meng-Kai Chen and I-Lung Chien*

Department of Chemical Engineering, National Taiwan University, Taipei 10617, Taiwan
ilungchien@ntu.edu.tw

Abstract

Production of dimethyl carbonate (DMC) has received much attention in CO_2 utilization recently. The conventional route of DMC synthesis from CO_2 is converting CO_2 with ethylene oxide into ethylene carbonate and followed by transesterification with methanol. However, the azeotrope of methanol-DMC mixture makes separation system energy intensive. In this paper, CO_2 utilization to instead produce diethyl carbonate (DEC) is investigated. Fortunately, because ethanol-DEC mixture would not generate azeotrope, the energy consumptions of this CO_2 utilization route to produce DEC is much less than the DMC route.

Three alternative processes to produce DEC are rigorously designed in this paper. The first one is the conventional two-step process to produce DMC and then followed by transesterification of DMC with ethanol to produce DEC. Two other alternative DEC synthesis routes are proposed: (1) converting CO_2 with propylene oxide into propylene carbonate and then transesterification of propylene carbonate with ethanol, (2) one-pot synthesis of converting CO_2, propylene oxide and ethanol into DEC. These two routes would not involve DMC, therefore, separation is easier. For these two alternative routes, propylene glycol is the by-product instead of ethylene glycol in the conventional process. Rigorous simulation of three routes and CO_2 emission calculation are performed in Aspen Plus V8.8. Due to lack of reaction kinetics of one-pot reaction, hypothetical simulation and discussion are made based on the experimental data from a recent paper.

Result shows that the first alternative route avoids 46.0 % of CO_2 emission per unit weight of DEC produced as compared with the conventional process. In addition, it saves 24.0 % for capital cost and 35.3 % for utility cost. The one-pot synthesis route also has potential to reduce CO_2 emission of 17.7 %.

Keywords: CO_2 utilization, diethyl carbonate, propylene oxide, process design.

1. Introduction

Transformation of captured CO_2 into valuable compounds is one way to reduce the global CO_2 emission. The utilization of CO_2 could avoid part of the usage of petrochemical, and the profit from the valuable product could compensate the cost of the capture process. Urea, polycarbonate, methanol, dimethyl ether and dimethyl carbonate (DMC) are usually considered as the potential target compounds. In the process of DMC synthesis from CO_2, there are two major options without involvement of toxic phosgene, direct synthesis and two-step transesterification. For direct synthesis process, methanol (MeOH) and CO_2 are converted in one reactor, producing water and

DMC. For two-step transesterification process, cycloaddition of CO_2 with epoxide to cyclic carbonate is followed by transesterification between cyclic carbonate and methanol, producing diol and DMC. There are several challenges of direct synthesis process, such as low yield and thermodynamic limitation. Compare with direct synthesis process, two-step transesterification process is more realizable recently. However, azeotrope of methanol-DMC mixture makes separation system energy intensive.

The major application of DMC are fuel additive, solvent and organic feedstock for other organic carbonate, such as diethyl carbonate (DEC), diphenyl carbonate and polycarbonate. Among of those derivative organic carbonates, DEC has similar or even better performance than DMC in the application field. In the view of fuel additive, although the oxygen content of DEC (40.7 wt %) is lower than DMC (53.3 wt %), it is still much higher than methyl tert-butyl ether (MTBE) (18.2 wt %). For octane number, both of DMC and DEC are higher than 100. The distribution of DEC in gasoline/water two-phase system is much better than that of DMC (Pacheco et al., 1997). In addition, DEC could not only be used as gasoline additive but also as diesel additive, which could reduce particulate emission about 30 % with lower than 10 % CO, hydrocarbon and NOx additional emission (Kozak et al., 2012).

Comparing with conventional DEC synthesis from transesterification of DMC with methanol, transesterification of cyclic carbonate with ethanol (EtOH) is an alternative way. The major advantage of this synthesis route is that there is no azeotrope between ethanol and DEC, as shown in Figure 1. Therefore, the separation is much easier than the separation of methanol/DMC mixture.

In recent study, DEC synthesis from propylene oxide (PO), CO_2, ethanol over a heterogeneous catalyst, cross linked poly(ionic liquid) integrated with MgO, has been investigated (Wang et al., 2017). This work demonstrates a one-pot synthesis composed of cycloaddition and transesterification in one reactor, which might make DEC synthesis process simpler.

In this study, three the above-mentioned DEC synthesis routes from PO are discussed and designed. Those process steps which have been studied and designed in the existing literature are reproduced in the same product yield basis to make a fair comparison. CO_2

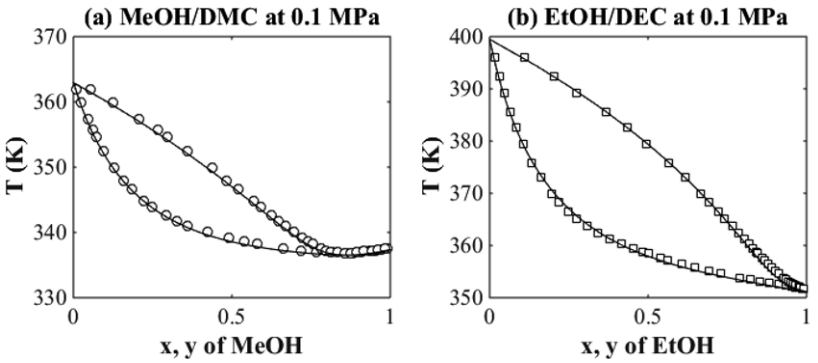


Figure 1. Boiling temperature (T) diagram for binary systems (a) methanol/DMC and (b) ethanol/DEC at 0.1 MPa. (-) UNIQUAC model, (○) experimental data from Rodríguez et al. (2002), (□) experimental data from Rodríguez et al. (2003).

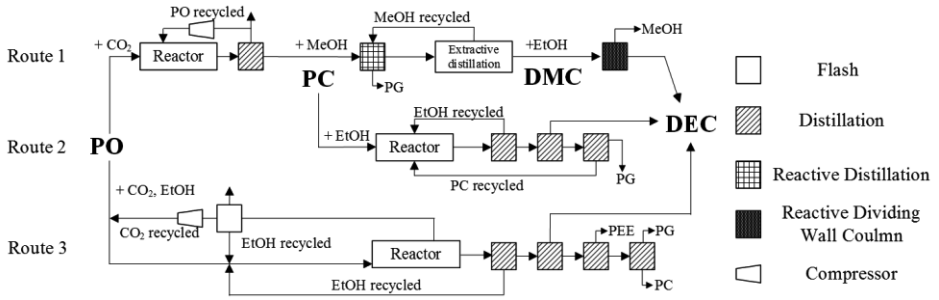


Figure 2. Illustration of three DEC synthesis routes.

emission calculation and economic analysis are applied. Due to lack of reaction kinetics of one-pot synthesis, hypothetical simulation and discussion are made based on the experimental data from Wang et al. (2017).

2. Process Description

Figure 2 shows the illustration of three DEC synthesis routes in this study. In all the three routes, PO is used as reacting agent to convert CO_2 and generate propylene glycol (PG) as valuable by-product. The process basis for this study is 20 kmol/hr CO_2 consumption as raw material (7.04 kt/y). All the product purity specifications are 99.5 mol%. All the operating pressures are 0.1 MPa if not specified.

2.1. Route 1

In Route 1, three steps are required to produce DEC. Among of them, the processes of transesterification of propylene carbonate (PC) with methanol (Holtbruegg et al., 2015) and transesterification of DMC with ethanol (Zheng et al., 2017) are already studied in the previous literature. The former applies reactive distillation first and pressure swing distillation to separate methanol and DMC. Reactive distillation could make almost 100% conversion of PC with excess methanol. However, the azeotropic composition of methanol/DMC mixture does not change obviously with pressure (93.7 mol% MeOH at 1 MPa and 86.5 mol% MeOH at 0.1 MPa), and Wang et al. (2010) shows that extractive distillation with phenol as entrainer is a better alternative to separate methanol/DMC azeotrope rather than pressure swing distillation. Furthermore, Hsu et al. (2010) found that extractive distillation with aniline as entrainer performs better than phenol. Therefore, a revised process consisting of reactive distillation and extractive distillation with aniline as entrainer is applied in this study.

Reactive dividing wall column is applied for transesterification of DMC with ethanol by Zheng et al. in 2017. The reactive distillation of DMC and ethanol is combined with the separation of methanol and ethanol. In this study, all key operating conditions follows Zheng et al.'s work (2017).

For cycloaddition of CO_2 and PO, kinetic model is studied with DEAE ion-exchange catalyst in a semi-batch slurry reactor (Jin et al., 2017). Modified Eley-Rideal type mechanism is used to describe this liquid phase reaction. First, PO and CO_2 (with 10% excess) are both compressed and sent to a tubular reactor at 95 °C with 10% bypass. The discharge from reactor is sent to a stripping column to purify PC, and unreacted CO_2 is compressed and sent back to reactor. A purge from recycled gas is needed to avoid

accumulation of CO_2. RK-SOAVE model is applied to describe phase equilibrium in this reaction system, and UNIQUAC model is applied for all other downstream systems.

2.2. Route 2

The first step in Route 2 is the same one in Route 1. As mentioned earlier, the process of converting PC and ethanol into DEC and PG in thermally-coupled reactive distillation column has been studied. In this quaternary system, two products are intermediate boiling between two reactants, which is the most difficult type to apply reactive distillation. Fortunately, the minimum azeotrope between methanol and DMC makes sure DMC could be withdrawn from distillate with excess methanol, and PG is obtained from bottom. The existence of minimum azeotrope breaks the bottleneck of reactive distillation for this type of boiling point rank. However, there is no azeotrope between ethanol and DEC (shown in Figure 1.), therefore, the reactive distillation structure should not be the one proposed by Wang et al. in 2014, which uses the same structure as transesterification between PC and methanol. Therefore, conventional reactor-distillation sequence is applied in this study rather than their proposed process.

PC and excess ethanol are sent to reactor at 30 ºC first and purified by the following separation section. Separation comprises direct sequence of three columns. Ethanol, DEC, PG are obtained respectively from three distillates, and PC is obtained from the bottom of the last column. Unreacted ethanol and PC are recycled back to reactor.

2.3. Route 3

One-pot synthesis between PO, ethanol, CO_2 has been studied by Wang et al. in 2017. Reaction conditions in the experiment are 0.510 mol ethanol, 0.035 mol PO, 3 MPa CO_2 initial pressure, 160 ºC and reaction time is 4.5 h. Under this condition, yields for DEC, PG, PC are 44.4 %, 56.6 %, 21.3 %, respectively, and 96.0 % PO conversion are reached. The difference between the yields of DEC and PG is due to hydrolysis of PO. In this system, propylene glycol monoethyl ether (PEE) with 7.4 % yield is another by-product from reaction between PC and ethanol.

Due to no kinetic model of this reaction system can be found in previous literatures, simulation of this synthesis route follows the same yields, selectivity and operating condition in the experiment. Therefore, it should be noticed that this process is merely a hypothetical simulation, and we could not alter the operating condition carelessly. Reactor outlet, ethanol and CO_2 in major, is flashed, and most of vaporized ethanol is condensed and sent back to the reactor. The remaining CO_2 is recompressed and then recycles to reactor. A purge from recycled gas is needed to avoid accumulation of CO_2. Separation also follows direct sequence of four distillation columns. Ethanol, DEC, PEE, and PG are obtained respectively from the four distillates. Excess ethanol is recycled back to reactor while PC is regarded as another by-product and not recycled in this route.

Table 1. Equivalent CO_2 emission amount and price of utility

Utility	Equivalent CO_2 emission (kg/GJ)	Price (USD/GJ)
Steam, 0.5 MPa, 160 ºC	72.86	13.28
Steam, 1.0 MPa, 184 ºC	76.60	14.19
Steam, 4.2 MPa, 254 ºC	91.14	17.70
Electricity	96.95	16.80
Cooling water	-	0.354

The yield of DEC is lower than the previous two routes because part of the intermediate, PC, inevitably remains unconverted in one-pot synthesis.

3. Carbon Emission and Economic Analysis

Carbon emission considered in this study is from steam and electricity. Three different grades of steam are used referring to the book from Turton et al. in 2009. The corresponding CO_2 emission amount follows the calculation method proposed by Gadalla et al. in 2005 and listed in Table 1.

The evaluations of capital cost and operating cost follow the book from Turton et al. in 2009. Only the first two synthesis routes are considered because Route 3 is a hypothetically scaled-up process. The cost of reactor could not be estimated because of the missing information of kinetic model.

4. Result and Discussion

In this study, only the significant variables are optimized, such as temperature in reactor, alcohol excess ratio, and operating pressure in the distillation column for alcohol separation. The result of CO_2 emission calculation and economic analysis for three alternatives DEC synthesis routes are summarized in Table 2. Even though all three routes emit CO_2 more than consumption, it should be noticed that the avoided CO_2 emissions in Route 2 and Route 3 revealing the potential of emission reduction.

The cycloaddition step accounts for minor CO_2 emission, and the following transesterification reaction emits tens of times more. In the transesterification step, alcohol is excess to make sure enough conversion of carbonate (PC or DMC). There is a tradeoff between conversion in reactor and energy consumption. The more alcohol used, the more alcohol should be vaporized and removed from product (DMC or DEC), which accounts for the major energy requirement. The optimal alcohol excess ratios in the transesterification reactions are 10.5 (PC to DMC), 6 (DMC to DEC) and 8 (PC to DEC). Additionally, the EtOH excess ratio in Route 3 is 14.6.

Because Route 2 uses one less transesterification step and lower alcohol excess ratio than Route 1, the advantage of Route 2 is obvious. Not only less amount of utility is required, but less capital investment in Route 2. Additionally, no azeotrope in EtOH/DEC mixture makes Route 2 more appealing. However, even though Route 3

Table 2. Result summary

	Route 1	Route 2	Route 3
Utility cost (kUSD/y)	6260 (0 %)	4050 (-35.3 %)	3870 (-38.2 %)
Capital cost (kUSD)	8170 (0 %)	6210 (-24.0 %)	-
CO_2 emission (kt/y)	Step 1: 0.97 Step 2: 19.75 Step 3: 10.79	Step 1: 0.97 Step 2: 19.30	20.67
CO_2 consumption (kt/y)	7.04	7.04	7.04
Net CO_2 emission (kt/y)	24.46	13.22	13.63
DEC production (kt/y)	18.89	18.89	12.79
Net CO_2 emission (kg CO_2/kg DEC)	1.30 (0 %)	0.70 (-46.0 %)	1.07 (-17.7 %)

further simplifies the synthesis process, large alcohol excess ratio and by-product generated still makes it not favourable in the reduction of CO_2 emission. Even so, if some catalyst could maintain the selectivity of one-pot reaction with less alcohol excess in the future, the potential for CO_2 emission reduction might be more significant.

5. Conclusions

In this study, three synthesis routes from PO to DEC are designed, and carbon emission for these three routes are compared to discuss the potential of CO_2 reduction in DEC production. The transesterification reaction accounts for the majority of energy requirement. Result shows that 46.0 % and 17.7 % net CO_2 emission reduction by Route 2 and Route 3, respectively, as compared with Route 1. In addition, Route 2 saves 24.0 % capital cost and 35.3 % utility cost, which shows that Route 2 not only has the potential of CO_2 emission reduction but is more economically favourable.

References

A. Rodríguez, J. Canosa, A. Domínguez, J. Tojo, 2002, Vapour–liquid equilibria of dimethyl carbonate with linear alcohols and estimation of interaction parameters for the UNIFAC and ASOG method, Fluid Phase Equilib., 201, 187–201

A. Rodríguez, J. Canosa, A. Domínguez, J. Tojo, 2003, Isobaric Phase Equilibria of Diethyl Carbonate with Five Alcohols at 101.3 kPa, J. Chem. Eng. Data, 48, 86-91

J. Holtbruegge, H. Kuhlmann, P. Lutze, 2015, Process analysis and economic optimization of intensified process alternatives for simultaneous industrial scale production of dimethyl carbonate and propylene glycol, Chem. Eng. Res. Des., 93, 411–431

K. Y. Hsu, Y. C. Hsiao, I. L. Chien, 2010, Design and Control of Dimethyl Carbonate-Methanol Separation via Extractive Distillation in the Dimethyl Carbonate Reactive-Distillation Process, Ind. Eng. Chem. Res., 49, 735–749

L. Wang, M. Ammar, P. He, Y. Li, Y. Cao, F. Li, X. Han, H. Li., 2017, The efficient synthesis of diethyl carbonate via coupling reaction from propylene oxide, CO2 and ethanol over binary PVEImBr/MgO catalyst, Catal. Today, 281, 360–370.

L. Zheng, W. Cai, X. Zhang, Y. Wang, 2017, Design and control of reactive dividing-wall column for the synthesis of diethyl carbonate, Chem. Eng. Process., 111, 127–140

M. A. Pacheco, C. L. Marshall, 1997, Review of Dimethyl Carbonate (DMC) Manufacture and Its Characteristics as a Fuel Additive, Energy & Fuels, 11, 2-29

M. A. Gadalla, Z. Olujic, P. J. Jansens, M. Jobson, R. Smith, 2005, Reducing CO2 emissions and energy consumption of heat-integrated distillation systems. Environ. Sci. Technol., 39 (17), 6860-6870

M. Kozak, J. Merkisz, P. Bielaczyc, A. Szczotka, 2012, Environmental Performance of Diesel Fuels Containing Oxygenated Additive Packages, Proceedings of the FISITA 2012 World Automotive Congress, 3, 227-238

R. Turton, R. C. Bailie, W. B. Whiting, J. A Shaeiwitz, 2008, Analysis, Synthesis, and Design of Chemical Processes. Pearson Education, Boston.

S. J. Wang, C. C. Yu, H. P. Huang, 2010, Plant-wide design and control of DMC synthesis process via reactive distillation and thermally coupled extractive distillation, Comput. Chem. Eng., 34, 361–373

S. J. Wang, S. H. Cheng, P. H. Chiu, K. Huang, 2014, Design and Control of a Thermally Coupled Reactive Distillation, Ind. Eng. Chem. Res., 53, 5982−5995

X. Jin, P. Bobba, N. Reding, Z. Song, P. S. Thapa, G. Prasad, B. Subramaniam, R. V. Chaudhari, 2017, Kinetic modeling of carboxylation of propylene oxide to propylene carbonate using ion-exchange resin catalyst in a semi-batch slurry reactor, Chem. Eng. Sci., 168, 189–203

Y. Zhang, Z. Ji, R. Xu, Y. Fang, 2012, Experimental and Kinetic Studies on a Homogeneous System for Diethyl Carbonate, Synthesis by Transesterification, Chem. Eng. Technol, 35, No. 4, 693–699

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50018-5

Superstructure-based Rigorous Simulation for Synthesis and Evaluation of Lignocellulosic Biofuels Processes

Paola Ibarra-Gonzalez[a], Carlo-Edgar Torres-Ortega[a], Ben-Guang Rong[a*]

[a]*Department of Chemical Engineering, Biotechnology and Environmental Technology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark. bgr@kbm.sdu.dk*

Abstract

In the present work, a process synthesis framework for the conversion of softwood biomass to liquid (BtL) transportation fuels was developed. We defined processing blocks out of promising thermochemical and upgrading technologies, and performed individual blocks rigorous simulations in Aspen Plus® V8.8. The simulations and experimental data taken from the literature were used to predict conversions, recovery factors, capital and energy costs of the processing blocks. From the simulations, it was found that Gasification-Low Temperature Fischer-Tropsch (LTFT) followed by fractional upgrading and fractionating units (BC-GLTUF), was the most cost-effective, in terms of total annual cost (TAC) and liquid fuels production. Then, given the preliminary results, the possibility of combining blocks between the thermochemical routes, as well as mass and energy integration were explored. A superstructure was proposed and defined as a Mixed Integer Non-Linear Programming (MINLP) problem coded in GAMS 24.5.6, which sets the objective to minimize the TAC´s of BtL fuels under different cases and integration scenarios. The results showed that a combined alternative of Gasification followed by simultaneous high and low temperature FT reactions (CA-GLTHT) could increase the liquids fuels production and product distribution from 87 % to 90 % (wt) and reduce the TAC in around 26 % if mass and energy integration are considered.

Keywords: thermochemical, biofuels, process synthesis, simulation, superstructure, MINLP.

1. Introduction

By the end of 2016 in the United States 95 % of petroleum usage went directly to transportation fuels and only 5 % of liquid transportation fuels production originates with renewable resources (EIA, 2017). Thermochemical conversion processes such as gasification, pyrolysis, supercritical fluid extraction and direct liquefaction can convert biomass into syngas, bio-oil, char and gaseous products (Demirbas, 2001). From these, bio-oil can be used as an intermediate step to produce transportation liquid fuels such as diesel and gasoline. However, the larger capital cost for BtL is a strong motivating factor for the development of new processes to refine the fuels at lower costs. For advanced biofuels the main factor is capital costs (35 % to 50 %), followed by feedstock costs (25 % to 40 %) (IEA, 2009).

Hence, a process synthesis framework for the conversion of softwood to gasoline and diesel was defined in order to reduce the capital and operating costs. This general strategy for BtL facilities development proposes a combined methodology of rigorous simulations

and a MINLP model. The optimization-based process synthesis strategy can simultaneously analyze distinct thermochemical-based process designs and determine the alternative that minimizes the overall softwood BtL fuels production system cost. For instance, mathematical modeling has been applied to a thermochemical-based refinery for the conversion of hardwood BtL (Baliban et al., 2013). However, to our knowledge there are not explicit works in the literature that consider simultaneously various thermochemical conversion routes, upgrading technologies and separation units using rigorous simulations as performance and cost sources.

2. Thermochemical-based superstructure base cases

We proposed through an optimization-based process synthesis framework, a thermochemical-based superstructure to convert softwood biomass (30 wt % moisture) to gasoline, diesel, and by-products. Norway spruce and Scots pine were selected as feedstocks. Two thermochemical routes were explored: Pyrolysis and gasification, since these are actively studied technologies and sufficient experimental data can be collected.

In the gasification routes, the pretreated biomass is sent to a gasification reactor, where it reacts with air, oxygen, or steam to produce a gaseous mixture known as raw gas (Mponzi, 2011). Then, the product raw gas is cleaned and the syngas obtained is sent to a Fischer-Tropsch reactor (FT). The FT process produces hydrocarbons from gasified biomass. Two main regimes have been used: LTFT that gives long chain molecules, and high temperature Fischer-Tropsch (HTFT), that gives shorter chain molecules (Maitlis and De Klerk, 2013). Before the upgrading, the FT products are split into fractions, since each fraction requires different upgrading (aromatic alkylation, gas oligomerization, naphtha aromatization, distillate hydrotreating, wax hydrocraking and aqueous products hydrogenation). Alternatively, the HTFT products can be upgraded by hydroprocessing: wax hydrocracking and distillate and naphtha product hydrotreating (Bechtel, 1998).

Regarding pyrolysis, biomass is heated in the absence of oxygen to produce pyrolysis oil also called bio-oil, which can be upgraded into chemicals and fuels. Upgrading bio-oil to a conventional transport fuel can be accomplished by two main routes: hydroprocessing or catalytic cracking over acidic zeolite catalysts (Czernik and Bridgwater, 2004).

In total five base cases consisting of different thermochemical conversion routes, upgrading technologies and separation units were proposed and evaluated: Pyrolysis-Hydroprocessing-Separation (BC-PHS), Pyrolysis-Catalytic Cracking-Fractionation (BC-PCCF), Gasification-LTFT-Fractional Upgrading-Fractionation (BC-GLTUF), Gasification-HTFT-Fractional Upgrading-Fractionation (BC-GHTUF), Gasification-HTFT-Hydroprocessing-Separation (BC-GHTH).

3. Methodology

3.1. Softwood BtL base cases: Synthesis and Conceptual Design

As initial analysis for the conceptual designs, each technological base case was primarily divided into three sections, namely, (1) thermochemical conversion: pyrolysis or gasification followed by LT and/or HT FT, (2) upgrading technology: hydroprocessing, catalytic cracking, oligomerization, aromatization, hydrocracking, isomerization, alkylation, and hydrotreating, and (3) separation section.

Section (1) converts softwood biomass into pyrolysis oil via pyrolysis or syngas via gasification. Then, the syngas from gasification is sent to the FT reactor for hydrocarbon

production via two types of FT reactions (HT and LT). In section (2), the pyrolysis oil is upgraded to biocrude via catalytic cracking or hydroprocessing. On the other hand, the FT product is sent to the fractional hydrocarbon upgrading or hydroprocessing units to produce syncrude. Finally, in section (3), the biocrude or syncrude is fractionated into gasoline, diesel, gases and aqueous products. The distinct process sections not only define the different thermochemical routes, but also determine the process technologies in terms of production costs as well as the final product profiles.

Then, the processing blocks conforming each technological section were defined and analyzed using the simulation software Aspen Plus® V8.8 to predict thermodynamic properties, conversion and separation factors, as well as capital and utilities costs of the unit operations. Simultaneously, liquid, solid and gas emissions were quantified, and from the total emissions, different possibilities of mass (water and gases) and energy (heat and power self-generation) integration were explored. From the data collected, process synthesis, integration and combination of processing blocks (performing similar tasks and presenting similar costs) into a superstructure were performed. Finally, an optimization model with an economic objective function solved through deterministic global optimization approaches evaluated the superstructure and selected the conversion route that reduces the TAC of BtL-fuels.

3.2. Superstructure formulation: definition of individual processing blocks and rigorous simulation.

3.2.1. Individual processing blocks definition

For the superstructure formulation, first the unit operations conforming the base cases´ technological sections were defined. Then, the unit operations were grouped into processing blocks (reaction and/or separation). A processing block is considered as a set of units where the mass composition of the feed stream changes. This is done to simplify the analysis of the processing blocks that can be combined into a superstructure. To exemplify the procedure, the formulation of reaction and separation blocks for the Pyrolysis-Hydroprocessing-Separation (BC-PHS) flowsheet is depicted in Figure 1.

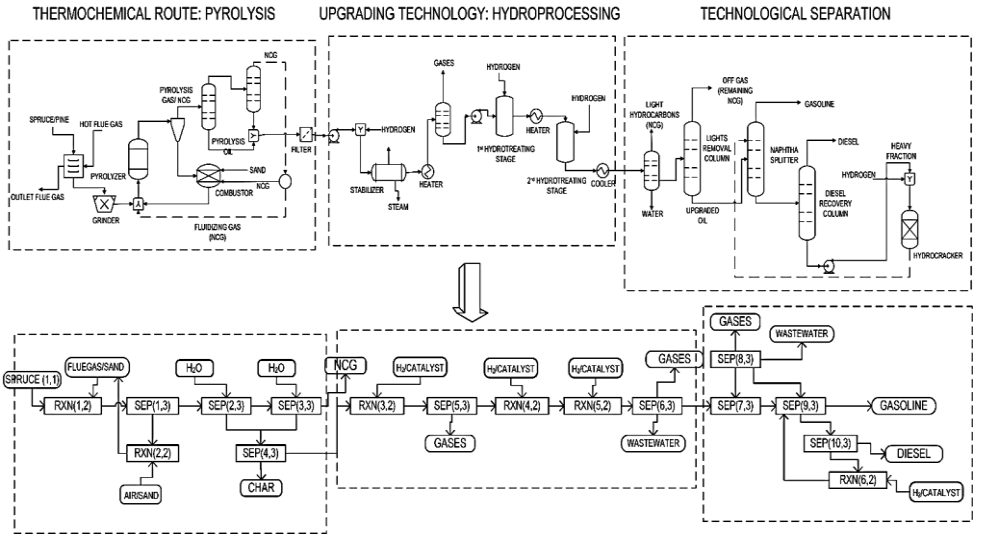


Figure 1. Formulation of processing blocks for BC-PHS: reaction (RXN) and separation (SEP)

3.2.2. Processing blocks simulation

The process simulator Aspen Plus® V8.8 was used to perform rigorous simulation of the pyrolysis, gasification, upgrading and separation blocks. The thermodynamic packages Soave-Redlich-Kwong equation of state with Kabadi-Danner mixing rules for the gasification route and Peng-Robinson for the Pyrolysis route were selected. The components not included in the Aspen database, such as, spruce, pine and ash were defined as non-conventional components based on their ultimate analysis including C, H, O, N, S, Cl and Ash elements, and proximate analysis (ECN, 1997). The biomass lower heating value (LHV) from the analyses was also specified with the HCOALGEN and DCOALIGT property models chosen to estimate the biomass enthalpy of formation, specific heat capacity and density. Each individual process setup was supported by literature and experimental data (operating conditions, conversions, representative components, and products obtained). Separation performance was based on the simulation´s thermodynamic models.

Aspen Process Economic Evaluator V8.8 was used to predict the energy consumption and equipment designs and costs of the reaction and separation blocks. The capital and utilities cost (electricity, heating and cooling) were calculated for a fixed capacity of 500 kg/h of spruce and pine.

3.2.3. Superstructure formulation

After the processing blocks definition and evaluation, it is possible to proceed with the formulation of the softwood BtL transportation fuels superstructure. The superstructure diagram defined is depicted in Figure 2.

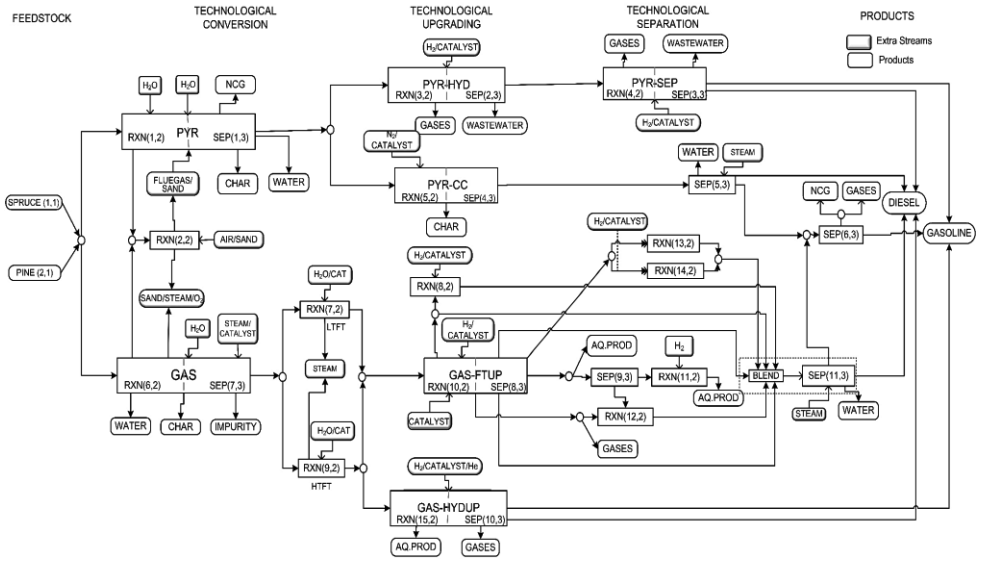


Figure 2. Thermochemical BtL process synthesis network superstructure

The possibilities of blocks combination between the base cases were analyzed by identifying blocks that perform similar tasks and present similar capital and energy costs. It was found that most of the blocks are sequential, so that we proposed lumped-process blocks performing major tasks like: biomass conversion to syngas via gasification RXN(6,2)-SEP(7,3); bio-oil via pyrolysis RXN(1,2)-SEP(1,3); pyrolysis oil catalytic

cracking RXN(5,2)-SEP(4,3); pyrolysis oil hydroprocessing RXN(3,2)-SEP(2,3); upgraded product separation: SEP(3,3), SEP(5,3), SEP(6,3) and SEP(11,3); and HTFT hydroprocessing RXN(15,2)-SEP(10,3). In addition, in the HTFT and LTFT hydrocarbons upgrading section, it was found that the reaction blocks aromatization and alkylation, as well as some separation units, perform similar tasks and present similar capital and energy costs. Therefore, these process blocks were combined into block RXN(10,2)-SEP(8,3), leaving outside reaction and separation blocks that are not in common between the base cases, as well as the ones that perform similar tasks but do not present similar costs.

3.3 Mathematical representation

We defined the synthesis network superstructure as a MINLP with the information collected from the rigorous simulations (thermodynamic properties, conversions, recovery factors, capital and energy costs). In previous works, it has been proven that MINLP can handle similar synthesis problems (Torres-Ortega et.al, 2017). The five base cases were studied with and without integrations, to produce diesel and gasoline simultaneously from one or two feedstocks. The objective was to analyze how these variables affect the product distribution and TAC. With this strategy, the MINLP problem explores the superstructure and proposes new pathways that increase the softwood conversion to transportation fuels and minimizes the TAC. The thermochemical-based processes´ TAC values were evaluated considering 10 years as time for recovering the investment. The objective function is depicted in Eq. (1):

$$minimize\ \ TAC = \sum \left[\left(\frac{Capital\ Cost}{Time\ of\ investment} \right) + Utility\ costs \right] \tag{1}$$

The advantages of the proposed methodology are consistency in the processing blocks parameters, individual processing block´s outputs and performance, as well as easier evaluation of the connectivity between processing blocks via MINLP.

4. Results

The superstructure was coded on GAMS 24.5.6 and the MINLP problem was solved by DICOPT and BARON solver. The five base cases described in previous sections were analyzed during the optimization and to test the performance of the model the desired final product requirements were specified. The MINLP solution is presented in Table 1 and additionally a combined alternative case consisting of Gasification-Simultaneous LT and HT FT-Fractional Upgrading-Separation (CA-GLTHT) was proposed and compared to the previous five cases. The individual process setup to reach the given profiles was confirmed by literature and experimental data. Likewise, each case was evaluated with and without mass and energy integration: gases and part of the water produced in each case were sent to a cogeneration plant for electricity and steam generation. The integration cases considered the cogeneration products, plant capital and energy costs.

From Table 1, we can observe that when the final biofuels profiles prefer more gasoline, BC-GHTH, BC-GHTUF and CA-GLTHT are the suitable technological routes. On the other hand, when the final biofuels profiles prefer more diesel, process routes BC-GLTUF and BC1-PHS are the suitable technological routes. Overall, the case with the highest liquid fuels production and the second lowest TAC was CA-GLTHT, achieving a TAC of 202,460.8 $/year, a product distribution of 69 % (wt) gasoline and 21 % (wt) diesel, and a TAC reduction of 26 % if mass and energy integration are considered.

Table 1. Case studies analyzed during the optimization

	BC-PHS	*BC-PCCF*	*BC-GLTUF*	*BC-GHTUF*	*BC-GHTH*	*CA-GLTHT*
Capital Cost [$]						
Conversion section	861,557.6	861,557.6	221,642.4	203,561.3	203,561.3	230,638.2
Upgrading tech.	904,703.2	198,875.5	161,743.8	313,825.2	226,000.0	182,077.6
Separation tech.	842,606.9	73,400.0	73,000.0	74,400.0	250575.6	74,400.0
Energy Cost [$/year]						
Conversion section	53,632.1	53,632.1	92,420.5	1,411,072	1,411,072	133,954.4
Upgrading tech.	141,082.3	50,748.6	38,078.4	39,598.8	39,103.4	39,598.8
Separation tech.	196,543.9	39,892.3	36,826.7	39,376.5	85,238.6	61,880.5
TAC [$/year]	652,145.1	257,656.3	212,964.2	1,549,225.8	1,603,427.7	284,145.3
TAC w/e-int. [$/year]	514,744.1	263,941.2	171,899.8	298,900.8	397637.2	202,460.8
WWD [kg/h]	18,932.0	19,154.4	9,953.0	2,552.2	2,098.0	9,601.9
Upgraded prod. [kg/h]	942	902	700	818	985	1,022
Total fuel prod. % (wt)	78	60	87	70	89	90
Gasoline % (wt)	54	45	38	64	84	69
Diesel % (wt)	24	15	49	6	5	21

5. Conclusions

In this work, a process synthesis framework for the conversion of woody BtL transportation fuels and simultaneous heat, power and mass integration, and blocks combination was developed. Multiple existing technologies including softwood gasification, pyrolysis, FT, upgrading and separation units were combined into a superstructure formulation. The rigorous simulations and MINLP approach solution were used to evaluate and compare the base cases and determine the alternative for liquid fuel production that minimizes the overall system cost. It was found that CA-GLTHT is the most favorable even when it presents a higher cost compare to BC-GLTUF. It produced more liquid fuels compare to the other cases and the TAC can be reduced if mass and energy integration are considered. It was demonstrated that this methodology applied to softwood BtL transportation fuels production can explore and propose optimal conversion pathways that reduce the costs of BtL-fuels and increase their viability.

References

R.C. Baliban, J.A Elia, C.A. Floudas, B Gurau, M.B. Weingarten, and S.D. Klotz, 2013. Hardwood Biomass to Gasoline, Diesel, and Jet Fuel: 1. Process Synthesis and Global Optimization of a Thermochemical Refinery. Energy & Fuels, 27, 4302-4324.

Betchel, 1998. Aspen Process Flowsheet Simulation Model of a Battelle Biomass-Based Gasification, Fischer-Tropsch Liquefaction and Combined-Cycle Power Plant. Topical report for U.S. Department of Energy.

S. Czernik and A.V.Bridgwater, 2004. Overview of applications of biomass fast pyrolysis oil. Energy & Fuels, 18(2), 590–598.

A. Demirbas, 2001. Biomass resource facilities and biomass conversion processing for fuels and chemicals. Energy Conversion and Management, 42 (11), 1357-1378.

EIA, 2017. Monthly Energy Review November 2017.

ECN Netherlands, 1997. ECN Phyllis classification: Untreated wood #113 & #166.

IEA, 2009. Transport, Energy and CO2: Moving toward Sustaintability.

P.M Maitlis and A De Klerk, 2013. Greener Fischer-Tropsch Processes for Fuels and Feedstocks. Somerset: John Wiley & Sons, Incorporated.

P. Mponzi, 2011. Master Thesis: Production of Biofuels by Fischer Tropsch Synthesis. Lappeenranta University of Technology.

C.E. Torres-Ortega, J. Gong, F. You, B.G. Rong, 2017. Optimal synthesis of integrated process for co-production of biodiesel and hydrotreated vegetable oil diesel from hybrid oil feedstocks. Computer Aided Chemical Engineering, 40, 673-678.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50019-7

Techno-economic Feasibility Study for Catalytic Production of 1,2-Pentanediol from Bio-renewable Furfural

Jaewon Byun[a,*], Yuchan Ahn[a], Juyeon Kim[a], Dongin Kim[a], Jeehoon Han[a,b]

[a]*School of Semiconductor and Chemical Engineering, Chonbuk National University, Jeonju 54896, Korea*
[b]*School of Chemical Engineering, Chonbuk National University, Jeonju 54896, Korea*
JWB@jbnu.ac.kr

Abstract

1,2-Pentanediol (1,2-PeD), which can be used as a monomer of polyester, has been produced from fossil resource by environmentally-unfriendly process using toxic solvents (propionic acid and benzene). We present the environmentally-friendly process for the production of 1,2-PeD from bio-renewable FF (furfural) using water solvent and shows its economic feasibility. This study consists of experimental investigation and simulation study. First, we determine a suitable process route to maximize the yield of 1,2-PeD through the experimental investigation. FF is converted to 1,2-PeD through two sequential catalytic conversion steps using Cu-MgO and Ru-MnO_x catalysts. Next, the simulation study including process design, heat integration, and economic analysis is performed to present the economic feasibility for the developed process. The minimum selling price of 1,2-PeD is $ 3,118 per tonnes, and it is close to the current market price. We minimized the separation costs for the integrated process by designing efficient separation subsystems and heat exchanger network. Despite of using an expensive raw material, the environmentally-friendly process can be cost-competitive with conventional process.

Keywords: furfural; 1,2-pentanediol; process design; economic evaluation

1. Introduction

Lignocellulosic biomass that consists of cellulose, hemicellulose, and lignin has a potential as a sustainable carbon source to replace fossil resources (Saxena et al., 2009). It can be converted to fuels and chemicals by catalytic conversion process (Chundawat et al., 2011; Zhou et al., 2011). Furfural (FF) is a valuable chemical obtained from hydrolysis of hemicellulose fraction of lignocellulosic biomass, and it is currently well produced an industrial scale (Huber et al., 2006). FF can be used as a precursor for synthesis of valuable furan-based chemicals as well as an additive in pharmaceutical and agrochemical industry. The previous studies for catalytic conversion process using metal catalysts to produce furan-based compounds such as furfuryl alcohol, furfuryl acetate, and furoylglycine from biomass-derived FF have been reported recently (Yan et al., 2014). Zhang et al. have reported a catalytic conversion process for production of 1,2-pentanediol (1,2-PeD) by a highly selective catalytic hydrogenolysis of furfuryl alcohol (FFA), which can be obtained from catalytic hydrogenation of FF (B. Zhang et al., 2012). 1,2-PeD is a valuable polyol that can be used as a preservative in cosmetics or a monomer of polyester (Antoce et al., 1998; Mizugaki et al., 2014). Especially,

polyester is a representative polymer used for engineering plastic, textile, etc. As the increment of polyester usage, the global demand for polyester polyol has been gradually increased (C. Zhang, 2015). In the fossil-based process for production of 1,2-PeD, pentene is converted to 1,2-PeD by epoxidation and saponification using propionic acid and benzene (Siegmeier et al., 1986). The conventional process can cause resource depletion and environmental toxicity problems, due to using fossil feedstock and toxic solvent.

This paper presents the environmentally-friendly process for the production of 1,2-PeD from bio-renewable FF using non-toxic water solvent and shows its economic feasibility. A suitable process route to maximize yield of product is determined by an experimental investigation that analyses reaction conditions (temperature, pressure, catalyst, solvent, reaction time, etc.) and results (conversion and yield) of each conversion technologies. Then, we develop a simulation model for the large-scale process based on the process route determined by the experimental investigation and evaluated the process economics based on the simulation results.

2. Method

2.1. Experimental investigation

FF-to-1,2-PeD conversion consisted of two catalytic conversion subsystems: (i) hydrogenation FF to FFA and (ii) hydrogenolysis of FFA to 1,2-PeD. The biomass derived FF can be converted to FFA through a fixed bed reactor containing Cu-MgO catalyst. Nagaraja et al. reported that the effect of Cu content in the catalyst and reaction temperature on the hydrogenation of FF to FFA (Nagaraja et al., 2007). They measured the yield of FFA by changing reaction temperature (450-530 K) and Cu loading in catalyst (5-80 wt %) while fixing pressure (1 bar) and molar ratio of FF and H_2 (1:2.5). The highest yield (96%) of FFA was observed over Cu-MgO catalyst containing 16wt% Cu at 453 K, and we applied this reaction condition to the integrated process. Next, FFA can be converted to 1,2-PeD through an autoclave reactor containing Ru-MnO_x catalyst (B. Zhang et al., 2012). Zhang et al. reported the optimized reaction condition that promotes the FFA hydrogenolysis to produce 1,2-PeD, while inhibiting the FFA hydrogenation. They analyzed the correlation between the major reaction condition (temperature and H_2 pressure) and results (1,2-PeD yield) and showed the high temperature and the low H_2 pressure increase the generation of 1,2-PeD. In the autoclave reactor using 10 wt % FFA in a water solution, FFA can be hydrogenolysed to 1,2-PeD with a 38 % molar yield at 423 K and 20 bar. Simultaneously, tetrahydrofurfuryl alcohol (THFA) was converted from FFA with a 48 % molar yield through FFA hydrogenation. This reaction condition was applied to design the integrated process. Consequently, the total experimental yield for FF-to-1,2-PeD calculated based on the process route determined by experimental investigation was 36 %.

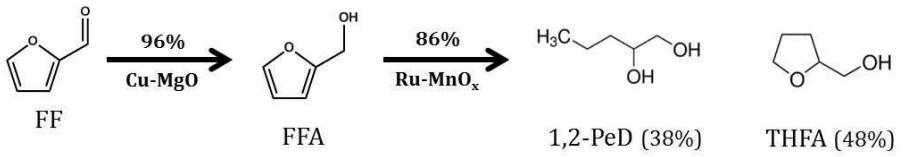


Figure 1. Reaction schematic view of FF-to-1,2-PeD and THFA

2.2. Simulation study

Simulation study followed three steps: (1) process design, (2) heat integration, and (3) economic analysis. First, we designed integrated process including conversion and separation subsystems based on the conversion route determined at experimental investigation by using ASPEN Plus Process Simulator. The processing rate of FF was 42,900 tonnes per year (t/y) which was produced from a large-scale process with a processing capacity of 350,000 dry t/y of white birch. Next, heat integration was performed to minimize the energy requirements for the integrated process by using ASPEN Energy Analyzer. Finally, we performed economic analysis including cost analysis and minimum selling price (MSP) calculation by using n^{th} plant analysis. The n^{th} plant analysis was assumed that the process could be scaled up without performance loss when the operating conditions are equal to experimental data. Capital cost was calculated by using ASPEN Economic Analyzer, while operating cost was determined by using cost data used in previous studies for biomass conversion process. The MSP of 1,2-PeD which makes revenue equal to cost was determined using a discounted cash flow analysis.

3. Results and discussion

3.1. Process design

We developed an integrated process to produce 1,2-PeD and THFA from FF. The integrated process included six subsystems: FFA production, 1,2-PeD production, 1,2-PeD purification, wastewater treatment (WWT), utilities, and storage (Figure 2).
Gaseous FF (42,900 t/y) was sent to a fixed bed reactor (R-1 in Figure 2) using Cu-MgO catalyst, and it was hydrogenated to FFA under H_2-rich condition. After FF hydrogenation, a resulting mixture containing FFA, unreacted H_2, and BR was sent to a flash tank (S-1 in Figure 2) at 393 K and 1 bar. Unreacted H_2 stream (14,700 t/y) obtained from the top of S-1 was recycled to R-1, while bottom liquid (58,600 t/y) was

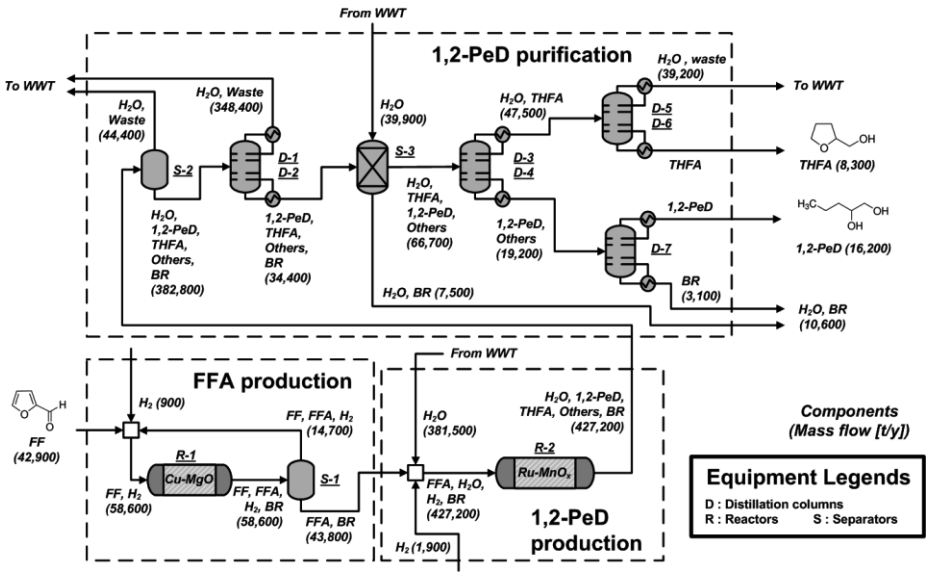


Figure 2. Simplified process flow diagram for the integrated process to produce 1,2-PeD and THFA from FF

sent to an autoclave reactor (R-2 in Figure 2) using Ru/MnO_x catalyst. After FFA conversion, the resulting mixture (427,200 t/y) containing water, 1,2-PeD, THFA, biomass residue (BR), and others was sent to 1,2-PeD purification subsystem to recover high purity (99 wt %) chemicals (1,2-PeD and THFA). First, 99 % of water (392,800 t/y) was separated from the mixture through a flash tank (S-2 in Figure 2) and two distillation columns (15-stage distillation column (D-1) and 10-stage distillation column (D-2) in Figure 2) connected in series. When the most of water was separated, 56 % of THFA was removed with water. Then, the bottom liquid obtained from D-2 was sent to a pressure filter (S-3 in Figure 2) to separate BR with the addition of water. BR can be used as a heat source by combustion, and it was considered to be sold as a fuel. After BR separation, THFA and water (47,500 t/y) was separated by two distillation columns (10-stage distillation column (D-3) and 20-stage distillation column (D-4) in Figure 2) and it was sent to two distillation columns (20-stage distillation column (D-5) and 3-stage distillation column (D-6) in Figure 2) to recover 99 wt% of THFA (8,300 t/y). The bottom liquid (19,200 t/y) obtained from D-4 was sent to a 40-stage distillation column (D-7 in Figure 2), and 99 wt % of 1,2-PeD (16,200 t/y) was obtained at the top of D-7.

Consequently, 95 % of 1,2-PeD and 39 % of THFA were recovered by the 1,2-PeD purification subsystem, respectively, and the total numerical yield for FF-to-1,2-PeD and THFA was 53 %.

3.2. Heat integration

When 30.4 megawatt-equivalent (MW_e) of FF and 11.1 MW_e of H_2 were processed to produce 24.3 MW_e of chemicals (1,2-PeD and THFA) and 6.6 MW_e of BR, the integrated process required 43.0 megawatts (MW) of heating energy, 75.9 MW of cooling energy, and 0.1 MW of electricity. The largest heat energy (35.7 MW) was required for 1,2-PeD purification subsystem, and 74 % of heat energy (26.4 MW) was required to remove water solvent. Because the total heating requirement was much larger than electricity requirement, we only conducted a heat integration to minimize the heating requirements. After heat integration, 57 % of heating energy (24.7 MW) was recovered, and the heating energy requirements were reduced to 18.3 MW. The energy flow in the integrated process after heat integration was shown in Figure 3. The energy efficiency which was the ratio of energy input (FF, H_2, primary energy) and energy output (1,2-PeD, THFA, BR). The energy of chemical species (FF, 1,2-PeD, THFA, BR) was calculated based on higher heating value, while the primary energy was calculated based on biomass-to-heat and electricity production efficiency (71 % and 30 %). The energy efficiency of the integrated process after heat integration was 45 %, which was 15 % point higher than the process without heat exchanger network.

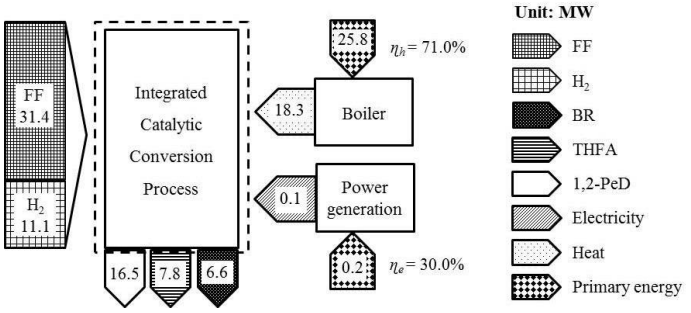


Figure 3. Energy flow diagram after heat integration

3.3. Economic analysis

Capital and operating costs for the integrated process were calculated based on simulation results, and all costs were converted on an annual basis to analyze significant cost drivers for the integrated process. Total cost (TC) for the integrated process was \$ 63.2 million per year (\$ million/y), and it was calculated by the sum of annualized capital cost (ACC) and annualized operating cost (AOC). Annualized costs for the integrated process were shown in Table 1.

Capital costs including direct and indirect costs, land cost, and working capital cost were annualized by capital charge rate calculated from a discounted cash flow analysis. ACC for the integrated process was \$ 7.4 million/y (12 % of TC), and the 1,2-PeD production (\$ 2.5 million/y, 4 % of TC) and purification (\$ 1.7 million/y, 3 % of TC) subsystems had the significant share of capital cost. Operating costs consisted of variable costs (material, utility, other variable operating costs) and fixed costs (labor, maintenance, insurance). AOC for the integrated process was \$ 55.9 million/y (88 % of TC), which was about 7.6 times that of ACC. FF cost (\$ 43.0 million/y, 68 % of TC) had the largest share of total AOC, followed by the H_2 cost (\$ 5.9 million/y, 9 % of TC). By-product revenue for the integrated process was \$ 12.9 million/y, and it consisted of THFA (\$ 12.4 million/y) and BR (\$ 0.5 million/y) revenues.

The MSP of 1,2-PeD was determined to ensure that TC and total revenue were the same by using discounted cash flow analysis, and it was expressed in terms of tonnes of 1,2-PeD ($/$t_{1,2\text{-PeD}}$) for the integrated process. The MSP of 1,2-PeD for the integrated process was \$ 3,118/$t_{1,2\text{-PeD}}$, and it was close to the market price of 1,2-PeD (\$ 3,000/$t_{1,2\text{-PeD}}$). In this process, high price of FF and loss of THFA in the separation process led to increase the MSP of 1,2-PeD. If a cost-effective process to produce FF from biomass and recover process to highly recover THFA is considered as a further study, the MSP of 1,2-PeD can be lower than the market price.

Table 1. Annualized capital and operating costs

	Value (\$ million/y)	Ratio (% of TC)
Total Cost (TC)	63.2	
Annualized capital costs (ACC)	7.4	12
FFA production	0.8	1
1,2-PeD production	2.5	4
1,2-PeD purification	1.7	3
WWT	0.7	1
Storage	0.4	1
Utilities	1.2	2
Annualized operating costs (AOC)	55.9	88
Variable operating costs	52.6	83
FF	43.0	68
H_2	5.9	9
Utility	3.4	5
Other variable operating costs	0.4	1
Fixed operating cost	3.3	5

* Other variable operating costs consist of water, WWT chemicals, and catalyst regeneration

4. Conclusions

In this paper, we developed a large-scale process for the production of 1,2-PeD from bio-renewable FF and showed the economic feasibility of the process. We designed the integrated process producing 16,200 t/y of 1,2-PeD from 42,900 t/y of FF based on experimental data. The developed process required large amount of heating energy, and heat integration was conducted to reduce the heating requirements. Fifty-seven percent of heating energy was recovered by heat integration, and the energy efficiency was increased from 30 % to 45 %. In this process, operating cost was significantly larger than capital cost. The FF cost was the largest cost driver among the operating costs, and it accounts for 71 % of total cost. The MSP of 1,2-PeD was \$ 3,118/$t_{1,2\text{-PeD}}$, and it was close to current market price of 1,2-PeD.

References

O. A. Antoce, V. Antoce, N. Mori, S. Yasui, A. Kobayashi, and K. Takahashi, (1998). Calorimetric evaluation of the antimicrobial properties of 1, 3-butanediol and 1, 2-pentanediol on various microorganisms. Netsu Sokutei, 25, 2-8.

S. P. Chundawat, G. T. Beckham,M. E. Himmel, and B. E. Dale, (2011). Deconstruction of lignocellulosic biomass to fuels and chemicals. Annual Review of Chemical and Biomolecular Engineering, 2, 121-145.

G. W. Huber, S. Iborra, and A. Corma, (2006). Synthesis of transportation fuels from biomass: chemistry, catalysts, and engineering. Chemical Reviews, 106, 4044-4098.

T. Mizugaki, T. Yamakawa, Y. Nagatsu, Z. Maeno, T. Mitsudome, K. Jitsukawa, and K. Kaneda, (2014). Direct transformation of furfural to 1, 2-pentanediol using a hydrotalcite-supported platinum nanoparticle catalyst. ACS Sustainable Chemistry & Engineering, 2, 2243-2247.

B. Nagaraja, A. Padmasri, B. D. Raju, an K. R. Rao, (2007). Vapor phase selective hydrogenation of furfural to furfuryl alcohol over Cu–MgO coprecipitated catalysts. Journal of Molecular Catalysis A: Chemical, 265, 90-97.

R. Saxena, D. Adhikari, and H. Goyal, (2009). Biomass-based energy fuel through biochemical routes: a review. Renewable and Sustainable Energy Reviews, 13, 167-178.

R. Siegmeier,G. Prescher, H. Maurer, and G. Hering, (1986). Continuous process for the production of pentanediol-1, 2. In: Google Patents.

K. Yan, G. Wu, T. Lafleur, and C. Jarvis, (2014). Production, properties and catalytic hydrogenation of furfural to fuel additives and value-added chemicals. Renewable and Sustainable Energy Reviews, 38, 663-676.

B. Zhang,Y. Zhu, G. Ding, H. Zheng,and Y. Li, (2012). Selective conversion of furfuryl alcohol to 1, 2-pentanediol over a Ru/MnO x catalyst in aqueous phase. Green Chemistry, 14, 3402-3409.

C. Zhang, (2015). Biodegradable Polyesters: Synthesis, Properties, Applications: Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany.

C. H. Zhou, X. Xia, C. X. Lin, D. S. Tong, and J. Beltramini, (2011). Catalytic conversion of lignocellulosic biomass to fine chemicals and fuels. Chemical Society Reviews, 40, 5588-5617.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50020-3

Methanol Production from High CO_2 Content Natural Gas

Benjamín Cañete[a], Nélida B. Brignole[a, b*], Carlos E.Gigola[a]

[a]*Planta Piloto de Ingeniería Química (Universidad Nacional del Sur - CONICET) Camino La Carrindanga km. 7 - 8000 B. Blanca – Argentina*
[b]*Lab. de Investigación y Desarrollo en Computación Científica (LIDECC)- Departamento de Ciencias e Ingeniería de la Computación (DCIC), Universidad Nacional del Sur (UNS), Bahía Blanca, Argentina.*
dybrigno@criba.edu.ar

Abstract

A process for methanol production from natural gas containing 30-80% CO_2 was analyzed. For this design a partial feed hydrogenation process is followed by a combined reforming operation and H_2 injection in order to obtain syngas of appropriate quality. A methanol plant of 400,000 t/y that uses a Lurgi reactor for the synthesis reaction, was simulated to obtain mass and energy balances. Although the total amount of required H_2 increases with the CO_2 content up to $H_2/CO_2 \approx 2$, it is considerably lower than the quantity required by direct CO_2 hydrogenation processes. It was also demonstrated that the reformer size decreases and the global CO_2 conversion to methanol increases leading to lower emissions. An economic analysis showed that this methanol process has lower investment costs, but higher operating costs in comparison with a methanol plant based on CH_4 steam reforming. This is related to the high H_2 cost, which was assumed to be obtained by water electrolysis. A preliminary financial analysis indicated that a positive net present value based on current H_2 and methanol prices is obtained when the CO_2 content is less than 50%.

Keywords: CO_2 hydrogenation, combined reforming; syngas; methanol synthesis

1. Introduction

Numerous studies are being conducted on methanol production processes based on CO_2 hydrogenation with the aim of reducing the emissions and the high energy consumption of the classical steam reforming process. One proposal is the CAMERE process (Joo et al., 1999), where syngas is produced by application of the RWGS reaction to pure CO_2 at 600 ^oC, followed by a methanol synthesis reactor working under standard conditions. However, CAMERE's industrial applications are mainly limited by the amount and high cost of H_2 produced by a carbon free process. The availability of a high flow of pure CO_2 introduces another technical and economic constraint. An alternative process that partially reduces these difficulties is based on the direct utilization of natural gas (NG) with a high CO_2 content, as available on numerous gas fields, followed by a combined reforming operation (Cañete et al., 2014). However, if the feed contains more than 30-35% CO_2, it becomes necessary to reduce its content by removal or hydrogenation. It has recently been demonstrated (Cañete et al., 2017) that partial hydrogenation of a feed containing 50 % CO_2 using a RWGS reactor operated at 600 ^oC, followed by a combined reforming process at 950 ^oC, leads to a syngas composition close to the one needed for methanol synthesis. In this way, the CO_2 emissions are

reduced and the global conversion of CO_2 is about 25 %. Moreover, an economic analysis demonstrated that the introduction of the RWGS reactor followed by a combined reforming furnace and a methanol synthesis reactor leads to lower capital investments for the methanol plant in comparison to the one based on CH_4 steam reforming. However, the operating costs are larger due to H_2 demand, although its incidence on the financial analysis is significantly reduced when compared to a plant based on the CAMERE process. In the present study we have analyzed in technical and economic terms the suitability of the mentioned process for methanol production from NG containing 30-80 % CO_2. The H_2 amount needed for partial hydrogenation of the feed mixture, the energy consumption in the reformer, the global CO_2 conversion to methanol and the amount of emitted CO_2 were estimated as a function of the feed composition. Based on this information, investment and operating costs were determined for a 400,000 ton/y methanol plant.

2. Process scheme and plant simulation

Figure 1 shows the proposed process scheme to produce methanol.

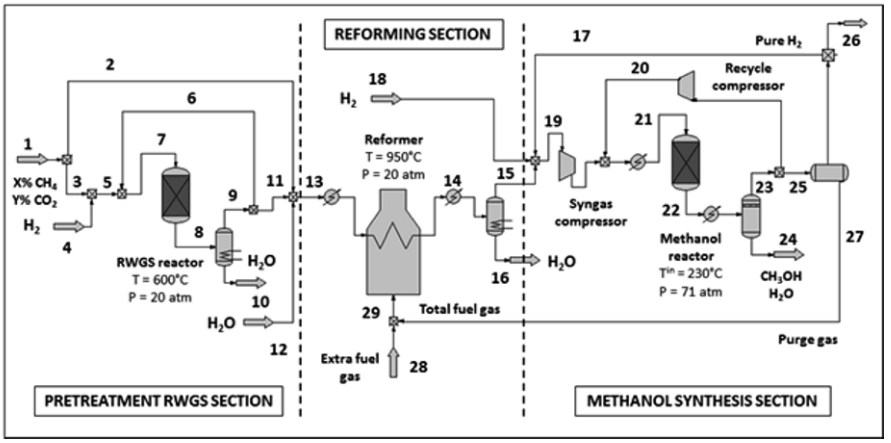


Figure 1. Methanol production from NG with a high CO_2 content. Flow diagram.

The main difference with a conventional methanol plant is the inclusion of a hydrogenation section prior to the reformer furnace. In this step an isothermal RWGS reactor operating at 20 bar and 600 ºC was used to process 40 % of the feed (whatever the CO_2 content), with a H_2/CO_2 ratio of 1.7 and a recycle ratio of 1.9 in order to increase the CO_2 conversion. In Cañete et al. (2017) the energy balances involving the RWGS and the reformer were coupled. The working conditions for the combined reforming furnace were 20 bar and 950 ºC with a H_2O/CH_4 ratio of 2, whatever the feed composition. The methanol synthesis reactor was simulated assuming a Lurgi-type unit operating at 230 ºC and 71 bar with a recycling ratio of 3. The Soave-Redlich-Kwong (SRK) equation of state was also employed for the simulation. The injection of pure H_2 (stream 18) from an external source and/or extracted with a membrane separator from the methanol-loop (stream 17) was considered in order to adjust the composition of the syngas to the value recommended for methanol production (M = 2). When the amount of removed H_2 exceeds what is necessary to adjust the syngas composition, it is used as fuel for the reformer (stream 26). The energy required for the reformer is mainly supplied by the purge gas (stream 27) and NG injection (stream 28). Kinetic

information of commercial catalysts was used for the RWGS (Joo, 2018) and Lurgi (Vanden Bussche and Froment, 1996) reactors simulation, while the reformer was assumed to be operating at equilibrium conditions. The RWGS reactor was simulated using a heterogeneous model and taking into account mass transfer limitations in the catalyst pellets. On the other hand, a pseudo homogeneous model was used for the methanol synthesis reactor. Reactor tubes of 1.5 inches in diameter and a length of 7 meters were adopted. A plant to produce 400,000 ton/y of methanol was simulated under steady state conditions using the gPROMS platform (PSE, 2017) in order to obtain the corresponding mass and energy balances.

3. Results and discussion

3.1. Simulation

The simulation work allowed us to obtain the flow rate and composition of all process streams identified in Figure 1. From this information, CO_2 conversion in the RWGS reactor, the conversion of CO_2 and CH_4 in the reformer, the methanol yield, the global CO_2 conversion into methanol and CO_2 emissions were calculated. In addition, the feed flow rate, the H_2 consumption and the NG amount that is necessary as reformer fuel were also estimated. It is important to remark that the extra H_2 requirement was considered to have been obtained from a zero emission source, such as water electrolysis. This consideration has a direct impact on the economy of the process, as it will be shown in Section 3.2. The results are presented in Table 1 as a function of the CO_2 content in the NG. Before proceeding with a detailed analysis of these results, it is interesting to note that the estimated size of the RWGS reactor was not significant when compared with the reformer or the methanol synthesis units. Regarding the size of the methanol reactor, it was found that, when the CO_2 concentration varied from 30 to 80 %, the number of catalyst tubes required in order to produce 400,000 ton/y of methanol increased slightly, from 3,912 to 4,117 units. Consequently, it may be assumed that the size of the methanol reactor is independent of the feed composition.

Table 1. **Main simulation results**

CO_2 %	Feed (kmol/h) CO_2+CH_4	X_{CO2} RWGS (%)	X_{CH4} REF (%)	X_{CO2} REF (%)	Yield CH_3OH* (%)	F_{CH4} (kmol/h)	CO_2 emission (ton/y)	X_{CO2}^{G}** (%)	F_{H2} (kmol/h)
30%	2,016	63.8	91.7	7.9	86.1	442	270,085	-28.1	411
40%	1,994	63.9	92.2	27.1	84.0	643	264,609	4.8	1,162
50%	1,923	66.5	92.8	35.3	86.6	478	249,984	25.4	1,416
60%	1,913	64.5	93.6	41.4	85.2	398	223,070	44.2	2,046
70%	1,882	63.2	94.3	43.7	84.0	335	190,914	58.4	2,712
80%	1,943	63.8	94.9	44.2	82.0	248	162,285	70.0	3,386

* Yield = $[F_{CH3OH}^{S22} - F_{CH3OH}^{S21}]/[F_{CO2}^{S19} + F_{CO}^{S19}]$

** Global CO_2 conversion = $[F_{CO2}^{S1} - (F_{CO2}^{S29} + F_{CO}^{S29} + F_{CH4}^{S29})]/[F_{CO2}^{S1}]$

According to the results in Table 1, the feed flow rate of NG +CO_2 to the plant is nearly constant because it exhibits a maximum variation of 7 %, as a function of CO_2 content. Assuming that the feed gas price is proportional to the heating value, a large reduction of this operating cost is expected as CO_2 concentration increases.

The overall CO_2 conversion in the hydrogenation reactor was nearly constant (63-66 %). Consequently, the CO_2 amount to be hydrogenated and the hydrogen required for the RWGS reaction (stream 4) increased with the CO_2 content in the feed. It went from 411 to 1057 kmol/h when the CO_2 content in the feed moved from 30 to 80 %.

Despite the CO_2 hydrogenation process, the feed to the reformer (stream 13) became richer in CO_2 as the concentration in the feed mixture increased. Therefore, the incidence of the dry reforming reaction in the CR process was higher and the CO_2 conversion in the reformer shifted from 7.9 to 44 %. Despite the increase in the extent of the reforming reactions, the energy required to heat the reactants and to maintain the reaction temperature decreased, mainly due to the fact that the amount of H_2O feed to the reformer decreased as the CO_2 concentration became higher. For this reason, the amount of NG that was necessary as extra fuel also decreased, as reported in Table 1. The dependence of the total energy required for the reformer as a function of the CO_2 content is shown in Table 2.

Table 2. Energy demand for the reformer as a function of the CO_2 content

%CO_2 (S_1)	30	40	50	60	70	80
Q (MW)	186	158	136	117	98	82

As the CH_4 conversion increases slightly, the quality of the syngas emerging from the reformer is characterized by a decreasing M module, as shown in Table 3. Therefore, increasing amounts of H_2 are necessary to satisfy the condition of M=2 for methanol synthesis, as shown in Table 3. Approximately 2 mol H_2 per mol CO_2 are required under the most unfavorable condition (80 % CO_2). This amount is considerably lower than the one required for methanol production by the CAMERE process.
In addition to an M module equal to 2, a common industrial practice in methanol production is to restrict CO_2/CO ratio to less than 0.5 and to limit CO_2 concentration at the loop entrance to 8 %. In the present study, these conditions are satisfied well for feed mixtures containing 30 to 60 % CO_2.

Table 3. Dependence of the syngas M module on the CO_2 content

% CO_2	30	40	50	60	70	80
M	1.94	1.56	1.26	0.96	0.65	0.36
H_2 (S18) (kmol/h)	0	620	762	1,265	1,816	2,329

In view of the results in Table 1, the methanol yield is higher than 80 % for all feed conditions. Accordingly, the methanol concentration at the synthesis reactor end (stream 22) is nearly constant (7.4 – 7.9 %).

Finally, it is important to point out that a large reduction in the amount of emitted CO_2 was achieved as a result of the diminution in the amount of NG employed as fuel and the parallel increase in CO_2 conversion in the reformer. The results in Table 1 indicate

that the global CO_2 conversion in the methanol plant increased continuously from a negative value up to 70 %, when the CO_2 feed content moved from 30 to 80 %.

3.2. Economic analysis

Taking into account that the proposed process presents advantages (lower feed cost; reduced energy demand in the reformer) and a main disadvantage (increased demand of high cost H_2), when it is compared to a classical methanol plant, an economic and financial analysis was performed to evaluate the feasibility of an industrial application. The RWGS reactor, the reformer, the turbine and the recycling compressors were considered in order to determine the capital costs. Equipment purchase and installation costs were estimated using Ulrich´s method (Ulrich & Vasudevan, 2004). These values were updated with The Chemical Engineering Plant Cost Index (2016). The total investment cost was found to be lower compared to a methanol plant based on CH_4 steam reforming. The consumption of NG with a high CO_2 content, pure NG as a fuel (0.139 US$/m^3), process and cooling water and H_2 produced by alkaline electrolysis (4 US$/kg), were considered in order to evaluate the operating costs. Regarding H_2 cost, the considered price is perhaps below the one resulting from water electrolysis when electricity from renewable energy sources is used. However, the chosen value (4 us$/kg) is much higher than the one corresponding to hydrogen produced by classical steam reforming (approximately 2 us$/kg). We have not included the electrolysis plant as a part of the process scheme and consequently, it was not present in the economic analysis. The price of NG with CO_2 was calculated as a function of the heating value. Mainly due to the high cost of H_2, the operating costs were in all cases higher than those expected for a typical methanol plant. In order to determine the feed conditions that make the proposed process viable, the net present value (NPV) was determined assuming a methanol price of 396 US$/t, which is an average price (Methanex, 2017). An interest rate of 10 % was considered for the NPV calculation. The additional information required for NPV calculation was taken from a previous study (Cañete et al., 2017).

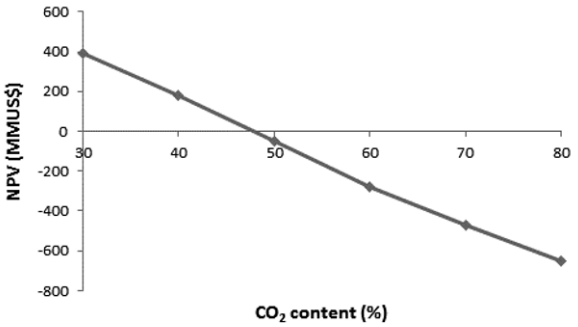


Figure 2. NPV vs CO_2 feed content.

Figure 2 shows that a positive NPV value is obtained when the CO_2 content in the feed is below 50 %. By using a similar procedure, it was determined that the H_2 price should decrease to 2.9, 2.5 or 2.4 US$/kg in order to obtain positive NPV values for feeds containing 60, 70 and 80 % CO_2, respectively.

4. Conclusions

Natural gas with CO_2 content in the 30 to 80 % range could be used to produce syngas and then methanol if partial hydrogenation of the feed is accomplished prior to a combined reforming furnace. It is shown that a substantial reduction in H_2 consumption is achieved in comparison with methanol processes based on total CO_2 hydrogenation. It is also demonstrated that the energy required for the reformer operation decreases as the amount of CO_2 in the feed increases. As a result, operational and investment costs for the reformer are reduced. Other favorable results are a high methanol yield (>80 %) and an increasing global CO_2 conversion, due to the higher incidence of the dry reforming reaction in the reformer. Consequently, CO_2 emissions are much lower. Despite the reduced H_2 consumption, an economic analysis demonstrated that operating costs are higher than those of a methanol plant based on CH_4 steam reforming. However, a financial analysis showed that the proposed process leads to a positive NPV value for a feed containing up to 50 % CO_2, provided that an H_2 cost of 4US$/kg is assumed. For a feed containing more than 50 % CO_2, H_2 price should be from 30 % to 40 % lower for the process to be competitive.

References

B. Cañete, C.E. Gigola, N.B. Brignole, 2014, Synthesis Gas Processes for Methanol Production via CH_4 Reforming with CO_2 , H_2O and O_2. Ind. & Eng. Chem. Res., 53 (17), 7103-7112.

B. Cañete, C.E. Gigola, N.B. Brignole, 2017, Enhancing the Potential of Methane Combined Reforming for Methanol Production via Partial CO_2 Hydrogenation. Ind. & Eng. Chem. Res., 56, 6480-6492.

O.S. Joo, Camere process for carbon dioxide hydrogenation to form methanol. ACS Fuel/Preprints Archives Washington DC, https://web.anl.gov/PCS/acsfuel/preprint%20archive/Files/45_4_WASHINGTON%20DC_08-00_0686.pdf Catalysis Laboratory, Korea Institute of Science and Technology, 45 (4), 686-689. (Accessed Feb. 2, 2018).

O. S. Joo, J.Kwang-Deog, I. Moon, A.Y. Rozovskii, G.I. Lin,S-H Han, S-J Uhm, 1999, Carbon dioxide Hydrogenation to Form Methanol via a Reverse-Water-Gas-ShiftReaction (the CAMERE process), Ind. & Eng. Chem. Res., 38, 1808-1812.

Methanex, Historical Methanex Posted Price, https://www.methanex.com/sites/default/files/methanol-price/MxAvgPrice_Oct%2027%2C%202017.pdf (Accessed Nov. 30, 2017)

PSE gPROMS, https://www.psenterprise.com/products/gproms (Accessed Nov. 30, 2017).

G. D. Ulrich, P. T. Vasudevan, 2004, Chemical Engineering Process Design and Economics: A Practical Guide, 2nd ed., Process Pub., Durham, NH.

K. M. Vanden Bussche, G.F. Froment, 1996, A steady-state kinetic model for methanol synthesis and the water gas shift reaction on a commercial $Cu/OZn/Al_2O_3$ catalyst. J. Catal. , 161, 1-10.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50021-5

Leveraging Atomistic Modeling during Precursor Design for Cobalt Film Deposition

Andrew J. Adamczyk[a,b*], Alan C. Cooper[b], Moo-Sung Kim[c] and Sergei V. Ivanov[b]

[a]*Department of Chemical Engineering, Auburn University, Auburn, AL, USA*
[b]*Materials Technology, Versum Materials Inc., Allentown, PA, USA*
[c]*Materials Technology, Versum Materials Inc., Republic of Korea*
aja0056@auburn.edu

Abstract

Atomic Layer Deposition (ALD) is used to fabricate ultrathin and conformal thin film structures for many semiconductor and thin film device applications. ALD is particularly advantageous for integrated circuit manufacturing which requires control of film structure in the nanometer or sub-nanometer scale. A growing area of interest is Cu interconnect encapsulation with thin Co films. Thin Co films have been shown to greatly improve electromigration performance of Cu interconnects in integrated circuits. Further shrinkage of electronic devices brings new challenges for more selective and controlled deposition of thin Co films. By bridging atomistic and reactor scales, this paper will focus upon applying the systems approach to precursor design to better understand the interaction of relevant substrates and process conditions with the organometallic precursors used to deposit thin Co films.

First-principles analysis of ALD mechanisms can significantly decrease experimental costs through initial virtual screening of both potential precursors and surface cleaning requirements. Simultaneously, these studies can also improve the fundamental understanding of precursor stability and surface reactivity. In this collection of studies, Density Functional Theory (DFT) at the BLYP/DNP level of theory was employed to study ligand dissociation energies, decomposition mechanisms, adsorption/desorption, and surface reactions of various commercially available Co precursors on substrates found in common integrated circuits. These studies have employed both gas-phase and periodic surface slab models in efforts to balance computational time and accuracy. Cobalt precursors considered include Co amides and Co carbonyls with cyclopentadienyl, allyl, and alkyne ligands. These studies have also investigated Co nucleation on various substrates by direct comparison of precursor adsorption and dissociation energies on small Co clusters relative to the underlying substrate surface.

Atomistic modeling results agreed with experiments and have provided insights on the parameters which should be considered in the systematic design of new ALD precursors for thin Co film deposition. This work demonstrates how theoretical investigations in parallel with experiments are directly impacting precursor discovery and deposition understanding in industry.

Keywords: Computational chemistry, Density Functional Theory (DFT), nonlinear eigenvalue problem, product design, surface reactions

1. Theoretical Background

1.1. Atomic Layer Deposition (ALD) reactor system

ALD is a thin-film manufacturing process in which the growth surface is exposed to an alternating sequence of gas-phase chemical precursor species separated by purge periods to prevent gas-phase reactions. ALD is characterized by self-limiting heterogeneous reactions between the gas-phase precursor species and surface-bound species, which, when allowed sufficient conditions to reach saturation, results in highly conformal thin films, on both planar and non-planar geometries, with atomic level control of film deposition. The rate of ALD depends strongly on the instantaneous state of the growth surface, and this state changes continuously through each exposure and purge period. Due to the heterogeneous nature of these ALD reactions, adsorption, desorption, and surface reaction kinetics play a crucial role in the deposition process. The overall ALD reactor material balance follows Eq.(1). Here the total moles of gas-phase species in the ALD reactor, $m_{rxr}(t)$, is a function of total molar flow rates of the feed gases, $\dot{m}_{in}$, and residual product gases, $\dot{m}_{out}$, and the instantaneous rates of precursor consumption, Γ^{react}, and leaving group production, Γ^{prod}, per unit area of the total growth surface, A_{rxr}, due to the deposition reactions.

$$\frac{dm_{rxr}}{dt} = \dot{m}_{in} - \dot{m}_{out} + (\Gamma^{react} + \Gamma^{prod})A_{rxr} \quad (1)$$

This paper will apply computational chemistry, specifically Density Functional Theory (DFT), to estimate the underlying thermodynamic driving forces for reaction rates, Γ, in Eq.(1) related to surface deposition. In turn detailed understanding of thermodynamic driving forces will allow for targeted product design and discovery of new precursors.

1.2. Density Functional Theory (DFT) - Molecular structure and energetics

DFT begins with a theorem by Hohenberg and Kohn (1964), which states that all ground-state properties of a polyatomic system are functionals of the charge density, n(**r**), in Eq.(2). For constrained nuclear positions (Born-Oppenheimer approximation), the modified Kohn-Sham (KS) DFT self-consistent field (SCF) procedure is used to solve the coupled set of KS equations in Eq.(3), which constitutes a nonlinear eigenvalue problem. Here ψ is an eigenvector of molecular orbitals (KS one-electron orbitals) dependent on electron position **r**, ε is a set of molecular orbital energy eigenvalues, and $\hat{h}_{KS}$ is the KS Hamiltonian matrix operator corresponding to the total electronic energy of the polyatomic system, where N is the total number of occupied molecular orbitals.

$$n(\boldsymbol{r}) = \sum_{i=1}^{N} |\psi_i(\boldsymbol{r})|^2 \quad (2)$$

$$\hat{h}_{KS}\psi_i(\boldsymbol{r}) = \varepsilon_i\psi_i(\boldsymbol{r}) \qquad i = 1, N \quad (3)$$

In practice, it is convenient to expand the molecular orbitals, ψ, as a linear combination of atomic orbitals (basis functions), χ, forming a basis set given in Eq.(4), where M is the total number of basis functions. Because atomic orbitals are not necessarily orthonormal, Eq.(3) is reformulated to Eq.(5), where **F** is the Fock matrix approximating the single-electron energy operator, **C** is an eigenvector of molecular orbital expansion coefficients, ϵ is a set of molecular orbital energy eigenvalues, and **S** is the overlap integral to ensure atomic orbital orthogonality.

$$\psi_i(\boldsymbol{r}) = \sum_{j=1}^{M} C_{ij}\chi_j(\boldsymbol{r}) \qquad i = 1, N \quad (4)$$

$$\boldsymbol{FC} = \epsilon\boldsymbol{SC} \quad (5)$$

The orbital coefficients, **C**, that are varied to minimize the total electronic energy are dependent on the energy through **F** and **S** so an iterative SCF procedure is used. This procedure requires solution of the secular determinant in Eq.(6). In each SCF cycle, new **F** and **S** are constructed. In Eq.(7), the matrix elements of **F** are given in *bra* (for ⟨|) and *ket* (for |⟩) notation which resemble outer and inner products in linear algebra. The **F** matrix elements are one-electron integrals for kinetic energy and external potential, and two-electron integrals for coulombic interaction and exchange-correlation (Saad et al., 2010). The eigenvector of coefficients, **C**, and set of energy eigenvalues, ϵ, to construct the density matrix and to calculate new total electronic energy, respectively, are obtained by solving Eq.(6). This procedure is repeated until the input and output charge densities are identical to within a given tolerance. Total electronic energy as a function of charge density is then defined in Eq.(8) as the sum of energy eigenvalues over occupied molecular orbitals and an energy correction, $E_{corr}(n)$. The energy correction can vary depending on the DFT level of theory as the Hohenberg-Kohn theorem provides no guidance as to the form of $E(n)$.

$$\sum_{\upsilon=1}^{M}\sum_{\mu=1}^{M}\left|F_{\upsilon\mu}-\epsilon_k S_{\upsilon\mu}\right|C_{\mu k}=0 \qquad \text{k = 1, M} \tag{6}$$

$$F_{\mu\upsilon}=\left\langle\chi_\mu\left|-\frac{1}{2}\nabla^2+\upsilon(\boldsymbol{r})+\int\frac{n(\boldsymbol{r}')}{|\boldsymbol{r}'-\boldsymbol{r}|}d\boldsymbol{r}'+\upsilon_{XC}\big(n(\boldsymbol{r})\big)\right|\chi_\upsilon\right\rangle \tag{7}$$

$$E(n)=\sum_{i=1}^{N}\varepsilon_i+E_{corr}(n) \tag{8}$$

To further minimize total electronic energy and optimize geometry, nuclear positions are relaxed after each completed SCF procedure. This is the process of finding a nuclear arrangement where, according to the DFT model of chemical bonding, the net inter-atomic force on each atom is acceptably close to zero, and the position on the potential energy surface is a local energy minimum. This geometry optimization problem follows the objective function definition in Eq.(9) for all nuclei.

$$\min E(\boldsymbol{R})\ \text{s.t.}\ \partial E/\partial\boldsymbol{R}=0\ \text{and}\ \partial^2E/\partial\boldsymbol{R}^2>0 \tag{9}$$

DFT is used to calculate the energy, E(**R**), the derivative of energy with respect to nuclear positions, $\partial E/\partial\boldsymbol{R}$, and the second derivative matrix of the system, $\partial^2E/\partial\boldsymbol{R}^2$, also known as the Hessian matrix, which describes the energy surface curvature at **R**. The optimization algorithm will minimize the atomic forces by a Quasi-Newton method, where $\partial E/\partial\boldsymbol{R}$ is used to estimate Hessian matrix.

2. Precursor Design with Computational Chemistry

2.1. Objectives

Knowledge of the existing ALD process conditions, including the state of the growth surface, is required for establishing the design space for new Co precursors. Computational screening will aid how the design space is mapped more efficiently. When benchmarked against known precursor performance, virtual screening with computational chemistry can be used to create molecular libraries for future experimental testing or to better understand the deposition of existing precursors. The primary objectives of this paper are to study:

- Gas-phase thermal stability and reactivity of Co precursors by comparing ligand dissociation energies.
- Adsorption and initial nucleation energies of Co precursors on various surfaces by comparing adsorption energies on both bulk and cluster-terrace surfaces.

- Surface reaction mechanisms by comparing decomposition energies of a selected precursor to Co metal on various surfaces.

2.2. Computational methodology

For DFT calculations, the Dmol^3 module in BioVia Materials Studio 7.0 software, from Dassault Systèmes was used. All electrons were included for each atom, and the BLYP functional, DNP numerical basis set, and 4.4 Å orbital cutoff were employed. For selected ligand dissociation energies, the more computationally intensive B3LYP functional was evaluated, and the relative ranking of electronic energy differences was the same as the BLYP functional. Global geometry optimization for gas-phase and surface slab systems was achieved by creating multiple input geometries of different conformations and spin state multiplicities. Experimental bulk crystal structures were cleaved at the most thermodynamically stable crystal planes to create surface slab models. The surfaces were then terminated based upon known experimental surface pre-treatment. Cluster-Terrace models were created by geometry optimization of a Co_4 cluster on Ta_3N_5 and SiO_2 substrates. Periodic boundary conditions were applied with atomic constraints on the bottom 2 layers of each 3-layer surface slab to minimize edge effects and improve geometry convergence.

To estimate thermodynamic driving forces, ligand removal, adsorption, and surface reaction energies were calculated by the difference in total electronic energies of reactants and products. For CO and Cp removals from Co-Cp (Figure 1), standard enthalpies of reaction were calculated to evaluate thermal effects. Reaction energies using only electronic energies were found to be within 5% of the standard enthalpies of reaction. Thermal effects were not incorporated for the remaining relative electronic energy differences due to the high correlation between reaction energies and enthalpies. This work instead focuses upon ranking relative electronic energy differences between elementary reaction steps of similar entropic changes (Stowasser and Hoffmann, 1999). This assumption allows for high-throughput screening of precursor candidates, where only promising leads are pursued for more computationally intensive analysis and experiments. The computational time for geometry optimization with electronic energy calculations for each polyatomic system scales with N_b^3 dependence, where N_b is the total number of basis functions. Each DFT calculation was performed on an Intel Xeon 2.2 GHz 16-core compute node.

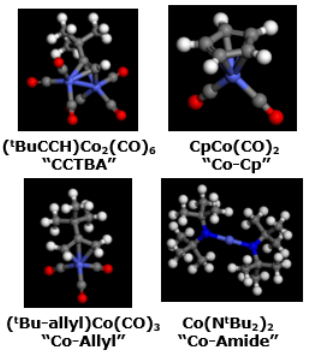


Figure 1: Optimized geometries of Co precursors in this study.

3. Case Studies – Design of Co Precursors for Selective Deposition

3.1. Gas-phase thermal stability and reactivity

Precursor delivery dynamics is an important aspect of the ALD process. Table 1 contains calculated ligand dissociation energies for the Co precursors in Figure 1 to better understand thermal stability and reactivity. It was found that compared to CCTBA, which has known poor gas-phase stability, Co-Cp, Co-Allyl, and Co-Amide showed better gas-phase stability because ligand removal energies are higher or less favorable to dissociate. Co-Cp is the most stable in the carbonyl series. CCTBA is expected to deposit the cleanest

films due to low dissociation energy of the organic ligand, but also it is the least stable. Cleavage of the organic ligand from Co-Cp, Co-Allyl, and Co-Amide requires high energy, which may result in carbon contamination during deposition. Co-reagents are then needed to remove the ligand during deposition.

Table 1: BLYP/DNP Co precursor ligand removal energies in kcal mol^{-1}.

Precursor	Co Oxidation State	Carbonyl Removal	Organic Ligand Removal
Co-Cp	1	28.8	42.3‡
Co-Allyl	1	19.5	30.4‡
CCTBA	0	15.7	1.4
Co-Amide	2	Not Applicable	143‡

‡ Denotes homolytic bond dissociation.

3.2. Surface adsorption and initial nucleation

Multiple growth surfaces were studied because selective deposition is critical for ALD applications such as integrated circuit manufacturing. This study ranks the relative adsorption and initial nucleation energies of Co precursors on various substrates. For the Co precursors in Figure 1, Table 2 contains calculated adsorption energies on bulk Cu, Ta_3N_5, SiO_2, and Co surfaces. CCTBA showed the least difference in adsorption energies between the surfaces. The differences in adsorption energies for Co carbonyl precursors on different surfaces are too small to explain experimentally observed selective deposition. The energy differences required to remove the organic ligand are likely responsible for Co deposition selectivity. Co-Amide showed favorable adsorption energy on OH-terminated SiO_2 with significant difference compared to naked Cu, suggesting preferential nucleation on the SiO_2 surface.

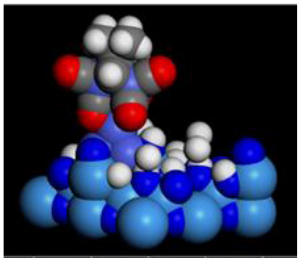


Figure 2: Optimized geometry for CCTBA adsorption on a Co_4 cluster with Ta_3N_5 terrace. Periodic boundary conditions are not shown for clarity.

Table 2: BLYP/DNP Co precursor adsorption energies on various surfaces in kcal mol^{-1}.

Precursor	Naked Cu111	NH-and NH_2 terminated Ta_3N_5 001	OH-terminated SiO_2 001	H-terminated Co001
Co-Allyl	-0.4	-2.1	-4.1	-1.6
Co-Cp	-0.9	-5.8	-3.7	-1.3
CCTBA	-0.3	-0.5	-1.4	0.4
Co-Amide	7.2	5.9	-8.6	-4.3

To quantify adsorption energies beyond the first ALD cycle, cluster-terrace surfaces were developed (Figure 2). Cluster-Terrace surfaces were modeled as a more realistic approach to study the initial stages of Co precursor nucleation, which is an ongoing challenge for the deposition of first-row transition metals (Knisley et al., 2013). Table 3 contains calculated adsorption energies of CCTBA on Co_4/Ta_3N_5 and Co_4/SiO_2 cluster-terrace surface models. In contrast to the adsorption energy on the bulk H-terminated Co surface in Table 2, CCTBA shows stronger adsorption on the naked Co cluster in the cluster-terrace surface model due to the presence of higher energy binding sites. During

adsorption on the metal cluster of the cluster-terrace surface, it was observed that CCTBA underwent spontaneous CO bridging suggesting a possible decomposition pathway consistent with *in situ* infrared absorption spectroscopy experimental observation in literature (Kwon et al., 2011). The differences in energies between cluster and terrace adsorption are lower for the SiO_2 surface than for the Ta_3N_5 surface. This result suggests that CCTBA will have better wetting on OH-terminated SiO_2 than on NH- and NH_2-terminated Ta_3N_5, which is consistent with experimental observations.

Table 3: BLYP/DNP CCTBA adsorption energies on cluster-terrace surfaces in kcal mol^{-1}, where Δ is the absolute difference between cluster and terrace adsorption energies.

Precursor	NH- and NH_2-terminated Ta_3N_5 001 terrace with naked Co_4 cluster			OH-terminated SiO_2 001 terrace with naked Co_4 cluster		
	Cluster	Terrace	Δ	Cluster	Terrace	Δ
CCTBA	-7.7	+4.5	12.2	-7.0	-4.2	2.8

3.3. Surface reaction mechanisms

Reaction energies for the complete decomposition mechanism of Co-Cp with H_2 co-reagent to Co metal on naked Cu and OH-terminated SiO_2 surfaces is shown in Figure 3. Ultimately, Co-Cp showed the best selectivity toward Co deposition on the naked Cu surface as opposed to the OH-terminated SiO_2 surface. The overall Cp ligand hydrogenation and CO removal energies are endothermic for both surfaces, but less favorable on the OH-terminated SiO_2 surface than on the naked Cu surface. Therefore, selective Co deposition on the naked Cu surface is preferred over the OH-terminated SiO_2, which is observed experimentally.

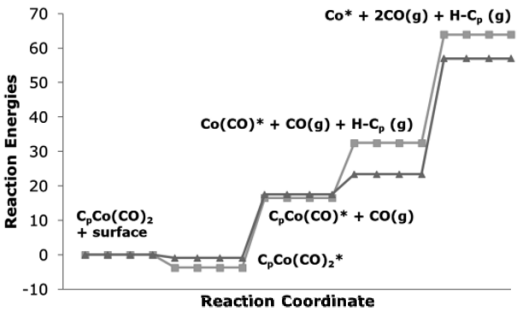


Figure 3: BLYP/DNP reaction energies for Co-Cp decomposition with H_2 co-reagent on naked Cu 111 (▲) and OH-terminated SiO_2 001 (■) in kcal mol^{-1}, where * denotes a surface species.

4. Conclusions

Computational chemistry techniques have been extended to Co precursor design and fundamental deposition understanding. This extension allows for improved mapping of the precursor design space by reducing experiments and providing new atomic-level insights to guide decision making. Future work for Co precursor design and deposition understanding will focus upon introducing thermal effects to current models (Eom et al., 2015) and estimating kinetic parameters for the development of multi-scale ALD reactor models.

References

Dassault Systèmes BIOVIA, Materials Studio 7.0, San Diego: Dassault Systèmes, 2016
T. Eom, T. Gwon, S. Yoo, B. J. Choi, M.-S. Kim, S. Ivanov, A. Adamczyk, I. Buchanan, M. Xiao, and C. S. Hwang (2015), J. Mater. Chem. C, 3, 1365
P. Hohenberg and W. Kohn (1964), Phys. Rev., 136, B864
T. J. Knisley, L. C. Kalutarage, and C. H. Winter (2013), Coord. Chem. Rev., 257 (23-24), 3222
J. Kwon, M. Saly, R. K. Kanjolia, and Y. J. Chabal (2011), Chem. Mater., 23 (8), 2068
Y. Saad, J. R. Chelikowsky, and S. M. Shontz (2010), SIAM Rev., 52 (1), 3
R. Stowasser and R. Hoffmann (1999), J. Am. Chem. Soc., 121, 3414

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50022-7

Methanol Production from Coke Oven Gas and Blast Furnace Gas

Lingyan Deng, and Thomas A. Adams II*

Department of Chemical Engineering, McMaster University, 1280 Main St. West, Hamilton, Ontario L8S 4L7, Canada
tadams@mcmaster.ca

Abstract

In this paper, a novel process is proposed which converts coke oven gas (COG) and blast furnace gas (BFG) from steel refineries into methanol. Specifically, we propose to use blast furnace gases (BFG) as a carbon source, which is a fuel with a low heating value that contains CO_2 and other gases. CO_2 is stripped from the BFG and blended with H_2-rich COG to adjust the $(H_2 - CO_2)/(CO + CO_2)$ molar ratio for methanol conversion. We also propose an advanced desulphurization process to remove certain sulphur compounds from the COG which can poison the methanol synthesis catalysts. The process design and simulation results using Aspen Plus are reported and used in a feasibility analysis to determine the potential environmental and energy benefits.

Keywords: Coke oven gas, blast furnace gas, COG desulphurization, CO_2 emission reduction, methanol production.

1. Introduction

In steel manufacturing, there are three major by-product off-gases: coke oven gas (COG), blast furnace gas (BFG), and basic oxygen furnace gas (BOFG). BOFG is produced in batch mode, which will not be considered in this particular paper. COG has the highest calorific value and a high H_2 content, while BFG has low calorific value but generally produced in the largest volumes (about 14 times that of COG). BFG consists of about 22 vol.% CO_2, 23 vol.% CO, and the rest mostly N_2. Detailed compositions of COG and BFG are shown in Table 1.

Table 1. COG and BFG compositions. * in PPMV. Source: Yang et al. (2009) and Uribe-Soto et al. (2015).

	HHV	Composition (vol.%)									
	MJ/kg	C_2H_2	CH_4	CO	CO_2	H_2	N_2	O_2	H_2S*	CS_2*	C_4H_4S*
COG	22.6-32.6	1.5-3.0	22-28	5.0-9.0	1.0-3.5	45-60	3.0-6.0	0.1-1.0	3420-4140	72-102	20-40
BFG	2.7	-	-	23.5	21.6	3.7	46.6	0.6	-	-	-

BFG is traditionally combusted for low-grade heat, leading to a very large amount of CO_2 emissions. Since steel manufacturing is a major source of CO_2 emissions, BFG valorization is an important pathway for CO_2 emissions mitigation.

As the world's largest producer of crude steel (Worldsteel Association, 2017), China has the most experience in dealing with steelmaking off-gases. According to the official Chinese Industrial Information website (2012), more than 57 commercialized methanol synthesis plants have been built and successfully operated in China, which use COG as the main raw material. However, the R parameter—the $(H_2 - CO_2)/(CO + CO_2)$ molar ratio—of COG is around 8 due to its large H_2 content. This is too high for methanol synthesis. According to Bermudez et al. (2013), the optimal R parameter is about 2.04. To bring it down, Bermudez et al. (2013) proposed that the reverse water gas shift reaction along with CO_2 injection can be used, or CO from coal gasification syngas can be blended with COG as Wu (2008) explained. However, refineries that co-produce COG and BFG are not taking advantage of the CO_2 and CO available in the BFG.

The idea for using BFG as a CO/CO_2 source is not new. Ghanbari et al. (2013) proposed and optimized a polygeneration system in which COG, BFG, and BOFG are used as raw materials to generate electricity, DME, or methanol. Later works by Ghanbari et al. (2015) studied replacing the coke used in blast furnaces with alternative reduction agents such as biomass, hydrogen, oil, natural gas, pulverized coal, and COG. Economic analyses of the polygeneration system were also performed (Ghanbari and Saxen, 2013). Although these initial studies showed promising results for the concept, they were based on the assumption that COG is sulphur free, and also did not use COG and BFG flow rates / production ratios that are comparable to commercial steel refineries. In fact, the methanol synthesis process is highly sensitive to sulphur: less than 0.1 ppmv in the syngas is permitted as Zhao et al. (2015) and Chen et al. (2009) stated according to commercialized MeOH synthesis experience. Some of the sulphur compounds such as thiophene (C_4H_4S) are particularly difficult to remove. Li (2009) stated that thiophene removal can be achieved with a complex two-stage hydro-desulphurization process which is commercialized in China for COG to Methanol but not for COG + BFG to methanol. Since the prior work did not consider thiophene or other sulphur compound removal, the proposed design structure and corresponding techno-economic analyses are likely to under-predict costs.

Therefore, it is important to consider many possible processes for thiophene removal. According to Wu (2008), two possible thiophene removal methods include (1) two-stage hydrodesulphurization and (2) non-catalyst partial oxidization (NCPO), which can crack methane and organic sulphur compounds simultaneously. Bermudez et al. (2010) and Cao et al. (2007) proposed CO_2 dry reforming (CDR) as a third approach. However, both the NCPO and CDR approaches have only been shown at the lab scale. CDR is a particularly promising technology because it can not only crack organic sulphur compounds and methane at the same time, but also, consume CO_2 to adjust the R parameter. Thus, from a systems perspective, the CDR approach has the potential for secondary benefits by reducing the complexity and size of the COG + BFG to MeOH synthesis process as a whole. Therefore, in this study, we present a new process and analysis which uses the CDR approach.

2. Methodology and Process Description

The proposed process, shown in Figure 1, was modelled in Aspen Plus v10 using the PR-BM physical property package for the gas related units and STEAMNBS for all water streams/operations. First, BFG (having a composition shown in Table 1) is fed to an MEA absorption-based unit (MEA) to extract the amount of CO_2 needed in the downstream process. The detailed MEA process was designed and simulated in ProMax according to the Tobiesen et al. (2007) study on blast furnace CO_2 capture, and the reader is referred

to that work for details. However, only a portion of the CO_2 in the BFG is extracted, since the available supply exceeds the downstream demand. The results of the ProMax simulation were implemented in Aspen Plus via a SEP block. The extracted CO_2 and the COG are fed to the CDR reactor, which is placed inside a furnace and maintained at high temperature (1200°C). In the CDR, the methane in the COG is reformed into synthesis gas, and the thiophene and carbon sulphide in the COG are reformed into H_2S, as demonstrated at the lab scale at this temperature by Zhang and Dong et al. (2010) and Cao et al. (2007). The CDR is modelled with RGIBBS, assuming chemical equilibrium. The results of the model show that organic sulphur is almost entirely converted to H_2S.

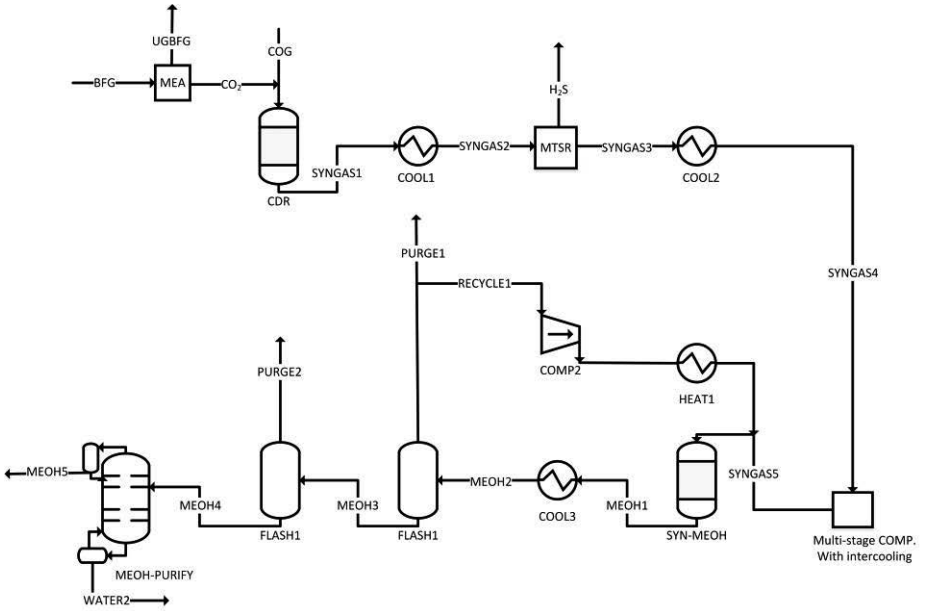


Figure 1. The proposed COG + BFG to methanol process using CDR for sulphur handling.

The reactor effluent is then cooled to 350°C and sent to a solvent-based middle temperature sulphur removal (MTSR) process. The model uses a SEP block for H_2S removal using recovery results from a previously published MDEA-based absorption process and ZnO catalyst-based middle temperature H_2S removal unit proposed by Li (2009), which reduces the H_2S content in the syngas to less than 0.1 PPMV. Note also that there are obvious heat integration opportunities that exist in this process (such as with off-gas recycle) or elsewhere in the steel refinery that were not considered in this work since this is a feasibility study.

The cleaned syngas is then feed to a multistage compressor to pressurize to 100 bar, using default compressor efficiencies (0.72) and air as the inter-stage coolant. The methanol synthesis reactor is a boiling-water shell-and-tube type, which helps maintain the temperature at 240°C. This is modelled using an isothermal RPLUG model with kinetics from Abrol et al. (2012) in order to account for the effects of recycle and inert content on reactor performance. Because economics were not considered in this preliminary study, the dimensions of the reactor (along with downstream recycle ratio) were manually adjusted to strike a balance between reactor yield and reactor weight. The reactor product

is cooled and flashed (using a FLASH2 block), with a percentage of the gas product recycled using a recycle compressor, with the rest being purged (flaring of this purge is not modelled). The liquid product contains methanol with higher-alcohol impurities, which are separated via distillation (modelled using RADFRAC).

The retentate BFG (mostly CO and N_2) could be used for heating in downstream steel manufacturing processes. Alternatively, the CO could also be extracted from the retentate after CO_2 separation via temperature swing adsorption (TSA). According to Ghanbari et al. (2013), up to 99 vol.% of CO could be extracted in this fashion. The high purity CO could then be recycled back to blast furnace (BF) to help reduce the coke requirement. However, the investigation of these options is out of the scope of this study.

3. Results and discussion

The results of the simulation are summarized in Table 2, with key stream conditions provided in Table 3. Overall, the system is effectively able to convert 99.90% of the CS_2 and 100% of the thiophene in the COG into H_2S, and 99.99% of the methane into CO and H_2. The conversion ratios of thiophene and methane increase with CDR reforming temperature, while the conversion ratio of CS_2 decreases with CDR reforming temperature. Approximately 0.58 kg CO_2 (in the BFG) is consumed per kg of MeOH produced.

Table 2. Important Aspen Plus / ProMax simulation results

CS_2 conversion rate (%)	C_4H_4S conversion rate (%)	CH_4 conversion rate (%)	COG production ability (kg-MeOH/kg-COG)	CO_2 consumption (kg-CO_2/kg-MeOH)
99.90	100	100	1.27	0.58

The carbon efficiency (defined as the number carbon atoms in the methanol divided by the number of carbon atoms in the BFG and COG) is 11.77%, which is low because only 10.86% of the CO_2 in the BOG was captured for methanol conversion. A more meaningful metric is the percentage of carbon atoms in the COG + captured CO_2 feed to the CDR that end up as methanol, which is much higher at 91.75%. The total direct CO_2 emissions for this process (through combustion of the two purge streams and of natural gas in the CDR furnace) is 0.85 tCO_2 per tMeOH produced. This is also 0.62 tCO_2 per tonne of raw material that goes into the CDR, which is about 43% lower than if the same amount of raw material are combusted directly (as they sometimes are in steel refineries).

The thermal efficiency—defined as the higher heating value (HHV) of the methanol divided by the sum of the HHV of the COG consumed, the HHV of the natural gas consumed in the furnace, and the utilities consumed in the process (including the MEA section)—is 45.40%. This is actually quite reasonable considering the primary feedstock is a waste source, and can be improved with proper heat integration. For example, the energy needed for heating is about 10.37 kJ/tMeOH, while the energy needed for cooling is about 15.80 kJ/tMeOH, demonstrating ample heat integration opportunities that is left for future work.

Table 3. Key stream conditions

	Temperature (°C)	Pressure (bar)	Flow rate (kg/kg MeOH)	HHV (MJ/kg)
COG	32.00	1.45	0.79	32.58
BFG	70.00	1.00	16.42	2.65
CO_2	70.00	1.00	0.58	0.00
UGBFG	70.00	1.00	11.82	3.67
Purge1	56.40	41.00	0.28	6.60
MEOH5	65.00	1.00	1.03	23.50

Although the CO_2 consumption potential of the process and a reduction of direct CO_2 emissions compared to ordinary combustion has been demonstrated, these results cannot be used to be conclude that this process is better for the environment than the status quo. Instead, a full cradle-to-grave life cycle analysis (LCA) is needed. This would largely be application specific, because although COG and BFG are usually ultimately combusted for one reason or another, how they are used varies from process to process. For example, displacing COG combustion for heat to produce methanol instead would probably require the combustion of some other fuel (likely natural gas) to provide the necessary heat, offsetting the reduction in CO_2 emissions somewhat. Or, displacing COG combustion for electricity generation to instead make methanol would then require increased electricity purchases to make up for the deficit in electricity. This could either increase or decrease the net lifecycle CO_2 emissions depending on the CO_2 intensity of the local grid. There are other secondary impacts which could cause an increase or decrease in net CO_2 emissions, such as the displacement of fossil-based methanol or the potential for increased use of fossil fuels. This is left for future work.

4. Conclusions

A process for converting COG and BFG to methanol has been presented, including the consideration of the removal of thiophene and other sulphur compounds. Overall, about 0.78 kg of COG (typically used for combustion) and 0.58 kg of CO_2 contained in BFG (typically destined for release to the atmosphere) are consumed per kg of methanol produced. The feasibility study demonstrated that the process is reasonably efficient and results in a net reduction in direct CO_2 emissions for a steelmaking process which would traditionally combust all of the COG and BFG for heat or other needs. This indicates that it is a good candidate to consider for a future techno-economic analyses or LCA in order to determine the potential economic and environmental benefits when considering the balance-of-refinery. However, the economic or environmental benefits of this process compared to business-as-usual will vary widely from existing refinery to refinery, and are likely to depend strongly on local factors such as carbon taxes, grid intensity, and the ways in which COG and BFG are currently used. For example, refineries that currently use COG for electricity production that have access to an electric grid with low CO_2 intensity are likely to gain the most environmental benefit by replacing the existing COG-to-electricity process with COG+BFG-to-methanol and purchasing additional grid electricity instead.

References

S. Abrol, C. M. Hilton, 2012, Modeling, simulation and advanced control of methanol production from variable synthesis gas feed, Computers & Chemical Engineering, 40, 117-131.

Worldsteel association, 2017, world steel in figures.

J.M. Bermúdez, B. Fidalgo, A. Arenillas, J.A. Menéndez, 2010, Dry reforming of coke oven gases over activated carbon to produce syngas for methanol synthesis, Fuel, 89.10, 2897-2902.

J.M. Bermúdez, N. Ferrera-Lorenzo, S. Luque, A. Arenillas, J.A. Menéndez, 2013, New process for producing methanol from coke oven gas by means of CO_2 reforming. Comparison with conventional process, Fuel processing technology, 115, 215-221.

China Industrial Information, 2012, Chinese State-of-the-art of Coke Oven Gas Synthesis into Methanol, http://www.chyxx.com/industry/201312/225545.html [accessed Oct. 4, 2017].

H.C. Cao, K.L. Bei, P.W., G.Z. Zhang, Y.F. Zhang, 2007, Discussion of the Desulphurization Process for Coke Oven Gas Synthesis to Methanol, Symposium on Methanol and Its Downstream Products New Technologies.

A.P. Chen, 2009, Analysis on the Purification Technology of Coke-Oven Gas to Methanol, Sci-Tech Information Development & Economy, 29, 214-215.

H. Ghanbari, H. Saxén, I. E. Grossmann, 2013, Optimal design and operation of a steel plant integrated with a polygeneration system, AIChE Journal 59.10, 3659-3670.

H. Ghanbari, F. Pettersson, H. Saxen, 2015, Sustainable development of primary steelmaking under novel blast furnace operation and injection of different reducing agents, Chemical Engineering Science 129, 208-222.

H. Ghanbari, H. Saxén, 2013, A Techno-Economic Analysis of Using Residual Top Gases in an Integrated Steel Plant, Chemical Engineering Transactions, 35, 169-174.

J.N. Li, 2009, The Process Improvements of Dry Desulphurization in the Production of Coke-oven Gas-to-methanol, Sci-tech Information Development & Economy,16, 180-182.

W. Uribe-Soto, J.F. Portha, J.M. Commenge, L. Falk, 2015, A review of thermochemical processes and technologies to use steelworks off-gases, ECOS 2015 - 28th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems.

F. A. Tobiesen, H. F. Svendsen, T. Mejdell, 2007, Modeling of blast furnace CO_2 capture using amine absorbents, Industrial & Engineering Chemistry Research, 46.23, 7811-7819.

C.M. Wu, 2008, Study on Methanol Production Technology from Coke Oven Gas, Gas & Heat, 1, 36-42.

Z.B. Yang, W.Zh. Ding, Y.Y. Zhang, X.G. Lu, Y.W. Zhang, P.J. Shen, 2010, Catalytic Partial Oxidation of Coke Oven Gas to Syngas in an Oxygen Permeation Membrane Reactor Combined with NiO/MgO Catalyst, International Journal of Hydrogen Energy, 35 (12), 6239–6247, doi:10.1016/j.ijhydene.2009.07.103.

G. Zhang, Y. Dong, M.Feng, Y. Zhang, W. Zhao, H.C.Cao, 2010, CO_2 reforming of CH_4 in coke oven gas to syngas over coal char catalyst, Chemical Engineering Journal, 156.3, 519-523.

Z.C. Zhao, Y.Y. Ding, 2015, Desulphurization and Purification Technology for Coke Oven Gas Synthesis into Syngas, Industrial Technology, 20, 40-40.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50023-9

Antibiotic Molecular Design Using Multi-Objective Optimization

Shweta Mapari[a*], Rex Gaumer[a], Kyle V. Camarda[a]

[a]*Department of Chemical and Petroleum Engineering, The University of Kansas, 1530 W. 15th St.,4132 Learned Hall, Lawrence, Kansas 66045, USA*
camarda@ku.edu

Abstract

This project aims to design a novel drug molecule via a computer aided molecular design approach to treat the bacterial infections caused by *Acinetobacter baumannii*. Research is acutely needed to develop novel therapies to treat resistant infections. *A. baumannii*, a pathogen causing infections such as pneumonia and bacteremia is on the highest priority. In order to treat this pathogen, a molecular design problem is formulated with pazufloxacin as a starting point. This multi-optimization problem has two objectives: minimizing MIC (Minimum Inhibitory Concentration) and maximizing the LD-50 value. Quantitative Structure Activity Relationships (QSAR) are created based on 3D-MoRSE descriptors and are used to predict these properties. The optimization problem is formulated using a multi-objective format in order to discover the trade-offs between various solutions. It includes both objectives as well as the structural feasibility constraints, resulting in a multi-objective, mixed integer non-linear program (MINLP). Because of the large number of potential chemical structures and the uncertainty in the structure-property correlations, stochastic algorithms are preferred to solve the resulting MINLP. The problem was solved using a stochastic algorithm, artificial bee colony (ABC), to find a set of near-optimal solutions, corresponding to novel candidate antibiotics. ABC's use of swarm intelligence provides a breadth of information which can be applied to the evaluation of pareto-optimal surface.

Keywords: stochastic optimization, multi-objective optimization, 3D structural descriptors.

1. Introduction

Antibiotics have been helping combat various bacterial infections for over a century. However, bacteria develop resistance towards these drugs, often by developing alternative metabolic pathways to circumvent inhibited ones. WHO recently published an article that describes a list of bacteria which have become resistant to multiple antibiotics, thus specifying the necessity for developing new drug molecules (WHO Report, 2017). *Acinetobacter baumannii* is a multidrug resistant bacteria that thrives in hospital environments and a major cause of nosocomial infections (Perez et al., 2007). In order to treat the infections caused by this bacteria effectively, a novel pharmaceutical is required. The traditional approach of developing a new drug is expensive, time consuming and requires much trial and error. However, computer aided molecular design (CAMD) is a tool that allows the creation of a molecule having certain property values, using comparatively less amount of time and money (Gani et al., 2002).

CAMD uses optimization techniques to form a new molecule. The optimization problem formulated for this work contains two objective functions, and structural feasibility constraints, which results in a mixed integer non-linear program. Due to the non-convex nature of this problem, stochastic algorithms are preferred. We have used the Artificial Bee Colony (ABC) algorithm to find near-pareto optimal solutions to the problem. This is the first time, CAMD approach has been used for the design of an antibiotic molecule using ABC algorithm.

When a molecule having more than one target property value is to be designed using a single objective, the properties are weighed against each other and combined into an equation. The solution of this problem is a point where one property is favored over the other. The goal of multi-objective optimization is to obtain a set of points which represent trade-off between the properties as all the properties are given equal importance. Hence, instead of getting a global or near-optimal value, we get a series points along the pareto-optimal curve. We can then choose one that meets our purpose from a range of points representing different set of property values.

2. Property prediction using 3D-MoRSE descriptors

Property prediction is implemented in this work using structure-property relationships. The property values used in developing such relationships are found experimentally or taken from published data and then correlated using structural descriptors. The structural descriptors used in this work are 3D-MoRSE descriptors, (Molecular Representation of Structure based on Electron diffraction). These descriptors contain information about the 3-D orientation of groups within. This information is obtained from the function called 'molecular transform'. The descriptors are calculated by using Equation 1.

$$I(s) = \sum_{i=1}^{A-1} \sum_{j=i+1}^{A} w_i w_j \frac{\sin(sr_{ij})}{sr_{ij}} \tag{1}$$

where *I(s)* is the MoRSE index, *w* is an atomic number, r_{ij} are the atomic distances between *i* th and *j* th atoms, s represents scattering of electrons by the group of *A* atoms at point r_i and *A* is the number of atoms (Todeschini and Consonni, 2008).

Antibiotics are used to treat the bacterial infection within host body, so it is very important that they are strong enough to inhibit the growth of bacteria. At the same time, they should not cause toxicity to the host which will produce undesired side effects. Hence, MIC and LD-50 value are used as objectives for this work. Both these properties are completely dependent on the biology of the organism. It is very difficult to predict such dependence with the help of topological/2-D descriptors, as they only provide information on the structural features of the molecule and hence work best for predicting physicochemical properties. However, when biological activity of a compound is to be predicted, it is important to consider the way that molecule binds with a biological target to create certain response. For example, in most of the cases for antibiotics, the response is generated when the drug molecule is attached to one of the

enzymes in a micro-organism (Saíz-Urra et al., 2006). Here, we are considering MIC and LD-50 values. MIC values represent the minimum concentration required to inhibit the growth of bacteria. LD-50 values represent the dose that causes death of half of the host population. Both these properties are strongly dependent on the way an antibiotic molecule interacts with receptors in body of the organism in question. Hence, 3D descriptors are needed for effective predictions of biological properties. Also, 3D descriptors have been successfully employed in developing structure-activity relationships in previous work (Schuur et al., 1996).

The quantitative structure activity relationships (QSARs) used in this work were developed for MIC and LD-50 values for an antibiotic class, fluoroquinolones as mentioned before.

Table 1: Correlations and their statistics

Property	**Correlation**	**Correlation Coefficient (R^2)**	**q^2**
MIC	Log(MIC) = 4.74 -0.01*Mor01u - 0.1*Mor02u - 0.59*Mor03u + 1.14*Mor04u + 0.87*Mor07u + 1.18*Mor08u + 3.35*Mor09u + 3.11*Mor10u	0.87	0.88
LD-50	Log(LD-50) = 8.46 +0.01*Mor01u + 0.25*Mor02u + 1.64*Mor03u – 0.72*Mor04u - 3.79*Mor07u + 1.42*Mor08u + 1.38*Mor09u	0.94	0.84

MIC values and LD-50 values were obtained from the literature. Mor01u, Mor02u, Mor03u, Mor04u, Mor07u, Mor08u, Mor09u, Mor10u are 3D-MoRSE descriptors evaluated by the online software e-Dragon (Tetko et al., 2005). Different combinations of these descriptors were used to get best fit models for MIC and LD-50 values using the method of least squares. These models are given in the Table 1. The statistics of the correlations are also shown in the table. q^2 is the prediction quotient that shows how well the correlation predicts the property value, for a similar compound that was not part of the data.

3. Formulation of the MINLP

Since, we are solving multi-objective optimization problem, we have two objectives. They are the QSARs for MIC and LD-50. The constraints include structural feasibility and connectivity constraints. Even though these constraints are not explicitly present in the problem, we make sure that they are accounted, by replacing a group with another group having the same valency. Pharmacophore is part of a pharmaceutical molecule that is responsible for getting the desired biological effect. So, another constraint was implicitly added, which does not allow the pharmacophore structure to change. As it can be seen from Equation 1, the information obtained from molecule is presented in the non-linear form, which introduces element of non-linearity in the optimization problem.

4. Artificial Bee Colony (ABC) algorithm

Many algorithms have been explored in the past to solve non-linear optimization problems. However, a comparative study (Karaboga and Akay, 2009) has shown that ABC, an algorithm based on swarm intelligence, can perform better than other stochastic algorithms. The ABC algorithm is based on the intelligent foraging behavior of a group of honeybees. ABC possesses three kinds of agents: employed bees, onlooker bees and scout bees. These bees are represented in the form of partitioned adjacency matrix of the molecule. The algorithm goes through three phases: the employed phase, the onlooker phase and the scout phase. In the employed phase, a set of feasible solutions are created by changing one group at a time. In the onlooker phase, these solutions are evaluated by onlooker bees and ranked from best to worst depending on their property values. Finally, in the scout phase, if a solution is not improved within a certain number of iterations, it is abandoned and replaced with a new solution. The algorithm repeats itself through these phases until a near-optimal solution is obtained (Karaboga and Basturk, 2007).

The ABC colony algorithm has been successfully used before to solve multi-objective problems (Akbari et al., 2012, Luo et al., 2017). However, this algorithm has not previously been used for multi-objective CAMD problems.

5. Result and conclusion

Drug discovery is a process consisting of many stages. The new molecule designed in this work can serve as a good starting point in that process. CAMD helps us design a molecule and predict its properties before any experiments are carried out. We are interested in getting the property values that are closer to the target property values and ABC algorithm, being stochastic in nature efficiently finds the near-optimal solution. A deterministic algorithm will find the global optimum of an optimization problem, if it exists, however that will definitely take more time than a stochastic algorithm.

The goal of this work was to design a molecule having minimum MIC and maximum LD-50 value. The solutions of the MINLP for both, single objective and multi-objective optimization problems, were pazufloxacin derivatives. The solutions of multi-objective problem were variety of structures offering wide range in MIC and LD-50 values. The use of multi-objective optimization allows to explore the pareto-optimal front, giving multiple solutions. The use of 3D-MoRSE descriptors to develop QSARs resulted in giving mathematical models with higher predictive power. For future work, the performance of ABC algorithm can be studied for different classes of antibiotics and different multi-objective problems.

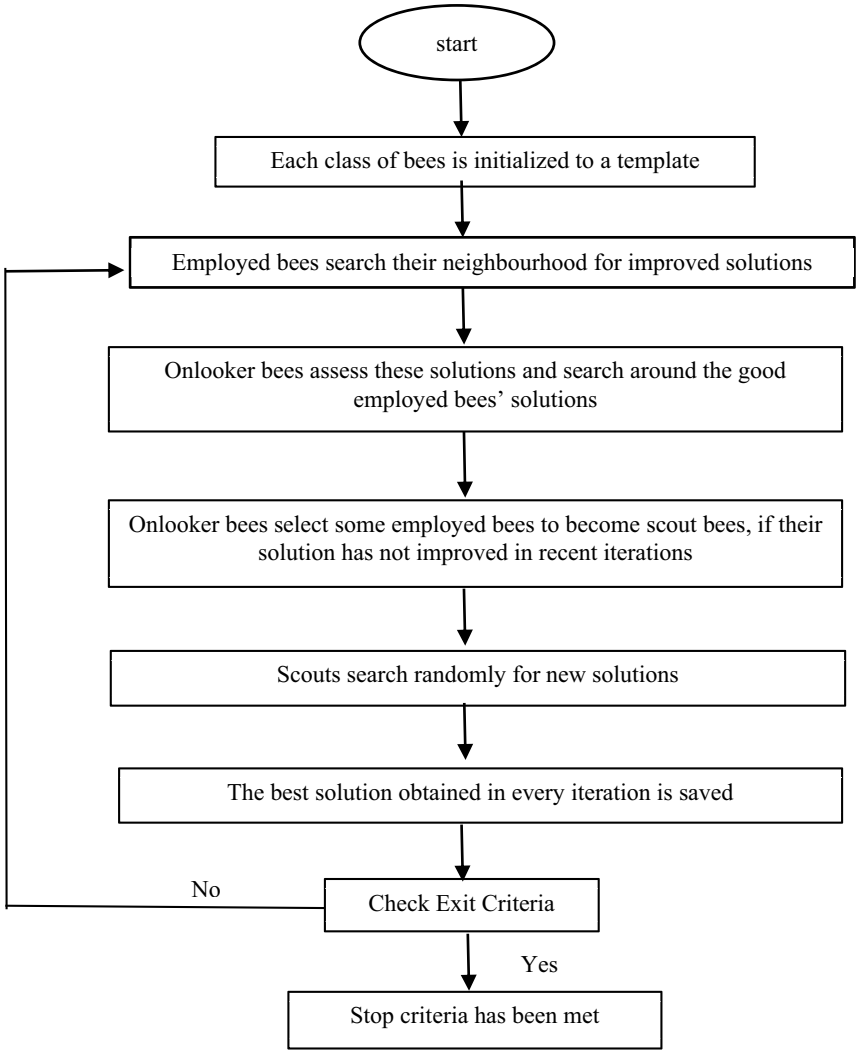


Figure 1. Artificial Bee Colony Algorithm Flowchart for CAMD

References

Akbari, R., Hedayatzadeh, R., Ziarati, K., & Hassanizadeh, B. (2012). A multi-objective artificial bee colony algorithm. *Swarm and Evolutionary Computation*, *2*, 39-52.

Achenie, L., Venkatasubramanian, V., & Gani, R. (Eds.). (2002). *Computer Aided Molecular Design: Theory and Practice*(Vol. 12). Elsevier.

Karaboga, D., & Akay, B. (2009). A comparative study of artificial bee colony algorithm. *Applied Mathematics and Computation*, *214*(1), 108-132.

Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *Journal of global optimization*, *39*(3), 459-471.

Luo, J., Liu, Q., Yang, Y., Li, X., Chen, M. R., & Cao, W. (2017). An artificial bee colony algorithm for multi-objective optimisation. *Applied Soft Computing*, *50*, 235-251.

Perez, F., Hujer, A. M., Hujer, K. M., Decker, B. K., Rather, P. N., & Bonomo, R. A. (2007). Global challenge of multidrug-resistant Acinetobacter baumannii. *Antimicrobial Agents and Chemotherapy*, *51*(10), 3471-3484.

Schuur, J. H., Selzer, P., & Gasteiger, J. (1996). The coding of the three-dimensional structure of molecules by molecular transforms and its application to structure-spectra correlations and studies of biological activity. *Journal of Chemical Information and Computer Sciences*, *36*(2), 334-344.

Sliwoski, G., Kothiwale, S., Meiler, J., & Lowe, E. W. (2014). Computational methods in drug discovery. *Pharmacological Reviews*, *66*(1), 334-395.

Saíz-Urra, L., González, M. P., & Teijeira, M. (2006). QSAR studies about cytotoxicity of benzophenazines with dual inhibition toward both topoisomerases I and II: 3D-MoRSE descriptors and statistical considerations about variable selection. *Bioorganic & Medicinal Chemistry*, *14*(21), 7347-7358.

Tetko, I. V., Gasteiger, J., Todeschini, R., Mauri, A., Livingstone, D., Ertl, P., ... & Tanchuk, V. Y. (2005). Virtual computational chemistry laboratory–design and description. *Journal of Computer-Aided Molecular Design*, *19*(6), 453-463.

Todeschini, R., & Consonni, V. (2008). *Handbook of Molecular Descriptors* (Vol. 11). John Wiley & Sons.

World Health Organization, Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics, Internal Report, (2017)

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50024-0

Design and Control of Diphenyl Carbonate Synthesis involving Single Reactive Distillation with the Feed-Splitting Arrangement

Hsiang-Ning Chang[a], Chien-Ying Chen[a], Hao-Yeh Lee[a*]

[a]*Department of Chemical Engineering, National Taiwan University of Science and Technology, Taipei 10607 Taiwan*
haoyehlee@mail.ntust.edu.tw

Abstract

This study presents an innovative diphenyl carbonate (DPC) process via diethyl carbonate (DEC) and phenyl acetate (PA) as reactants applying single reactive distillation (RD) with the feed-splitting arrangement. This paper uses the boiling point ranking of reactants and feed-splitting to attain the equivalent design. The optimal design was determined by the minimum total annual cost (TAC). In process dynamics, all the product and byproduct can achieve the specifications by using temperature controllers during $\pm$10 % throughput and 10 % composition disturbances. In contrast to Chen et al. in 2015, the results show proposed design can have high energy saving and good control performance.

Keywords: Reactive distillation; Feed-splitting; Diphenyl carbonate; Diethyl carbonate

1. Introduction

DPC is an essential and nontoxic precursor for the polycarbonate (PC) production. With the development of carbonate industry, the need for DPC will be increasing. Kim and Lee (1999) suggested a transesterification way of dimethyl carbonate (DMC) and phenol to manufacture DPC. However, the equilibrium constant for transesterifications of DMC and phenol was only 3×10^{-4} at 180 °C in the research of Tundo and Selva (2002). To overcome this problem, diethyl carbonate (DEC) and phenyl ester (PA) transesterification to produce the product DPC and the byproduct ethyl acetate (EtAc) was proposed. The advantages of this process include higher equilibrium constant and no azeotropes.
The study for the innovative design of DPC process via one RD with the feed-splitting arrangement is divided into five sections. Section 2 presents the thermodynamic and kinetic models. Next, Section 3 explains the DPC process of transesterification reaction from DEC by applying RD and feed-splitting concept. The process dynamics with throughput and composition disturbances are presented in Section 4. The final section is the conclusion of this paper.

2. Thermodynamic and kinetic models

2.1. Thermodynamic model

In this study, the DPC synthesis via DEC as reactant is called DEC configuration. In DEC configuration, thermodynamic properties are described by NRTL model. The experimental vapor-liquid equilibrium (VLE) data were given by Ho et al. (2014). The boiling point ranking is in order of: DPC (302.0 °C) > EPC (239.5°C) > PA (195.7 °C) > DEC (126.8 °C) > EtAc (77.2 °C).

2.2. Kinetic model

The kinetics of DEC configuration were referred to Lin (2014). DPC process is two-step reaction. First step reaction is transesterification of reactants DEC and PA to intermediate ethyl-phenyl carbonate (EPC) and byproduct EtAc. Next step, EPC and PA generate the main product DPC and byproduct EtAc. The two-step reactions are shown in Eq.(1) and Eq.(2).

$$DEC + PA \leftrightarrow EPC + EtAc \tag{1}$$

$$EPC + PA \leftrightarrow DPC + EtAc \tag{2}$$

The DEC configuration used titanium (IV) ethoxide as the homogeneous catalyst. The equilibrium constant rate is described by the Arrhenius equation which is presented in Eq.(3). In addition, the kinetic parameters are listed in Table 1.

$$k_i = k_0 e^{-E_a/RT} \tag{3}$$

Table 1: The kinetic parameters of DPC synthesis

	k_0 (m^3/kmol·s)	E_a (kJ/kmol)
k_1	2.8×10^5	83,973
k_{-1}	6.2×10^0	49,960
k_2	2.4×10^1	49,979
k_{-2}	1.0×10^{-1}	22,676

3. Process design

This section is divided into two parts: conventional/thermally coupled design, and proposed design.

3.1. Conventional RD process and thermally coupled design (Chen et al.,2015)

Chen et al. (2015) investigated DEC configuration with excess PA in thermal coupling design. Figure 1 compared the conventional design (Figure 1(a)) with thermally coupled design (Figure 1(b)). The thermally coupled arrangement is applied between the RD2 and C3 by removing the condenser of RD2. In the thermally coupled arrangement, the vapor stream from the top of RD2 was fed to the distillation column (C3). C3 extracted a liquid stream from stage 5, and this stream was recycled back to the top of RD2. The results show that the thermally coupled configuration can save around 10 % energy requirement than the conventional one.

3.2. Proposed design

This research offered the equivalent design by using the feed-splitting arrangement to adjust PA mole flow rate. According to the boiling point ranking, the heavier component feeds in the upper section of RD. Identically, the lighter component feeds in the lower section of RD. Therefore, the transesterification reaction occurs in the middle section of RD.

This work illustrates the proposed design to achieve 99.5 mol % DPC and 99.5 mol % EtAc for industrial use. There are seven design variables in both configurations which are total stages (N_T), reactive section (N_{rxn}), column pressure (P), the spilt ratio (SR) of PA feed, the 1st PA feed stage ($NF1_{PA}$), the 2nd PA feed stage ($NF2_{PA}$), and DEC feed stage (NF_{DEC}). SR is defined as the 1st PA feed flow rate over total PA feed flow rate. In the terms of catalyst, the decomposition temperature is 312 °C. Hence, the stage temperature over 312 °C was presumed that there is no reaction.

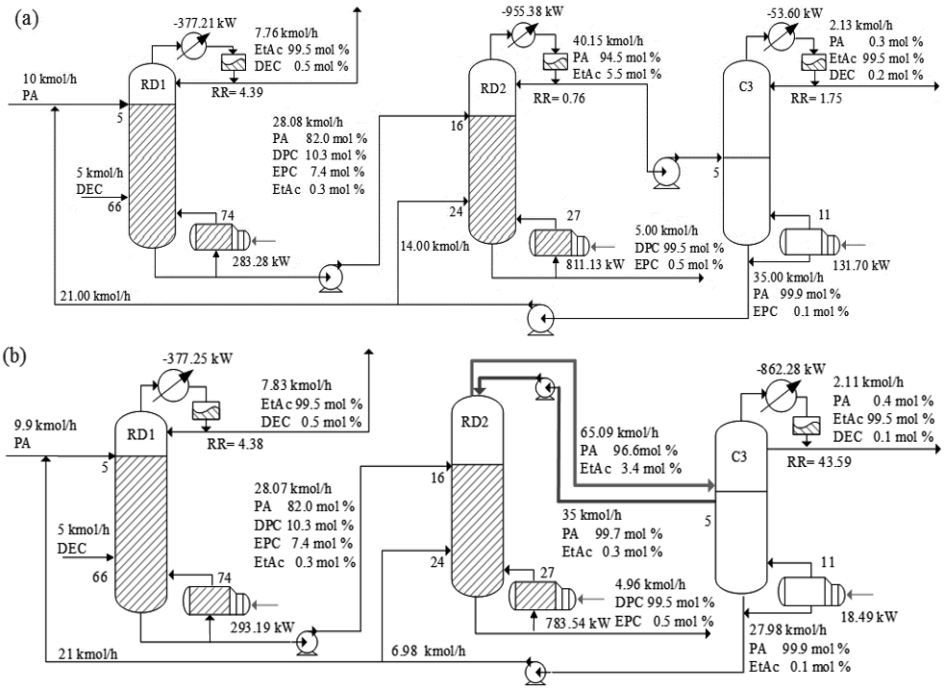


Figure 1: (a) Conventional RD configuration; (b) Thermally coupled design (Chen et al., 2015)

As shown in Figure 2, $NF1_{PA}$ must be arranged upper than NF_{DEC} because DEC has the lower boiling point than PA. The first-step reaction occurred in the top section of the reactive zone to generate EPC and EtAc. To achieve the counter-current flow for reactants in the second-step reaction, the other PA feed location, $NF2_{PA,}$ should be specified at the lower section of the reactive zone to produce DPC and by-product. Hence, the high purity by-product and DPC can be obtained and drained out from the top and bottom of RD column, respectively. Besides, the optimal is determined by the minimum TAC, which is shown in Eq.(4).

$$TAC = TAC_{operating} + TAC_{capital}/3 \tag{4}$$

The optimal results are shown in Figure 2(a). RD involves 59 stages in the total, and reactive section is from stage 2 to stage 47. In the RD column, the fresh PA stream was set 10 kmol/h and split into two streams. $NF1_{PA}$ is located at the second stage of RD. $NF2_{PA}$ is set in stage 47. The SR is 0.35 which means the flow rate of $NF1_{PA}$ and $NF2_{PA}$ are 3.5 and 6.5 kmol/h, respectively. In addition, PA: DEC is 2: 1 in the overall reaction. Therefore, the flow rate of DEC is set as 5 kmol/h and fed in stage 46.

Figure 2 (b) and (c) indicate the composition and temperature profiles of the proposed configuration. As shown in Figure 2 (b), DEC was consumed in the top reactive section because the first-step reaction occurs at the top of the RD column, whereas PA reacted in the middle of the column. Therefore, the mole fraction of PA had a downward trend in the center of the column. The PA mole fraction tended to increase in the bottom portion of the column because the other PA stream was fed in stage 47. Meanwhile, EPC had a slight increase in the bottom column due to the higher boiling point. Furthermore, EPC and PA reacted at the bottom of the reactive section. In the same pattern, EtAc decreased from the top to the bottom of the RD column. In contrast, DPC rapidly increased at the bottom of the RD column. In the temperature profile, the temperature gradually rose from the top to the bottom of the RD column since the transesterification reactions are endothermic reactions.

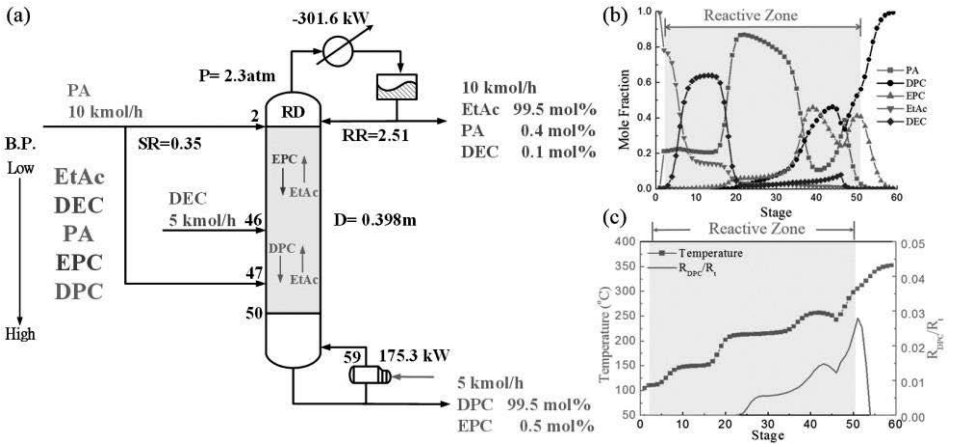


Figure 2: The DEC proposed design of (a) steady-state results (b) composition profile (c) temperature profile

Compared with the study of Chen et al. (2015), the proposed configuration could reduce energy consumption by 75.66 % for condenser duty, 83.97 % for reboiler duty, resulting in a 79.57 % lower TAC.

4. Process dynamics

In the dynamic simulation, two control loop level designs were employed: Inventory control loop and quality control loop. The objective of the inventory control loop scheme is to maintain the system safety. The main purpose of the quality control loop is to attain the product specifications for industrial usage. To be more precise, the quality control loop scheme adopts the temperature controller to meet the product specifications. This work investigated the open-loop sensitivity analysis with 0.05 % change in reboiler duty (QR), feed ratio (FR), reflux ratio (RR), and split ratio (SR). On the basis of sensitivity analysis, the tray temperatures are chosen by finding the location where the temperature profile is steep: stage 6, 18, 37 and 54.

The RGA analysis results of dual and triple temperature control are summarized in Figure 3(a). In CS1, FR and QR were as two manipulated variables for dual-point temperature control. FR was selected to control the stage 18, and QR was selected to control the stage 54. In addition, the tray temperature of stage 18 is controlled by manipulating FR, and the tray temperature of stage 54 is controlled by manipulating RR, named CS2. From triple

RGA analysis, the suggested pairings are 6th stage temperature to pair with RR, 18th stage temperature to pair with FR, and 54th stage temperature to pair with QR in CS3. The proposed overall control strategy, CS3 is presented in Figure 3(b).

The processing throughput was affected by the economic situations in reality. To simulate the actual situations, this research did the ±10 % throughput disturbance and 10 % composition disturbance, and the results are shown in Figure 4 and 5, respectively. Furthermore, this study explores the feasibility of control schemes.

In the proposed configuration, CS3 is the best control structure among three different control strategies. According to the results demonstrated in Figure 4 and 5, the deviation (%) of product DPC and by-product EtAc purities is the smallest in CS3 because CS3 adopted triple temperature control to control the sensitivity trays. CS3 settles to steady state within 6 hours in both throughput and composition disturbance.

(a)

$$\Lambda = \begin{matrix} \text{FR} & \text{QR} \\ \begin{bmatrix} 0.90 & 0.10 \\ 0.10 & 0.90 \end{bmatrix} & \begin{matrix} T_{18} \\ T_{54} \end{matrix} \end{matrix}$$

$$\begin{matrix} \text{FR} & \text{RR} \\ \begin{bmatrix} 0.90 & 0.10 \\ 0.10 & 0.90 \end{bmatrix} & \begin{matrix} T_{18} \\ T_{54} \end{matrix} \end{matrix}$$

$$\begin{matrix} \text{FR} & \text{QR} & \text{RR} \\ \begin{bmatrix} 0.11 & -4.63 & 5.52 \\ 0.85 & -0.66 & 0.81 \\ 0.05 & 6.28 & -5.33 \end{bmatrix} & \begin{matrix} T_6 \\ T_{18} \\ T_{54} \end{matrix} \end{matrix}$$

(b)

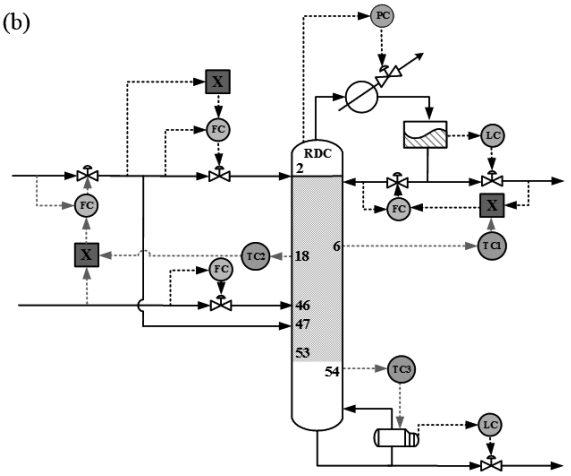


Figure 3: (a) RGA analysis results (b) The overall control scheme of CS3.

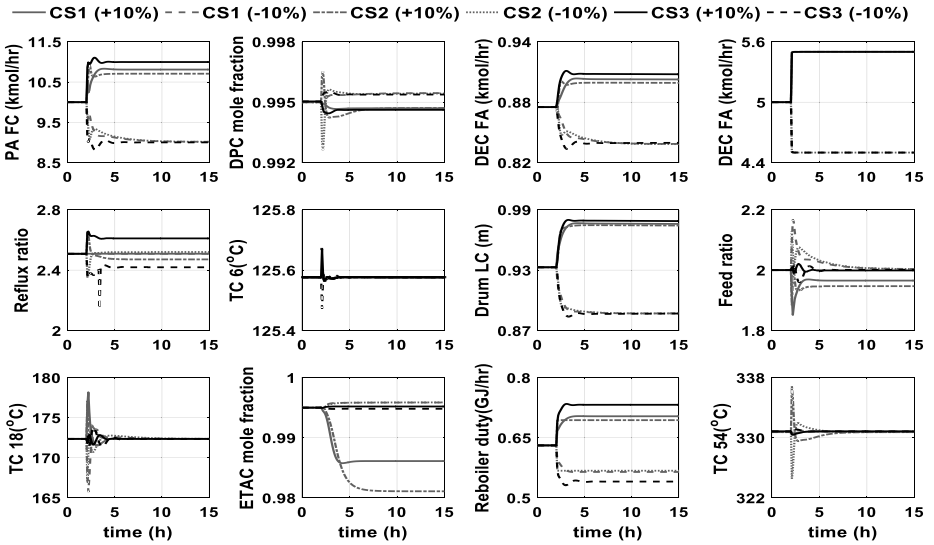


Figure 4: The results of ±10 % throughput disturbance in CS1 to CS3.

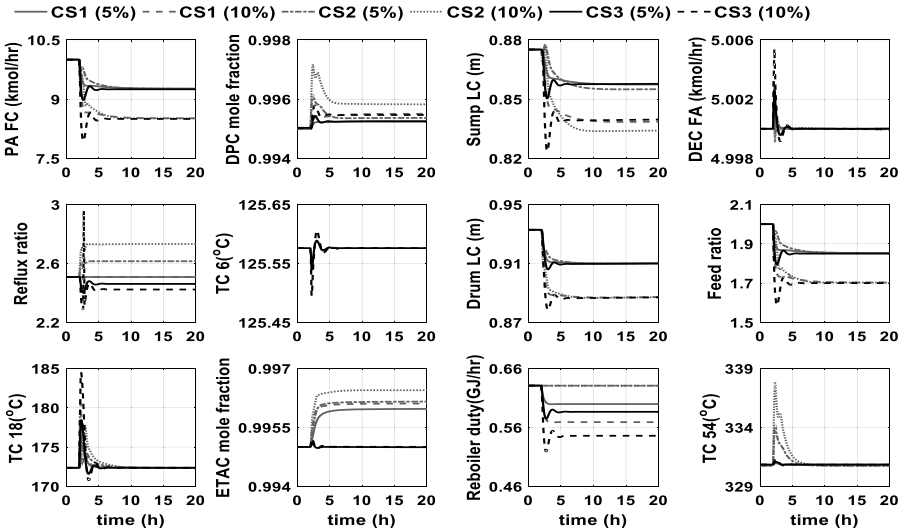


Figure 5: The results of 10 % composition disturbance in CS1 to CS3.

5. Conclusion

This research produced some major contributions: First, in comparison with the studies of Chen et al. (2015), the proposed design combines the RD column and distillation column into single RD column. Besides, there is no remixing effect was observed in the proposed design. Second, compared to the aforementioned design, the DEC proposed configuration with equivalent design can save 75.66 % in condenser duty, 83.97 % in reboiler duty, and 79.57 % in TAC. Third, dynamic performance results reveal that CS3 is the best control scheme in three different control strategies.

Acknowledgement

The authors gratefully acknowledge the Ministry of Science and Technology, Taiwan (project number: 106-2221-E-011-128-) for providing financial support.

References

D.Y. Chen, K. C. Weng, S. J. Wang, H. Y. Lee, 2015, Design of Diphenyl Carbonate Process via Reactive Distillation Configuration, Chem. Eng. Trans., 45, 1195-1200.

H.Y. Ho, S. S. Shu, S. J. Wang, M. J. Lee, 2015, Isothermal (Vapor + Liquid) Equilibrium (VLE) for Binary Mixtures Containing Diethyl Carbonate, Phenyl Acetate, Diphenyl Carbonate, or Ethyl Acetate, J. Chem. Thermodynamics., 91, 35-42.

W. B. Kim, J. S. Lee, 1999, A New Process for the Synthesis of Diphenyl Carbonate from Dimethyl Carbonate and Phenol over Heterogeneous catalysts, Catal Lett, 59, 83-88.

S. S. Lin, 2014, Kinetics Behavior Study of the Synthesis of Diphenyl Carbonate at Elevated Temperatures and Pressures. Master Thesis, National Taiwan University of Science and Technology.

P. Tundo, M. Selva, 2002, The Chemistry of Dimethyl Carbonate, Acc. Chem. Res, 35, 706-716.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50025-2

Scale-up Modeling of a Pharmaceutical Crystallization Process via Compartmentalization Approach

Merve Öner[a], Christian Bach[a], Tannaz Tajsoleiman[a], Getachew S. Molla[a], Michael F. Freitag[b], Stuart M. Stocks[b], Jens Abildskov[a], Ulrich Krühne[a], Gürkan Sin[a*]

[a]*Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads, Building 227m Lyngby DK-2800, Denmark*
[b]*LEO Pharma A/S, Industriparken 55, Ballerup DK-2750, Denmark*
**gsi@kt.dtu.dk*

Abstract

In this work, a scale-up model of a pharmaceutical batch cooling crystallization process was developed. The model is based on a compartmentalization approach of a large-scale crystallizer. The compartmental properties (*e.g.* volume, location and flux connections) were extracted by using two different, manual and automatic zoning, strategies. Each compartmentalization technique is based on a mixing study in the large-scale reactor with the help of computational fluid dynamics (CFD). The scale-up model was simulated in MATLAB/Simulink environment. The simulation results show the significant role of the fluid dynamics on the simulated process performance in terms of crystal size distribution and final average crystal diameter by means of comparing the multi-compartment model with a single-compartment well-mixed model.

Keywords: Scale-up modeling, pharmaceutical crystallization, compartmental modelling, computational fluid dynamics.

1. Introduction

Prediction of the influence of the crystallizer scale on the process behavior and performance (*e.g.* crystal quality attributes) is one of the major challenges in the design of industrial crystallization processes. Fluid dynamic conditions of large-scale crystallizers are far from well-mixed behavior compared to lab scale systems. This leads to spatial variations of critical process variables such as temperature, super-saturation, particle concentration within crystallizer geometry (Kramer et al., 1999). Crystallization process models based on the assumption of well-mixed behavior are not representative for a scaled-up process, therefore not credible for the use in supporting optimal design of industrial crystallizers. A more detailed insight into mixing behavior and its consequences for local crystallization phenomena – such as a spatially distributed parameter model - must be taken into account in order to achieve reliable process design and scale-up (Kramer et al., 2000, Myerson, 2002). The main objective of this work is to develop a predictive scale-up model of a pharmaceutical crystallization process based on a compartmentalization approach. Fully developed computational fluid dynamics (CFD) models (*e.g.* including the crystallization kinetics, population balance etc.) could be used for modeling the large-scale system, as opposed to compartmental modeling. However,

the major drawback of the full CFD models in the industrial practice is the high computational demand. In comparison, 3 min. transient CFD simulation of a non-reactive mixing in a reactor with 218 L liquid volume and 100 rpm stirrer speed took about 6 days to be completed on a computer cluster with 2x Intel Xeon 2.80 GHz processor with 128 GB RAM. Further implementation of the crystallization kinetics and population balance equation (PBE) into CFD simulation will increase this computational demand from days to weeks. On the other hand, a compartmental model simulation of the same system including crystallization kinetics and PBE for a 50 min. of crystallization, which was run in MATLAB/Simulink on computer with Intel Core i5-6200U 2.30 GHz processor with 8 GB RAM, took 10 min. to be completed. Compartmental modeling is a reduced-model approach, in which the detailed mixing and fluid dynamic information obtained from non-reactive CFD simulation is coupled with the process kinetics in a simpler simulation environment (Wells and Ray, 2005). Besides, this approach provides a unique opportunity to take into account the effects of non-homogeneous mixing on the process performance in many other applications as well *e.g.* process control and sensitivity analysis. While compartmental modeling is a well-known and successfully applied technique in other areas of process engineering, there is a significant lack of study on the full application of a compartmental modeling to scaling-up of a pharmaceutical crystallization processes.

2. Development of The Scale-up Model

2.1. Crystallizer configuration

A pilot scale crystallizer as shown in Figure 1 was used for the development of the scale-up model. The diameter of the vessel, T is 0.64 m and the liquid height, H is 0.74 m, corresponding to a liquid volume of 218 L. A marine type of impeller with impeller diameter, D 0.30 m is located 0.27 m above the bottom of the vessel (impeller clearance, C). Four baffles are located with 9 mm clearance from the vessel wall.

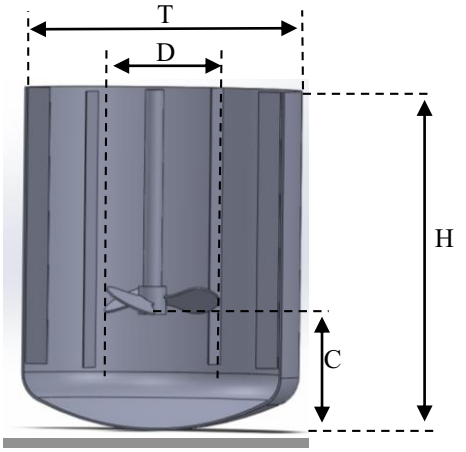


Figure 1. Pilot scale crystallizer.

2.2. CFD simulation

Transient CFD simulations of the crystallizer were performed with ANSYS CFX release 17.1. The representative Navier-Stokes equations were solved to predict fluid dynamics behavior in a single-phase volume. The standard k-ε model was used as the turbulence model. Rotation speed, N was 100 rpm that creates a turbulent flow within the studied

volume. The crystallizer vessel was divided into two domains: a rotating zone (rotating domain) and a stationary zone (stationary domain), which correspond to a cylindrical zone around impeller rotating at the impeller speed and a zone covering the rest of the crystallizer, respectively. The rotation of impeller was modelled with frozen rotor approach as frame change/mixing model between the intersections of two domains. The stationary domain consists of hexahedron mesh of 1,037,970 elements, while the rotating domain consists of tetrahedron mesh with total 118,895 elements. The time step used in the simulations was equal to 0.001 s. When the velocity field was fully developed, or in other terms, the steady state behavior of the velocity components was observed, the simulation was stopped and transient-average value of the flow velocity components was recorded.

2.3. Compartmentalization

Application of compartmental modelling to pharmaceutical batch cooling crystallization requires the division of the crystallizer into a finite number of compartmental volumes. Minimized or negligible gradients *e.g.* in temperature, crystal distribution, super-saturation and energy dissipation should exist within individual compartmental volumes. However, the fundamental challenge is to characterize the network of compartments with respect to volume, location, mass and energy fluxes between adjacent zones as well as the affected properties due to non-ideal fluid mixing (Kramer et al., 1999, Bezzo and Macchietto, 2004). To determine the compartmental zones within the volume of pilot-scale crystallizer equipment, primarily the fluid dynamics, mixing and heat transfer were studied by means of CFD simulations. The results of CFD simulations can be processed through a manual or automatic zoning technique in order to identify the compartments (Delafosse et al., 2010). Manual zoning is a user-dependent method and it can be troublesome and time-consuming especially when handling the complex 3D structures. On the other hand, automatic zoning technique is computationally inexpensive, reliable and practical for 3D geometries (Bezzo et al., 2004, Bezzo and Macchietto, 2004). In the automatic zoning technique, one or several property of the system can be selected as a base criterion for creating the compartments. A tolerance value for this specific property should be specified. The compartments are built-up by the agglomeration of the cells, in which the value of the specific property of the system is uniform with respect to the given tolerance (Delafosse et al., 2010). In this work, both manual and automatic zoning techniques were utilized to analyze results of CFD simulations. In both zoning approaches, three components of the average flow velocity in 3D space were chosen as base criterion, since the mixing in the system significantly is linked to the velocity fields. The compartmental volumes, location and flux connections were extracted from data analysis to define the interconnected compartment network.

2.4. Crystallization process simulation

Seeded-batch cooling crystallization of paracetamol from ethanol (solvent) is chosen as a case study in this work. Related solubility and kinetic data can be found in the literature (Fernandes, 1999, Mitchell, 2012). In the model, a linear cooling rate of 0.4 ^{o}C/min is applied to the initial solution, which has an initial concentration of 0.26 kg paracetamol / kg ethanol and an initial temperature of 40 ^{o}C, for 50 min. The overview of the model is shown in Figure 2. Even though the process is in batch mode, every individual compartments are in continuous mode with multiple inlet and outlet streams. Therefore, the batch process model was extended to continuous crystallizer model for the compartments. The crystallization process modeling based on compartmentalization approach is implemented in MATLAB/Simulink. The same set of model equations and

model parameters are defined in order to solve the conservation balance equations for crystallization process system (population, mass, etc.) for every compartment. However, spatial variations in the crystallizer volume are modelled by temperature gradients between compartments.

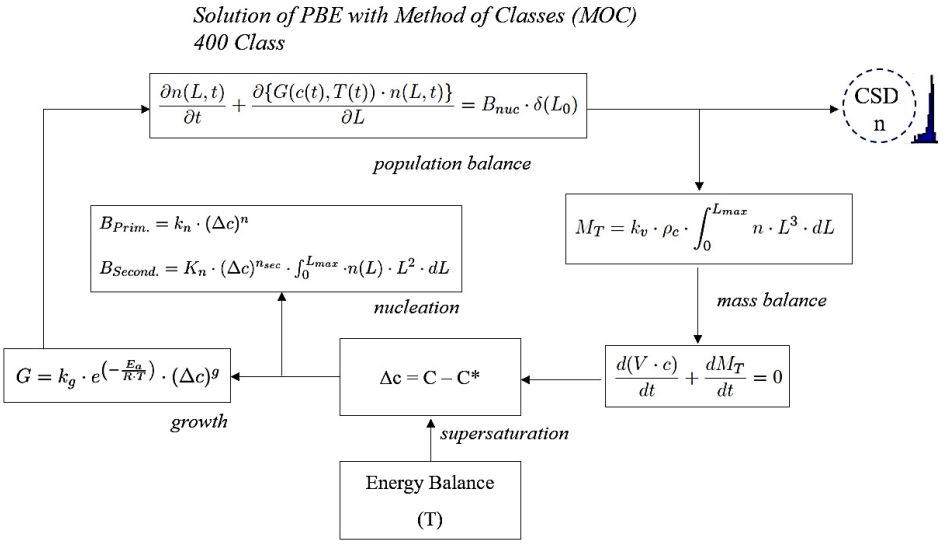


Figure 2. Batch cooling crystallization model (Adapted from Giuletti et al., 2001).

3. Results and Discussion

3.1. CFD simulation and compartmentalization

CFD simulations were performed to study primarily the fluid dynamic behavior in the large-scale crystallizer as mentioned previously in Section 2.1 and 2.2. Besides, heat transfer from a cooling jacket to the liquid in the vessel was also implemented in order to monitor the temperature gradients in the system. The overall flow pattern was extracted in order to be used as a base criterion for creation of the compartments. Total six compartments were detected by manual zoning technique, which is shown in Figure 3. On the other hand, an automatic zoning algorithm was developed in MATLAB environment to analyze the results of the CFD simulations. The network of compartments were identified by accepting the tolerance limit of 0.3 m/s in axial and 0.15 m/s in radial flow vector, at each compartment. The achieved compartmental map with respect to each individual target parameters, radial and axial flow, is illustrated in Figure 4. The results of compartmentalization techniques were used as input in the simulation of scale-up model of pharmaceutical crystallization process in the next section 3.2.

3.2. Simulation of the scale-up model of pharmaceutical crystallization process

The performance of the pharmaceutical batch cooling crystallization process was analyzed in terms of crystal growth rate, crystal size distribution and final average crystal size. The multi-compartment model was compared with a single-compartment well-mixed model to examine the influence of the mixing and corresponding temperature gradients on the process performance. CFD simulations showed that there is a difference in average temperature over time between different locations in the crystallizer vessel, so

called in the different compartments. Hence, crystals are subjected to different growth and nucleation rates in different compartments. This leads to wider final crystal size distribution and less growth (as illustrated in Figure 5) with a final average crystal diameter of 113 µm compared to the single-compartment model (final average crystal diameter is 135 µm).

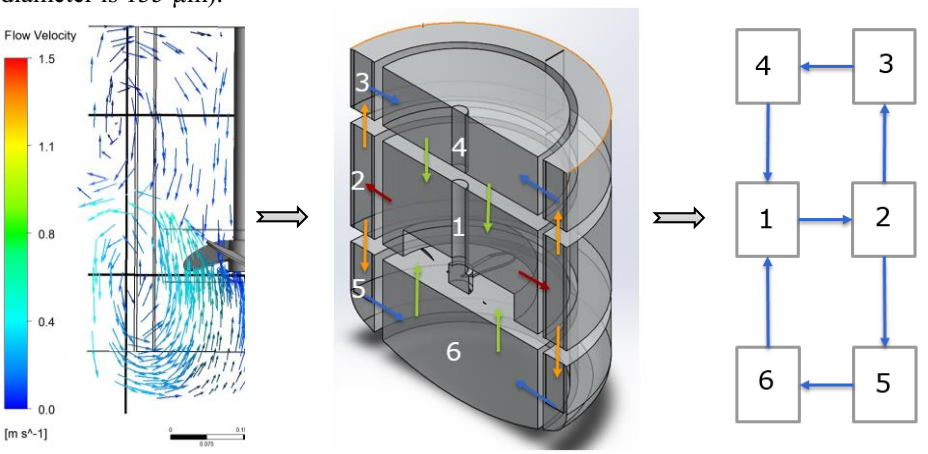


Figure 3. Characterization of compartmental zone network via the manual zoning.

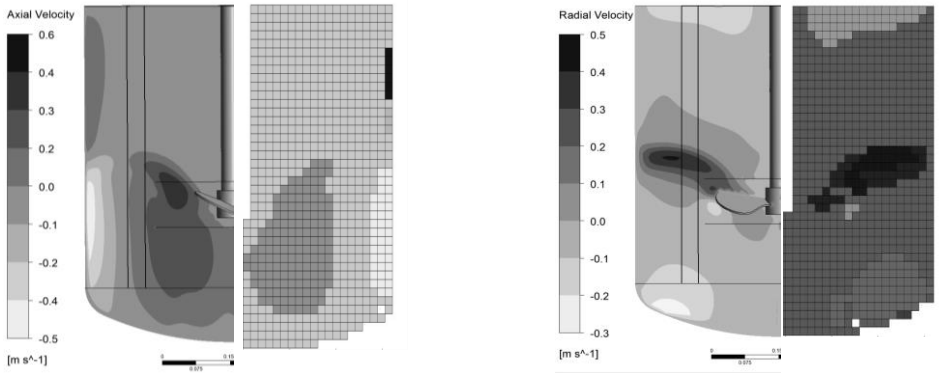


Figure 4. Characterization of the compartmental zone network via the auto zoning with 0.3 m/s tolerance of flow velocity: (left) axial direction, (right) radial direction.

4. Conclusions

A scale-up model of pharmaceutical batch cooling crystallization process was developed based on the compartmentalization approach. Fluid dynamics, mixing and heat transfer in the large scale system were studied by means of CFD simulations. Based on CFD simulations results, compartmental zones in the crystallizer vessel were determined via using both manual zoning and automatic zoning techniques. The simulated process performance in multi compartment model was compared with the single compartment well-mixed model. In this study, it was clearly observed that fluid dynamics and mixing have a significant impact on the final crystal size distribution as well as final average crystal diameter. It plays an important role in scale-up systems and, therefore, it should be taken into account in the modeling large systems. Moreover, future work will focus on

the experimental validation of the results in lab and pilot scale studies as well as compartmental model-based control and monitoring for improved process control of quality attributes.

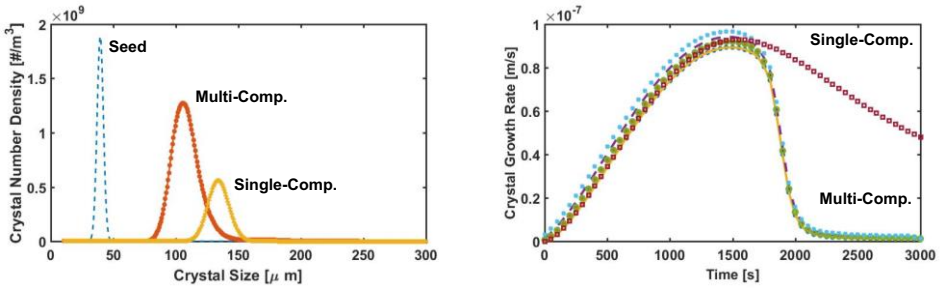


Figure 5. Final crystal size distribution (left) and crystal growth rate (right).

Acknowledgement

We would like to thank the Danish Council for Independent Research (DFF) for financing the project with grant ID: DFF-6111600077B.

References

F. Bezzo, S. Macchietto and C.C. Pantelides, 2004, A general methodology for hybrid multizonal/CFD models Part I. Theoretical framework, *Computers & Chemical Engineering*, 28, 501-511.

F. Bezzo and S. Macchietto, 2004, A general methodology for hybrid multizonal/CFD models Part II. Automatic zoning, *Computers & Chemical Engineering*, 28, 513-525.

A. Delafosse, F. Delvigne, M.L. Collignon, M. Crine, P. Thonart and D. Toye, 2010, Development of a compartment model based on CFD simulations for description of mixing in bioreactors, *Biotechnol. Agron. Soc. Environ*., 14, S2, 517-522.

C. Fernandes, 1999, Effect of nature of the solvents on the crystallization of paracetamol, accepted for *14th International Symposium on Industrial Crystallization (ISIC 14).*

M. Giulietti, M.M. Seckler, S. Derenzo, M.I. Re and E. Cekinski, 2001, Industrial crystallization and precipitation from solutions: State of the technique, *Brazilian Journal of Chemical Engineering*, 18, 4, 423-440.

E. Kougoulos, A.G. Jones and M. Wood-Kaczmar, 2006, A hybrid CFD compartmentalization modeling framework for the scaleup of batch cooling crystallization process, *Chem. Eng. Comm*., 193, 1008-1023.

H.J.M. Kramer, S.K. Bermingham and G.M. van Rosmalen, 1999, Design of industrial crystallizers for a given product quality, *Journal of Crystal Growth,* 198/199, 729-737.

H.J.M. Kramer, J.W. Dijkstra, P.J.t. Verheijen and G.M. van Rosmalen, 2000, Modeling of industrial crystallizers for control and design purposes, *Powder Technology,* 108, 185-191.

A. Myerson, 2002, Handbook of Industrial Crystallization, Butterworth-Heinemann, Oxford, UK

N.A. Mitchell, 2012, Numerical Modeling of Cooling Crystallization: Process Kinetics to Optimization, PhD Thesis, University of Limerick.

G.J. Wells and W.H. Ray, 2005, Methodology for modeling detailed imperfect mixing in complex reactors. *AIChE Journal*, 51:1508-1520.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50026-4

Optimization of extractive distillation – integrated solvent selection and energy integration

Thomas Waltermann[a*], Tamara Grueters[a], Mirko Skiborowski[a]

[a]*TU Dortmund University, Department of Biochemical and Chemical Engineering, Laboratory of Fluid Separations, Emil-Figge-Strasse 70, 44227 Dortmund, Germany Thomas.Waltermann@tu-dortmund.de*

Abstract

The separation of azeotropic mixtures is of particular importance for bio-based processes in the chemical industry. Extractive and heteroazeotropic distillation are oftentimes the favored solution for medium to large scale processes due to their proven robustness and the economics of scale. The feasibility as well as economic efficiency of these processes depends strongly on the selection of a suitable mass separating agent (MSA). Furthermore, energy integration can increase the energy and economic efficiency of these thermal processes. Since, MSA selection, process design and energy integration are usually performed as consecutive steps in process design, potential synergies are easily missed, resulting in sub-optimal choices. In order to determine an optimal process design, including solvent selection and energy integration, an efficient optimization-based approach for the design of extractive distillation processes is proposed. The application of the method is illustrated for the separation of the azeotropic mixture of acetone and methanol. The results highlight that the optimal MSA choice under consideration of energy integration differs from the selection without energy integration.

Keywords: Extractive distillation, optimization, solvent selection, energy integration

1. Introduction

The separation of azeotropic mixtures is a complex task that is frequently encountered in chemical process design and of special importance for bio-based processes. Despite the availability of potentially less energy intensive separation techniques such as membrane processes, the separation by extractive or heteroazeotropic distillation is often the preferred separation technique due to the proven robustness and favourable economics, particularly for large scale processes (Kiss and Suszwalak, 2012).

The design of extractive and heteroazeotropic distillation processes is particularly challenging due the numerous design degrees of freedom (DDoF) and the integration of the separation and the solvent recovery column. In order to enable a systematic process design often a hierarchical design procedure, such as the one illustrated in Figure 1 is followed (e.g. by Douglas, 1985).

At first possible process variants are generated based on heuristic rules or expert knowledge. This especially includes the selection of a suitable mass separating agent (MSA), which has a significant effect on the feasibility as well as the cost and energy efficiency of the separation process (Julka et al., 2009). Nowadays, computer-aided molecular design methods are readily available and can be used to identify suitable MSA based on a purely computational screening (Gani et al., 2006).

Figure 1 Hierarchical design procedure for process design

Next, the feasibility of the generated process variants is evaluated by means of simplified methods. For extractive distillation the influence of the MSA on the relative volatility between the azeotropic components and the residue curve map of the mixture with the MSA can be analysed (Julka et al., 2009). Such classical methods for the pre-selection of MSA focus only on the general feasibility of the separation. However, a more meaningful analysis already includes the effort for the recovery of the MSA. Such pre-selection in terms of the process performance can already provide an estimate of the economic performance, but requires appropriate shortcut methods, like the rectification boundary method (RBM), which are able to account for the non-ideal thermodynamic behaviour of the mixtures (Brueggemann and Marquardt, 2004; Kossack et al., 2008).

Only a small selection of promising MSA candidates is further evaluated based on detailed process models. In case of extractive distillation this is normally done based on equilibrium-stage stage models in combination with rigorous thermodynamics, while the objective function represents the economic performance. Since each MSA entails its specific optimal process design a simulation-based design approach is very tedious due to the numerous DDoF and the integration of both columns. Optimization-based methods allow for a more efficient economic optimization of the process variants and evaluation of the best MSA (Kossack et al., 2008).

Finally, additional options for energy integration are considered in a consecutive step to further optimize the selected process variant. Energy integration can be achieved by conventional heat integration between the columns (HI), vapour recompression for each column (VRC), or thermal coupling in separate columns (TC) respectively in a divided wall column (DWC). However, the investigation of energy integration options is usually restricted to the previously selected MSA and process design, although the saving potential of the different options depends strongly on the selected MSA. Therefore, a consideration of the different means for energy integration for all promising MSA candidates is desirable, in order to exploit possible synergies and identify the most economic solution.

Since this requires the evaluation and optimization of a large number of process options with a significant number of DDoF, a computationally efficient and reliable design approach is required. Conventional simulation-based process design approaches applying manual procedures (Brito et al., 2016) or the combination of a process simulator and a metaheuristic to optimize the design (Tututi-Avila et al., 2014) are only of limited use. Models and appropriate initialization routines for strongly integrated processes, such as DWC's are mostly missing (Waltermann and Skiborowski, 2017a). Therefore, even if a metaheuristic could efficiently handle all DDoF of the different process configurations, a feasible initial solution would first need to be developed manually for each process configuration and every MSA. The exploitation of equation-based modeling and mixed-integer nonlinear programming (MINLP) enables the creation of tailored optimization models and automated initialization procedures, such that the manual demand can be reduced to a minimum. The current contribution illustrates the potential of such an approach for the design of extractive distillation processes.

2. Optimization-based design approach

The evaluation of the different MSA candidates for extractive distillation based on the process performance of optimized process configurations has previously been addressed by Kossack et al. (2008). The proposed approach considered a shortcut-based evaluation of the minimum energy requirements and a subsequent economic optimization of the most-promising candidates. In order to extend this approach towards different energy integration concepts we combine the approach with our previously proposed optimization-based design approach for the design of azeotropic distillation processes (Waltermann and Skiborowski, 2016; Waltermann and Skiborowski, 2017b). This approach builds on the optimization of superstructure models based on rigorous equilibrium-stage models including non-ideal thermodynamics, embedded in external functions. Different g^E models such as Wilson, NRTL and UNIQUAC, as well as some cubic equations of state are readily available in the implementation. For each specific case only the necessary property parameters need to be provided. The resulting MINLP problems are solved by means of a sequentially relaxed series of nonlinear programming (NLP) problems in combination with an embedded solution of thermodynamic equilibrium computations in from of implicit functions (Kraemer et al., 2009; Skiborowski et al. 2015a).

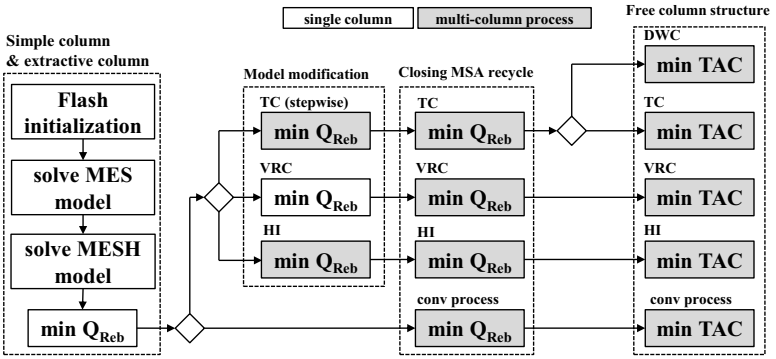


Figure 2 Automatic initialization & optimization routine

The special solution approach considers a set of initialization and optimization models (cf. Figure 2), which enables an automated evaluation of the conventional as well as energy integrated process variants. Therefore, only the separation problem, the constraints and appropriate bounds need to be provided, while no initial solution in terms of a feasible simulation result is required. This enables a numerically efficient evaluation with computational times of less than 20 minutes for each process variant (executed on a PC with a 3.2 GHz quad-core CPU).

3. Case Study – Separation of the azeotropic mixture acetone / methanol

3.1. Problem statement

In order to illustrate the influence of the MSA selection and energy integration on the optimal process design and to demonstrate the developed method, the separation of the azeotropic mixture of acetone and methanol is investigated. A saturated liquid feed with a molar flow rate of 150 mol/s and a composition of 75 mol-% acetone and 25 mol-% methanol is to be separated into product streams with at least 99.5 mol-% purity at a

recovery of at least 99 mol-%. The considered MSA are based on the pre-selection considered by Kossack et al. (2008) and include chlorobenzene, DMSO, ethanol, p-xylene and water. The evaluation of the best process configuration is based on the total annualized costs (TAC). To account for different utility choices low (3 bar), medium (15 bar) and high pressure steam (25 bar) at costs of 12, 14 and 16 €/t, cooling water at 0.05 €/t and electricity for a price of 0.06 €/kWh are considered. An operating time of 8000 h/y and a depreciation period of 10 years with an interest rate of 6 % are assumed to determine the TAC. Column sizing and costing is performed assuming packed columns with Sulzer BX^{TM} structured packings, considering a HETP value of 0.24 m. The non-ideality of the liquid phase is modeled based on the UNIQUAC model and the extended Antoine equation, while an ideal gas phase is assumed. The specific enthalpies are determined based on DIPPR correlations for the calculation of specific heat capacities and heat of evaporation.

3.2. MSA selection based on the conventional process

Taking into account the described economic parameters, the TAC values of the optimized conventional process configurations for each investigated MSA are shown in Figure 3.

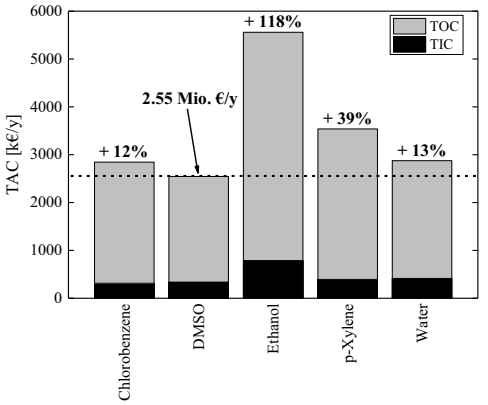


Figure 3 Annualized costs of the optimal designs for the MSA candidates based on the conventional extractive distillation process

Based on the conventional extractive distillation process DMSO is identified as the MSA option with the lowest TAC.

3.3. Influence of energy integration on the optimal process

While the use of ethanol and p-xylene as MSA result in significantly larger costs, the application of chlorobenzene and water results in only marginally more expensive processes. Consequently the latter may be able to outperform the DMSO-based process, if energy integration is considered. The optimal designs of the energy-integrated processes are illustrated in Figure 4. As a reference to the previous results, the costs of the conventional DMSO-based process design are indicated by the dotted line. It becomes apparent that none of the means for energy integration are able to improve the DMSO-based process. While the option to apply compression of the top vapour streams is just discarded in the optimization of the VRC case, the increasing use of high pressure steam in the single reboiler of the TC process and the DWC results in significantly increased operating costs. For DMSO the HI variant is not considered due

to the necessity of either using steam with a pressure of about 100 bar or a fuel heated reboiler.

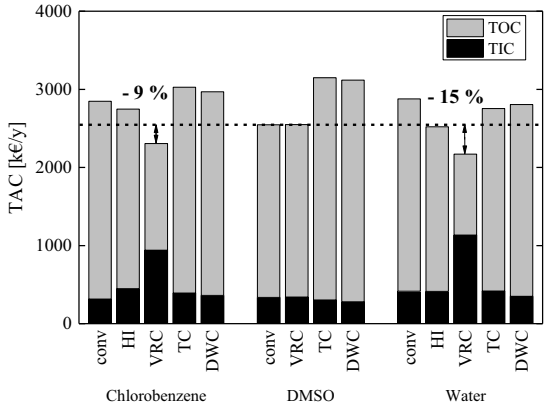


Figure 4 Annualized costs of the optimal, energy-integrated process designs for the MSA candidates chlorobenzene, DMSO and water

For chlorobenzene the TC and DWC options also result in increased operating costs due to similar reasons as for DMSO. However, the HI and the VRC variant allow cost savings compared to the conventional chlorobenzene process. The VRC process is even superior to the conventional DMSO process, allowing for savings of about 9% in TAC.

All considered means of energy integration are able to reduce the TAC of the water-based process. The main reason for this difference to the other two MSA is the lower boiling point of water, which allows for the use of low pressure steam in all cases, except for the HI process. However, the HI water process can still save about 1% of the TAC in relation to the DMSO-based process. Overall, the water-based VRC process results in the most economic process variant, saving about 15% of the TAC of the DMSO-based process. The design of this process is illustrated in Figure 5.

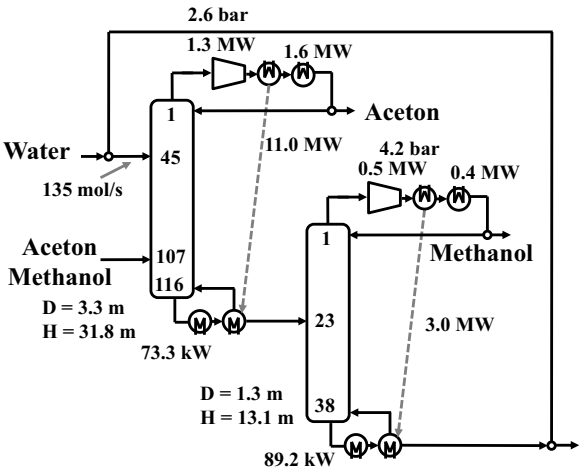


Figure 5 Optimal design of the VRC water process

VRC is used for both columns and thus, reduces the demand for external thermal energy from 13.9 MW to only 0.2 MW. In order to drive the compressors about 1.8 MW of

electrical energy is required. While the capital investment for the additional compressors results in a large increase in the overall capital costs, the operational costs are reduced by more than 50%. Thus, considering an operation beyond the depreciation time, the water-based VRC process will operate at less than half of the cost of the DMSO-based process.

4. Conclusions

In this contribution the influence of an integration of solvent selection and energy integration on the optimal process is investigated. Based on the investigated case study it is demonstrated that the optimal MSA choice under consideration of energy integration can differ from the selection without energy integration. The illustrated advantages require a consideration of the economic performance, since an energy-based evaluation cannot account for the alternating utility requirements and an exergy-based evaluation cannot correctly account for the capital investment of the compression stages. The proposed optimization-based approach allows for a computationally efficient evaluation of the increased number of process variants with the various DDoF, taking into account a selection of promising MSA, means for energy integration as well as the column sizing and operating conditions. The resulting complex optimization problems are solved effectively without the need of an initial feasible solution based on the developed solution approach. In order to further automate and improve the design process future work will investigate the combination of the approach with a hybrid evolutionary-deterministic optimization approach (Skiborowski et al., 2015b).

References

K. D. Brito, G. M. Cordeiro, M. F. Figueirêdo, L. G. S.Vasconcelos, R. P. Brito, 2016. Comput. & Chem. Eng., 93, 185–196.

S. Brueggemann, W. Marquardt, 2004. AIChE journal, 50(6), 1129–1149.

J. M. Douglas, 1985. AIChE Journal, 31(3), 353–362.

R. Gani, C. Jimenez-Gonzalez, A. ten Kate, P. A. Crafts, M. Jones, L. Powell, J.H. Atherton, J. L. Cordiner, 2006. Chem. Eng., 113(3), 30–44.

V. Julka, M. Chiplunkar, L. O'Young, 2009. Chem. Eng. Prog., 105(3), 47–53.

A. A. Kiss, D. J.-P. Suszwalak, 2012. Sep. Purif. Technol. 86, 70–78.

K. Kraemer, S. Kossack, W. Marquardt, 2009. Ind. & Eng. Chem. Res., 48 (14), 6749–6764.

S. Kossack, K. Kraemer, R. Gani, W. Marquardt, 2008. Chem. Eng. Res. Des. 86 (7), 781–792.

M. Skiborowski, A. Harwardt, W. Marquardt, 2015a. Comput. & Chem. Eng., 72, 34–51.

M. Skiborowski, M. Rautenberg, W. Marquardt, 2015b. Ind. & Eng. Chem. Res. 54 (41) 10054–10072.

S. Tututi-Avila, A. Jimenez-Gutierrez, J. Hahn, 2014. Chem. Eng. Proc.: Proc. Int. 82, 88–100.

T. Waltermann, M. Skiborowski, 2016. Comput. Aided Chem. Eng. 38, 571–576.

T. Waltermann, M. Skiborowski, 2017a. Chem. Ing. Tech., 89(5), 562–581.

T. Waltermann, D. Muenchrath, M. Skiborowski, 2017b. Comput. Aided Chem. Eng. 40, 1045–1050.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50027-6

An Industrial Park Layout Design Method Considering Pipeline Length Based on FLUTE Algorithm

Ruiqi Wang[a], Yan Wu[b], Yufei Wang[a*], Xiao Feng[b], Mengxi Liu[a]

[a]*College of Chemical Engineering, China University of Petroleum - Beijing, No.18, Fuxue Road, Beijing, 102249, China*
[b]*School of Chemical Engineering and Technology, Xi'an Jiaotong University, No.28, Xianning West Road, Xi'an, 710049, China*
wangyufei@cup.edu.cn

Abstract

The layout of an industrial park can significantly impact its economic benefit and safety. In terms of economy, a good layout means shorter pipeline length, leading to lower pipeline cost and material transport cost. In practice, the layout of an industrial park is manually determined based on experience according to the relationship of upstream and downstream. In academia, few systematic programming method has been proposed, and most of them only consider one-to-one connection pattern, pipeline networks with multi-branch are ignored. For systematic programming method, it has always been difficult to calculate the minimum length of pipeline network with multi-branch in an industrial park. FLUTE (Fast Lookup Table Based Wirelength Estimation Technique) algorithm is an accurate and fast method to solve rectilinear Steiner minimum tree problem based on a lookup table. In this paper, a novel method based on FLUTE algorithm is came up to calculate the minimum length of steam pipeline network in an industrial park and determine the layout. The total pipeline cost, which includes steam pipeline cost and material pipeline cost, is set as an objective function. Genetic algorithm is employed to optimize the layout of several plants in an industrial park. In the case study, it can be seen that, the calculation time is significantly reduced, and a better layout is obtained.

Keywords: Industrial park layout; Pipeline network; Rectilinear Steiner minimum tree; FLUTE; Genetic algorithm

1. Introduction

Layout design is a key step which must be passed in the process of chemical engineering design and construction. A good layout can reduce the capital cost for the pipeline and operation cost for material transportation, and enhance safety. Economy and safety are the main factors considered in the layout problem. Researches in chemical engineering layout can be divided into two aspects: one is to study the facility layout within a plant, and the other is to study the plant layout within an industrial park.

Many researches about the facility layout within a plant are reported. Penteado and Ciric (1996) proposed a MINLP approach to determine the coordinates of facilities in a plant, considering land cost, piping cost, protection devices cost and risk cost. In their work, protection devices can prevent accidents or minimize the damage to the other nearby facilities, resulting in a protection devices cost at the same time. Patsiatzis et al. (2004)

quantified the risk cost using FEI (Dow Fire & Explosion Index) system, while piping cost and protection cost were also involved to determine the layout of a plant. Latifi et al. (2017) worked out an optimal layout of a plant considering uncertainty and domino effects, and the allowance of multi-layer layout in this work makes this model more practical.

Very few works have been done for designing the plant layout in an industrial park. A noteworthy difference between the facility layout and plant layout is pipeline network with branches. Pipeline networks, such as steam pipeline work, existed widely within an industrial park. Studies about facility layout mentioned above only calculated the lengths and costs of pipelines with point-to-point structure. It has always been difficult to calculate the minimum length of pipeline network with multi-branch in an industrial park. Wu and Wang (2017) firstly came up with an algorithm based on Kruskal algorithm to work out the minimum length and arrangement of pipeline network within an industrial park. Then Wu et al. (2016) combined this algorithm with risk evaluation to determine the layout of plants. But their algorithm is expensive in calculation time.

A rectilinear Steiner minimal tree (RSMT) is a tree with minimum total edge length in Manhattan distance to connect a given set of nodes possibly through some extra nodes (Chu and Wong, 2008). Many algorithms have been come up to solve RSMT problem. GeoSteiner is an exact RSMT software developed by Warme et al. (2015) , but is relatively expensive in calculation time. Chen et al. (2002) proposed an algorithm called Refined Single Trunk Tree (RST-T), which is very effective for low-degree networks but not for high-degree networks. FLUTE (Fast Lookup Table Based Wirelength Estimation Technique) (Chu and Wong, 2005) is a fast and relatively accurate algorithm for RSMT. It achieves a good balance between calculation time and accuracy (Chu, 2004b).

In this work, an industrial park layout design method is proposed, in which the problem of minimum length of pipeline network is converted to RSMT problem and calculated based on FLUTE algorithm. In the case study, the proposed method was compared with Wu and Wang's work (2017). From the result, it can be seen that, when solving the same case, the calculation time of the proposed method is 38.9 s and Wu's method is 69215.0 s. Calculation time is reduced sharply, and a better layout is obtained.

2. Methodology

2.1 Problem statement

The objective function and constraints in this work are based on Wu and Wang's work (2017). The free space in the industrial park is divided into several grids with the same size. Plants are treated as points. Each plant can only occupy one grid, which is put in the center of the grid. The distance between two plants is the distance between centers of the two grids and the distance is absolutely safe for plants. The pipeline must be arranged horizontally or vertically. Moreover, some plants should be put in some specific locations, according to the geographical and meteorological condition of the area where the industrial park sits. For example, the railway transportation plant should be located as near as possible to the railway line, and the air separation plant should be located in the upwind direction. Based on the assumption mentioned above, the aim of this work is to determine the location of plants in an industrial park to minimize the capital cost of the pipeline, including material pipeline and steam pipeline.

2.2 Objective function

The objective function is set as the sum of material pipeline cost and steam pipeline cost, as is shown in Eq. (1).

$$\min\ C = \sum_{i=1}^{n} a_i K_i + \sum_{j=1}^{3} b_j L_j \tag{1}$$

Where C is the total pipeline cost, n is the quantity of material pipeline, i represent different kinds of material flow, a_i is the unit prize of i-th material pipeline, K_i is the length of i-th pipeline; j represent 3 different kinds of steam, b_j is the unit prize of j-th steam pipeline, L_j is the length of j-th steam pipeline network. a_j and b_j come from the market prize.

As the pipeline must be arranged horizontally or vertically, the K_i can be calculated by Eq. (2)

$$K_i = | x_{in} - x_{out} | + | y_{in} - y_{out} | \tag{2}$$

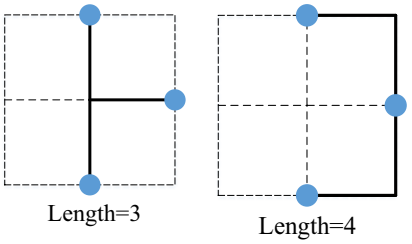
Where (x_{in}, y_{in}) and (x_{out}, y_{out}) is the coordinates of input and output plants respectively.

The calculation of L_j is very complex that will be expressed in the next section.

2.3 Minimum length of steam pipeline network

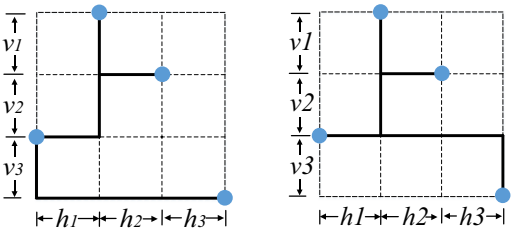
As mentioned in Introduction, the calculation of the minimum length of pipeline network has always been a thorny problem. In this work, this problem has been converted into RSMT problem. RSMT problem is to find a way to connect several given points by only horizontal and vertical lines. As is shown in figure 1, figure 1 (a) is the RSMT of the 3 given points, and figure 1 (b) is not. The Pipeline network connecting the plants of the steam generator or steam user, is the RSMT of those plants.

Length vector is used to represent a certain connection. Figure 2 (a) is a connection of 4 given plants. The pipeline length of this connection is L_1=2*h_1+2*h_2+h_3+v_1+v_2+v_3, so the length vector of this connection is p_1= (2,2,1,1,1,1); figure 3 (b) is the shortest connection for the 4 plants, and the length is L_2=h_1+2*h_2+h_3+v_1+v_2+v_3. The relevant length vector is p_2= (1,2,1,1,1,1), which is called the optimal length vector. We can know which connection is shorter by comparing the length vector directly, instead of calculating the total length. For instance, length vector of (1,2,1,1,1,1) is shorter than (1,2,2,1,1,1) certainly.



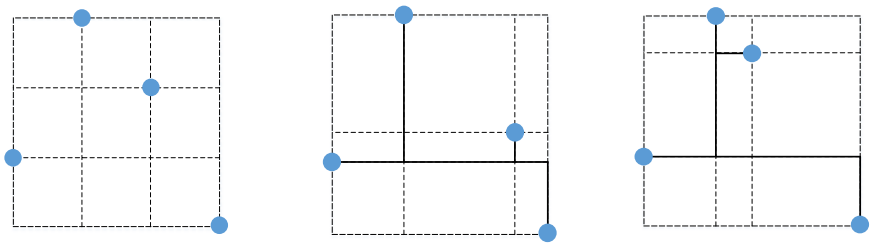
(a) (b)

Figure 1 Examples of RSMT and non-RSTM



(a) (b)

Figure 2 Two connection patterns of 4 given plants



(a) (b) (c)

Figure 3 Different relative locations of 4 plants

If the relative location of several plants is given and the edge lengths of grids are not given, many connections may be the shortest connection. Figure 3 (a) shows a kind of relative location of 4 plants (assuming that the lengths of edges are not determined). Both p_3= (1,1,1,1,2,1) and p_4= (1,2,1,1,1,1) may be the optimal length vector. Which one is optimal depends on the lengths of grid edges. In figure 3 (b), p_3 is shorter, and in figure 3 (c) p_4 is shorter. So p_3 and p_4 are called potential optimal length vector. For a certain relative location of plants, only few length vectors are potentially optimal comparing with all feasible connections. The authors of FLUTE previously calculated all the potential optimal length vectors for each condition of the relative location of points and saved the results. When the coordinates of plants are given, the relevant potential optimal length vectors for the relative location of those plants can be found. Then we calculate the lengths of each potential optimal length vector, and choose the shortest one. The minimum length of steam pipeline network and the arrangement can be obtained. This algorithm is very fast.

3. Case study

3.1 Basic data

The case from Wu and Wang's work (2017) is used, so that the proposed method can be compared with theirs. The case comes from a real petrochemical industrial park, which contains 20 plants and they should be arranged as 4×5 form. The distance between adjacent plants is set as 400 m. The plants and steam requirement are shown in Table 1, in which 1 means the plant needs steam and 0 means the plant does not need steam. This case contains 76 material flows. The type and diameter of a pipeline can be determined according to the property, flow rate and flow velocity of the transported material. In this work, the steam pipeline for the same degree has the same diameter.

According to the geographical and meteorological condition of the area where the industrial park sits, 5 plants should be placed in the specific locations. They are shown in Figure 4.

Both Wu and Wang's method and the proposed method are implemented respectively to determine the layout of this case. The program of Wu and Wang's method is provided by them. MATLAB is used to run the two programs. Genetic algorithm, which is also adopted in Wu and Wang's work, is employed to solve the model for a fair competition. Genetic algorithm is generally regarded as one of the most effective algorithms. The generation is set as 1000, and the individual is set as 50. FLUTE is supplied by the authors in their official website (Chu, 2004).

Table 1 The plants and steam requirement of industrial park

Number	Name	High pressure steam	Middle pressure steam	Low pressure steam
1	Power station	1	1	1
2	Crude oil fractionation plant	0	1	1
3	Gas separation plant	1	1	1
4	Hydrogenation plant	0	1	1
5	Residue and wax hydrodesuifurization plant	1	1	1
6	Air compression and separation plant	0	1	1
7	Fluid catalytic cracking plant	1	1	1
8	Light hydrocarbon recovery plant	1	1	1
9	Liquefied petroleum gas desulfuration and demercaptan plant	1	1	1
10	Sulfur recovery plant	0	1	1
11	Aromatic combine plant	1	1	1
12	Hydrogen production	1	1	0
13	Continuous reforming plant	1	1	0
14	Naphtha hydrotreating	0	1	0
15	Polypropylene and polyester	1	1	1
16	Delayed coking plant	0	1	0
17	Sewage treatment area	0	0	1
18	Tank farm	0	1	0
19	Railway transport department	0	1	0
20	Central control room	0	0	1

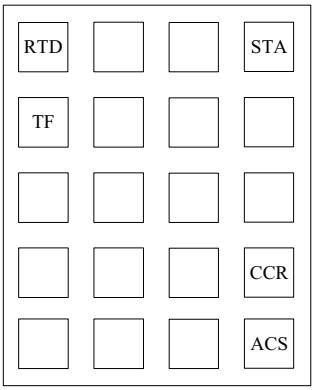
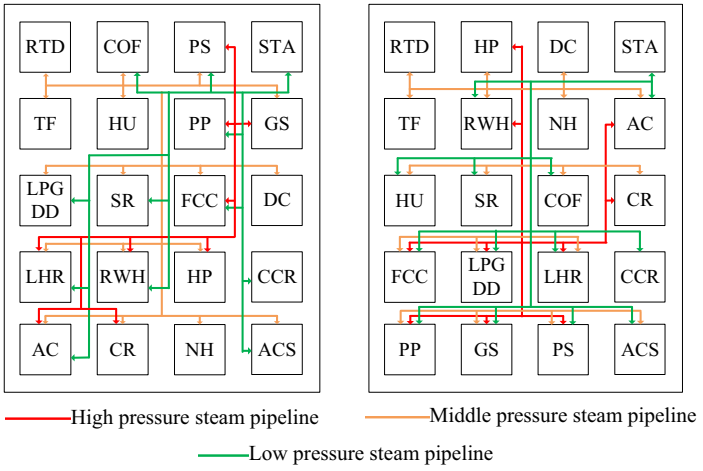


Figure 4 The plants with fixed location



(a) The layout form Wu's method (b) The layout from the proposed method

Figure 5 The layouts from two methods

3.2 Results

The numerical results are presented in Table 2. The final layouts from Wu and Wang's method and the proposed method are shown in figure 5 (a) and figure 5 (b) respectively.

Table 2 Numerical results of the two methods

	Wu and Wang's method	The proposed method
Calculation time	69,215.0 s (19.2 h)	38.9 s
Steam pipeline cost (10^3 ¥)	2,475	2,660
Material pipeline cost (10^3 ¥)	10,581	9,818
Total pipeline cost (10^3 ¥)	13,056	12,478

From the result, we can see that the calculation time has been reduced sharply. The proposed method is 1779 times faster than Wu's method. And a better layout with lower total pipeline cost is obtained.

4. Conclusion

The calculation of the minimum length of pipeline network has always been a thorny problem in the process of industrial park layout design. In this work, FLUTE algorithm is firstly applied in the calculation of the minimum length of a pipeline network. And a mathematical model is established to obtain the layout of an industrial park with the lowest pipeline cost. Both the previous method and the proposed method are used to solve the same case. The result illustrates that, with the proposed method, the calculation time is reduced sharply while a better layout is obtained.

References:

H. Chen, C. Qiao, F. Zhou, C. Cheng, 2002. Refined Single Trunk Tree: A Rectilinear Steiner Tree Generator For Interconnect Prediction, In Proc. ACM Intl. Workshop on System Level Interconnect Prediction, pp. 85-89.

C. Chu, 2004, FLUTE, http://home.eng.iastate.edu/~cnchu/flute.html, accessed 29 November 2017

C. Chu, 2004b. FLUTE: Fast Lookup Table Based Wirelength Estimation Technique, In Proc. International Conference on Computer Aided Design, pp. 696-701.

C. Chu, Y. Wong, 2005. Fast and Accurate Rectilinear Steiner Minimal Tree Algorithm for VLSI Design., In Proc. International Symposium on Physical Design, pp. 28-35.

C. Chu, Y. Wong, 2008, FLUTE: Fast Lookup Table Based Rectilinear Steiner Minimal Tree Algorithm for VLSI Design, IEEE Transactions on Computer-aided Design of Intergrated Circuits and Systems, 1, 27, 70-83.

S.E. Latifi, E. Mohammadi, N. Khakzad, 2017, Process plant layout optimization with uncertainty and considering risk, Computers & Chemical Engineering, 106, 224-242.

D.I. Patsiatzis, G. Knight, L.G. Papageorgiou, 2004, An MILP approach to safe process plant layout, Chemical Engineering Research and Design, 82, A5, 579-586.

F.D. Penteado, A.R. Ciric, 1996, An MINLP Approach for Safe Process Plant Layout, Industrial & Engineering Chemistry Research, 4, 35, 1354-1361.

D. Warme, P. Winter, M. Zachariasen, 2015, GeoSteiner, http://geosteiner.net/, accessed 29 November 2017.

Y. Wu, Y. Wang, 2017, A Chemical Industry Area-wide Layout Design Methodology for Piping Implementation, Chemical Engineering Research & Design, 118, 81-93.

Y. Wu, Y. Wang, X. Feng, S. Feng, 2016, A Genetic Algorithm Based Plant Layout Design Methodology Considering Piping and Safety, Chemical Engineering Transactions, 52, 25-30.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50028-8

Influence of epistemic uncertainty in the selection of flowsheet structures

Edelmira D. Gálvez[a, *], Luis A. Cisternas[b], Freddy A. Lucay[b], and Renato Acosta-Flores[b]

[a] *Department of Mines and Metallurgical Engineering, Universidad Católica del Norte, Av. Angamos 0610, Antofagasta 124000, Chile.*
[b] *Department of Mineral Process and Chemical Engineering, Universidad de Antofagasta, Av. Angamos 601, Antofagasta 124000, Chile.*
egalvez@ucn.cl

Abstract

The optimal design of flowsheet is a complex task, making it difficult to obtain a global optimum solution. Furthermore, in several processes, there are epistemic uncertainties in the evaluation of the performance of unit operations. This epistemic uncertainty is generated due to a lack of knowledge of the behavior of unit operations. This lack of knowledge generates the need to use semi-empirical models, or define some operational conditions a priori based on some approximation. Usually, these empirical models use parameters that are a function of the flowsheet design. Therefore, it is not possible to adjust these models based on experimental data as the experimental conditions are not known at the design stage. This work analyzes the effect of the epistemic uncertainties in the selection of flowsheet structures. Flotation circuit without grinding, flotation circuit with grinding units, and fractional crystallization were considered as case studies. As a result, this paper demonstrated that few structures exist that are optimal for a wide range of values in the performance of the process units. To demonstrate that there are few optimal structures when there is high uncertainty in the process units, two methods of proof were used: proof by construction, and proof by exhaustion.

Keywords: flow sheet design; epistemic uncertainty; grinding; flotation; crystallization

1. Introduction

The optimal design of flowsheet based on optimization is a complex task as the mathematical model is nonconvex and includes binary variables (usually representing the structure of the flowsheet), making it difficult to obtain a global optimum solution. Examples of these situations are mineral concentration plants (Cisternas et al., 2016), heat exchanger networks and utility systems (Martelli et al., 2016), reverse osmosis desalination processes (Sassi and Mujtaba, 2011), and reactive distillation (Ropotar et al., 2009), among many other cases. The determination of the optimal design comprises three steps: (a) postulation of a superstructure that represents a set of process alternatives, (b) its formulation as a mathematical programming model, and (c) determination of the optimal configuration and operation conditions by solving the model with an optimization algorithm (Chen and Grossmann, 2017).

Furthermore, in several processes, there are epistemic uncertainties. The epistemic uncertainty is associated with the lack of knowledge of the behavior of a unit process.

This lack of knowledge is generated because the exact value of some parameters is not known, the model is not able to adequately represent the unit process, or for both reasons. This lack of knowledge generates the need to use semi-empirical models, or some operational conditions are defined a priori based on some approximation. Furthermore, these empirical models use parameters that are a function of the flowsheet design. Therefore, it is very probable that these semi-empirical models are not capable of adequately representing the unit process generating uncertainty.

This work analyzed the effect of the epistemic uncertainties in the selection of flowsheet structures, but not on operational conditions. A hypothesis was postulated that few structures exist that are optimal for a wide range of values in the performance of the process units.

2. Procedure

As indicated above, the hypothesis postulated was that few structures exist that are optimal for a wide range of values in the performance of the process units. The performance of process units are recoveries, conversion, equilibrium concentration, among other indicators. To demonstrate that there are few optimal structures when there is uncertainty in the process units, an a priori approach or first principles was used. Two methods of proof were used: proof by construction, and proof by exhaustion. Proof by construction is the construction of a concrete example with a property to demonstrate that something having that property exists. In this work, the design of the flotation circuit without grinding, flotation circuit with grinding units, and fractional crystallization were considered in the proof by construction. In proof by exhaustion, the conclusion was established by dividing it into a finite number of cases and proving each one separately. Here, the performance of operation units was represented by uniform distribution functions from which a large dataset formed by random values was obtained. The optimal structure design was obtained for all the values in the dataset. The idea was that given a reasonably large dataset, the set of optimal structure design is equal to all optimal designs in the region represented by the uniform distribution functions. In all cases, one structure design was different from another if it contained a different stream and/or equipment. Table 1 presents the proof approach.

3. Examples for proof by construction

3.1. Flotation circuit without grinding

The example of flotation circuit without grinding considered the concentration of copper ore. The feed to the circuit corresponded to 6 t/h of chalcopyrite (33% of copper), 12 t/h of chalcopyrite slow (16% of copper), and 300 t/h of gangue. The superstructure considers five flotation stages. If all interconnection was allowed, there were over 3 million circuit structure alternatives. However, if origin-destination matrices were used to eliminate nonsense and redundant alternatives, the number of feasible flotation circuits was 6912. The procedure utilized for the postulation of a superstructure and the formulation of the mathematical programming model was the one utilized by Calisaya et al. (2016), which corresponded to a MINLP. The variables with uncertainties corresponded to the stage recoveries of the chalcopyrite, chalcopyrite slow, and gangue. The stage recoveries were difficult to model as there is not a model that can be used under all flotation circuit

structures included in the superstructure. Here, the stage recoveries were represented by values obtained from the uniform distribution. Under these conditions, the design problem is a MILP.

3.2. Flotation circuit with grinding

The postulation of the superstructure and the formulation of the mathematical programming model for the example of flotation circuit with grinding was generated using the procedure described by Cisternas et al. (2006). Nine species were considered to represent the ore including three components (Chalcopyrite, Chalcopyrite slow, and silica) and three particle sizes. Three flotation stages and one grinding stage were included in the superstructure that represented twelve process structures. These twelve process structures included six with and six without grinding stages. Note that, if all interconnection was allowed there were 4096 alternatives, but by eliminating nonsense and redundant alternatives, the twelve process structures were obtained. In addition to the stage recoveries of all species, the "conversion" from one particle class to another particle class was considered as uncertain variables. Forty-eight variables were represented using a uniform distribution function; twenty-seven corresponded to stage recovery and twenty-one corresponded to conversion in the grinding stage. In addition to stage recovery, grinding is also difficult to model as it is a function of feed which is not known.

It is important to indicate that particle size is an important material property to be considered in the design and operation of various types of chemical and mineral processes. For the flotation process, particle size affects energy, water consumption, and mineral liberation.

Table 1. Description of proof approach.

Phase 1: Proof by construction	For the design of flotation circuit without grinding, the design of flotation circuit with grinding, and the design of fractional crystallization, the first two steps of the determination of the optimal design is performed. This is: (a) The postulation of a superstructure that represents a set of process alternatives, and (b) Its formulation as a mathematical programming model, usually a MINLP.
Phase 2: Proof by exhaustion	For each design problem of phase 1, the unit operation performance is represented by a uniform distribution. The unit operation for flotation stages is stage recoveries; for grinding, it is the "conversion" from one particle class to another particle class; for fractional crystallization, the equilibrium composition in multiple saturation points. Then, the following steps are performed: (a) The uniform distribution is sampled, and the values are used as parameters in the mathematical programming model, which usually became a MILP problem. (b) Determination of the optimal configuration by solving the model with an optimization algorithm. (c) If the solution is new, it is saved. (d) Steps (a) to (c) are repeated several times.

3.3. Fractional crystallization

The fractional crystallization problem corresponds to the separation of sodium sulfate and magnesium sulfate from astrakanite ($Na_2SO_4.MgSO_4.4H_2O$). This problem was formulated by Cisternas (1999). Eleven equilibrium states at four temperatures were considered in the superstructure. The superstructure represented 3456 process structures. The variables with uncertainties corresponded to the equilibrium compositions. The equilibrium compositions were a function of the temperature and composition. A thermodynamic model can be used to predict the conditions, but the prediction is not always accurate and adds computational cost to the mathematical programming model. For this reason, in this study, a hundred and five compositions were represented by a uniform distribution.

4. Results and discussion

4.1. Flotation circuit without grinding

Only three flotation structures were obtained from the 30,000 cases studies analyzed. The 30,000 cases were generated by random sampling the uniform distribution function that represented the stage recoveries. These uniform distributions are $\sim U(0.8,0.9)$ for chalcopyrite, $\sim U(0.5,0.6)$ for chalcopyrite slow, and $\sim U(0.02,0.1)$ for gangue. Only 0.04% (three flotation structures) of the 6912 feasible circuit structures were optimal in a wide range of stage recoveries, which is a remarkable result. Figure 1 shows the regions where each optimal solution was obtained.

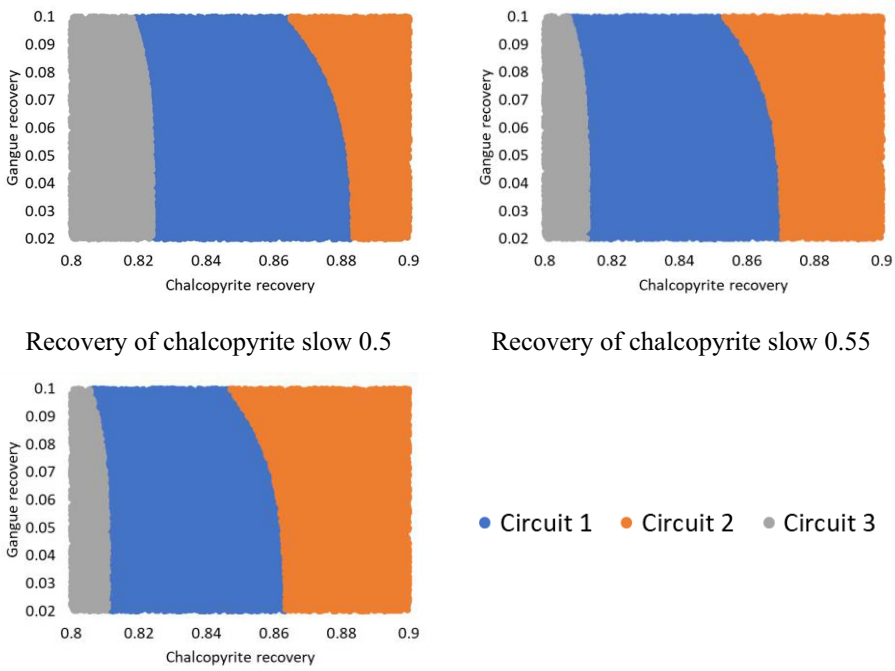


Figure 1. Regions of optimal design circuit structures.

Figure 1 clearly shows that there were regions where each flotation circuit was optimal, and that within each region only one structure was optimal. Additionally, Figure 1 can assist in identifying the range of recovery values a structure will remain as optimal.

4.2. Flotation circuit with grinding

This example was solved 60,000 times for different combinations of the stage recoveries and grinding conversions. Only two process structures were obtained from the twelve feasible structures. These structures are shown in Figure 2. The structure without grinding was obtained in 82% of the cases, whereas the structure with grinding was observed in 18%. The fact that a structure was optimal in 82% of the cases is remarkable considering that 46 variables were represented by uniform distributions.

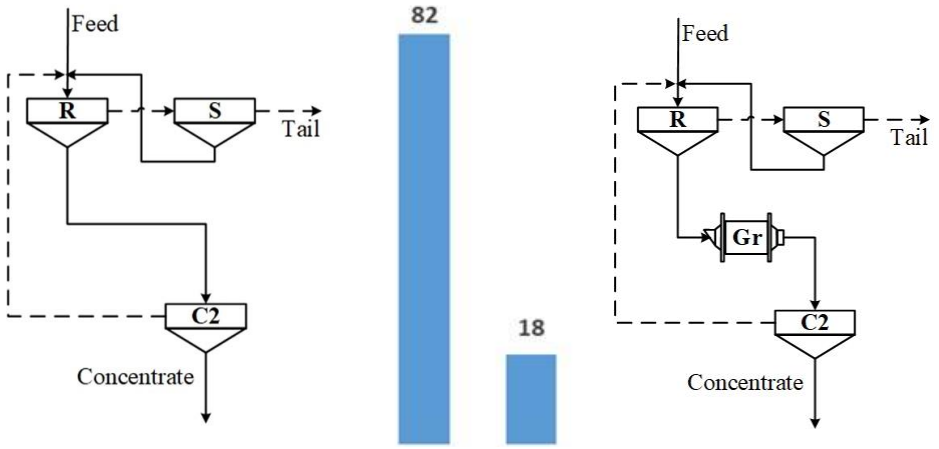


Figure 2. Optimal process structures obtained after 60,000 cases studies.

4.3. Fractional crystallization

This example was solved 30,000 times for different combinations of the equilibrium compositions. Sixteen process structures were obtained from the 3456 feasible options. Furthermore, only five process structures were obtained in more than 5% of the 30,000 cases analyzed (see Figure 3).

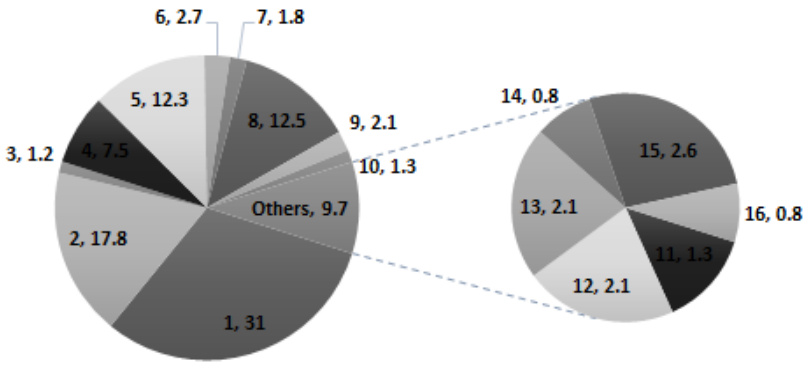


Figure 3. The percentage that each fractional crystallization structure was obtained.

The fact that only 0.46% of the feasible process structures were optimal in a wide range of stage recoveries is a remarkable result as it allows the selection of a set of optimal process structures without a process unit model by using uniform distribution functions.

5. Conclusions

In all the examples studied, it was observed that there were few process structures that were optimal for a wide range of values in the performance of the process units. A remarkable aspect of this conclusion is the possibility of separating the flowsheet design into two steps. First, a set of optimal flowsheet processes can be identified using uniform distribution functions to represent the unit operation performance. Then, after selecting a set of optimal processes, the equipment design parameters and operating conditions can be defined based on simulation and laboratory tests. In addition, these optimal structures can be evaluated from different points of view—economic, environmental, control, flexibility— to make a final design decision.

Acknowledgments: The author thanks CONICYT for funding through Fondecyt 1171341.

References

L. Cisternas, R. Acosta-Flores, F. Lucay, E. Gálvez, 2016, Mineral Concentration Plants Design Using Rigorous Models, Computer Aided Chemical Engineering, 38, 1461-1466

L. Cisternas, 1999, Optimal Design of Crystallization-Based Separation Schemes, AIChE J., 45, 7, 1477-1487.

L. Cisternas, D. Méndez, E. Gálvez, R. Jorquera, 2006, A MILP Model for Design of Flotation Circuits with Bank/Column and Regrind/no regrind selection, International Journal of Mineral Processing, 79, 253-263.

D. Calisaya, A. López-Valdivieso, M. de la Cruz, E. Gálvez, L. Cisternas, 2016, A strategy for the identification of optimal flotation circuits, Minerals Engineering, 96-97, 157-167.

Q. Chen, I. Grossmann, 2017, Recent Developments and Challenges in Optimization-Based Process Synthesis, Annual Review of Chemical and Biomolecular Engineering, 8, 249-283.

E. Martelli, C. Elsido, A. Mian, F. Marechal, 2016, Synthesis of Heat Exchanger Networks and Utility Systems: sequential initialization procedure and simultaneous MINLP algorithm, Computer Aided Chemical Engineering, 38,1449-1454

M. Ropotar, Z. Pintarič, J. Reneaume, Z. Kravanja, 2009, MINLP Synthesis of Reactive Distillation Using a Disjunctive, Hybrid Model, Computer Aided Chemical Engineering, 27, 543-548.

K. Sassi, I. Mujtaba, 2011, Optimization of Design and Operation of Reverse Osmosis Based Desalination Process Using MINLP Approach Incorporating Fouling Effect, Computer Aided Chemical Engineering, 29, 206-210.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50029-X

Optimal Synthesis of Batch Water Networks with a Flexible Scheduling Framework Using Dynamic Programming

Zhiwei Li and Thokozani Majozi*

School of Chemical and Metallurgical Engineering, University of the Witwatersrand, 1 Jan Smuts Ave, Johannesburg 2001, South Africa
thokozani.majozi@wits.ac.za

Abstract

Batch processes have received considerable attention from both industry and academia because of their flexibility to produce low volume, high value-added products. Sustainable water management and strict environmental regulation drive the development of systematic methods for water integration in batch plants. In batch processes, scheduling plays a critical role towards resource conservation. In some cases, production rescheduling can further reduce freshwater consumption and wastewater generation. Although many papers about simultaneous optimization of batch water network integration and product scheduling have been published, the insight-based method is scarce. This work proposes a dynamic programming (DP) method for the optimal design of batch water networks with a flexible product scheduling. The proposed methodology is explained as follows. Firstly, water-using units are ranked based on the ascending of inlet limiting concentration. The whole process is divided into N stages and each concentration locates at a specific stage. Secondly, the target of freshwater consumption of the process can be determined using the recursive function of DP model. Thirdly, once the target is determined, the optimal match between water sources and sinks are identified. The timing of each operation could be further determined. If the batch water network is feasible, the final batch water network is determined. If the obtained water network is infeasible, it will return to the second step to relax the target or chose another alternative with the same target until obtaining feasible batch water network. Two illustrative examples from literature were considered to demonstrate the feasibility of the proposed method. In the first case study a regenerator with a fixed outlet concentration is incorporated in batch water network. In the second case study is presented a batch plant with multiple contaminants in which a fixed removal ratio model is employed.

Keywords: Water network, Process integration, Flexible scheduling, Batch processes.

1. Introduction

Wastewater minimization has gained significant attention from both academia and industry due to the stringent legislation on effluent and the scarcity of freshwater. Reduction of freshwater consumption not only makes the plant save operation costs, but also discharges less wastewater to the environment. Batch processes have been widely used in chemical industry because of their suitability for the production of small volume, high value added products, as well capability of adjusting to fast market changes. Hence, it is urgent to develop a systematic approach to minimize wastewater consumption by

exploiting opportunities for water reuse and removing hazardous contaminants through effective regeneration operations.

Graphical methods and mathematical programming have been effectively utilized to minimize freshwater consumption in batch processes (Gouws et al., 2010). Especially, much effort has been placed on water integration of batch processes with the predefined schedule. However, results obtained from these studies could be suboptimal. In the work of Adekola and Majozi (2011), the optimal product schedule can result in further reduction of freshwater consumption.

Recently, wastewater minimization and optimal production scheduling have been considered simultaneously based on mathematical programming. The formulations could identify the minimum freshwater consumption together with the corresponding production schedule. Chen and Chang (2007) proposed a general mixed integer nonlinear programming model to synthesize water networks in batch processes. In their work, better overall designs can be generated. Zhou et al. (2009) developed a systematic design methodology to optimize batch process scheduling and water allocation networks simultaneously. A continuous-time based time model is introduced and an improved state-space superstructure with state-event network and state-task network representations is adopted to capture the structural characteristics of the integrated water-allocation network for batch processes. Li et al. (2010) used the mathematical technique to deal with single- and multiple-contaminant batch water allocation networks. A flexible scheduling model was presented by integrating batch production scheduling and water allocation networks. Adekola and Majozi (2011) addressed the problem of wastewater minimization with a flexible product schedule. Results showed that the proposed model can achieve a wastewater reduction of 19.2 % and 26 % for the two case studies considered. Chen et al. (2011) developed a mathematical model based on resource-task network representation for simultaneous scheduling and water minimization in multipurpose batch plants. The objective function is to maximize the profit by taking into account the net income from production and water-related costs. Chaturvedi and Bandyopadhyay (2014) proposed a mathematical formulation to minimize the operational cost of water allocation networks in a flexible schedule plant by utilizing multiple freshwater resources. The results of two examples showed that reductions of 17 % and 32 % in operating costs are observed when multiple water resources are used. Chaturvedi et al. (2016) analyzed the effect of multiple water resources in a flexible schedule water network. The findings indicated that once the optimum schedule for a single resource that results in the minimum operating cost is obtained, the same schedule can be applied for problems involving multiple resources. Lee and Foo (2017) considered simultaneous process scheduling and water minimization using pinch-based automated targeting model and a state-task network based discrete-time scheduling model. Li and Majozi (2017) used the match ranking matrix approach to prioritize the matches between water sources and sinks. Based on the ranking, batch water networks can be designed considering the time-dependent nature of batch processes. Some progress of optimization of water integration and production scheduling is presented in the text book (Majozi et al., 2017).

These methods are all based on mathematical programming. However, insight-based methods to deal with these issues are rare. The contribution of this work is to consider the optimization of water utilization in flexible batch processes using dynamic programming (DP). This research aims to investigate the opportunity to further reduce freshwater consumption and wastewater generation by optimizing product scheduling.

2. Problem statement

This work addresses the problem of optimizing water resources for flexible schedule batch processes. The problem can be stated as follows.

Given:

(i) the production recipe for each product, including mean processing times in each unit operation,

(ii) the contaminant mass load of each contaminant,

(iii) water requirement and the cleaning duration for each unit to achieve the required cleanliness

(iv) maximum inlet and outlet concentrations of each contaminant,

(v) the performance of the regenerator and

(vi) the time horizon of interest

Determine the production schedule that achieves the minimum freshwater consumption by exploring opportunities of water recycle/reuse and regeneration.

3. Methodology

The dynamic programming (DP) method is used to determine the target of freshwater consumed in the process. DP is generally used to reduce a complex problem with many variables into a series of optimization problems with one variable in every stage. It is characterized fundamentally in terms of stages and states. Each stage constitutes a new problem to be solved in order to find the optimal result. In each stage the problem can be described by a relatively small set of state variables. The states in this work are decisions that are made on whether to use freshwater and/or reuse wastewater or regenerated water. The dynamic programming equation is updated using the chosen state of each stage. At the last stage, it thus obtains the target of freshwater for the whole problem. The dynamic programming equation can not only assure in the present stage the optimal solution to the sub-problem is chosen, but it also guarantees the solutions in other stages are optimal through the minimization of recurrence function of the problem. The principle of optimality of DP is explained in Bellman (1957). The details of DP approach are introduced in Li and Majozi (2017).

The detailed procedure for design of flexible batch water network is shown in Figure 1. As shown in Figure 1, the first step is to divide the process into many stages. The stages can be determined based on the inlet concentration of each operation. Next, the target of freshwater consumption for the whole process, as well as the specific freshwater consumption for each stage can be identified using DP method. Since the information of freshwater consumption, reused water in each stage is determined, the sequence of operation can be subsequently identified. In other words, the receiving unit should start immediately after the wastewater generating unit finishes. It can thus design the initial water network of batch processes with the constraint of time. If the initial water network is feasible, it will obtain the final batch water network.

If the process requires considering water regeneration scenario, the timing of operation for water reuse/recycle scheme can be used as the basis for further investigation. In every stage, regenerated water as a water resource is incorporated into the analysis and the

match with minimum freshwater and/or minimum quantity of regenerated water is selected as the optimal strategy. The same procedure of water reuse/recycle is repeated to get the final batch water network.

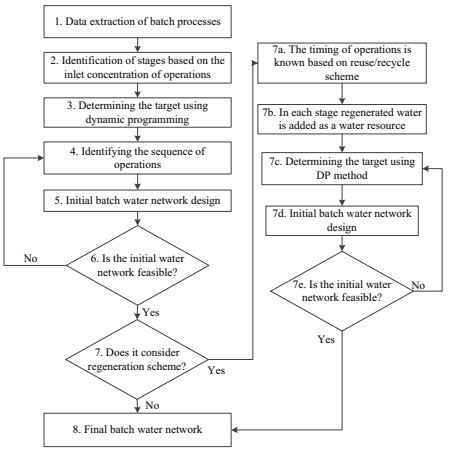


Figure 1 The design procedure for batch water network

4. Illustrative examples

4.1. Example 1

A literature example adopted from Liu et al. (2009) is introduced to illustrate the design procedure for batch water network. Based on the inlet concentration of operations, the whole process can be divided into 5 stages, as shown in Figure 2. The target of freshwater consumption for the whole process can be determined using the DP method (Li and Majozi, 2017). Therefore, the optimal match between water sources and sinks can be determined in each stage with the objective function of minimizing freshwater consumption. For example, wastewater from processes A and B can be reused in processes C, D and E. Thus, when processes A and B are finished, processes C, D and E should start at the same time, this information can assist to determine the sequence of operations achieving the goal of minimum freshwater consumption. On the other hand, the time horizon of interest is fixed as 8.5 h. The water storage profile for Example 1 is shown in Figure 3. The resultant water networks with reuse/recycle and regeneration are shown in Figures 4 and 5, respectively. The freshwater targets of Liu et al. (2009) and Adekola and Majozi (2011) are 68.594 t and 64.07 t, respectively.

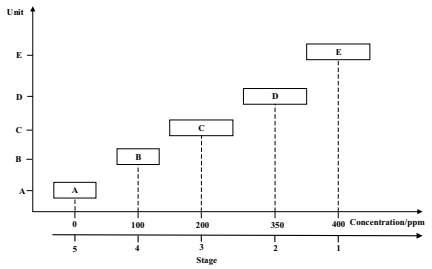


Figure 2 The identification of stage

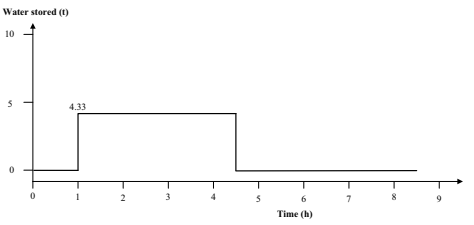


Figure 3 Water storage profile of Example 1 (reuse)

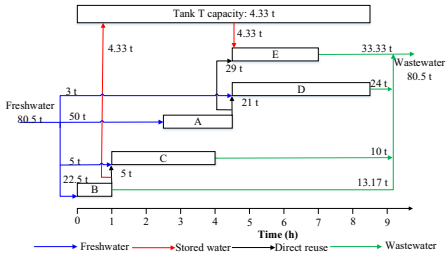


Figure 4 Results of Example 1 (reuse)

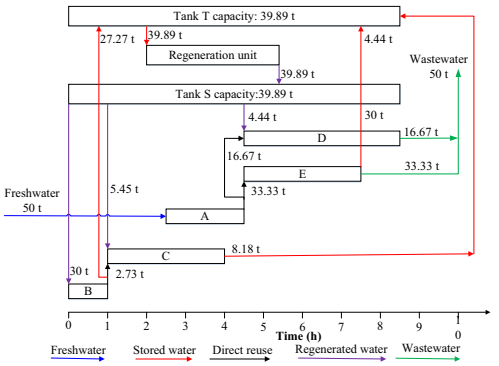


Figure 5 Results of Example 1 (regeneration)

4.2. Example 2

This example taken from Adekola and Majozi (2011) is based on a section of a pharmaceutical production plant, which produces four types of products, i.e. shampoos, deodorants, lotions and creams. Each product was produced by a specific mixer. Mixer 1 was dedicated to the mixing of shampoos, mixer 2 was dedicated to the mixing of deodorants, mixer 3 was dedicated to the mixing of lotions and mixer 4 was dedicated to the mixing of creams. Raw material is charged to a mixer. The raw material is then mixed until the required physical characteristics are obtained. Once a product is mixed, it is removed and stored. The mixers are then washed. Freshwater consumption could be further reduced by changing the sequence of batch operations. The resultant water network with water reuse/recycle scheme is shown in Figure 6, while in Figure 7 it shows the batch water network with regeneration scheme. The result of regeneration scenario in this work is less than that of Adekola and Majozi (2011), i.e. 2653 t. The solution time of Adekola and Majozi (2011) was 464 s. However, the computational time of this work is incomparable. This is because in this work the complex problem has been decomposed into a few small problems, which can be easily solved using direct search algorithm. It is also inappropriate to compare the summation of solution time of sub-problem with that of mathematical programming. The solution time of each sub-problem is substantially reduced compared with the gradient-based algorithm.

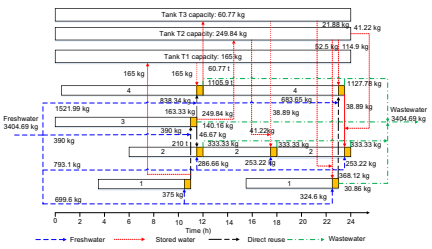


Figure 6 Results of Example 2 (reuse)

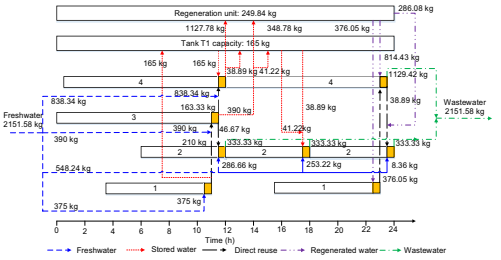


Figure 7 Results of Example 2 (regeneration)

5. Conclusions

A DP approach is used to identify the optimal water using policy that could achieve the target of minimum freshwater consumption and minimum storage capacity. In every stage, the optimal matches between water sources and sinks are determined and the

sequence of operation can also be identified. Based on the obtained sequence of operations, the product can be rescheduled to minimize freshwater target and storage facilities with the constraint of time horizon of interest. Two different examples from literature are used to demonstrate the applicability of the proposed approach. The obtained results match with published literature. Future work should focus on the robust design of batch processes with minimum freshwater consumption and optimal product scheduling.

References

O. Adekola, T. Majozi, 2011. Wastewater minimization in multipurpose batch plants with a regeneration unit: Multiple contaminants. Computers & Chemical Engineering 35, 2824-2836.

R. Bellman, 1957. Dynamic Programming. Princeton, New Jersey, USA: Princeton University Press.

C.L. Chen, C.Y. Chang, J.Y. Lee, 2011. Resource-task network approach to simultaneous scheduling and water minimization of batch plants. Industrial & Engineering Chemistry Research 50, 3660-3674.

N.D. Chaturvedi, S. Bandyopadhyay, 2014. Optimization of multiple freshwater resources in a flexible-schedule batch water network. Industrial & Engineering Chemistry Research 53, 5996-6005.

N.D. Chaturvedi, Z.A. Manan, S.R. Wan Alwi, Bandyopadhyay S., 2016. Effect of multiple water resources in a flexible-schedule batch water network. Journal of Cleaner Production 125, 245-252.

K.F. Cheng, C.T. Chang, 2007. Integrated Water Network Designs for Batch Processes. Industrial & Engineering Chemistry Research 46(4), 1241-1253.

J.F. Gouws, T. Majozi, D. C. Y. Foo, C.L. Chen, J. Y. Lee, 2010. Water Minimization Techniques for Batch Processes. Industrial & Engineering Chemistry Research 49(19): 8877-8893.

J.Y. Lee, E.R. Seid, T. Majozi, 2016. An improved model for Heat Integration of intermittent process streams in multipurpose batch plants. Applied Thermal Engineering 105, 822-838.

J.Y. Lee, D.C.Y. Foo, 2017. Simultaneous targeting and scheduling for batch water networks. Industrial & Engineering Chemistry Research 56, 1559-1569.

J. Li, C.A. Floudas, 2010. Optimal event point determination for short-term scheduling of multipurpose batch plants via unit-specific event-based continuous-time approaches. Industrial & Engineering Chemistry Research 49, 7446-7469.

Z. Li, T. Majozi, 2017. Optimal Design of Batch Water Network with a Flexible Scheduling Framework. Chemical Engineering Transaction 61: 139-144.

Z. Li, T. Majozi, 2017. Dynamic Programming for Optimal Synthesis of Water Networks in Batch Processes. Computer Aided Chemical Engineering 40, 919-924.

Y. Liu, , G. Li, L. Wang, J. Zhang, K. Shams, 2009. Optimal design of an integrated discontinuous water-using network coordinating with a central continuous regeneration unit. Industrial & Engineering Chemistry Research 48, 10924-10940.

T. Majozi, E.R. Seid, J.-Y. Lee, 2017. Understanding batch chemical processes: Modelling and case studies. CRC Press, Boca Raton, Florida, United States.

R.J. Zhou, L.J. Li, W. Xiao, H.-G. Dong, 2009. Simultaneous optimization of batch process schedules and water-allocation network. Computers & Chemical Engineering 33, 1153-1168.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50030-6

A Novel Method of Integrating Flexibility and Stability for Chemical Processes under Parametric Uncertainties

Ying Chen, Zhihong Yuan, Bingzhen Chen*

Department of Chemical Engineering, Tsinghua University, Beijing, 100084, China
dcecbz@mail.tsinghua.edu.cn

Abstract

Process optimization under parametric uncertainties is a challenging topic of practical importance in the Process Systems Engineering. Flexibility and stability are two crucial components of chemical process operability. In previous research, flexibility and stability usually have been studied separately, resulting in flexible regions containing unstable areas that are difficult to operate and control. The major difficulty of integrating flexibility and stability lies in how to convert the Lyapunov stability conditions into specific constraints embedded within the flexibility analysis model. This paper proposes a novel method for the integration of process flexibility and stability, which incorporates stability constraints obtained by using singularity theory based stability analysis method into the MINLP model for calculating flexibility index. As a result, a stable flexible region is obtained, which not only adapts to variations in uncertain parameters, but also ensures stable operation and process inherent safety. Two cases: 1,3-propanediol fermentation system and methyl methacrylate polymerization process, have been studied to demonstrate the effectiveness of the proposed methodology. The obtained results illustrate that the novel method for considering flexibility and stability simultaneously possesses distinct computational advantages over the eigenvalue optimization algorithm, so it has great potential for applications in large-scale systems.

Keywords: parametric uncertainties; flexibility; stability analysis; singularity theory

1. Introduction

Parametric uncertainties are inevitable in all stages of chemical production, ranging from process design and synthesis, to operational planning, production scheduling and supply chain optimization. Uncertainties in feedstock's properties, kinetic constants and transfer coefficients widely exist in chemical process design and operation. Since the solution of an optimization problem generally exhibits high sensitivity to the parameter variations, it is obviously significant to account for the effect that uncertainties can have on both the optimality and feasibility of the chemical process operation (Halemane and Grossmann, 1983).

Flexibility and stability are two crucial components of chemical process operability (Swaney and Grossmann, 1985). Flexibility guarantees feasible steady-state operation over a range of uncertain operating conditions and denotes static resiliency. Stability reflects the tolerance for slight perturbations in the open-loop system and emphasizes the dynamic characteristics of a system. Stability is a basic feature that can significantly decide the fate of a design and the inherent safety of a system (Wang et al., 2009).

In previous research, flexibility and stability usually have been studied separately, resulting in flexible regions containing unstable areas that are difficult to operate and control, as shown in Figure 1. Although the flexible regions can satisfy the production requirements, underlying instability limits the degree of inherent safety of a system. Within the unstable areas, the system may experience a dramatic change or even run away under very slight disturbance. Therefore, we want to find an operating region that is not only robust to parametric uncertainties, but also ensures stable operation without complex control system.

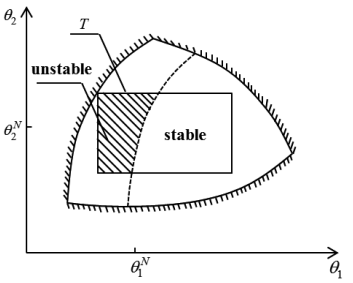


Figure 1. Stable and unstable part of flexible region(Jiang et al., 2014)

The more recent trend has been to integrate various indicators of operability. The major difficulty of integrating flexibility and stability lies in how to convert the Lyapunov stability conditions into specific constraints embedded within the flexibility analysis model. Jiang et al. (2014) first integrated flexibility with stability through an eigenvalue optimization algorithm, in which the necessary condition of stability where the real part of the eigenvalues is smaller than zero is transformed into the positive definiteness of a real symmetric matrix so that the stability condition is expressed explicitly, as illustrated in Figure 2. The limitation of this approach is that it can only deal with the problem in which the inner optimization is convex. Moreover, the positive definiteness of a real symmetric matrix is judged by the positive definiteness of the successive principal minors of matrix, which leads to great computational expense. Hence, it is not suitable for practical applications in large-scale systems.

$$\mathrm{Re}(\lambda_i(J(h))) < 0 \xrightarrow[\text{Lyapunov equation}]{} \begin{matrix} J^T P + PJ + I = 0 \\ P \text{ - Positive definite} \end{matrix} \rightleftharpoons \begin{matrix} J^T P + PJ + I = 0 \\ \det(P_k) > 0 \end{matrix}$$

Figure 2. Eigenvalue optimization algorithm

This paper presents a singularity theory based stability analysis method with the aim to obtain concise stability constraints which could be easily embedded into the flexibility analysis model. The proposed method possesses distinct computational advantages over the eigenvalue optimization algorithm and has great potential for applications in more complex systems.

2. Novel method for integrating flexibility and stability

The proposed method is described as follows. First of all, by adjusting the value of uncertain parameters and control variables within a given range, a series of steady-state solution curves is depicted, which reflects how process stability changes with the variations of operating conditions. It is essential to identify singularity points where the transformation of system's stability characteristics may take place. Then the dynamic trajectories of singularity points extracted from the series of curves are regressed into functions of uncertain parameters and control variables. Finally, based on the stability region partition, the obtained functions are transformed into inequality constraints, which are further embedded into the optimization model for calculating flexibility index. After adding the stability constraints, the boundary of feasible region becomes stable, the flexible region shrinks and the flexibility index reduces accordingly. The flowchart of the novel method for integrating flexibility and stability is shown in Figure 3.

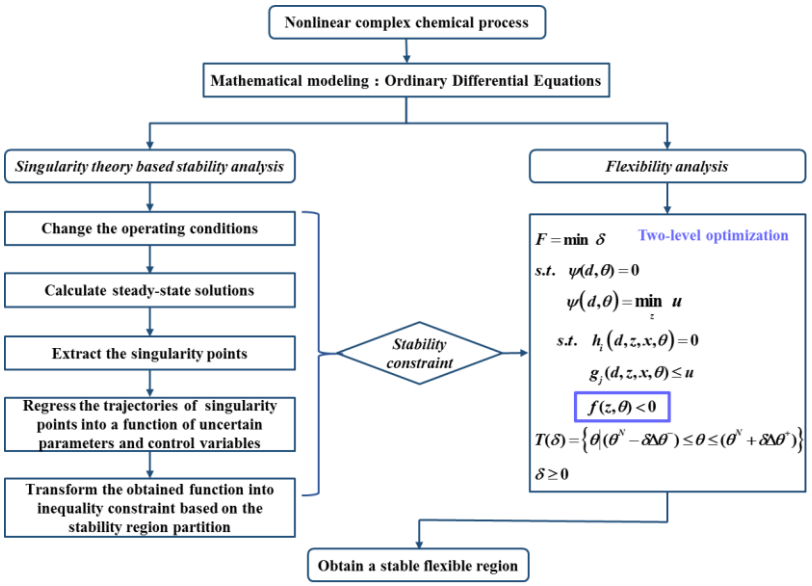


Figure 3. Flowchart of the novel method for integrating flexibility and stability

The general form of calculating flexibility index through active constraint algorithm (Floudas et al., 2001) with consideration of stability is presented as follows:

$$
\begin{aligned}
&F = \min \delta \\
&\text{s.t. } \psi(d,\theta) = 0 \\
&\quad \psi(d,\theta) = \min_{z} u \\
&\quad \text{s.t. } h_i(d,z,x,\theta) = 0,\ i \in I \\
&\qquad g_j(d,z,x,\theta) \le u,\ j \in J \\
&\qquad f(z,\theta) < 0 \\
&\qquad T(\delta) = \left\{\theta \middle| (\theta^N - \delta\Delta\theta^-) \le \theta \le (\theta^N + \delta\Delta\theta^+)\right\} \\
&\qquad \delta \ge 0
\end{aligned}
\tag{1}
$$

This is a two-level optimization problem which can be further solved by applying the necessary and sufficient KKT optimality conditions to transform the flexibility index problem into a single optimization problem. However, it is worth noting that the application of the KKT conditions to reformulate the inner level model may not be correct in cases where regularity conditions do not hold.

In this way, the integration of flexibility and stability is formulated as a MINLP problem that could be solved by extensively used GAMS. Consequently, we obtain a stable flexible region that not only adapts to variations in uncertain parameters, but also ensures stable operation and inherent safety.

$$
\begin{aligned}
&F = \min_{\theta,x,z,\delta,s_j,\mu_j,\lambda_j,y_j} \delta \\
&\text{s.t. } g_j'(d,z,x,\theta) + s_j = 0,\ j \in J \\
&\quad h_i(d,z,x,\theta) = 0,\ i \in I \\
&\quad \sum_{j=1}^{J} \lambda_j = 1 \\
&\quad \sum_{j=1}^{J} \lambda_j \frac{\partial g_j'}{\partial z} + \sum_{i=1}^{I} \mu_i \frac{\partial h_i}{\partial z} = 0 \\
&\quad \sum_{j=1}^{J} \lambda_j \frac{\partial g_j'}{\partial x} + \sum_{i=1}^{I} \mu_i \frac{\partial h_i}{\partial x} = 0 \\
&\quad \lambda_j - y_j \le 0,\ j \in J \\
&\quad s_j - U(1 - y_j) \le 0,\ j \in J \\
&\quad \sum_{j=1}^{J} y_j \le n_z + 1 \\
&\quad \theta^N - \delta\Delta\theta^- \le \theta \le \theta^N + \delta\Delta\theta^+ \\
&\quad \delta \ge 0, y_j = \{0,1\}, \gamma_j \ge 0, s_j \ge 0, j \in J
\end{aligned}
\tag{2}
$$

g_j' denotes the inequalities, including the product quality requirements g_j and the stability constraint $f(z,\theta) < 0$

3. Case studies

3.1. 1,3-Propanediol Fermentation

The anaerobic fermentation system is a highly nonlinear complicated chemical process, which generally exhibits oscillation and multi steady-state characteristics. These characteristics severely affect the stable process operation and the quality of desired products. The mathematical model of 1,3-propandiol fermentation process is presented as follows (Xiu et al., 1998) and the range of process parameters is listed in Table 1.

$$\begin{aligned}
\frac{dx}{d\tau} &= x(u-d)\\
\frac{dy}{d\tau} &= d(y_0-y)-x\varphi_1\left(\beta_1+\gamma_1 u+\frac{y}{y+\alpha_1}\right)\\
\frac{dz}{d\tau} &= x\varphi_2\left(\beta_2+\gamma_2 u+\frac{y}{y+\alpha_2}\right)-zd\\
u &= \frac{y}{y+\alpha_3}(1-y)(1-z)
\end{aligned} \tag{3}$$

Table 1. Ranges of process parameters (1,3-Propanediol Fermentation)

Parameters	Ranges
x	0.29-0.6
z	≥0.58
d	0.3-0.4

x, y and z are the state variables which denote dimensionless biomass concentration, substrate concentration and product concentration respectively. The control variable is dimensionless dilution rate d and the two uncertain parameters are dimensionless feed substrate concentration y_0 and dimensionless critical parameter α_3. Given the nominal operating point, $y_0^{\mathrm{N}} = 0.228$, $\alpha_3^{\mathrm{N}} = 1.4 \times 10^{-4}$ and the maximum positive and negative deviations of the two uncertain parameters are $\Delta\theta_1^{+} = 0.02$, $\Delta\theta_1^{-} = 0.02$ and $\Delta\theta_2^{+} = 2 \times 10^{-5}$, $\Delta\theta_2^{-} = 2 \times 10^{-5}$.

The singularity theory based stability analysis method can characterize regions in parameter space over which different kinds of stability characteristics may exist (Wang et al., 2009). As shown in Figure 4 and Figure 5, the steady-state solution curve is divided by singularity points into three sections which have different stability.

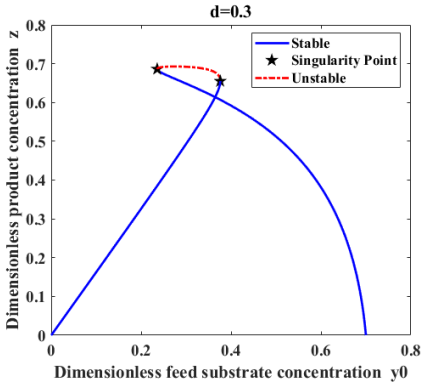


Figure 4. Different stable regions separated by singularity points when $d = 0.3$

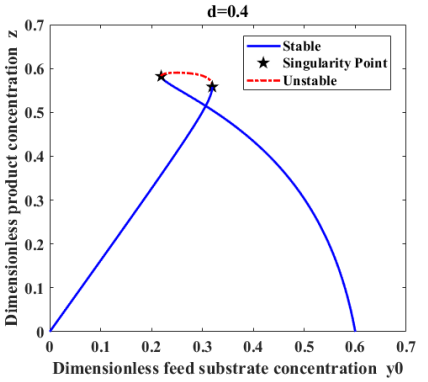


Figure 5. Different stable regions separated by singularity points when $d = 0.4$

Moreover, the stability characteristics of the system will change with the variations of parameters. A series of steady-state solution curves is depicted in Figure 6 by changing the value of control variable d and uncertain parameter α_3 within a range. The dynamic trajectories of the left singularity points and the right singularity points can be regressed

into functions of uncertain parameters and control variables respectively. By connecting the left and the right moving singularity points, a linear equation, which distinguishes the stable regions from the unstable one, can be obtained. Based on the stability region partition, we can get the concise form of stability constraint, $f(z, y_0, \alpha_3, d) < 0$, which can be further embedded into the optimization model for calculating flexibility index.

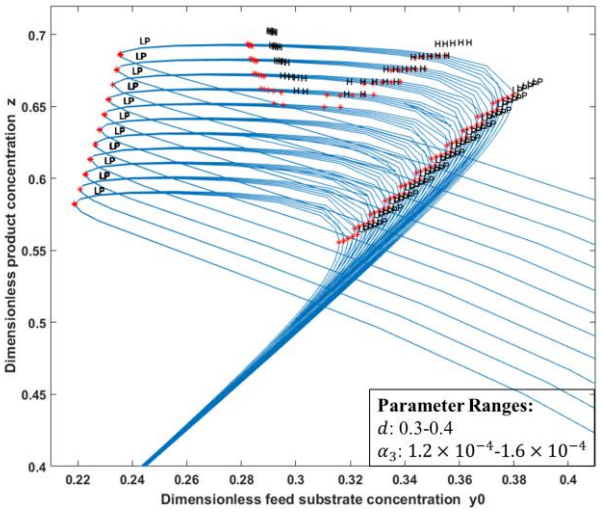


Figure 6. Stability varies with the control variable and uncertain parameters within a given range

The comparison of flexibility index without/with stability is presented in Table 2.

Table 2. Comparison of flexibility index without/with stability

Method	Flexibility Index	Computation Time (BARON)
Traditional flexibility analysis without stability	0.453	1003.30 s
Integrating flexibility and stability	0.343	1003.64 s

3.2. MMA Polymerization

The dynamic behaviour of free-radical continuous stirred tank polymerization reactors has received extensive attention. The MMA polymerization reaction takes place in a CSTR with a cooling jacket to remove extra heat, as shown in Figure 7. Six differential equations constitute the mathematical model of the MMA polymerization reactor, which can be acquired from the previous literature (Silva-Beard and Flores-Tlacuahuac, 1999).

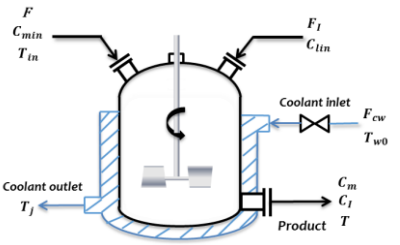


Figure 7.Polymerization reactor flow sheet

Here, the state variables are C_{m}, C_{I}, T, D_0, D_1 and T_{j}, and the control variable is the flow rate of cooling water F_{cw}. There are two uncertain parameters: monomer feed concentration C_{min} and monomer feed flow rate F.The nominal operating points and the maximum positive and negative deviations of uncertain parameters are given as follows: $C_{\mathrm{min}}^{N} = 6.4678\ \mathrm{kmol/m^3}, F^N = 1\mathrm{m^3/h}, \Delta\theta_1^+ = 0.3, \Delta\theta_1^- = 0.2, \Delta\theta_2^+ = 0.2, \Delta\theta_2^- = 0.2$. The use of singularity theory based stability analysis method to distinguish regions with different dynamic characteristics is illustrated in Figure 8 and Figure 9.

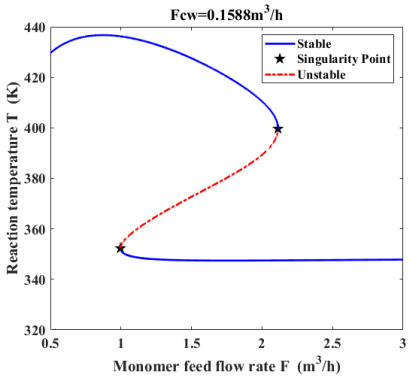


Figure 8. Different stable regions separated by singularity points when F_{cw}=0.1588m^3/h

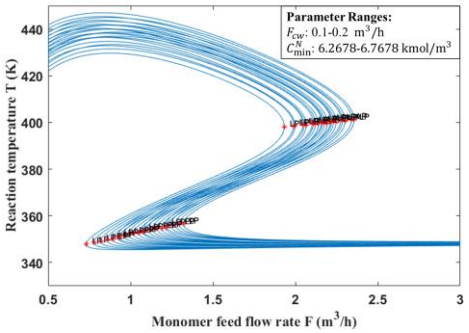


Figure 9. Stability varies with the control variable and uncertain parameters within a given range

The comparison of flexibility index without/with stability is presented in Table 3.

Table 3. Comparison of flexibility index without/with stability

Method	Flexibility Index	Computation Time (BONMIN)
Traditional flexibility analysis without stability	0.495	3.83 s
Integrating flexibility and stability	0.260	5.36 s

4. Conclusions

This paper presents a novel method for the integration of process flexibility and stability, in which singularity theory based stability analysis method is utilized to convert the Lyapunov stability conditions into stability constraints embedded within the flexibility analysis model. Due to the concise form of stability constraints, the proposed method possesses distinct computational advantages over the eigenvalue optimization algorithm. Accordingly, it has great potential for applications in large-scale systems. Attractive as it is, it also has a certain limitation. It can only deal with the uncertainties within a specific range in which process stability observes a certain changing trend.

References

C.A. Floudas, Z.H. Gümüş, and M.G. Ierapetritou, 2001, Global optimization in design under uncertainty: feasibility test and flexibility index problems, Ind. Eng. Chem. Res., 40, 4267-4282.

K.P. Halemane, and I.E. Grossmann, 1983, Optimal process design under uncertainty, AIChE J., 29, 3, 425-433.

H. Jiang, B. Chen, H. Wang, T. Qiu, and J. Zhao, 2014, Novel Method for Considering Process Flexibility and Stability Simultaneously, Ind. Eng. Chem. Res., 53, 38, 14765-14775.

A. Silva-Beard, and A. Flores-Tlacuahuac, 1999, Effect of Process Design/Operation on the Steady-State Operability of a Methyl Methacrylate Polymerization Reactor, Ind. Eng. Chem. Res., 38, 12, 4790-4804.

R. E. Swaney, and I. E. Grossmann, 1985, An index for operational flexibility in chemical process design. Part I: Formulation and theory, AIChE J., 31, 4, 621-630.

H. Z. Wang, B. Z. Chen, X. R. He, T. Qiu, J. S. Zhao, 2009, Singularity theory based stability analysis of reacting systems, Computer Aided Chemical Engineering, 27, 645-650.

Z. L. Xiu, A. P. Zeng, W. D. Deckwer, 1998, Multiplicity and stability analysis of microorganisms in continuous culture: effects of metabolic overflow and growth inhibition, Biotechnology and bioengineering, 57, 3, 251-261.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50031-8

Analysis and evaluation of a heat integrated horizontal distillation system

Emmanouil Papadakis, Seyed Soheil Mansouri, Jakob Kjøbsted Huusom, Jens Abildskov*

Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads, Buidling 229, DK-2800 Kgs. Lyngby, Denmark
ja@kt.dtu.dk

Abstract

A model-based approach is demonstrated, to analyze a heat integrated horizontal distillation system for separation of light alcohol mixtures from second-generation biomass. The objective is to develop an engineering tool, based on process systems engineering concepts. This tool is used to analyze and evaluate the proposed separation system to provide guidelines with respect to specific design and process variables. The model is simulated and comparative studies for different mixtures containing alcohols are carried out. It is shown that separation of mixtures with low concertation of light alcohol content requires less energy compared to conventional distillation systems.

Keywords: Horizontal distillation system, Energy recovery, Process modelling, Process analysis and evaluation.

1. Introduction

The production of chemicals from second-generation (2G) feedstock requires gathering adequate feedstock for one central plant (Maity et al. 2015), which can be challenging. This makes small-scale factories based on local flexible production an attractive alternative. Thus, one may explore a small-scale modular separation system tailor-made for mobile refineries, to enable fast and cost-effective manufacturing of customized products at various locations. Therefore, a container format that is easily configured for different products and processes is considered here. The container format allows on-site transportation to provide manufacturing anywhere, enabling the benefits of localized service delivery without duplication of equipment at multiple locations. We focus on the development of separation technology for the separation of mixtures containing light alcohols from small-scale production facilities of alcohols from 2G biomass (such as straw and wood chips). The paper deals with development of a steady-state model using process systems engineering tools. Its purpose is to describe sufficiently the separation system and perform analysis and evaluation in terms of energy requirements. The proposed horizontal separation system aims at reduction of the costs (such as capital cost and operating cost) associated with conventional distillation systems and novel intensified and highly integrated designs such as reactive distillation, diabatic distillation, internally heat integrated distillation column (HIDiC) and the divided wall column (Kiss et al. 2013). The feasibility of other horizontal distillation systems, in separating binary liquid mixtures, has been demonstrated by experimental and numerical simulation studies (Seok et al. 1985, Ramirez-Gonzalez et al. 1992). Recently, Kim et al. (2013) and Jang et al. (2015) have shown experimentally and numerically the potential energy savings using a diabatic rectangular horizontal column in liquid binary mixtures. Here, the separation

system is also horizontal. It consists of two sections, i.e. a stripper and a rectifier. The system has no trays. Therefore, the separation is rate-based. It operates in continuous mode, it can handle viscous liquids (or liquids containing solids), high-value temperature sensitive components and offers opportunities for heat integration. The steady state model is based on the two-film theory of mass transfer. It assumes phase equilibrium at the vapor-liquid interface and counter-current flow of the vapor and liquid phases. It is applied to different mixtures and is compared with conventional separation technologies.

2. Horizontal heat integrated separation system

The system is shown in Figure 1. It consists of two, co-axially arranged tubes. The inner tube operates as a stripper and the outer one as a rectifier. The feed consists of light alcohols, water and temperature sensitive compounds and enters the stripping section at reduced pressure preventing decomposition of the temperature sensitive compounds. The heavy product of the stripper consists of water and valuable compounds and the light product consists of more light alcohols. The stripper light product is compressed and charged into the rectifying section, where the alcohols are concentrated and collected as light product, whereas the bottom product is returned to the stripper after expansion. Heat integration is thereby made possible between the rectifying and the stripping chambers.

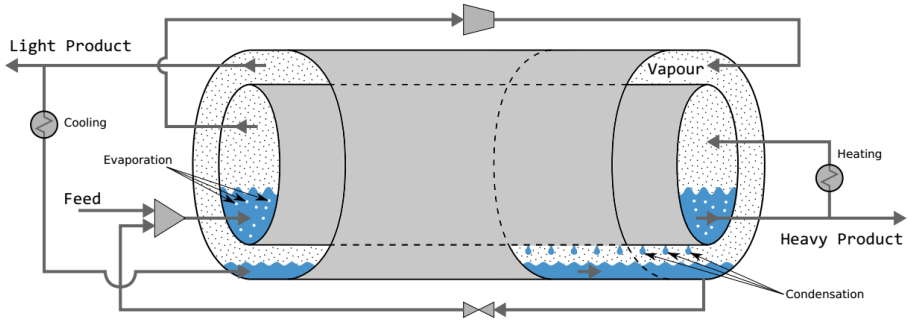


Figure 1. The heat integrated distillation system. The stripper is represented by the inner chamber and the rectifier is represented by the outer chamber. Dimensions: Length = 2 m, stripper diameter = 1.8 m and rectifier diameter = 2 m.

3. Model development

3.1. Modelling objective

The objective is to develop a model to describe the fundamental phenomena taking place in the diabatic horizontal distillation system. The model needs to predict, qualitatively, for both chambers the internal flow rate profiles, the temperature and the composition profiles in the vapor and liquid phases. The model, at steady-state, is applied to simulation studies, sensitivity analysis, evaluation and analysis in terms of economics and energy and for design studies. The model is used as a tool, which in the absence of experimental data and low technology awareness may assist in analysis and evaluating preliminary process designs, and for comparative and optimization studies.

3.2. System phenomena and assumptions

The following phenomena have been considered to model the system.

- Mass transfer controlled separation

- Counter-current flow of vapor and liquid phases in both chambers
- Heat transfer between rectifying and stripping section
- Compression of the strippers exiting vapor and pressure reduction of the rectifiers (outlet) liquid

The following assumptions have been made to develop the model:

- Vapor-liquid equilibrium is assumed only at the interface
- The pressure of the rectifying and stripping section is perfectly controlled
- The reflux ratio and the boil-up ratio are perfectly controlled
- Heat and mass transfer coefficients are calculated using known correlations (Bravo et al. 1985 and Chilton and Colburn (Taylor, R., Krishna, 1993) respectively).

3.3. Model construction

The basis of the rate-based model is shown in Figure 2a, where a segment, j, together with mass and heat transfer phenomena as well as the vapor-liquid equilibrium at the interface is shown. The model for compartment j has been used as a building block for N consequtive segments and has been combined with reboiler and condenser models as well as models for compression and valve (see Figure 2b). The model consists of mass and energy balance equations for the vapor and liquid phase, mass and heat transfer rates, equilibrium equations, and summation equations. In addition, the steady-state rate-based model employs activity coefficient models, fugacity coefficient model and the required physical properties such as densities and vapor pressures. For the calculation of the activity coefficient, the Wilson model has been selected as a thermodynamic model. The resulting equation system consists of algebraic implicit equations that requires non-linear solvers and it can be implements in various programming languages.

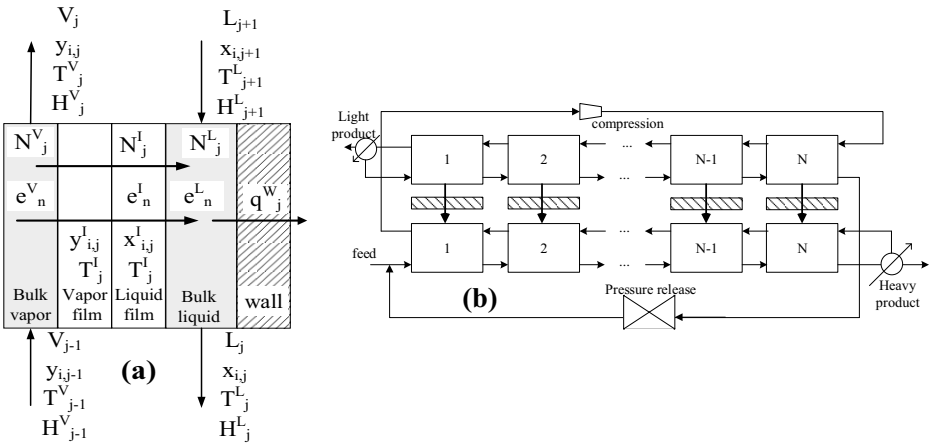


Figure 2. Model representation. (a) Single segment *j* and (b) integration of the N segments, together with compression and expansion.

3.4. Heat Integration-calculation sequence

Heat integration takes places between the rectifying and the stripping sections. The rectifying section operates at higher pressure, to ensure sufficient temperature difference. The heat transferred from the rectifying to the stripping section per segment j is:

$$Q_j = UA_j \Delta T_j \tag{1}$$

Here U is the overall heat transfer coefficient in $\frac{kW}{m^2K}$, A_j is the heat transfer area in m^2, and ΔT_j is the temperature difference between the rectifying and the stripping section $\Delta T_j = T_{R,j} - T_{S,j}$ in K. For the overall heat transfer coefficient a value of 1 $\frac{kW}{m^2K}$ has been chosen. Heat transfer area has been calculated considering that the minimum heat transfer area per segment j is the surface of the stripping segment j. The temperature profiles for the rectifying and stripping section are calculated through simulations, considering no heat integration. Once all the right-hand side variables of Eq. (1) are known, heat integration is added in the model calculations. Here, the value of the product UA_j is uncertain due to the low technology readiness level of the system; however, it can be analyzed as a sensitivity variable to investigate its effect on the separation performance. For example, a maximum value of UA_j might exist and above that limit, the system might may not have further improvements or it might lower the performance, leading to over-design and higher capital costs that are prohibited for such applications.

3.5. Sensitivity analysis

Sensitivity analysis has been performed in order to gain a deeper understanding of the effects of certain design and process variables. First, sensitivity analysis is used to investigate the effect of variables such as mass, heat transfer and interfacial area coefficients and heat transfer area on process variables such as reflux and boilup ratio in order to achieve the high separation performance in terms of product purity, flow rate and recovery. Changes above 50-60% may increase dramatically the energy requirements.

4. Model application and results

Case 1: Ethanol-water. The mixture of ethanol-water has been selected to highlight the application of proposed design using the developed steady-state model. That mixture is common after fermentation of 2G biomass. Table 1, lists the values of the variables specified to solve the model.

Table 1. Process variables used for the simulation of the heat integrated horizontal system.

Process Variables	Values	Process Variables	Values
Inlet flow rate	0.1 kmol/sec	Rectifier pressure	0.3 atm
Inlet Temperature	40 °C	Reflux ratio	2.5
Inlet Pressure	0.1 atm	Boil-up ratio	0.15
Ethanol composition	10 wt.-%	Phase contact area, S	50 m^2/m^3
Stripper pressure	0.1 atm	Phase contact area, R	50 m^2/m^3

Solving the heat integrated distillation system over the tube length, the internal flowrate, temperature and composition profiles have been obtained, as shown in Figure 3a-c, respectively. Figure 3a shows the internal flow rate of the liquid (subscript "L") and vapor (subscript "V") phase in the stripping (S) and rectifying (R) sections. The feed is at length equal to zero (0) where the flow rate has the maximum value for both vapor and liquid phases. The flowrate of the liquid phase reduces towards the tube's end due to evaporation taking place in the stripping section while the flowrate of the vapor phase increases towards the start of the tube due to the evaporation of the liquid. Similarly, in the rectifying section, the vapor flow rate decreases from the end towards the start of the tube while the liquid flow rate increases towards the end of the column due to the condensation. In Figure 3b, the temperature profiles in the stripping and rectifying sections are shown, in stripping section the temperature profiles are nearly constant, close the boiling point of the mixture and lower than 60°C (maximum temperature limit in the

stripper). In the rectifying section, the temperature is high at the inlet of the rectifying section (length at 6 m) due to the vapor compression and reduces towards the end of the rectifier due to heat transfer to the stripping section. In Figure 3c, the ethanol composition in the liquid and vapor phase is shown. In the stripping section, the composition reduces towards the end of the tube meaning that no ethanol is in the heavy product due to evaporation. In the rectifying section, the composition of ethanol increases from the inlet towards the end of tube due to the condensation process. The obtained profiles shown in Figure 3 are in good agreement with the two-film theory of mass transfer.

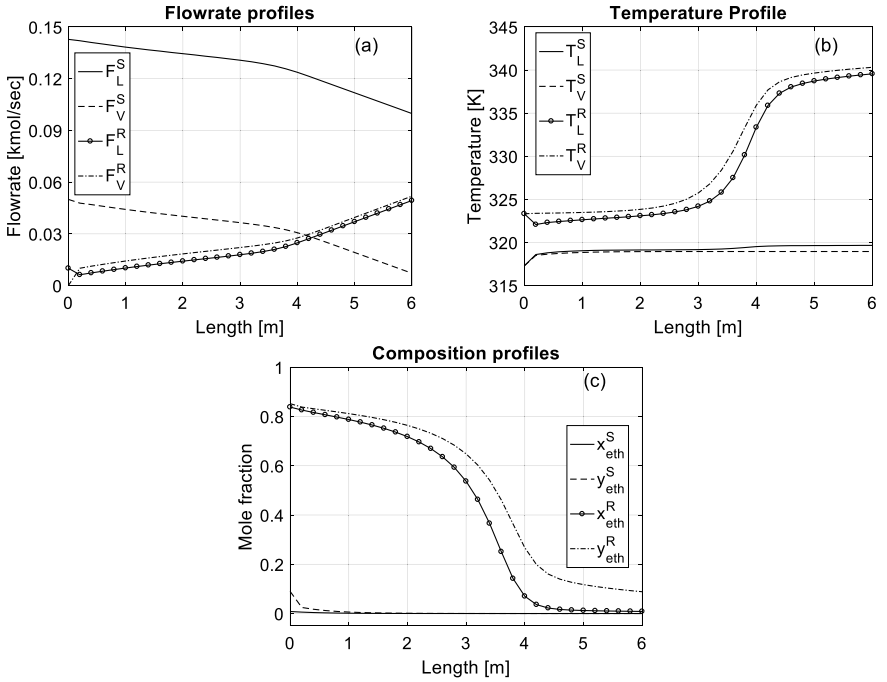


Figure 3. Simulation results for the heat integrated horizontal system. (a) Internal flow rate profiles, (b) internal temperature profiles and (c) internal composition profiles. Note that in the liquid flows direction [0→6 m] and the vapor flows at the opposite direction.

The process performance of the adiabatic horizontal distillation system has been compared to a conventional distillation column and a non-heat integrated horizontal system. For the analysis, process performance criteria such as product flow rate (0.19 kg/sec), purity (93 wt.-%) and recovery (93%) were kept constant while the energy requirements were evaluated. For the analysis, an equilibrium-based distillation column has been considered with the same feed conditions (Table 1), 15 equilibrium stages, feed location 5th stage from the top and reflux ratio 3.94. In Figure 4, the process performance criteria for the three different cases are shown, and it can be seen that the potential of reducing the energy requirements is very high (approximately 55% energy reduction).

Case 2. Mixtures of isopropanol-methanol (52 mol% methanol, Kim et al. 2013). The model has been applied to the mixture of isopropanol-methanol and it has been shown that for such a mixture the heat and mass transfer in the stripping section should not be very low, otherwise, the separation is not efficient (separation performance: 89 mol% methanol as light product, 85% product recovery, Kim et al. 2013) and the system is not suitable for this separation. However, in case of enhanced mass and heat transfer, the

separation compared to conventional technologies might be beneficial (energy savings app. 20%). The application in this mixture shows that the specific design should be used in mixtures where the light component is in small quantity (approximately 15 mol%).

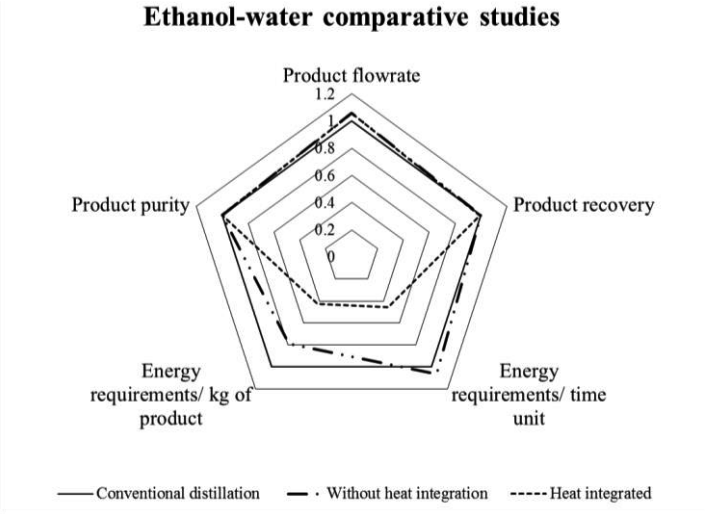


Figure 4. Ethanol-water separation comparative studies.

5. Conclusions

In this paper, an adiabatic horizontal distillation system, intended for small-scale separations of chemicals from 2G bio-refineries, has been described. The development of a process model to provide a meaningful description, in absense of experimental data, of the separation system using process systems engineering tools, has been demonstrated. The model predicts the internal flow rate, temperature and composition in stripping and rectifying sections as well as the energy demands. It serves as a tool to assist in conceptually design and to evaluate the separation performance of the system. Simulation studies and comparative studies demonstrate the potential improvements in reduction of the energy requirements over conventional separation schemes in mixture containing small amounts of light alcohols. Complementary work to investigate process economics, dynamics and operability is planned.

Acknowledgement

The authors would like to acknowledge the financial support of the SYNFERON (journal no. 4106-00035B by Innovation Fund Denmark) project and Morten J.G. Larsen for designing Figure 1.

References

J.L. Bravo, J.A. Rocha, J.R. Fair, 1985, Hydrocarb. Process., 64, 91-95.
D. J. Jang, and Y. H. Kim, 2015, Korean J. Chem. Eng. 32, 2181–2186.
B. C. Kim, H. H.Chun, Y. H. Kim, 2013, Ind. Eng. Chem. Res. 52**,** 14927–14935.
A. A. Kiss, 2014, J. Chem. Technol. Biotechnol., 89, 479-498.
S. K. Maity, 2015, Renew. Sust. Energ. Rev., 43, 1427-1445.
E. A. Ramirez-Gonzalez, C. Martinez, J. Alvarez, 1992, Ind. Eng. Chem. Res., 31, 901-908.
D. R. Seok, S. T. Hwang, 1985, AIChE J., 31, 2059-2065.
R. Taylor AND R. Krishna, 1993, Multicomponent Mass Transfer, New York.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50032-X

Optimal Multicomponent Distillation Column Sequencing: Software and Case Studies

Radhakrishna Tumbalam Gooty[a], Parham Mobed[a], Mohit Tawarmalani[b], Rakesh Agrawal[a*]

[a]*Davidson School of Chemical Engineering, Purdue University, West Lafayette, IN-47907, USA*
[b]*Krannert School of Management, Purdue University, West Lafayette, IN-47907, USA*
agrawalr@purdue.edu

Abstract

Distillation is one of the most widely used unit operation for separations in chemical and petrochemical industries. It is well-known that the number of distillation configurations available for the separation of an n-component mixture increases combinatorially with n. In this article, we describe a tool we have developed that screens through the entire search space to identify a handful of distillation configurations that are attractive for an application. Towards our goal, we formulate a novel Mixed Integer Nonlinear Program (MINLP) using Underwood's method to estimate the vapour duty in each column. The MINLP formulation is integrated with DISTOPT: an easy-to-use in-house visualization software that takes feed properties as input, solves the optimization problem, and displays attractive configurations pictorially. The capabilities of the developed tool are illustrated with two case studies.

Keywords: Multicomponent distillation, column sequencing, MINLP

1. Introduction

Separations are ubiquitous in chemical, biochemical and petrochemical industries. Majority of the separations are energy intensive, and account for 40-70 % of the total plant energy requirement. Of the numerous unit operations, 90-95 % of the liquid separations are carried out using distillation (Humphrey and Keller, 1997). Even for gas separations, distillation is the only alternative for larger throughput and high purity products. According to one estimate more than 40,000 distillation columns are in operation in the US. These columns consume energy equivalent of 1.2 million bbl oil per day (Humphrey and Keller, 1997). Given the scale of operation, within given cost constraints, it is essential to make the processes as efficient as possible in order to reduce the overall energy consumption.

Majority of the industrial distillations involve the separation of three or more components. Such separations require a sequence of columns known as a configuration. The number of configurations available for a multicomponent separation increases combinatorially with increase in number of components in the process feed. For example, there are over half a million configurations available for a six-component system (Shah and Agrawal, 2010). All these configurations differ in the composition of distillate and residue from each split in a configuration and thermal qualities of internal transfer streams (referred to as submixtures). Consequently, the extent of remixing losses differ among different configurations, making some configurations more energy

efficient compared to others. This raises the following questions: (a) which configurations consume less energy for a given separation task? (b) under what operating conditions, will these configurations require the least amount of energy? We address these questions in this article. We use the total vapour duty; defined as the sum of vapour flows in all the reboilers, as a proxy for energy consumption.

Traditionally, owing to the size of the search space, the design of configurations is made using heuristics, trial-and-error and experience of the practitioner. Such practices often lead to suboptimal solutions resulting in high energy penalty (Shah and Agrawal, 2010). Even with sophisticated process simulators, like ASPEN PLUS, screening through the entire search space is not feasible in a reasonable amount of time. Besides, our experience shows that the rigorous simulators face convergence issues with thermally coupled and satellite-column configurations. Therefore, we develop an easy-to-use software that enables an industrial practitioner to identify a handful of attractive configurations after considering the entire search space. Towards our goal, we formulate and solve a novel Mixed Integer Nonlinear Program (MINLP). The solution of the MINLP gives the optimal configuration and its optimal operating conditions. As a first step, we focus only on non-azeotropic mixtures. Consequently, Underwood method (Underwood, 1948) can be used to estimate column vapour duties, providing a mechanism to screen the configurations. The screening process results in a handful of candidate configurations. Rigorous simulations can then be performed only on the candidate configurations to determine the configuration appropriate for an application. Finally, we integrate our formulation with DISTOPT, a software developed by our group, to aid visualization of distillation configurations.

2. Summary of MINLP formulation

In this section, we briefly describe the MINLP formulation. The first step in the problem formulation is the isolation of the search space of feasible configurations. We define binary variables indicating the presence/absence of streams and heat exchangers. Constraints on these variables are formulated on the lines of the method proposed by Shah and Agrawal (2010). The second step in the problem formulation is mass balance equations. We adapt the superstructure in Caballero and Grossmann (2004) by proposing some modifications. We observe that, by appropriately bypassing streams through pseudocolumns, we can embed the entire search space in the superstructure. We treat the problem as a multicommodity network flow problem, where each pseudocolumn is analogous to a node on which component mass flows are modelled. The rectifying and stripping sections of a pseudocolumn are analogous to edges that carry material from one node to another. When a pseudocolumn carrying out the split of a stream is absent, the component, vapour and liquid flows are bypassed (Caballero and Grossmann, 2004). On the other hand, when the pseudocolumn is present, the component flows are redistributed between the rectifying and stripping sections while satisfying Underwood constraints to minimize column vapour flow. Underwood's method is a shortcut to reasonably estimate the minimum vapour flow for a separation, while avoiding tray-by-tray equilibrium calculations. Unfortunately, the resulting problem is nonconvex, even for a fixed configuration, and can be challenging to solve to global optimality.

Nonconvex problems are typically solved by iteratively refining the search region using bounding schemes. Since simple factorable relaxations for the nonconvex problem do not produce reasonable bounds for this process to be effective, we improve these bounds

by adding valid cuts that improve the bound quality. For example, we exploit the monotonic nature of the Underwood equations. Though these constraints have not been used in formal MINLP models, they have been used to argue that Underwood roots 'flow' or 'carryover' from one column to another (Carlberg and Westerberg, 1989; Halvorsen and Skogestad, 2003). These constraints play a vital role during the range reduction process for thermally coupled columns. Other constraints are derived using Reformulation-Linearization Technique (RLT), a simple yet powerful convexification technique (Sherali and Adams, 1992). All these constraints aid the global solvers in navigating to global optimum by improving the bounds and linear relaxations. We remark that our constraints on binary variables are at least as tight as the intersection of the current formulations in the literature (Caballero and Grossmann, 2004; Giridhar and Agrawal, 2010; Shah and Agrawal, 2010). To measure the vapour duty of a configuration, we sum the vapour flows of all reboilers in the configuration. The resulting MINLP is solved to ε - global optimality (with 1% as relative tolerance for convergence) using BARON (Tawarmalani and Sahinidis, 2005).

3. DISTOPT: Configuration visualization tool

To aid the visualization and the interpretation of the results from MINLP, we have integrated our formulation with DISTOPT: a software developed by our group for the optimization of distillation configurations. The software takes the process feed composition and the relative volatility information as input, formulates and solves the MINLP using BARON, interprets the output from BARON, and displays the configurations pictorially. It also displays the optimal component, vapour and liquid flow rates in each stream. In addition to the optimal solution, the software can determine up to the 'K^{th}' best solution, where K is a user specified positive integer. The solutions are obtained by successively eliminating configurations as they are obtained using binary cuts and/or branch-and-bound. Furthermore, through the user-friendly interface, practitioners can customize the search space to only include those configurations with certain desired properties. Few of the options currently available include specification of maximum number of sloppy splits, thermal couplings, submixtures, thermal quality of the side-draw, etc. Other features such as enforcing the presence/absence of specific streams and/or heat exchangers, are being built into the software. Since MINLP problems can be computationally intensive, narrowing the search space by excluding the undesirable configurations could improve the computational speed. When no customization option is specified, the program optimizes over the entire search space. In short, the software enables identification of attractive configurations for a given application with just a click of a button!

4. Case studies

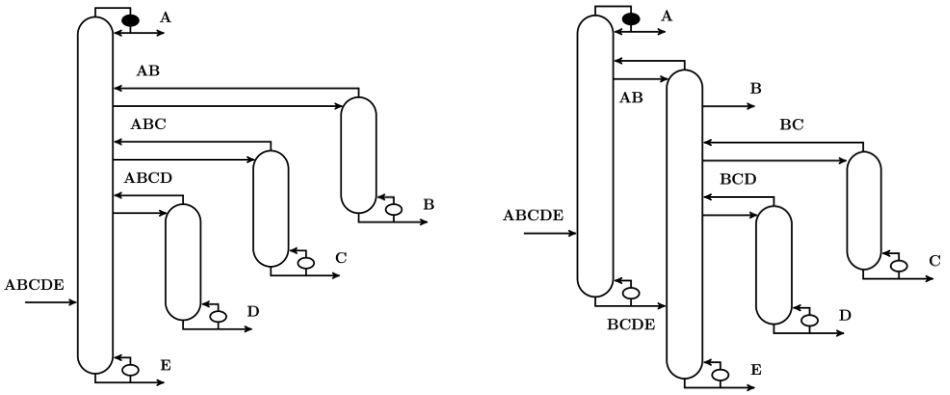
It has been hypothesized for a general n-component system that the Fully Thermally Coupled (FTC) configuration requires the least vapour duty (Fidkowski and Agrawal, 2001; Halvorsen and Skogestad, 2003). Nevertheless, with maximum number of column sections and transfer streams, FTC might be unattractive from the perspective of capital cost and operation. However, is it always necessary to construct FTC to achieve the maximum energy benefits/savings? The answer is no (Shah and Agrawal, 2010), and we will show via a case study that examples of this sort are easy to find with our tool.

4.1. Determination of attractive configurations: Mixture of Alcohols

Here, we consider the separation of an example mixture of alcohols from literature (Caballero and Grossmann, 2004): Ethanol (A), Isopropanol (B), 1-Propanol (C), Isobutanol (D) and 1-Butanol (E). The process feed flow is taken to be 200 kmol/h at saturated liquid state. The process feed consists of 20 % of A, 30 % of B, 20 % of C, 20 % of D and 10 % of E. The relative volatilities of the components w.r.t component E are 4.1, 3.6, 2.1 and 1.42, respectively. The data can be fed to the software, and the underlying algorithms can determine the optimal solution over the entire search space with default options. This often leads to configurations which have several submixtures. The presence of each submixture in a configuration increases the number of column sections, and thus the capital cost. Thus, we will seek a configuration that simultaneously minimizes the number of submixtures/column sections (measure of capital cost) and the vapour duty (measure of operating cost).

With minimum number of submixtures/column sections, sharp-split configurations are attractive in terms of capital cost. Moreover, sharp-split configurations have been preferred in industry due to their simplicity and ease of operation for decades. To determine the optimal sharp-split configuration and vapour duty, we tailor the search space by adding a constraint that enforces exactly n-2 submixtures in a configuration. This constraint characterizes sharp-split configurations. By specifying the maximum number of submixtures as n-2, the software automatically adds the constraint to the MINLP model before solving it. Figure 1(a) shows the optimal sharp split configuration for the separation of the mixture of alcohols. This configuration is known in the literature as the indirect split configuration. The vapour duty of the configuration scaled w.r.t the vapour duty of the FTC is 1.479 i.e., the best sharp-split configuration requires 47.9% more vapour duty compared to the FTC. Clearly, while potentially attractive in terms of capital cost, the optimal sharp split configuration is not attractive in terms of operating cost. Therefore, we will focus on the next best alternative: configurations with n-1 submixtures. The optimal solution for this case is shown in Figure 1(b).

The solution in Figure 1(b) is attractive w.r.t to both capital and operating costs for three reasons. First, the configuration requires only two additional sections, and thus the



(a) Vapour duty is 47.9 % higher than FTC (b) Vapour duty is the same as that of FTC

Figure 1: Candidate configurations for the separation of alcohols: (a) Optimal sharp-split configuration (b) Optimal configuration with one sloppy split

capital cost is not too far off compared to a sharp split configuration. Second, each column has a heat exchanger allowing for a better control of reflux, and makes the start-up and shutdown process easier. Finally, the vapour duty of the configuration is the same as that of FTC. We point out that energy savings of the order of 48% are estimated by just allowing one sloppy split.

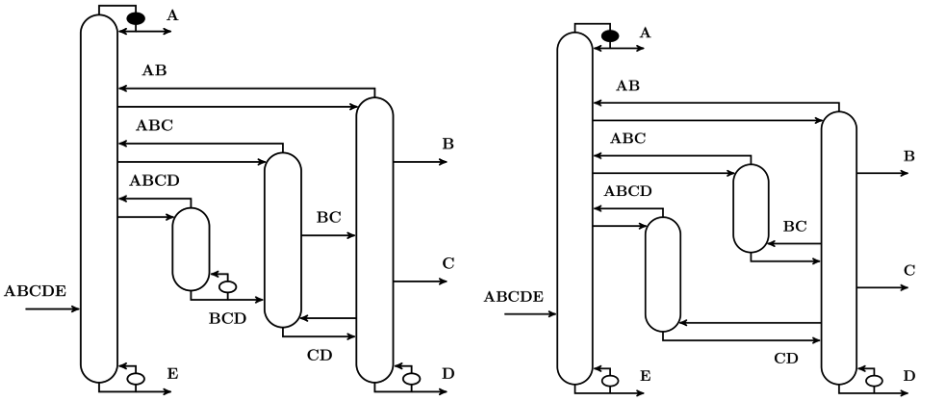


Figure 2: Configurations amenable for retrofitting of indirect split configuration, for light crude fractionation example. (a) and (b) require 19.4 % and 16 % less vapor duty compared to indirect split configuration

4.2. Retrofitting applications: Light crude distillation

The MINLP can also aid in identifying novel configurations that are amenable for retrofitting. We demonstrate this using the example involving separation of light crude (Shah and Agrawal, 2010). We treat light crude as a mixture of Naphtha (A), Kerosene (B), Diesel (C), Gas Oil (D) and Residue (E). The feed flow is taken to be 100 kmol/h, and it consists of 46.1 % A, 19.5 % B, 7.3 % C, 11.4 % D and 15.7 % E. The relative volatility of components w.r.t E is taken to be 45.3, 14.4, 4.7 and 2, respectively.

Ignoring details such as liquid pump-around, the indirect split configuration shown in Figure 1(a) is used for crude fractionation currently. We highlight that our formulation also identifies this configuration as the optimal sharp-split configuration. This shows that the assumptions and the simplifications made in the problem formulation are justifiable. As shown in the previous example, sharp-split configurations, in general, require significantly more energy. Clearly, there is a scope and need for better configurations. Since ab initio redesign of an existing plant is unlikely to be economical, practitioners would be interested in configurations which are amenable for retrofitting. We tailor the search space with such configurations by adding structural constraints that are desired in a configuration.

For issues associated with fouling, residue is preferred to be separated in the first column. Hence, we enforce that all the submixtures which contain residue viz., BCDE, CDE and DE must be absent. In addition, if any other transfer streams are desired, they can be enforced to be present. The optimal and the second best configurations are shown in Figure 2. These configurations require 19.4 % and 16 % less vapour duty compared to the indirect split configuration. By comparing with Figure 1(a), these configurations look attractive for retrofitting.

5. Conclusion

The objective of this work is to briefly describe an easy-to-use tool for the optimization of multicomponent distillation configurations that we have recently developed. The tool formulates and solves a novel MINLP, with minimization of vapour duty as the objective function. By appending appropriate cost or exergy models, the MINLP can be extended to minimize the total annualized cost or exergy loss. The MINLP has been integrated with our easy-to-use software, DISTOPT, to aid visualization and selection of the desired configurations to consider during optimization. Finally, different features of the software are illustrated through two five-component examples. This tool enables the industrial practitioners to systematically design distillation configurations for an application, and eliminates the dependence on the heuristics and trial-and-error methods.

Acknowledgements

We express our gratitude to the U.S. Department of Energy, award number: DE-EE0005768, for the financial support. T.G.R. would like to thank Tony Mathew for his help in the integration of the MINLP with DISTOPT, and Zheyu Jiang for valuable discussions.

References

J. A. Caballero and I. E. Grossmann, 2004, Design of distillation sequences: from conventional to fully thermally coupled distillation systems, Computers & Chemical Engineering, 28(11), pp.2307-2329.

N. A. Carlberg and A. W. Westerberg, 1989, Temperature-heat diagrams for complex columns. 2. Underwood's method for side strippers and enrichers, Industrial & Engineering Chemistry Research, 28(9), pp.1379-1386.

Z.T. Fidkowski and R. Agrawal, 2001. Multicomponent thermally coupled systems of distillation columns at minimum reflux, AIChE journal, 47(12), pp.2713-2724.

A. Giridhar and R. Agrawal, 2010, Synthesis of distillation configurations. II: A search formulation for basic configurations, Computers & chemical engineering, 34(1), pp.84-95.

I. J. Halvorsen and S. Skogestad, 2003, Minimum energy consumption in multicomponent distillation. 3. More than three products and generalized Petlyuk arrangements, Industrial & Engineering Chemistry Research, 42(3), pp.616-629.

J. L. Humphrey and I. I. Keller, 1997, Separation Process Technology, McGraw-Hill.

V. H. Shah and R. Agrawal, 2010, A matrix method for multicomponent distillation sequences, AIChE journal, 56(7), pp.1759-1775.

H. D. Sherali and A. Alameddine, 1992. A new reformulation-linearization technique for bilinear programming problems, Journal of Global optimization, 2(4), pp.379-410.

M. Tawarmalani and N. V. Sahinidis, 2005, A polyhedral branch-and-cut approach to global optimization, Mathematical Programming, 103(2), pp.225-249.

A. J. V. Underwood, 1948, Fractional distillation of multi-component mixtures, Chemical Engineering Progress, 44, pp.603-614.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50033-1

A Hybrid Approach for Process Optimization of Distillation Reflux Condition using First Principle Models and Least Squares Regression

Tomoyuki Taguchi[a*], Yoshiyuki Yamashita[b]

[a] *Chiyoda Corporation, Advanced Process Engineering Unit, 4-6-2, Minatomirai, Nishi-ku, Yokohama-shi, Kanagawa 220-8765, Japan*
[b] *Department of Chemical Engineering, Tokyo University of Agriculture and Technology, 2-24-16 Naka-cho, Koganei-shi, Tokyo 184-8588, Japan*
taguchi.tomoyuki@chiyodacorp.com

Abstract

Distillation columns are conventionally controlled at fixed reflux ratio to maintain the quality of the overhead product. If the cooling temperature of condenser becomes cold, the reboiler heat increases to cope with the internal reflux flow. Also, the loss of product recovery happens at the bottom of the column during the control delay of maintaining the bottom temperature. The reflux ratio is adequately adjusted by based on a proposed hybrid model analysis using the fist principle model (FPM) and least squares regression. According to the sensitivity case analysis with FPM on benzene / toluene binary distillation column, the reflux rate is substantially estimated by the multivariate least squares regression using the following explanatory variables; 1) feed flow rate, 2) column temperature in certain of stages, 3) column pressure, 4) sub-cooling temperature of condenser.

To confirm the effectiveness of the hybrid approach, the dynamic simulation study cases of benzene / toluene binary distillation column are carried out. The simulation results show the effectiveness of the hybrid approach to reduce the energy consumption of reboiler heat duty and loss prevention of product.

Keywords: Distillation operation optimization, Hybrid model analysis, Least squares regression, Steady state and dynamic process simulation, Column reflux optimal control, Low calculation cost, Model reduction.

1. Introduction

The industrial importance of distillation column is widely recognized. As the high purity products can be recovered, it still occupies a dominant part of the process industry. On the other hand, it requires the large amount of heat consumption so that the internal vapor flow is generated to operate the separation utilizing the vapor-liquid equilibrium condition. With respect to the composition and temperature control of distillation column, the control of reflux flow rate becomes slower than that of reboiler boiling up. Since the liquid mass holdup existing in the column is much higher than the vapor one, temperature change in liquid holdup is delayed due to the large amount of liquid mass and mass transfer in column internal. Boiling up through reboiler is often used to maintain the required condition of column operation according to the readiness response (Shinskey, 1984). If the high purity products are recovered from column overhead, the reflux control is not expected to maintain the specification of the overhead. Therefore, the reflux control

is conventionally fixed with constant flow rate including the margin to compensate the uncertainty of dynamics characteristic. The control response, the unexpected process disturbance or the equipment performance change are the representative dynamic uncertainty.

To adapt the reflux rate with the minimum required condition to maintain the product specification of the overhead, RTO-MPC (real time optimization – model predictive control) software system is widely utilized. This is the combination approach of the steady state optimization of RTO and the transmission of optimized set point according to MPC. Darby *et. al.* (2007) describes the RTO-MPC basic methodology and its performance assessment regarding the unsuccessful supporting and maintaining, complication of modelling, and calculation waiting time of steady state model. Also, the Grey-box model including both physical (first principle) knowledge and empirically determined constitutive equations are mentioned.

A presented hybrid approach proposes the solution as the alternative methodology of RTO-MPC using the model reduction technique.

2. Methodology - Hybrid model analysis

The proposed hybrid model analysis includes the following three steps;

1) To obtain *a priori* knowledge of optimized target value within certain constraints, the sensitivity study using the steady state simulation model is carried out (refer to section 2.2).

2) The multivariable inputs-outputs relationship extracted from the sensitivity study results of the steady state model is statistically transformed to the reduced variable model (refer to section 2.3).

3) The statistical reduced variable model is applied to the optimized set point estimator (refer to section 3.2).

The hybrid model analysis is a gray-box approach which enables to combine *a priori* knowledge regarding the system characteristic with various operation condition change and the fast calculation environment due to the reduced variable model.

2.1. The steady state simulation model

As an example distillation model, the simple binary system of benzene and toluene is considered. The basic condition of this example is shown as follows;

Feed throughput:	2,000kmol/h
Column diameter:	4.0m
Number of theoretical plate:	30

Condenser type is assumed as air fin cooler. The condenser temperature is varied as the intent of day-change. The representative process variable such as flow rate, pressure and temperature is shown in Fig.1. Benzene product recovery of 99.8% and toluene product recovery of 99.0% are assumed at the overhead and the bottom of distillation column, respectively.

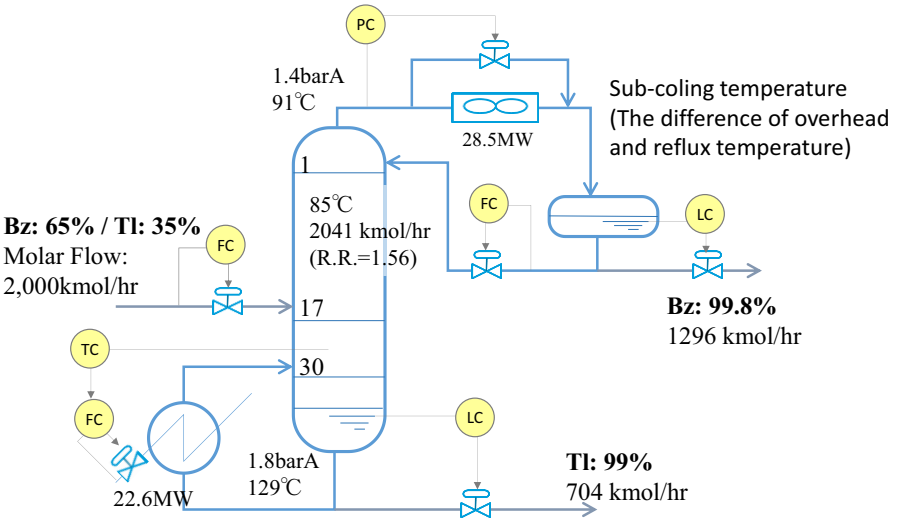


Fig.1 Example distillation column model and representative process variables

2.2. Sensitivity study cases

Sensitivity of the following process variables are investigated to confirm the required reflux rate; (1) feed compositions, (2) feed flow rate, (3) overhead pressure, (4) sub-cooling temperature according to condenser temperature change. From the sensitivity study, inputs-outputs relationship shown in Table 2 is obtained.

Table 1 Inputs-outputs relationship achieved from sensitivity study

No. of Case Study	Feed Flow Rate	Feed Comp.	Over-head Pressure	Sub-cooling Temp.	Column Temperature in Each Stage				Reflux Flow Rate	Reboiler Boiling Up Flow Rate
Run 1	F_1	x_1	$P_{1,1}$	ΔT_1	$T_{1,1}$	$T_{2,1}$	$\cdots$	$T_{n,1}$	R_1	V_1
Run 2	F_2	x_2	$P_{1,2}$	ΔT_2	$T_{1,2}$	$T_{2,2}$	$\cdots$	$T_{n,2}$	R_2	V_2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Run i	F_i	x_i	$P_{1,i}$	ΔT_i	$T_{1,i}$	$T_{2,i}$	$\cdots$	$T_{n,i}$	R_i	V_i

Variables as input of sensitivity study

Variables as output of sensitivity study and observable value in real operation

Variables as output of sensitivity study and objectives of control

To consider the on-line measurement problem, feed composition is difficult to select as the explanatory variable of statistical estimator. It is supposed that the column internal temperature has an inferential characteristic with respect to the feed composition change. Feed benzene compositions are changed from 0.5 to 0.75, and feed flow rates are changed from 1,400 kmol/h to 2,400 kmol/h. Feed compositions and flow rate are varied with 15 data points. Overhead pressure is changed from 1.0 to 2.0 bar with 5 data points.

Subcooling temperature is changed from 0 to 20 ℃ with 6 data points. Totally, 450 data points are provided for the expression of operation design space.

2.3. *Statistical model to predict the optimum reflux rate*

The relationship of multivariable inputs-outputs relationship is statistically expressed by the least squares. As the explanatory variable are **X** and objective variables are **Y**, the following equation (1) is the least squares basic form.

$$\mathbf{Y} = \mathbf{X\theta} \quad (1)$$

Where, $\boldsymbol{\theta}$ is the parameter of determining the relationship of **X** and **Y**, and $\boldsymbol{\theta}$ is calculated as follows;

$$\boldsymbol{\theta} = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{Y} \quad (2)$$

According to the least squares regression, the required reflux flow rate can be estimated well by the four types of variables mentioned in section 2.2. Estimated results are shown in Fig.2.

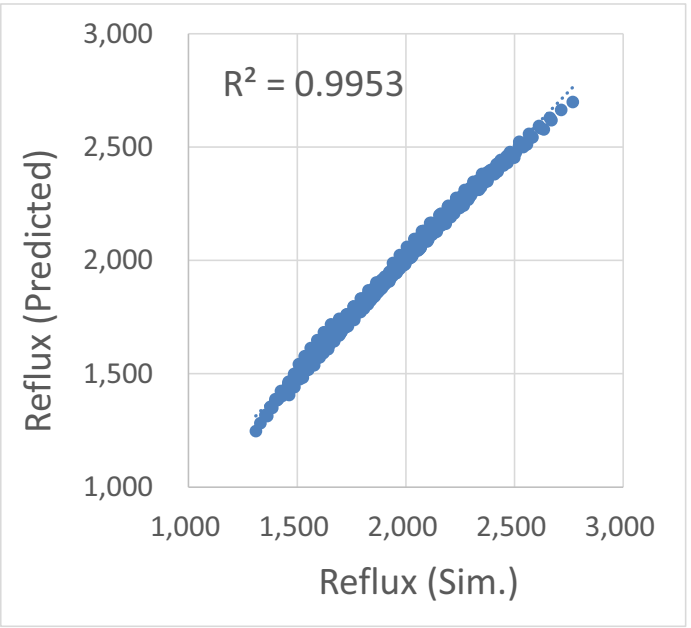


Fig. 2 Estimated results of required reflux flow rate by least squares regression

3. Verification of dynamic simulation model

The effectiveness of hybrid model estimator for required reflux flow is verified in the dynamic simulation model. The design information is achieved by the steady state model. Basic control configuration is shown in Fig.1 (refer to section 2.1).

3.1. Connection of the statistical model to process controller

The hybrid model estimator and the set point of reflux flow rate are combined without consideration of dynamics effect. An amount of margin is incorporated in that connection of statistical model to process controller shown as equation (3).

$$F_R = \hat{Y}_R + m \quad (3)$$

Where, F_R is the set point of reflux flow,
$\hat{Y}_R$ is the estimated value of required reflux flow, and m is the margin to compensate the error occurred by dynamic effect especially caused by the delay of column internal liquid distribution and reboiler heat up by PID control.

3.2. Case Study & Simulation results

The ambient temperature change shown in Fig. 3 is applied to the dynamic simulation model.

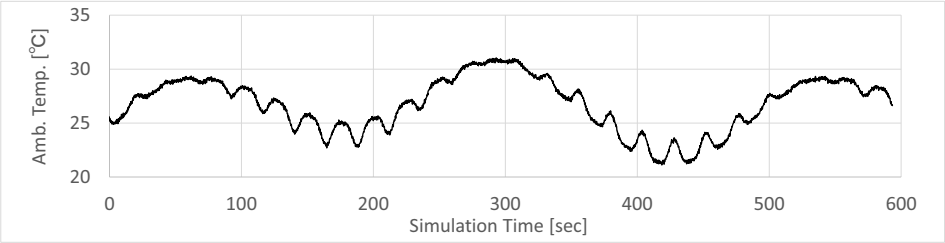


Fig.3 The ambient temperature change applied to the condenser inlet

The cycle of simulated temperature change is adjusted to the 240 seconds instead of the 24 hours cycle of day change (Fig.3). The reason of setting the shorten cycle of temperature change is as follows;

1) To confirm the dynamic effect in the severer condition.
2) Reduce calculation cost of dynamic simulation.

Two study cases are verified;

Type-A: Fixed reflux flow rate

Type-B: Variable reflux flow rate with fixed benzene composition

The simulation results for the comparison of type-A and type-B are shown in Fig. 4, Fig. 5, Fig. 6, Fig. 7.

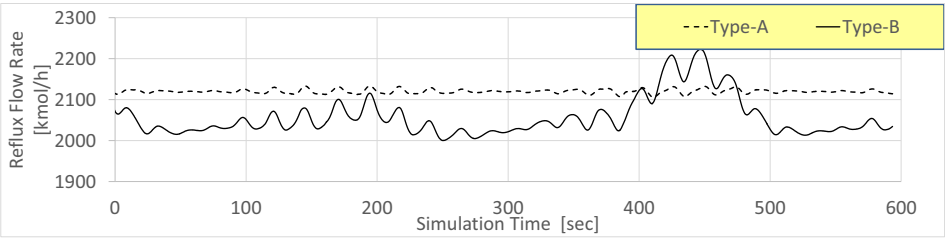


Fig. 4 Dynamic simulation results of the reflux flow rate

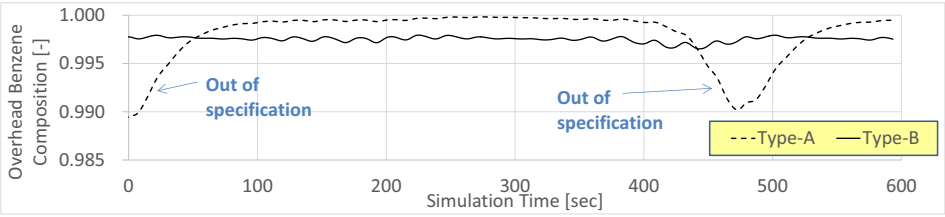


Fig. 5 Dynamic simulation results of the overhead benzene composition

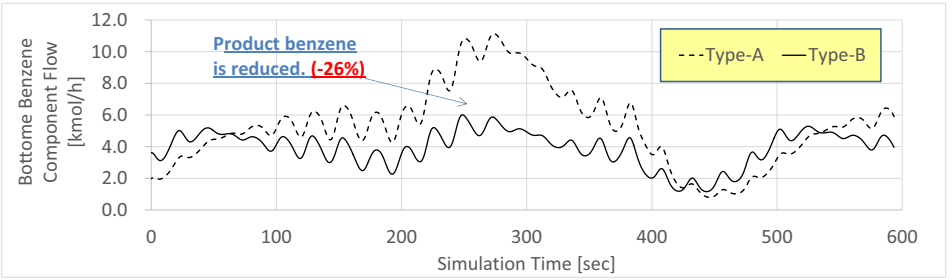


Fig. 6 Dynamic simulation results of the bottom benzene component flow rate

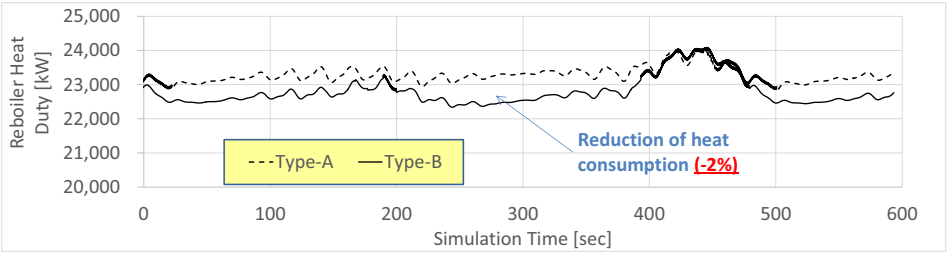


Fig. 7 Dynamic simulation results of reboiler heat duty

The reflux rate (type-A) is stable, on the other hand, the reflux rate (type-B) is changed (Fig. 4). According to the adjustment of the reflux rate (type-B), the overhead benzene composition is maintained to the specified composition. In the results of type-A, duration of the out of specification is found (Fig. 5). Since the reflux rate of type-B is maintained with lower amount than that of type-A, the benzene slip amount at the column bottom is reduced by 26% during the simulation verification period (Fig. 6). Also, the reboiler heat duty in type-B is reduced by 2% from that in type-A (Fig. 7).

4. Conclusions

Hybrid model analysis of the steady state simulation and the regression model is proposed. *A priori* knowledge such as the system changes in operational process variables can be extracted by the sensitivity study conducting in the steady state process simulator. An illustrative case study of a distillation column with process control is successfully demonstrated. Application of the hybrid model approach to the more complicated process systems will be the future work. According to the increase of process model complexity, the problem regarding to overfitting and robustness of regression model will be expected to emerge. Future work will not only focus the regression model building but its appropriate variable selection method and robustness.

References

M. L. Darby, M. Nikolaou, J. Jones, D. Nicholson (2011), RTO: An overview and assessment of current practice, Journal of Process Control, 21, 874–884

F. G. Shinskey (1977), Distillation control: for productivity and energy conservation, McGraw-Hill

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50034-3

GC-COSMO based Reaction Solvent Design with New Kinetic Model using CAMD

Qilei Liu[a], Lei Zhang[a*], Linlin Liu[a], Jian Du[a], Xinyuan Liang[a], Haitao Mao[a], Qingwei Meng[b]

[a]*Institute of Chemical Process Systems Engineering, School of Chemical Engineering, Dalian University of Technology, Dalian 116024, China*
[b]*State Key Laboratory of Fine Chemicals, School of Pharmaceutical Science and Technology, Dalian University of Technology, Dalian 116024, China*
379785365@qq.com

Abstract

Reaction solvents have been playing an important role in liquid-phase organic synthesis. To identify the optimal solvents for accelerating reaction rates, a GC-COSMO based computer-aided molecular design model is formulated, in which a new reaction kinetic model is proposed with the help of parameter reformulation strategy based on the conventional transition state theory. The high-performance solvent with the maximum reaction rate constant is generated using the proposed design model. This solvent design method can be applied to various reaction types. The validity of our method is demonstrated for the selected *Menschutkin* reaction.

Keywords: solvent design, computer-aided molecular design, kinetic reaction, solvation effect, group contribution.

1. Introduction

With the development of process industry, solvents are widely used in physicochemical processes. When involved in liquid-phase organic synthesis, solvents can have significant impacts on rates and selectivity to facilitate the existing synthetic routes or even develop new routes (Struebing et al., 2013). Thus, the choice of reaction solvents is particularly important to product yield and economic benefit. However, it is impossible to perform thousands of kinetic experiments for screening reaction solvents. Therefore, efficient methods should be developed to help the solvent selection.

In the last few years, Computer-Aided Molecular Design (CAMD) technique has been successfully adopted for the optimal design of solvents. Although it has contributed a lot to solvent design in the separation processes, there are few contributions in the reactions comparatively. Gani et al. (2008) proposed a systematic method for the selection of solvents, which shows great applicability for a wide range of reaction systems. However, their method requires abundant priori knowledge about the reaction systems for the construction of solvent score table and can only give qualitative predictions on the solvent choice. For quantitative predictions, the pure Quantum Chemical (QC) calculations (e.g., ab initio or DFT computations) can be employed to make the reaction rate constants prediction available without any experimental data. But it is time-consuming and the obtained results are usually inaccurate.

Another popular way to study the reaction rates is the Quantitative Structure-Property Relationships (QSPR) method, which is often represented as a multiple linear

relationship between experimental rate constants and a set of solvent descriptors. Folić et al. (2008) have employed a solvatochromic equation to correlate the rate constants with solvent properties. Although this method can generally provide satisfactory predictions with some prespecified solvents, it has poor abilities in extensive solvent screening as the solvent properties are extremely dependent on considerable experiments. Zhou et al. (2015) have introduced new solvent descriptors based on the COSMO-RS (COnductor like Screening MOdel for Real Solvents) model. The reaction kinetic model avoids large numbers of experiments for the solvent properties, and it shows excellent prediction results in the Diels–Alder reaction. But this approach has little supporting theoretical basis. Most importantly, it may only be adapted to a certain reaction type, which means it lacks extensive adaptability in other kinetic reactions.

In the second part of this paper, a new reaction kinetic model is proposed based on Conventional Transition State Theory (CTST). Then, a CAMD model is developed for the reaction solvent design, in which the COSMO-SAC (Segment Activity Coefficient) model is employed for the calculation of infinite dilution activity coefficients, Group Contribution (GC) method is integrated for the prediction of other solvent properties. Next, an MINLP problem is formulated, and followed by a decomposition-based solution algorithm to reduce the difficulty in dealing with the nonlinear COSMO-SAC equations. In the third part, the validity of our method is demonstrated by a selected *Menschutkin* reaction in the case study.

2. The CAMD model for the reaction solvent design

CAMD uses mathematical modelling to determine the molecule or molecular structure that matches the predefined target properties with a given set of building blocks. The development of CAMD methods started in 1980s, and nowadays have been successfully applied in various type of products, such as solvents, refrigerants and polymers. Here, a CAMD based framework is developed for the reaction solvent design, which consists of three sub-models as shown in Figure 1. In the reaction model, the rate constant (Eq.(4)) is derived with the help of parameter reformulation strategy and the knowledge based descriptor selection.

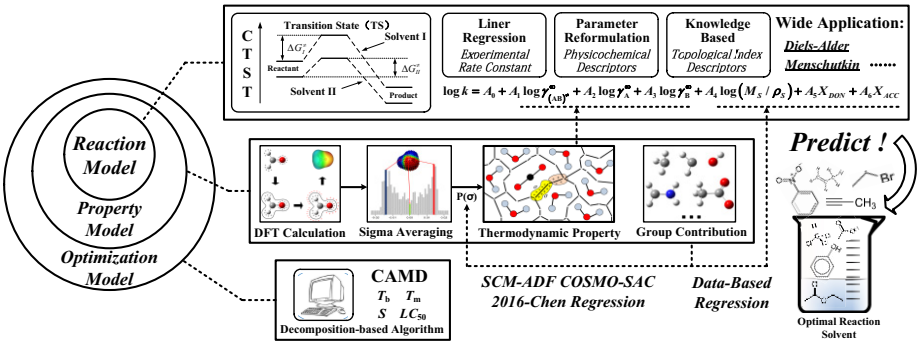


Figure 1. Reaction solvent design framework

In the property model, the key descriptors, infinite dilution activity coefficients, are calculated inexpensively based on the state-of-the-art COSMO-SAC model (Chen et al., 2016) in SCM-ADF (Velde et al., 2001), which guarantees the accuracy compared to the previous COSMO versions. Other descriptors are available from the database. Then, a GC method has been developed to associate molecular structure with theoretical

solvent descriptors. In the optimization model, the CAMD problem is formulated as an MINLP model and then solved by a decomposition-based approach. Finally, the optimal reaction solvent with the highest rate constant is identified.

2.1. Product attributes identification

Nowadays, kinetic controlled reactions have played important roles in organic synthesis. Reaction rate constant, which has been regarded as the objective product attribute, is one of the most crucial variables to control the reaction rate under the conditions of constant reactant concentration, temperature and pressure. Besides, to ensure normal operations and environmental sustainability, other solvent attributes, such as melting point, boiling point, solubility and LC_{50}, are also taken into account in the CAMD solvent design.

2.2. Reaction model

Solvents are widely used as reaction media in liquid-phase synthesis. The variation of solvents can dramatically change the reaction rate and mechanism. This phenomenon is explained as solvation effect, which can be quantitatively assessed through the solvation free energy. Here, a strategy of parameter reformulation and the knowledge based descriptor selection are employed to develop the kinetic equation. Firstly, the reaction rate constant (Eq.(1)) has been derived based on CTST. Since the solvation free energy can be expressed as Eq.(2) in the COSMO model, Eq.(3) is subsequently formulated.

$$k = \kappa \frac{k_B T}{h} \prod_i \left(c_i^{\theta,L}\right)^{\nu_i} \prod_i \left(q_i'^{IG}\right)^{\nu_i} \exp\left(-\frac{\Delta^{\neq} \underline{E}_i^{el,IG}}{RT}\right) \cdot \exp\left(-\frac{\Delta^{\neq} \Delta \underline{G}_i^{p\to c,IG} + \Delta^{\neq} \Delta \underline{G}_i^{\theta,solv}}{RT}\right) \tag{1}$$

$$\Delta \underline{G}_i^{\theta,solv} = RT \ln \gamma_i^{\infty} + RT \ln P_i^S + RT \ln \frac{M_S}{\rho_S V_{gas}} \tag{2}$$

$$\begin{aligned} k = \kappa \frac{k_B T}{h} \prod_i \left(c_i^{\theta,L}\right)^{\nu_i} \prod_i \left(q_i'^{IG}\right)^{\nu_i} \exp\left(-\frac{\Delta^{\neq} \underline{E}_i^{el,IG}}{RT}\right) \cdot \exp\left(-\frac{\Delta^{\neq} \Delta \underline{G}_i^{p\to c,IG}}{RT}\right) \\ \cdot \exp\left(\Delta^{\neq} \ln \gamma_i^{\infty} + \Delta^{\neq} \ln P_i^S + \Delta^{\neq} \ln \frac{M_S}{\rho_S V_{gas}}\right) \qquad \Delta^{\neq} = (AB)^{\neq} - A - B \end{aligned} \tag{3}$$

$$\log k = A_0 + A_1 \log \gamma_{(\mathrm{AB})^{\neq}}^{\infty} + A_2 \log \gamma_{\mathrm{A}}^{\infty} + A_3 \log \gamma_{\mathrm{B}}^{\infty} + A_4 \log\left(M_S / \rho_S\right) + A_5 X_{DON} + A_6 X_{ACC} \tag{4}$$

The assumptions of constant temperature T and transmission coefficient κ is proposed for model simplification. Then, it's easy to find out that most of other parameters, such as Planck constant h, Boltzmann constant k_B and gas molar volume V_{gas} are constants, while the standard state concentration $c_i^{\theta,L}$, partition function q'^{IG}, electronic energy $\underline{E}_i^{el,IG}$, Gibbs free energy change $\Delta \underline{G}_i^{p\to c,IG}$ and saturated vapor pressure P_i^S are all solute related, which means that the rate constant is only correlated with infinite dilution activity coefficient γ_i^{∞}, solvent relative molecular mass M_S and density ρ_S. On the basis of the analysis above, the solvent-independent parameters can be regarded as solute-determined constants. In this way, a multiple linear expression can be obtained through logarithmic treatment on the both sides of Eq.(3). In addition, as hydrogen-bond interaction has great impacts on the kinetic reaction rates, hydrogen bond

donor/acceptor are also chosen as topological index descriptors based on knowledge, and Eq.(4) is obtained for the reaction solvent design.

2.3. Property model

Property models are need for the estimation of solvent descriptors in Eq.(4), among which the infinite dilution activity coefficients between solutes (reactants and transition state) and solvents can be calculated inexpensively by sigma-profiles and COSMO volumes using COSMO-SAC model. Unlike the UNIFAC model, COSMO-SAC is a continuum solvation model used in QC calculation. It doesn't require any experiments for regression of binary interaction parameters and can be easily applied to the transition state. Other descriptors including solvent relative molecular mass, density and hydrogen bond donor/acceptor are obtained from the database. Later, a GC method is developed to associate functional groups with solvent descriptors, which are regressed for 75 groups with a sample of 151 solvents.

2.4. Optimization model

Integrated with reaction and property model, the CAMD problem can be formulated as an MINLP model. It includes an objective function, molecular structural constraints and the property constraints. Reaction rate constant is selected as the objective function to be maximized. The molecular structural constraints restrict the combination of numerous functional groups to develop a feasible chemical structure. Property constraints are set to ensure the application of the solvent and low environmental impact. More details can be found in Zhang et al. (2015). The boiling point T_b, melting point T_m, solubility S and LC_{50} are obtained using GC method (Marrero and Gani, 2001).

2.5. Solution algorithm

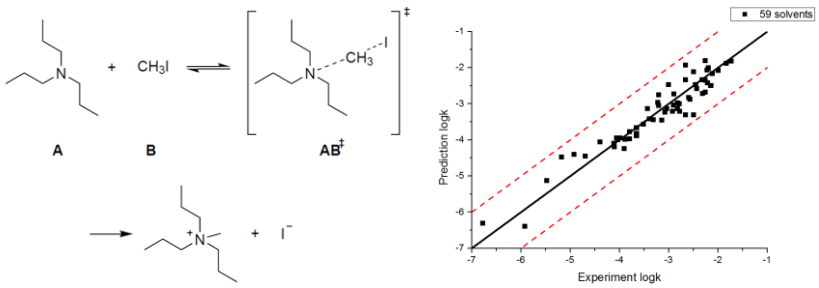
The GC-COSMO based CAMD problem is finally formulated as an MINLP model. Although the discrete choices and conditional constraints in the COSMO-SAC model can be reformulated using GDP and Big-M method, the large number of nonlinear implicit equations can hardly be solved by BARON solver. Thus, a decomposition-based algorithm (Karunanithi et al., 2005) is employed. In this algorithm, the MINLP model is decomposed into an ordered set of sub-problems, in which linear constraints and equations are first considered to generate a certain number (N) of feasible solutions. Then, nonlinear parts are iteratively performed with the calculation of reaction rate constants. Finally, the candidate molecules can be sequentially obtained in the order of objective function maximization, and the solution of this method is equivalent to the original problem as long as N is large enough.

2.6. Verification

In this step, the GC-estimated descriptors like activity coefficients and other solvent properties can be verified from the Density Functional Theory (DFT) and database search separately. As for reaction rate constants, rigorous models or kinetic experiments need to be performed for further demonstration.

3. Case study

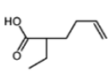
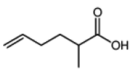
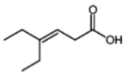
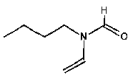
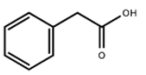
A case study of *Menschutkin* reaction for reaction solvent design is presented. The mechanism of *Menschutkin* reaction is shown in Figure 2 (a). To demonstrate the reaction model, a sample of 59 solvents with experimental rate constants is employed for linear regression. We have applied Zhou et al. (2015)'s model to the *Menschutkin* (R^2/0.291, MAPE/20.29 %), our result is excellent with R^2/0.905 and MAPE/8.07 %. Another case study (Diels-Alder reaction) is also performed in our work with good results (R^2/0.918, MAPE/2.80 %), which indicates that the proposed reaction model is more reliable in different types of kinetic controlled organic synthesis and conducive to the design of optimal reaction solvents, including ionic liquids. The predicted rate constants are shown in Figure 2 (b). As all the other researchers have adopted the expression log$k$ in their results, we thus choose log$k$ rather than k for a better comparison. Besides, the current existing studies are still difficult to predict k in high accuracy. Further corrections are needed in the CTST and COSMO model.



(a) The mechanism of *Menschutkin* reaction (b) The predicted logk vs experimental logk

Figure 2. The mechanism and regression results of *Menschutkin* reaction

Table 1. The candidate molecules of the case study

Formula	$C_8H_{14}O_2$	$C_7H_{12}O_2$	$C_8H_{14}O_2$	$C_7H_{13}NO$	$C_8H_8O_2$
Groups	1 CH_3 3 CH_2 1 CH 1 CH_2=CH 1 COOH	1 CH_3 2 CH_2 1 CH 1 CH_2=CH 1 COOH	2 CH_3 3 CH_2 1 CH=C 1 COOH	1 CH_3 3 CH_2 1 CH_2=CH 1 HCON	5 ACH 1 $ACCH_2$ 1 COOH
Molecular structure					
log$k$	0.0408	-0.2382	-0.3114	-0.4468	-0.6672

Based on the established reaction model, the MINLP model is built and decomposition-based algorithm is adopted to obtain the solution. 1525 feasible solutions are first generated by using the property constraints of T_b, T_m, S and LC_{50} as these properties are linear equations. Then, all the generated solutions are tested by the COSMO-SAC model and reaction model using GC method. Five best solvents are selected in Table 1. It should be noted that the candidate $C_7H_{13}NO$ has similar groups with N,N-dimethylformamide, which is already proved to be a great reaction solvent (Lassau and Jungers, 1968). Based on the analysis of the optimal results, molecules containing

alkenyl group with carboxyl or acylamino may have good performance, which can be verified in further experiments.

4. Conclusions

In this paper, a GC-COSMO based CAMD model for the design of reaction solvent is developed, in which a new reaction model is proposed. Then the MINLP model is established and solved using the decomposition-based algorithm. A case study is presented with the excellent results of R^2/0.905 and MAPE/8.07 % in the reaction model, which demonstrates the effectiveness of our reaction solvent design model.

Acknowledgment: The authors would like to thank the financial support of the Fundamental Research Funds for the Central Universities (DUT17RC(3)008), NSFC (21576036) and NSFC (21776035).

References

W. L. Chen, C. M. Hsieh, L. Yang, C. C. Hsu, & S. T. Lin, 2016, A Critical Evaluation on the Performance of COSMO-SAC Models for Vapor-Liquid and Liquid-Liquid Equilibrium Predictions based on Different Quantum Chemical Calculations, Industrial & Engineering Chemistry Research, 55, 9312-9322.

M. Folić, C. S. Adjiman, & E. N. Pistikopoulos, 2008, Computer-Aided Solvent Design for Reactions: Maximizing Product Formation, Industrial & Engineering Chemistry Research, 47, 15, 5190-5202.

R. Gani, P. A. Gómez, M. Folić, C. Jiménez-González, & D. J. C. Constable, 2008, Solvents in organic synthesis: Replacement and multi-step reaction systems, Computers & Chemical Engineering, 32, 10, 2420-2444.

A. T. Karunanithi, L. E. K. Achenie, & R. Gani, 2005, A New Decomposition-Based Computer-Aided Molecular-Mixture Design Methodology for the Design of Optimal Solvents and Solvent Mixtures, Industrial & Engineering Chemistry Research, 44, 13, 4785-4797.

C. Lassau, & J. Jungers, 1968, L'Influence du Solvant sur la Réaction Chimique. La Quaternation des Amines Tertiaires par l'Iodure de Méthyle, Bull. Soc. Chim. Fr, 7, 2678-2685.

J. Marrero, & R. Gani, 2001, Group-Contribution Based Estimation of Pure Component Properties, Fluid Phase Equilibria, 1, 183-208.

H. Struebing, Z. Ganase, P. G. Karamertzanis, E. Siougkrou, P. Haycock, P. M. Piccione, A. Armstrong, A. Galindo & C. S. Adjiman, 2013, Computer-aided molecular design of solvents for accelerated reaction kinetics, Nature Chemistry, 5, 11, 952-957.

G. te Velde, F.M. Bickelhaupt, E.J. Baerends, C. Fonseca Guerra, S.J.A. van Gisbergen, J.G. Snijders, T. Ziegler, 2001, Journal of Computational Chemistry, 22, 9, 931-967.

T. Zhou, K. Mcbride, X. Zhang, Z. Qi, & K. Sundmacher, 2015, Integrated solvent and process design exemplified for a Diels–Alder reaction, AIChE Journal, 61, 1, 147-158.

L. Zhang, S. Cignitti, & R. Gani, 2015, Generic Mathematical Programming Formulation and Solution for Computer-Aided molecular Design, Computers & Chemical Engineering, 78, 79-84.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50035-5

Optimal sizing of selective catalytic reduction reactor considering the position of urea injector

Sanha Lim, Yeonsoo Kim, Taekyoon Park and Jong Min Lee*

School of Chemical and Biological Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea
limmtrv@snu.ac.kr

Abstract

SCR, the selective catalytic reduction using reduction agent, is essential in diesel vehicles in order to meet the tightening NOx emission regulations. This study proposes a first-principle model for the SCR with the urea injector with a consideration of the effect of catalyst deactivation. In order to determine an optimal reactor sizing, sensitivity analysis is used. The optimal design performance is analysed with the following factors: the NOx emission, the amount of ammonia slip, and the capital cost. The optimal length is found to be 25 ~ 30 cm, under the premise that the total length is restricted to 40 cm.

Keywords: Urea-SCR, NOx emission, ammonia slip, cost, optimal sizing

1. Introduction

With global environmental issues emerging, regulations on diesel automobile NOx emissions are also being tightened. In particular, the Euro 6 emissions standards, which began in 2014, are far more stringent than those of the previous Euro 5. In order to meet the Euro 6 standards, an exhaust aftertreatment system must be added, which may incur additional costs. As a result, the future of environmentally friendly vehicles faces the challenge of achieving high fuel efficiency and low pollutant emissions at a reasonable cost.

Continuous efforts have been made towards reducing NOx from diesel automobiles, and one of the exemplary technologies is Urea-SCR (Selective Catalytic Reduction). In the Urea-SCR, injected urea is decomposed by the high temperature exhaust gas. The generated ammonia is then used as a reducing agent of NOx, converting NOx into nitrogen gas. Copper-Zeolite is mainly used as the SCR catalyst, and the kinetic model is available from previous studies (L. Olsson et al. 2008).

Depending on the location of the urea injector, the amount of produced ammonia can vary. This in turn affects the NOx conversion in the SCR. Considering the limited available space for aftertreatment system in a vehicle, the location of urea injector and the length of SCR need to be carefully designed. For example, as the length of SCR increases, the NOx conversion increases and the amount of ammonia slip decreases. On the other hand, if the distance between the urea injector and SCR is too short, the urea may not go through enough thermal decomposition.

2. Urea-Selective Catalytic Reduction System

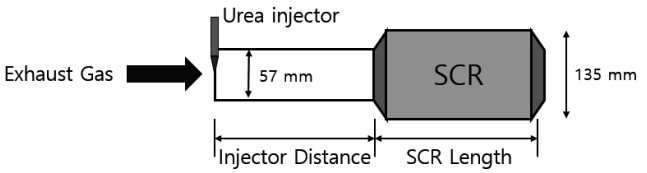


Figure 1. Configuration of SCR with urea injector (not in scale)

The configuration of urea-SCR system is shown in Figure 1. The diameter of the urea decomposition pipe is 57 mm, while the diameter of the SCR system is 135 mm. This nominal design is adapted from Y. Cho et al. (2014). The total length is limited to 40 cm.

In this study, 1-dimensional model suggested by C. Depcik et al. (2007) is used for sensitivity analysis. The governing equations derived from mass balance and energy balance equations are described as follows:

$$\frac{\partial C_{g,j}}{\partial t} + u\frac{\partial C_{g,j}}{\partial x} = \frac{-k_{m,j}G_a}{\varepsilon}(C_{g,j} - C_{s,j}) \tag{1}$$

$$\frac{dC_{s,j}}{dt} = \frac{k_{m,j}G_a}{\varepsilon_{wc}(1-\varepsilon)}\left(C_{g,j} - C_{s,j}\right) - \frac{G_{ca}\sum_{i=1}^{Rxn}\nu_{j,i}R_i}{\varepsilon_{wc}(1-\varepsilon)} \tag{2}$$

$$\Gamma\frac{d\Theta_m}{dt} = \dot{s_m} \tag{3}$$

$$\rho_g\tilde{c_p}\frac{\partial T}{\partial t} = \frac{U_0G_a}{\varepsilon}(T_{wc} - T) - u\rho_g\tilde{c_p}\frac{\partial T}{\partial z} \tag{4}$$

$$\left(\rho_g\tilde{c}_{p,wc} + \frac{\rho_s\tilde{c}_{p,s}}{\varepsilon_{wc}\varepsilon_s}\right)\frac{\partial T_{wc}}{\partial t} = \frac{\alpha(T_e - T_{wc})}{\varepsilon_{wc}\varepsilon_s} + \frac{U_oG_a}{\varepsilon_{wc}(1-\varepsilon)}(T - T_{wc}) \tag{5}$$

$$+\frac{G_{ca}}{\varepsilon_{wc}(1-\varepsilon)}\sum_{i=1}^{Rxn}\nu_{j,i}R_i\Delta H_{Rxn,j,i}(T_{wc})$$

It is also necessary to consider the washcoat porosity (ε_{wc}) in order to express the system more accurately. When the equations represent the behavior of a system containing chemical reactions, a forward integration of the equations becomes difficult (H. Robertson, 1966). MOL (Method of Lines) is appropriate for such stiff differential equations. Another method, Rothe's method, which allows choosing a different mesh for each time step, is too complex to use in this system. Therefore, in this study, MOL shall be used to numerically solve the PDE model.

Before the SCR reaction occurs, injected urea is decomposed into ammonia and isocyanic acid. The chemical reactions involved in the system are reported in S. Yim (2004). The produced isocyanic acid is hydrolyzed into ammonia, which plays the role of a reducing agent in SCR reactions. Since exhaust gas has high temperature, the injected urea solution is immediately vaporized and thermally decomposed. The reactions of urea decomposition into isocyanic acid and hydrolysis of isocyanic acid are as follows:

$$CO(NH_2)_2 \rightarrow NH_3 + HNCO \tag{6}$$

$$HNCO + H_2O \rightarrow NH_3 + CO_2 \tag{7}$$

The SCR reactions and their reaction rates are shown in Table 1. Since ammonia adsorption reaction is reversible, the storage amount of ammonia in SCR depends on the temperature as well as the size of SCR. The kinetic parameters are adapted from L. Olsson et al. (2008). All SCR reactions are inserted in the governing equations (1) ~ (5) as R_i.

Table 1. Reactions and reaction rate for SCR

Reaction	Reaction rate expressions
$NH_3 + S1 \leftrightarrow NH_3 - S1$	$r_1 = k_{1,f} c_{NH_3} \theta_{S1-vacant} - k_{1,b} \theta_{NH_3-S1}$
$2NH_3 - S1 + \frac{3}{2}O_2 \rightarrow N_2 + 3H_2O + 2S1$	$r_2 = k_2 c_{O_2} \theta_{NH_3-S1}$
$NO + \frac{1}{2}O_2 \leftrightarrow NO_2$	$r_3 = k_{3,f} {c_{O_2}}^{1/2} c_{NO} - k_{3,b} c_{NO_2}$
$4NH_3 - S1 + 4NO + O_2 \rightarrow 4N_2 + 6H_2O + 4S1$	$r_4 = k_4 c_{NO} \theta_{NH_3-S1}$
$2NH_3 - S1 + NO + NO_2 \rightarrow 2N_2 + 3H_2O + 2S1$	$r_5 = k_5 c_{NO} c_{NO_2} \theta_{NH_3-S1}$
$4NH_3 - S1 + 3NO_2 \rightarrow 3.5N_2 + 6H_2O + 4S1$	$r_6 = k_6 c_{NO_2} \theta_{NH_3-S1}$
$2NH_3 - S1 + 2NO_2 \rightarrow N_2 + 3H_2O + 2S1$	$r_7 = k_7 c_{NO_2} \theta_{NH_3-S1}$

3. Sensitivity Analysis Results

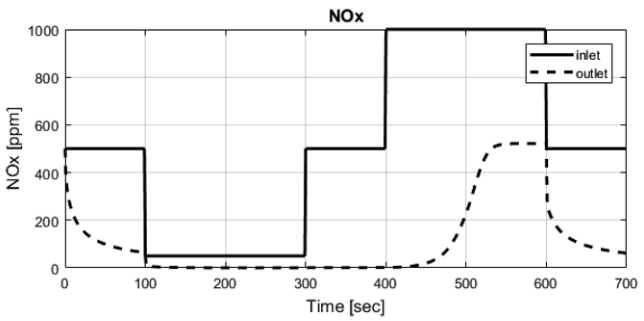


Figure 2. Inlet and outlet concentration of NOx

It is assumed that the exhaust gas has 8 % oxygen. The ratio of NO_2/NOx is assumed to be 0.1 (J. Heywood, 1988). As shown in Figure 2, sensitivity analysis shall be made under the premise that the NOx concentration of inlet gas changes over the time horizon

5 times, in a step-like manner. From 0 to 100 seconds, the inlet concentration of NOx is 500 ppm. Then the NOx concentration in the outlet decreases as a result of the SCR reactions. In the second step, from 100 to 300 seconds, the NOx concentration of engine exhaust gas is lowered to the level of 50 ppm. Since the NOx concentration is very low, the ammonia produced by urea decomposition is stored in catalyst or exits the exhaust pipe in the form of ammonia slip. In the following 400 to 600 seconds, the inlet concentration of NOx is relatively high, 1,000 ppm, which means the amount of ammonia is deficient. Since unreacted ammonia is stored in the catalyst as a result of the previous step, NOx is converted into nitrogen for a while. The amount of stored ammonia increases as the size of the module becomes large. That is to say, the longer the SCR, the more amount of ammonia shall be stored when there is no NOx. In turn, the ammonia stored in increased amount is then used to convert more NOx when its concentration becomes high.

Using the data of 700 seconds in Figure 2, the cumulative amounts of NOx emission, ammonia slip, and the capital cost are calculated according to the length of SCR and distance of urea injector. The capital cost of SCR is calculated by using data from J. GerMan (2012). The SCR cost includes not only the cost of the catalyst but also the cost of urea pump, injector, and tank. However, since these are constant regardless of the reactor size, only substrate and washcoat costs and canning costs are considered in this study. The objective function is defined as a function of SCR length (L).

$$f_0(L) = w_a NOx(L)/NOx^o + w_b NH_3(L)/NH_3{}^o + w_c COST(L)/COST^o \quad (8)$$

In this study, the total length from urea injector to SCR exit is limited to 40 cm. The superscript ^{o} in Equation (8) indicates the order of magnitude of its nominal value. Since the amount of NOx is the most important for a system's performance, w_a shall be set to 3 whereas the others shall be set to 1. If NOx emission is controlled more strictly, w_a can be even bigger, and the optimal length will change accordingly.

As the length of SCR increases, the NOx conversion increases and the amount of ammonia slip decreases. On the other hand, if the distance between the urea injector and SCR is shortened, the urea may not go through enough thermal decomposition. In addition, the capital cost increases as the module becomes larger. Therefore, it is clear that there exists an optimal SCR length which minimizes the objective function.

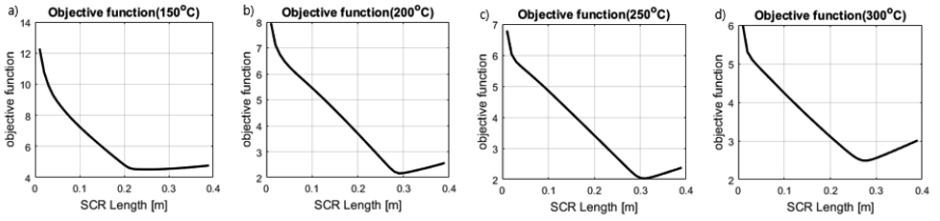


Figure 3. Objective function with sensitivity analysis of a) 150 ℃, b) 200 ℃, c) 250 ℃, d) 300 ℃

The results of sensitivity analysis are shown in Figure 3. The optimal length of SCR is between 25 cm and 30 cm. Therefore the distance of urea injector should be between 10 cm and 15 cm in order to respect the constraint of the total length, 40 cm.

Within the given temperature range in Figure 3, NOx conversion is the highest at 250 ℃. In the low temperature range (150 ℃), the optimal length of SCR is shorter than that of

250 ℃. This does not imply that there is no need for large amount of catalyst to minimize NOx emission. In this low temperature, the catalytic reaction rate is slow, therefore the catalyst efficiency per unit length is also low. A large module only increases the total cost. On the other hand, in the high temperature range (300 ℃), substrate desorption reaction of ammonia is accelerated and thus the amount of ammonia that can be stored in the module is reduced. Therefore, the optimal SCR length is shorter than that of 250 ℃.

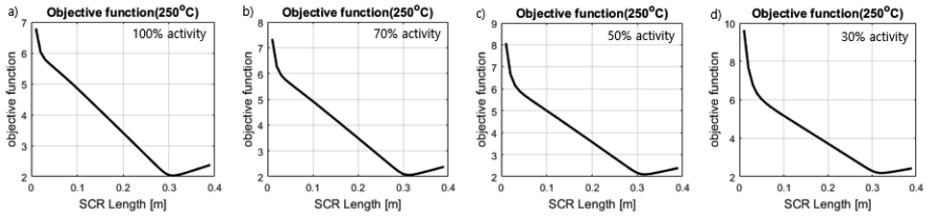


Figure 4. Sensitivity analysis of the objective function of a) 100 % activity, b) 70 % activity, c) 50 % activity, d) 30 % activity at 250 ℃

In this study, deactivation of SCR catalyst is also considered. One problem related with automobile exhaust gas catalysts is the problem of reduced activity due to aging of catalyst. Contaminants such as Sulfur adhere to the catalyst surface and interfere with the catalytic reactions. Hot temperature also causes sintering which results in significant loss of catalyst activity. In this study, the catalyst deactivation was performed by multiplying reaction rate constants by a fixed number, between 0 and 1. In all, cumulative amount of NOx emission and ammonia slip increased, but the optimal SCR length at which the objective function is minimized remained practically unchanged.

Commercial urea-SCR size is shown in Y. Cho et al. (2014). General geometry of urea-SCR for 2.7 L engine consists of 150 mm urea decomposition pipe and 234 mm SCR system. With this geometry and inlet gas shown in Figure 2, the cumulative amounts of NOx emission and ammonia slip are 288.9 mg and 88.41 mg, respectively. In addition, the capital cost of SCR is $117.23. Through sensitivity analysis with the objective function, the optimal configuration of total 384 mm urea-SCR system is calculated as 77 mm injector distance and 307 mm SCR catalyst. With this geometry, the cumulative amounts of NOx emission and ammonia slip are 145.0 mg and 5.40 mg, respectively, and the capital cost of SCR is $153.83. A mere additional cost of $36.6 reduced the cumulative amount of NOx emission by 143.9 mg and ammonia slip by 83.01 mg, respectively.

4. Conclusion

In this study, optimal urea injector position and optimal SCR size are determined by sensitivity analysis. The objective function has three factors; cumulative NOx emission, cumulative ammonia slip, and capital cost. The total length is limited to 40 cm, and the simulation results showed that the optimal SCR length is found between 25 cm and 30 cm. The catalyst deactivation, which is problematic in automotive catalysts, is also considered. Compared to the conventional commercial SCR, the cumulative amounts of NOx emission and ammonia slip are reduced by 143.9 mg and 83.01 mg respectively, under the inlet data used in this study. By creating a new objective function in combination with the weights, we can observe the change of the objective function value under various scenarios. It will be dealt with more deeply through Pareto front observation in our future work.

(novelty) This study is the previous stage to observe the change of the objective function considering various factors and to derive the optimal condition for each case. Pareto front observation will be continued as a future research.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (MSIP) (NRF-2016R1A5A1009592).

This research was respectfully supported by Engineering Development Research Center (EDRC) funded by the Ministry of Trade, Industry & Energy (MOTIE). (No. N0000990)

References

Y. Cho, S. Lee, W. Choi, and Y, Yoon, 2014, Urea-SCR system optimization with various combinations of mixer types and decomposition pipe lenghts, International Journal of Automotive Tchnology, 15, 723-731

C. Depcik, D. Assanis, and K Bevan, 2007, A one-dimensional lean NOx trap model with a global kinetic mechanism that includes NH_3 and N_2O, International Journal of Engine Research, 9, 57-77

J. GerMan, 2012, Estimated Cost of Emission Reduction Technologies for Light-Duty Vehicles, The International Council on Clean Transportation

J. Heywood, 1988, Internal combustion engine fundamentals

L. Olsson, H. Sjövall, and R. J. Blint, 2008, A kinetic model for ammonia selective catalytic reduction over Cu-ZSM-5, Applied Catalysis B: Environmental, 81, 203-217

H. Robertson, 1966, The solution of a set of reaction rate equations, Numerical analysis: an introduction

S. Yim et al, 2004, Decomposition of Urea into NH_3 for the SCR Process, Industrial & engineering chemistry resarch, 43, 4856-4863

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50036-7

Novel Symbolic Regression-Mathematical Programming based Predictions of the Molecular Cetane Number with Small Sampling Data

Jiawen Wei, Zhihong Yuan*

Department of Chemical Engineering, Tsinghua University, 100084 Beijing, China
zhihongyuan@mail.tsinghua.edu.cn

Abstract

In order to conduct the high-throughput screening of molecules and subsequently speedup the process of exploring latent new fuels with desirable properties such as high CN (Cetane Number) which is one of the most vital signatures to assess the fuels combusted in the engine, a novel machine learning method named hybrid Symbolic Regression-Mathematical Programming based approach for extracting surrogate model from data sets is proposed. We formulated the Symbolic Regression Tree (SRT) as a Generalized Disjunctive Programming (GDP) model. Via the Big-M, the GDP model is transferred to a mixed integer nonlinear programming (MINLP) model. BARON is then utilized to solve the MINLP. The established model in this paper demonstrates a good advantage in training efficiency, model representation and prediction accurate.

1. Introduction

Concerns about diminishing fossil fuel reserves along with climate change and national security have promoted considerable research activities on exploring alternative, environmentally friendly processes for producing liquid transportation fuels. During the past decades, industrial and academia have doing their bests to pursue the proper biofuels, which have desirable combustion properties and strong competitiveness, to partially or completely replace traditional petroleum-driven gasoline and diesels.

The CN (Cetane Number) is one of the most important signatures to evaluate the fuels combusted in the engine, which is a correlation based on ignition delay from the start of fuel injection. In general, the higher a fuel's cetane number is, the shorter ignition delay period the fuel has. Cooperative Fuel Research (CFR) and Ignition Quality Tester (IQT) are two mainly used methods in determining CN. The American Society for Testing and Materials (ASTM) Standard D613 (ASTM D613, 2015) uses the single-cylinder CFR for determination of CN, while ASTM Standard D7170 (ASTM D7170-16, 2016) and ASTM Standard D6890 (ASTM D6890, 2015) utilize the IQT. All the methods provide accurate CN measurements, while the CFR can mostly reflect a fuel's actual combustion behaviour in the engine and the IQT offers a faster measure process with lower volumetric requirements.

Clearly, if a series of new biofuel molecules are experimentally synthesized one by one to test the potential CN, a huge amount of time and money will be spent inefficiently. To some extent, a proper experimental data driven predictive model representing the structure-CN relationships of molecules approximately, which can help chemists screen the molecules and subsequently speed up the process of exploring latent new fuels with good performance, is desirable.

Motivated by the tremendous advancement of machine learning, predicting molecules' properties from their structures is an obvious way and has been extensively investigated. Yang used backpropagating neural networks to predict the CN of iso-paraffins and diesel fuels based on quantitative structure property relationship (QSPR) (Yang H, 2001). Taylor utilized the QSARIS software to calculate more than 100 molecular descriptors of 275 compounds that may be used to build the predictive model of the CN. Then the software determined which descriptors are most relevant in modelling the CN using a genetic algorithm and built a predictive model (RMSE = 9.1CN units) (Taylor J, 2004). Although this model is not accurate enough, Taylor's work did provide QSPR inputs for later research to predict CN and indicated which molecular descriptions are crucial. For example, Kessler retained 15 molecular descriptors to build a backpropagation neural network based prediction model of cetane number for furanic biofuel additives, with a total RMSE of 5.97 for the core data set of 284 molecules (Kessler T, 2017). Some other types of models such as utilizing an inverse function method (Smolenskii E, 2008) and using the genetic function approximation (GFA) (Creton B, 2010) have also been used to predict CN. Apparently, these methods have made it possible to describe the relationship between molecular descriptors and properties accurately in a single chemical family of similar chemical properties within test range.

Although the aforementioned methods have pretty good robustness and can fit the relationship between CNs and molecular descriptors quite accurately, it is easy to overlook that in choosing which molecular description to be included in a predictive model of CN. To some extent, humans' choices may limit the performance of the predictive model since these choices are made following humans' knowledge. Alpha Go Zero, which beat Alpha Go by removing the constraints of human knowledge, is an implicational instance. Under such circumstance, it is very important and interesting to formulate a method for surrogate model building, which can automatically select the molecular descriptors while guaranteeing the prediction accurate, to predict molecular CN. Considering the excellent characteristics of Symbolic Regression and Mathematical Programming, the objective of this paper is to propose a novel machine learning method to explicitly correlate the relationship between CNs and molecular descriptors.

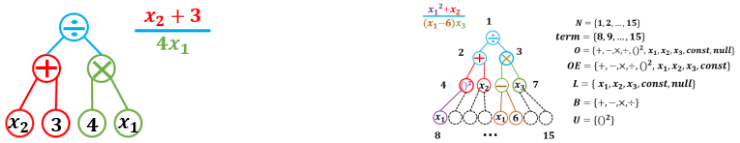
2. Method Description

2.1. Symbolic Regression

Existing literature demonstrated that symbolic regression can give concise analytical expressions without humans' prior knowledge so that it may help scientists to discover laws behind some focused phenomenon (Schmidt M, 2009). An expression tree such as shown in Figure 1(a) can illustrate how symbolic regression runs. It is obvious that any explicit function can be represented by expression trees. What a symbolic regression algorithm does is to combine the pre-set operators such as $\{+,-,\times,\div,\exp\ldots\}$ and operands including variables and parameters to form an explicit function which can represent the mathematical link among the input/output data sets. Due to the absence of any information except basic data set, symbolic regression aims to form proper functional forms and calculate corresponding parameters simultaneously.

Conventionally, most symbolic regression algorithms are based on evolutionary computation which is approached as an application of Genetic Programming (GP). Genetic operations like crossover and mutation would generate more proper expressions, some of which are redundant ones. On the other hand, the expressions may not be concise and may not be even be local optimal due to the GP's stochasticity. To our knowledge,

there is no strict global optimization algorithm to solve the problem of Symbolic Regression.



(a) An instance of expression tree (b) Description of an expression tree

Figure 1. Description of Symbolic Regression Tree.

2.2. The mathematical programming formulation of symbolic regression

To describe an expression tree, several sets and parameters are introduced. As shown in Figure 2(b), the set N is proposed for the number (1 to 15) representing the corroding node of the tree in a sequence. The set O is the compilation of operators and operands. Obviously, there are some absent nodes, so an operator "null" is introduced to represent them. For convenience, some subsets are also defined. OE refers to the compilation of operators and operands except "null", L refers to operands and "null", U refers to unary operators while B refers to binary operators. Furthermore, the variable v_n is given to each node and the link function f^{oe} is introduced to make the tree computable. For convenience, we set v_{ln} to be the left child node's value of node "n", and v_{rn} to be the right one's value. j_n is introduced to represents a constant. Necessarily, for some operators and operands, the constraint $g^{oe} \leq 0$ is enforced to avoid math error or meaninglessness such as the dividend is zero. The binary variable y_n^o describes which operator or operand the node "n" takes. Logical constraints such as each node taking only one operator or operand are also added. I represent the number of data sets. We can measure the fitness of the expression tree by the error between v_{1i} and the sample dependent variable z_i. Therefore, a symbolic regression tree can be formulated as the following Generalized Disjunctive Programming (GDP).

$$\min s = \sum_i (z_i - v_{1i})^2 \tag{1}$$

$st.$

$$\bigvee_{o \in O} \begin{bmatrix} y_n^o \\ f_{ni}^o(v_{lni}, v_{rni}, x_i, , j_n, v_{ni}) = 0 \\ g_{ni}^o(v_{lni}, v_{rni}, x_i, , j_n, v_{ni}) \leq 0 \end{bmatrix} \vee \begin{bmatrix} y_n^{null} \\ v_{ni} = 0 \end{bmatrix} \qquad \forall n \notin term \tag{2}$$

$$\bigvee_{o \in L} \begin{bmatrix} y_n^L \\ f_{ni}^L(x_i, j_n, v_{ni}) = 0 \\ g_{ni}^{const} = \varepsilon - |j_n| \leq 0 \end{bmatrix} \vee \begin{bmatrix} y_n^{null} \\ v_{ni} = 0 \end{bmatrix} \qquad \forall n \in term \tag{3}$$

$$\sum_o y_n^o = 1 \qquad \forall n \notin term \tag{4}$$

$$\sum_{o \in L} y_n^o = 1 \qquad \forall n \in term \tag{5}$$

$$\sum_{o \in BVU} y_n^o = \sum_o y_{ln}^{oe} \qquad \forall n \notin term \tag{6}$$

$$\sum_{o\in B} y_n^o = \sum_o y_{rn}^{oe} \qquad \forall n \notin term \qquad (7)$$

$$y_n^o \in \{0,1\} \qquad (8)$$

$$\underline{V} \le v_{ni} \le \bar{V} \qquad (9)$$

$$\underline{J} \le j_{ni} \le \bar{J} \qquad (10)$$

$$y_{ln}^{cst} + y_{rn}^{cst} \le 1 \qquad \forall n \notin term \qquad (11)$$

$$y_n^{\div} + y_{rn}^{cst} \le 1 \qquad \forall n \notin term \qquad (12)$$

$$y_n^{-} + y_{rn}^{cst} \le 1 \qquad \forall n \notin term \qquad (13)$$

$$\sum\nolimits_{o\in U} y_n^o \le 1 - y_{ln}^{cst} \qquad \forall n \notin term \qquad (14)$$

$$y_{ln}^{x} + y_{rn}^{x} \le 1 \qquad \forall n \notin term \qquad (15)$$

If equations and constraints in (2) and (3) are overwritten with Big-M constraints, the GDP model can easily be transformed into a MINLP model.

3. Results and Discussions

To intuitively test the efficiency of the approach, 13 n-alkanes (C_3~C_{15}) are chosen as the training data sets and then 5 n-alkanes (C_{16}~C_{20}) are predicted. Six descriptors listed in Table 1 are fed to the above GDP model. The final descriptors appeared in the explicit expression will be optimally decided through solving GDP model by BARON (Tawarmalani M, 2004).

Table 1. Definitions of the chosen molecular descriptors (Kessler T, 2017).

Descriptor	Definition
Mor32e	3D-MoRSE - signal 32 / weighted by atomic Sanderson electronegativities
EEig08x	Eigenvalue 08 from edge adj. matrix weighted by edge degrees
CIC1	Complementary information content
RDF090m	Radial Distribution Function - 9.0 / weighted by atomic masses
RDF020p	Radial Distribution Function - 2.0 / weighted by atomic polarizabilities
L/Bw	Length-to-breadth ratio by WHIM

In addition to the set of GDP, a Mixed Integer Nonlinear Programming (MINLP) is also directly built to represent the symbolic regression tree. An Artificial Neural Network (ANN) with six inputs (six descriptors), one output (CN), two hidden layers of 32 neurons which is actually as same as Kessler's work (Kessler T, 2017) is also set. The aim for setting MINLP and ANN models is to conduct the comparisons with GDP to demonstrate its efficiency.

The surrogate model driven by ANN is traditionally inexplicit. Equations (16) and (17) offer the explicit expression between CN and descriptors from GDP and MINLP, respectively.

$$\text{GDP: CN} = (\text{Mor32e} + 0.125) \times \left(7.687 + 2 \cdot \text{CIC1} - 4.964 + \frac{\text{CIC1} - 2.419}{0.59 \cdot \text{Mor32e}}\right) \times \frac{\text{RDF090m} - 2.8 \cdot \text{EEig08x} - 21.595}{\text{CIC1} - 2.19} \quad (16)$$

$$\text{MINLP: CN} = (\text{RDF020p})^2 + \frac{(\text{RDF090m})^2}{\text{CIC1}} + 19.032 \cdot \text{CIC1} - \frac{284.364}{\text{L/Bw}} \quad (17)$$

Clearly, GDP chosen optimally More32e, CIC1, RDF090m, and EEig08x as the independent variables, while MINLP chosen RDF020p, RDF090m, CIC1, and L/Bw. The overall results are shown in Figure 2. The MINLP model showed the worst fitness in the train set. The ANN demonstrated the worst prediction performance. Three models' average relative errors of the prediction are 20.32%(ANN), 8.20%(MINLP), 1.11%(GDP). The GDP model performs the best CN predictions of new molecules which are not listed in the training sets.

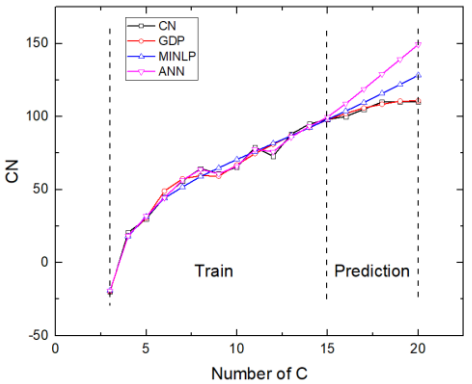


Figure 2. The predicted and original CNs of n-alkanes (C_3~C_{20}).

A 31-node tree containing $\{+, -, \times, \div, ()^2\}$ is used in both the GDP and MINLP model. The MINLP model has 15,504 equations, 6,210 continuous variables and 372 binary variables. The GDP model's size is smaller with 4,653 equations, 838 continuous variables and 403 binary variables. The main factor to cause the different results of the GDP model and the direct MINLP model is the complexity of the calculation. Each constraint, even the loose constraint is concerned when the MINLP problem is solved, such consideration may lead to very long computational time and each cause the wrong results. While the GDP problem can select necessary constraints according to the logical variable y_n^o. Therefore, the GDP model can find a better result. Obviously, excellent training and prediction capacities may be achieved if more operators and nodes are introduced.

Although the prediction capacity of ANN can also be improved through adding more layers and more nodes (Deep learning), a very large data sets are required. In other words, the accuracy of deep learning with ANN may not be high enough when facing small sampling data. However, our proposed method can efficiently tackle small sampling data than ANN through automatically and optimally selecting the equation structure and parameters. Apparently, the mechanism of the proposed method is totally different from the trail-and-error approach of ANN.

The properties of liquid fuels such as CN should be considered during the process design stage under the pressure of clean energy manufacturing. The computational intractable associated with the optimal design/synthesis for liquid fuels production may happen if the complex inexplicit model for CN representation is introduced, as one kind of constraints, to the overall optimization framework. Clearly, the explicit model can accelerate the process design for quality-oriented liquid fuels production. Our computational experiments (including the tests of large training-testing data sets) suggest that the computational burden for the proposed method grows rapidly following the increases on the problem size. Therefore, the tailored algorithm for the proposed method is being developed.

Conclusion

We established a novel Symbolic Regression-Mathematical Programming based method to predict n-alkanes' CNs. The descriptors were selected spontaneously to form the explicit expressions. The proposed method demonstrated better prediction performance when compared to ANN with multiple hidden layer especially when facing small sampling data. If more molecules are put in the training set, more operators and nodes need to be calculated to form the expression tree. Needless to say, the calculation burden would be significantly heavier. Therefore, a tailored solving algorithm is under the exploration. The optimal design/synthesis and operation can benefit from the proposed method due to its capacity to generate explicit surrogate model.

References

ASTM D613. 2015, Standard Test Method for Cetane Number of Diesel Fuel Oil. West Conshohocken, PA: ASTM International.

ASTM D7170-16. 2016, Standard Test Method for Determination of Derived Cetane Number (DCN) of Diesel Fuel Oils—Fixed Range Injection Period, Constant Volume Combustion Chamber Method. West Conshohocken, PA: ASTM International.

ASTM D6890. 2015, Standard Test Method for Determination of Ignition Delay and Derived Cetane Number (DCN) of Diesel Fuel Oils by Combustion in a Constant Volume Chamber. West Conshohocken, PA: ASTM International.

Yang H, Fairbridge C, Ring Z. 2001, Neural network prediction of cetane number for iso-paraffins and diesel fuel. Pet Sci Technol, 19:573–586.

Taylor J, McCormick R, Clark W. 2004, Report on the relationship between molecular structure and compression ignition fuels, NREL Technical Report.

Kessler T, et al. 2017, Artificial neural network based predictions of cetane number for furanic biofuel additives. Fuel, 206:171-179.

Smolenskii EA, et al. 2008, Cetane numbers of hydrocarbons: calculations using optimal topological indices. Russ Chem Bull, 57:461–467.

Creton B, et al. 2010, Prediction of the cetane number of diesel compounds using the quantitative structure property relationship. Energy Fuels, 24:5396–5403.

Schmidt M, Lipson H. 2009, Distilling Free-Form Natural Laws from Experimental Data. Science, 2009, 324: 81-5.

Tawarmalani M, Sahinidis NV. 2004, Global optimization of mixed-integer nonlinear programs: A theoretical and computational study. Mathematical Programming, 99: 563-591.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50037-9

Modelling under Uncertainty for Process Design and Scale-up of an Industrial AACVD

Pedro I. O. Filho, Panagiota Angeli, Eric S. Fraga*

Centre for Process Systems Engineering, Department of Chemical Engineering, University College London, WC1E 7JE, London, United Kingdom
e.fraga@ucl.ac.uk

Abstract

Aerosol-Assisted Chemical Vapour Deposition (AACVD) can be used to manufacture, for example, transparent conducting oxides (TCOs), liquid crystal displays (LCDs) and light emitting diodes (LEDs). This paper presents an integrated model to support the scale-up of the AACVD process, which is comprised of three modules: aerosol generation, transport and delivery. Latin Hypercube Sampling (LHS) method is used to generate input data for the sensitivity and uncertainty analysis. Based on industrial ranges of process parameters, the incorporation of AACVD into current TCO industrial plants, without replacing their existing facilities, was found to be mainly impacted by the flow rates, the length of the transport system and the temperature of the chemical deposition site.

Keywords: AACVD, aerosol, process scale-up, LHS, machine learning.

1. Introduction

Solar cells exploit the photovoltaic effect to convert incident light into electricity. They provide clean and renewable energy, which has the potential to reduce our dependence on fossil fuels. The demand for solar cells has been increasing steadily and the advancement of their manufacturing process is crucial. Solar cells rely on the production of Transparent Conducting Oxide (TCO) films, which is usually accomplished by Chemical Vapour Deposition (CVD). This technique involves the vaporisation of volatile precursors and their transport to the reaction site, where the deposition and film formation take place. A modification of the CVD technique results in the Aerosol-Assisted Chemical Vapour Deposition (AACVD), which generates aerosol droplets from the precursor solution instead of vaporising it. Hence, for AACVD, the precursors do not need to be volatile and a new range of them becomes available. The aerosol is then transported to the reaction site for the synthesis of films such as the TCOs, as well as coatings, powders and composites for different final applications.

As examined by Grossmann and Westerberg (2000, 2017), the computational tractability of complex problems is challenging. To mitigate this concern, we avoided the use of Computational Fluid Dynamics (CFD) packages and tried to build a tractable model for the whole AACVD process, which allowed us to repeatedly run simulations to perform a global sensitivity analysis. It takes seconds to run hundreds of simulations, which would not be possible using CFD. This will be very useful in the optimisation of the entire process, which is part of the future work. For the moment, we present an integrated model to support process scale-up, comprised of three modules: aerosol

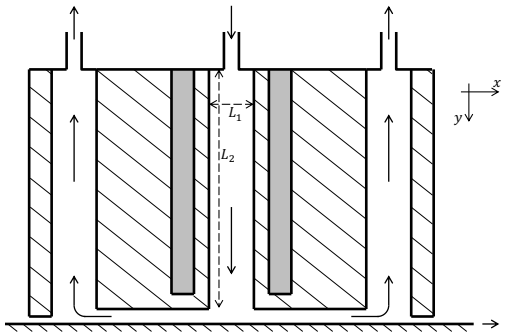


Figure 1. Cross-section of the deposition site in the direction of the flowing glass.

generation, transport and delivery. There is a range of droplet sizes obtained when generating aerosol, which needs to be properly represented. Using models from the literature, the generation of aerosol by ultrasonic vibration is therefore described by a distribution of droplet sizes. Such a distribution is the input to a transport model that incorporates the impact of aerosol losses, which depend on the properties of the droplets, the flow and the piping system. The output distribution provides sufficient information to predict the amount and sizing of aerosol reaching the deposition site. The transport model has been applied to different scenarios, using horizontal, inclined and vertical pipes of different lengths connected by bends. The delivery module includes the droplet evaporation after leaving the transport system and it accounts for uncertainties in the temperature profile of the deposition site. This is a key factor for the droplet evaporation, which sets the precursors free to react and form the desired products. The scale-up of an AACVD process has not been studied or performed yet, to our best knowledge. Additionally, switching from conventional CVD, commonly used in the industry, to AACVD is challenging. Ideally, this transition could happen without having to replace any equipment, aiming at lower costs and the achievement of the same final products or improved ones.

2. Modelling of Aerosol Generation, Transport and Delivery

The generation of aerosol droplets can be done by ultrasonic vibration, using a piezoelectric transducer. The sizes of such droplets are known to be approximately log-normally distributed (Ramachandran and Cooper, 2011). Therefore, assuming a log-normal distribution for the diameter of the aerosol droplets generated, the median diameter $\overline{d_d}$ [m] can be predicted by:

$$\overline{d_d} = f(\boldsymbol{x_n}) \cdot (\sigma_d \cdot \rho_d^{-1} \cdot f^{-2})^{1/3} \tag{1}$$

where σ_d [N·m^{-1}] and ρ_d [kg·m^{-3}] are, respectively, the precursor solution surface tension and density; f [Hz] is the ultrasonic frequency; $f(\boldsymbol{x_n})$ is a function of the vector $\boldsymbol{x_n}$, which takes into account the atomised fluid flow rate, viscosity, surface tension, density and the properties of the ultrasonic atomiser, which are the ultrasonic frequency and power surface intensity. The present work uses Equation 5 from Ramisetty et al. (2013) to describe $f(\boldsymbol{x_n})$, given the good agreement with experimental data.

After generating the aerosol, its transport is done via horizontal, inclined or vertical pipes and bends. Correlations from the literature were used to model the losses of aerosol during transport. Such losses are mainly caused by gravitational settling, diffusional deposition and turbulent inertial deposition, as described by Brockmann (2011). Details about modelling and experimental validation of the aerosol transport system were previously presented by Filho et al. (2017).

Table 1. Range of parameters for the aerosol generation (G), transport (T) and delivery (D) integrated modules. The last column describes which variables are the input for each of the three modules.

Symbol	Parameter	Range	Unit	Module
$c_{p,d}$	Droplet specific heat	$[2.0, 4.0] \cdot 10^{3}$	$J \cdot kg^{-1} \cdot K^{-1}$	D
c_p	Carrier fluid specific heat	$[1.0, 1.1] \cdot 10^{3}$	$J \cdot kg^{-1} \cdot K^{-1}$	D
d	Pipe inner diameter	$[0.5, 2.0] \cdot 10^{-2}$	M	T
d_d	Droplet diameter	$[5.0, 9.0] \cdot 10^{-6}$	M	T, D
f	Ultrasonic frequency	$[1.0, 3.0] \cdot 10^{6}$	Hz	G
$h_{vap,d}$	Droplet specific heat of evaporation	$[1.0, 1.4] \cdot 10^{6}$	$J \cdot kg^{-1}$	D
I	Power surface intensity	$[1.0, 3.0] \cdot 10^{3}$	$W \cdot m^{-2}$	G
k	Carrier fluid thermal conductivity	$[3.5, 4.8] \cdot 10^{-2}$	$W \cdot m^{-1} \cdot K^{-1}$	D
L	Pipe length	$[0.5, 8.0] \cdot 10^{0}$	m	T
N	Number of 90° pipe bends	0 to 4	−	T
Pr	Carrier fluid Prandtl Number	$[6.8, 7.2] \cdot 10^{-1}$	−	D
Q	Carrier fluid flow rate	$[0.1, 1.0] \cdot 10^{-4}$	$m^3 \cdot s^{-1}$	T, D
Q_p	Precursor solution flow rate	$[0.1, 1.0] \cdot 10^{-11}$	$m^3 \cdot s^{-1}$	G
Q_T	Reciprocating nozzles total flow rate	$[0.1, 1.0] \cdot 10^{-2}$	$m^3 \cdot s^{-1}$	D
T_w	Wall temperature	$[5.0, 7.0] \cdot 10^{2}$	K	D
μ	Carrier fluid dynamic viscosity	$[1.8, 3.7] \cdot 10^{-5}$	$N \cdot s \cdot m^{-2}$	T, D
μ_d	Droplet dynamic viscosity	$[0.3, 1.1] \cdot 10^{-3}$	$N \cdot s \cdot m^{-2}$	G, T
ρ	Carrier fluid density	$[0.4, 1.2] \cdot 10^{0}$	$kg \cdot m^{-3}$	T, D
ρ_d	Droplet density	$[0.7, 1.0] \cdot 10^{3}$	$kg \cdot m^{-3}$	G, T, D
σ_d	Droplet surface tension	$[2.0, 5.0] \cdot 10^{-2}$	$N \cdot m^{-1}$	G
ϕ	Pipe inclination angle	$[0.0, 0.5] \cdot 10^{0}$	rad	T

Once the aerosol droplets reach the deposition site, they are heated. The solvent evaporates and the reactions take place, resulting in the film formation. The deposition site shown in Figure 1 is based on the International Patent 96/11802 (Soubeyrand, 1996). The diagram shows a cross-section of the device, where the aerosol arrives from the top middle part and reaches the surface of a moving glass in the x-axis direction on the bottom part. The formation of a thin film with specific optoelectronic properties on the glass is the final objective. The grey part in the diagram is a heat exchanger, which allows the temperature of the deposition site walls to be controlled. The glass width, in the direction perpendicular to the xy-plane, is 3 m, while $L_1 = 0.03$ m and $L_2 = 0.3$ m. The temperature of the carrier fluid travelling between the walls of the reaction site can be modelled as shown in Equation 2:

$$dT/dy = (h_1 \cdot (T_w - T) \cdot P)/(\dot{m} \cdot c_p) \quad (2)$$

where T [K] is the carrier fluid temperature; y [m] is the vertical distance travelled by the fluid; h_1 [$W \cdot m^{-2} \cdot K^{-1}$] is the heat transfer coefficient; T_w [K] is the wall temperature; P [m] is the surface perimeter; $\dot{m}$ [$kg \cdot s^{-1}$] and c_p [$J \cdot kg^{-1} \cdot K^{-1}$] are, respectively, the carrier fluid mass flow rate and specific heat at constant pressure.

The modelling of the droplet evaporation is based on mass and energy balances. Two differential equations are numerically solved simultaneously:

$$h_{vap,d} \cdot \dot{m}_v + c_{p,d} \cdot m_d \cdot dT_d/dt = h_2 \cdot (T_w - T_d) \cdot \pi \cdot d_d^2 \quad (3)$$

$$dd_d/dt = -2\,\dot{m}_v/(\rho_d \cdot \pi \cdot d_d^2) \quad (4)$$

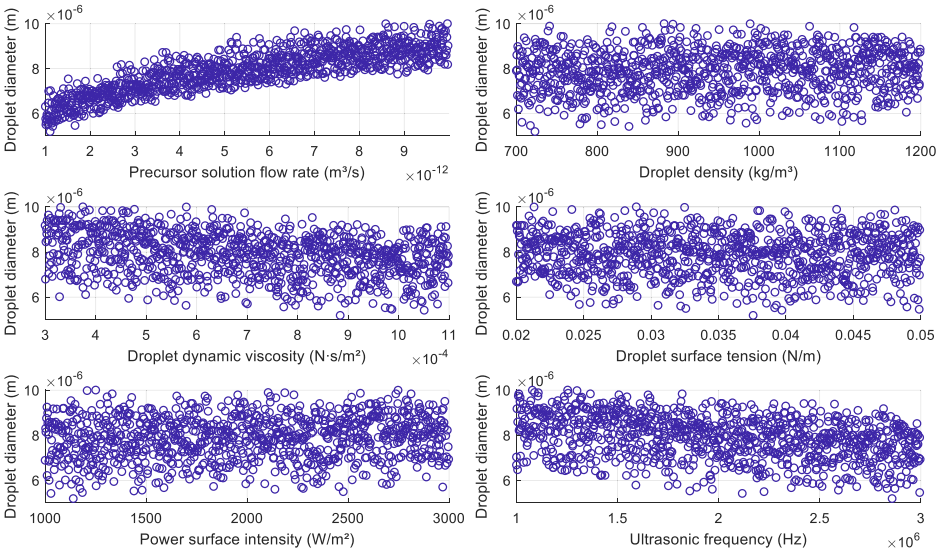


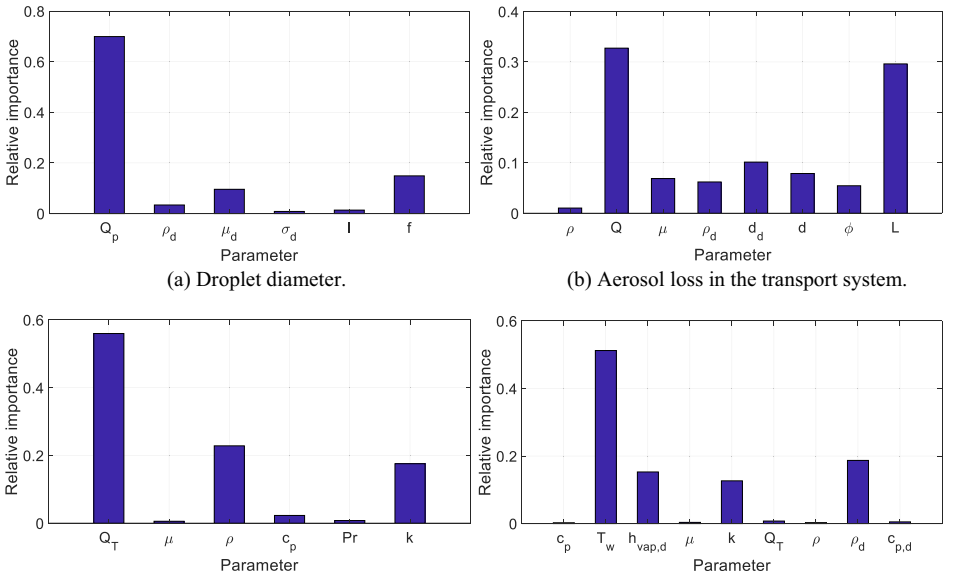
Figure 2. Scatter plots of parameters affecting the droplet diameter obtained using an ultrasonic vibration atomiser. Each parameter is described in Table 1.

where $h_{vap,d}$ [J·kg^{-1}] is the droplet specific heat of evaporation; $c_{p,d}$ [J·kg^{-1}·K^{-1}], m_d [kg], ρ_d [kg·m^{-3}], T_d [K] and d_d [m] are, respectively, the droplet specific heat at constant pressure, mass, density, temperature and diameter; t [s] is time; h_2 [W·m^{-2}·K^{-1}] is the heat transfer coefficient; T_w [K] is the wall temperature; $\dot{m}_v$ [kg·s^{-1}] is the mass transfer rate, given by $\dot{m}_v = k_c \cdot (\rho_{v,s} - \rho_{v,\infty}) \cdot \pi \cdot d_d^2$, where k_c [m·s^{-1}] is the mass transfer coefficient; $\rho_{v,s}$ [kg·m^{-3}] and $\rho_{v,\infty}$ [kg·m^{-3}] are, respectively, the partial vapour densities over the droplet surface and far from it. Correlations for the heat and mass transfer coefficients were found in Mezhericher et al. (2008). The thermophysical properties of possible precursor solutions and carrier fluids were found in Bergman et al. (2011). The aerosol droplets, the reactive solute and the evaporating solvent, are assumed to be in thermal equilibrium with the carrier gas. Therefore, the droplet temperature is a function of time and indirectly of the vertical distance travelled by the carrier fluid.

3. Methodology, Results and Discussion

The model presented by Filho et al. (2017) focused solely on aerosol transport and presented a local sensitivity analysis. This is now expanded to include the aerosol generation and delivery, along with a global sensitivity analysis. The AACVD process can be operated using a number of different precursors, with different thermophysical properties. The generated aerosol can also be carried by different fluids to the reaction site. Therefore, the global sensitivities of the model provide understanding of which parameters have the greatest relative importance to the objective variables. Additionally, uncertainty analysis is performed by propagating the uncertainties through the entire process.

Based on lab-scale experiments and extrapolating to the industrial-scale process, for different precursors and settings, a range for each parameter of interest is described in Table 1. The generation of samples for all the variables described in Table 1, based on



(a) Droplet diameter.

(b) Aerosol loss in the transport system.

(c) Carrier gas temperature.

(d) Droplet drying time.

Figure 3. Relative importance of different parameters given their impacts on each output variable from the regression trees, as described in the legends of each subfigure. The parameters are described in Table 1.

the range column, is done by implementing the Latin Hypercube Sampling (LHS) method. The range of each parameter is divided into equally probable subintervals and a random sample is taken from each permuted subinterval. This method created 1000 samples fully stratifying the initial range of all parameters. Figure 2 shows scatter plots obtained for the droplet sizes of the generated aerosol. It can be seen how the precursor solution flow rate has the greatest impact on the droplet sizes. Some impact can also be observed for the ultrasonic frequency and droplet dynamic viscosity.

Using the dataset generated by the simulations, regression trees are grown for the response variables using the parameters in Table 1 as predictors. Surrogate splits are used with the CART method (Loh, 2014). The use of regression trees can be justified by their accuracy and computational speed. However, the main reason for their use in this work is the interpretability of the tree structures, making it easy to rank parameters by importance. This is done by summing changes in the mean squared error due to splits on every predictor and dividing the sum by the number of branch nodes. For example, Figure 3 (a) confirms what can be visually seen from Figure 2 regarding the most impactful parameters on the droplet size: the precursor flow rate has the greatest impact and, to a lesser extent, the ultrasonic frequency and droplet dynamic viscosity. Density, surface tension and power intensity have minimal impact. Overall, Figure 3 shows that the aerosol generation, transport and delivery are mainly impacted by the flow rates, the length of the transport system and the wall temperature in the reaction site. These results are based on the ranges of parameters that can be used in the industrial-scale process. Therefore, they serve as a guide to the AACVD scale-up. Additionally, the most impactful parameters will be chosen to optimise the reactions and film formation, which are the next modelling steps.

4. Conclusions and Future Work

The use of CVD to manufacture TCO materials is sometimes limited by high costs, environmental impacts and scarcity of precursors. Therefore, it is important to look for cheaper and more sustainable precursors, which becomes less challenging when a new range of them is unlocked by the use of AACVD instead of CVD. Preliminary simulations worked as a proof of concept for the feasibility of keeping the current CVD equipment used in the industry to operate using the AACVD technique instead. The modules for aerosol generation and transport have already been validated using experimental data. Future work includes further model validation, covering the delivery module. Finally, the feasibility study will be completed by modelling the chemistry in the reaction site. Then, it will be possible to optimise the film growth rate and its optoelectronic properties.

The integrated model presented is suitable for application on the scale of industrial processes. It can also be used for any range of the parameters studied and it is suitable for applications that rely on atomisation and transport of particles, for example, spray drying or cooling, ink-jet printing, agricultural sprays and fuel combustion. Furthermore, lessons learned in modelling uncertainties and their impact on process scale-up motivates research into formulation, modelling and solution methods for such applications. The aim should be to ease the procedure of design under uncertainty for process scale-up. The representation and impacts of these uncertainties form the basis of future research for the AACVD case study and the treatment of uncertainty in general, through the development of a novel modelling framework focused on the treatment of uncertainty in process design and optimisation.

Acknowledgements: The authors acknowledge the support provided by CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico, and EPSRC, Engineering and Physical Sciences Research Council (EP/L017709/1).

References

T. L. Bergman, A. S. Lavine, F. P. Incropera, D. P. Dewitt, 2011, Fundamentals of Heat and Mass Transfer, John Wiley & Sons, 7, 995-999.

J. Brockmann, 2011, Aerosol Transport in Sampling Lines and Inlets, in: P. Kulkarni, P. A. Baron, K. Willeke, editors, Aerosol Measurement, John Wiley & Sons Inc., 68-105.

P. I. O. Filho, D. B. Potter, M. J. Powell, C. J. Carmalt, P. Angeli, E. S. Fraga, 2017, Probability Density Functions for Droplet Sizing in Aerosol Transport Modelling, Computer Aided Chemical Engineering, 40(1), 2245-2250.

I. E. Grossmann, 2017, Evolution of Process Systems Engineering and Future Trends in Research, Computer Aided Chemical Engineering, 40(1), 1-2.

I. E. Grossmann, A. W. Westerberg, 2000, Research Challenges in Process Systems Engineering, AIChE Journal, 46(9), 1700-1703.

W. Y. Loh, 2014, Fifty Years of Classification and Regression Trees, International Statistical Review, 82(3), 329-348.

M. Mezhericher, A. Levy, I. Borde, 2008, Modelling of Particle Breakage During Drying, Chemical Engineering and Processing: Process Intensification, 47(8), 1404-1411.

G. Ramachandran, D. W. Cooper, 2011, Size Distribution Data Analysis and Presentation, in: P. Kulkarni, P. A. Baron, K. Willeke, editors, Aerosol Measurement, John Wiley & Sons Inc., 479-506.

K. A. Ramisetty, A. B. Pandit, P. R. Gogate, 2013, Investigations into Ultrasound Induced Atomization, Ultrasonics Sonochemistry, 20(1), 254-264.

M. J. Soubeyrand, 1996, Glass Coating Method and Glass Coated Thereby, International Patent Application WO 96/11802.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50038-0

A Fluidized Bed Process Model of a Chemical Looping Combustion Fuel Reactor

Chinedu O. Okoli, Andrew Lee*, Anthony P. Burgard and David C. Miller

National Energy Technology Laboratory, 626 Cochrans Mill Road, Pittsburgh, PA, 15236, USA
andrew.lee@netl.doe.gov

Abstract

The development of a bubbling fluidized bed chemical looping combustion (CLC) fuel reactor model to support process synthesis and large-scale optimization is presented in this work. The model is built using the Institute for the Design of Advanced Energy Systems (IDAES) PSE framework and is implemented in Pyomo. It is equation-oriented and consists of a set of differential algebraic equations, which describe the gas and solid phase reactions between the fuel gas and oxygen carrier (OC), as well as the mass and heat transfer phenomena occurring in the different regions of the fluidized bed. In addition, the model provides axial profiles of the hydrodynamic and state variables. The capabilities of the model are demonstrated for the simulation of an industrial scale CLC fuel reactor which uses a hematite-based solid as the OC for natural gas combustion. The impact of select design variables on the performance of the model is evaluated, and the results show that methane conversion is more sensitive to changes in bed diameter than to the OC/fuel ratio. In addition, negligible performance differences are seen between the co-current and counter-current configurations indicating that good mixing between the gas and solid phases occurs within the bed.

Keywords: chemical looping combustion, fluidized bed, fuel reactor, process model.

1. Introduction

Chemical looping combustion (CLC) is a promising approach to provide cost effective, low carbon energy from fossil fuels. It creates a CO_2 rich flue gas with less equipment than other technologies such as oxy-combustion, which requires an air separation unit, or post-combustion carbon capture in conjunction with a conventional fossil-based power plant (Adanez et al., 2012). CLC is an active area of research and development in the areas of materials, process design, optimization, and scale up.

The CLC process consists of two interconnected reactors, a fuel reactor and an air reactor, with a metallic oxygen carrier (OC) circulating between them. In the fuel reactor, the OC is reduced, oxidizing the fuel to produce nitrogen free flue gas consisting mainly of CO_2 and H_2O. In the air reactor, the reduced OC is oxidized with air. Fluidized bed reactors have been investigated as reaction vessels for CLC because of their excellent gas/solid contacting. The overall performance of the CLC process is dependent on efficient and effective gas and solid contacting in the fuel reactor.

Computational models used in conjunction with an experimental development program can help to accelerate the development of new energy technologies. The models provide a representation of complex physical phenomena that might be difficult or expensive to measure in physical prototypes. In addition, the models can be used to optimize a

Table 1. Feed Conditions and fluidization properties

Gas Feed Conditions		OC Feed Conditions		OC Fluidization properties	
Flow rate (mol/s)	298.11	Flow rate (kg/s)	1,235	Min Fluid. Vel. (m/s)	0.04
Temperature (K)	800	Temperature (K)	1186	Min. Fluid. Void. (-)	0.45
Composition (%):		Composition (%):		Particle Diam. (mm)	0.28
CH_4	45.82	Fe_2O_3	45	Solid Density (kg/m3)	3,252
CO_2	47.72	Fe_3O_4	0		
H_2O	6.46	Al_2O_3	55		

complete, integrated process to identify the design and operating conditions that should be studied during bench and pilot scale testing. Finally, the models can be used in conjunction with statistical design of experiments to maximize the learning from expensive pilot scale testing.

A particular research need when developing a new technology such as CLC is to consider systems integration and optimization. A systems level CLC process will comprise of phenomena ranging from the micro-scale, such as mass transfer associated with gas diffusion into the OC particle, to the macro-scale, such as the interaction of different unit operations in the flowsheet. Understanding and balancing these interactions can be best accomplished by employing large-scale optimization models of the complete system. Because such systems typically have nonlinear interactions, rigorous optimization approaches are able to identify important trade-offs and process synergies, which cannot be readily observed from simple sensitivity experiments.

To effectively employ these large-scale optimization methods requires the development of suitable models that are sufficiently rigorous to capture all the important underlying physics of the process, from micro to macro scale, while also formulated in a way that is computationally efficient. Models built using an equation-oriented framework are particularly suitable for large-scale optimization since they provide first and second derivative information which is required by advanced optimization solvers capable of exploiting large, sparse matrices.

To address these challenges, the Institute for the Design of Advanced Energy Systems (IDAES) is developing an open-source PSE framework to enable the rapid development and optimization of next generation advanced energy systems. The framework is based on the Pyomo algebraic modeling language (Hart et al. 2017). This paper describes the development of a bubbling fluidized bed (BFB) CLC fuel reactor within the IDAES framework. The model is developed in a modular fashion from two standalone models – a reaction kinetics and physical properties model and a fluidized bed hydrodynamic model. The capabilities of the overall model are demonstrated for the simulation of an industrial scale CLC fuel reactor in which natural gas is combusted using a hematite-based OC.

2. Methodology

2.1. Bed hydrodynamic model

The fundamentals of the hydrodynamic model developed in this work are based on the three-region BFB model described by Kunii and Levenspiel (1968). The three-regions are the bubble, emulsion, and cloud-wake regions. In this model gas bubbles formed at the distributor plate rise through a dense emulsion phase carrying along a cloud of gas with suspended solids. The rest of the gas is in the emulsion phase, which also contains most of the solids. The Kunii and Levenspiel model assumes that the solids are

isothermal and well-mixed, thus neglecting the impact of axial variations in the solid phase. Furthermore, they assume an average bubble diameter to estimate the bed hydrodynamic properties.

Lee and Miller (2013) utilized the Kunii and Levenspiel approach in their modeling of a BFB for CO_2 adsorption from flue gas, extending the approach by considering axial variations in both the gas and solid phases, because adsorption processes are strongly impacted by temperature and pressure variations. They also use a bubble growth equation to model the bubble diameter as a function of height. The resulting model is a steady-state one-dimensional system of differential algebraic equations (DAEs). This model has been extended for the hydrodynamic model of the CLC fuel reactor with changes made to the relevant model equations to account for the different reaction kinetics.

2.2. Reaction kinetics and physical property model

It is assumed that natural gas contains only methane, with the higher hydrocarbons treated as methane equivalents. In the fuel reactor, the Fe_2O_3 based OC is reduced to Fe_3O_4 by reaction with CH_4. It is assumed that methane reforming within the reducer is negligible; thus, the overall reaction considered here is given by Eq.(1):

$$CH_4 + 12Fe_2O_3 \rightarrow 8Fe_3O_4 + 2H_2O + CO_2 \tag{1}$$

The reaction kinetics used in this model for the reaction of a hematite based OC with methane are as reported by Abad et al. (2007) and shown in Eq.(2):

$$\frac{dX}{dt} = \frac{3bkC^n}{\rho_m r_g (1-X)^{\frac{2}{3}}} \tag{2}$$

where X is the fraction conversion of the metal oxide in the OC, C (mol/m^3) is the reaction gas concentration (CH_4 in this case), b is the reaction stoichiometric coefficient (12 moles of Fe_2O_3 converted per mole of CH_4), k (m$^{1.9}$mol$^{-0.3}$s^{-1}) is the kinetic rate constant, ρ_m is the molar density of the carrier particle (32,811 mol/m^3), and r_g is the particle grain radius (2.6e-7 m) within the OC particle. The value of n is 1.3.

The kinetic constant k follows an Arrhenius type relationship with temperature, T (K).

$$k = 0.0008\exp\left(\frac{-49}{RT}\right) \tag{3}$$

Since the system operates at high temperature and low pressure, the ideal gas law was used for the state variables in the gas phase. The heat capacity of both the gas and solid phases are calculated using the standard Shomate correlations (Chase, 2008). Gas phase diffusivity (Wilke method), thermal conductivity (Wassilijewa equation) and viscosity (Wilke method) correlations are obtained from Poling et al., 2001.

2.3. Model implementation

The CLC fuel reactor model is built by integrating two standalone models: the bed hydrodynamic model, and the reaction kinetics and physical property model. This approach allows the fuel reactor model to be adapted to study other OCs by changing the reaction and physical property model. It also allows the CLC fuel reactor model to be integrated into large scale flowsheets for systems optimization.

The resulting system of DAEs is discretized using Pyomo's automatic discretization tool, pyomo.dae (Nicholson et al., 2017), with orthogonal collocation on finite elements using the Lagrange-Radau method. 16 finite elements with 3 collocation points per element were used. The resulting discretized model consists of 12,710 variables and

12,710 equations. The model is solved using IPOPT on an Intel Xeon CPU with 31 GB RAM. The solution times (including initialization) are 15.52 s for the co-current configuration and 15.70 s for the counter-current configuration.

3. Results

The primary application for this model is the development and optimization of advanced CLC systems. Feed conditions and fluidization properties used for this simulation case study (see Table 1) are obtained from a techno-economic study (Keairns et al., 2014) for a commercial scale industrial, natural gas-fueled, steam generation plant. It is assumed that the natural gas contains only methane and that the combined natural gas and fluidization gas (steam and CO_2) streams are pressurized to 400 kPa and preheated to 800 K before entering the fuel reactor. The fuel reactor is assumed to have a diameter of 5.57 m, which is comparable to the industrial scale values used in Keairns et al. 2014, and the bed height (22.14 m) is selected such that the gas outlet pressure is close to atmospheric conditions.

The profiles of select variables of a simulated co-current configuration are shown in Figure 1. In the co-current configuration both the gaseous fuel and solid OC are fed at the bottom of the bed ($x = 0$) and exit at the top of the bed ($x = 1$). From Figure 1(a), it can be seen that the reaction is very fast in the first 10 % of the bed since most of the methane conversion takes place in the lower regions. There is also a noticeable variation between the mole fraction and temperature profiles (Figure 1(b)) in the bubble and emulsion regions of the bed suggesting that there are mass transfer limitations and inefficient mixing of gases between those regions. This can be partly explained by an increase in gas bubble size from 0.02 m at the bottom of the bed to 3.05 m at the top of the bed as shown in Figure 1(c). As bubble size increases mass transfer resistances

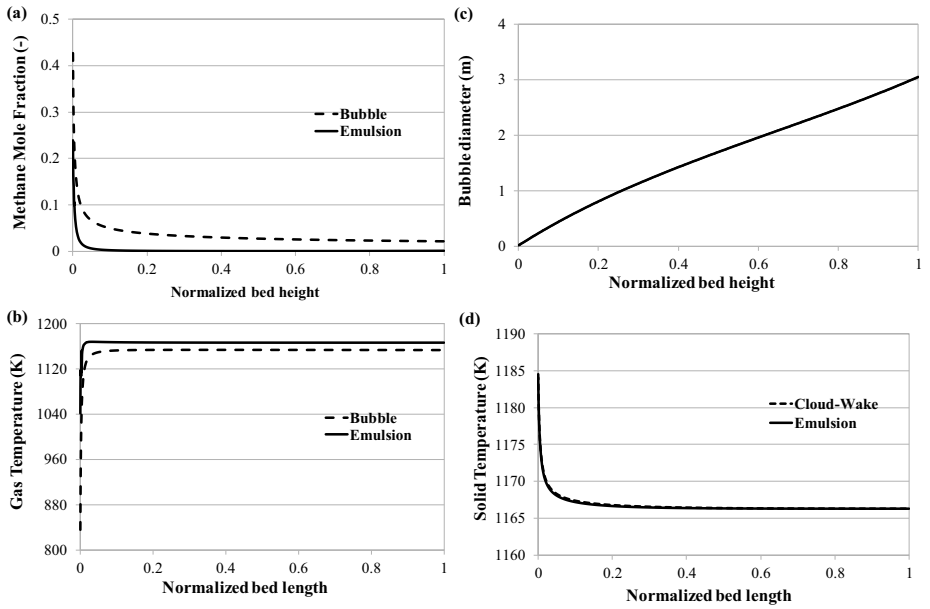


Figure 1. Profile data of selected variables vs. bed height (a) Methane mole fraction (b) Gas velocities (c) Gas Temperatures (d) Solid Temperatures

between the bubble and the cloud-wake and emulsion regions increase; thus gas mixing between regions becomes poorer. The temperature profile of the solid phase (Figure 1(d)) shows that the temperature drops from 1186 K at the inlet to 1166 K at the outlet because the reaction is endothermic. Note also that the temperature profiles of the cloud-wake and emulsion regions are virtually identical, suggesting that good mixing occurs between these two regions of the solid phase.

The sensitivity of the co-current configuration's methane conversion to some design parameters was also evaluated. It was found that increasing the bed diameter by 9 % leads to a 2.1 percentage point increase in methane conversion while reducing the bed diameter by 9 % leads to a 2.7 percentage point decrease in methane conversion. Furthermore, a 24 % increase in OC flow rate results in a 1.1 percentage point increase in methane conversion, while a 24 % decrease in OC flow rate results in a 2.3 percentage point decrease in methane conversion. These results suggest that the methane conversion is more sensitive to changes in bed diameter than in OC flow rate, indicating that residence time is more important than the OC/fuel ratio for improving fuel conversion.

Table 2 summarizes the model results for co-current and counter-current configurations. In the counter-current configuration, the solid OC is fed at the top of the bed ($x = 1$) and exits at the bottom of the bed ($x = 0$), thus both configurations will have different solid temperature and composition profiles. For the given design conditions, 91.4 % methane conversion is achieved, with 43 % of the OC converted for the co-current configuration while 93.28 % methane conversion and 44 % OC conversion is achieved for the counter-current configuration. The close similarity between the values for both configurations suggest that they have similar gas/solid residence times, and that the impact of good mixing within the bed as a result of fluidization dominates compared to the impact of the solid feed/exit locations. Thus, other design considerations such as controllability or ancillary unit setup need to be evaluated before any configuration selection decisions can be made.

The kinetic model has been validated at small scales (Abad et al. 2007), and the hydrodynamic trends of the model such as the bubble diameter profile are similar to those available in literature (Lee and Miller 2013). On-going future work is focusing on using pilot scale data to validate the model's hydrodynamic parameters and the kinetic parameters of oxygen carriers of interest for a wide range of operating conditions. The resulting robust, validated model will be used in the IDAES PSE framework to optimize the design and performance of a complete CLC system consisting of the fuel and air reactors, as well as ancillary equipment such as turbines and heat exchangers.

Table 2. Comparison between co-current and counter-current configurations

	Co-current configuration	Counter-current configuration
CH_4 Conversion (%)	91.38	93.28
OC Conversion (%)	43.03	43.93
Exit Gas Pressure (kPa)	107.58	108.41
Exit Gas Flowrate (mol/s)	547.75	552.93
Exit Solid Flowrate (kg/s)	1,227.17	1,227
Exit Gas Temperature (K)	1,153.33	1,165.77
Exit Solid Temperature (K)	1,166.29	1,165.72

4. Conclusions

This work discusses the development of a fluidized bed CLC fuel reactor model suitable for use in the large scale optimization of CLC systems using the IDAES PSE framework. The model was applied to the simulation of a CLC fuel reactor and shows that in a co-current configuration most of the fuel conversion occurs in the lower regions of the bed due to very fast reaction rates. Furthermore, methane conversion is more sensitive to changes in bed diameter than to the OC/fuel ratio. Another important finding was that good gas/solid mixing within the bed from the fluidization leads to negligible performance differences between co-current and counter-current configurations.

Acknowledgements and Disclaimer

This research was supported in part by an appointment to the National Energy Technology Laboratory Research Participation Program, sponsored by the U.S. Department of Energy and administered by the Oak Ridge Institute for Science and Education. KeyLogic Systems, Inc.'s contributions to this work were funded by the National Energy Technology Laboratory under the Mission Execution and Strategic Analysis contract (DE-FE0025912) for support services. This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

References

A. Abad, J. Adanez, F. Garcia-Labiano, L. F. de Diego, P. Gayan, and J. Celeya, 2007, Mapping of the range of operational conditions for Cu-Fe-, and Ni-based oxygen carriers in chemical-looping combustion, Chemical Engineering Science, 62, 533 – 549.

J. Adanez, A. Abad, F. Garcia-Labiano, P. Gayan, and L. F. deDiego, 2012, Progress in Chemical-Looping Combustion and Reforming Technologies. Progress in Energy and Combustion Science, 38, 2, 215 – 282.

W. E. Hart, C. D. Laird, J. P. Watson, D. L. Woodruff, G. A. Hackebeil, B. L. Nicholson and J. D. Siirola, 2017, Pyomo – Optimization Modeling in Python, 2nd ed., Springer, Vol. 67.

M. W. Chase, 1998. NIST – JANAF Thermochemical Tables. Journal of Physical and Chemical Reference Data, Monograph 9, 1 – 1951, 4th ed.

D. Keairns, R. Newby, V. Shah, M. Turner, M. Woods, A. Zoelle, S. Ostheim, S. Carpenter, G. Richards, D. Straub, and R. Breault, 2014, Industrial Carbon Management Initiative: Chemical Looping Combustion Techno-Economic Study, U.S. National Energy Technology Laboratory. URS-RES-1100564/002.

D. Kunii and O. Levenspiel, 1968. Bubbling Bed Model: Model for the Flow of Gases through a Fluidized Bed, Industrial & Engineering Chemistry Fundamentals, 7, 3, 446 – 452.

A. Lee, and D. C. Miller, 2013, A One-Dimensional (1-D) Three-Region Model for a Bubbling Fluidized-Bed Absorber. Industrial & Engineering Chemistry Research, 52, 469 – 484.

B. L. Nicholson, J. D. Siirola, J. P. Watson, V. M. Zavala, L. T. Biegler (2017). Pyomo.DAE: A modeling and automatic discretization framework for optimization with differential and algebraic equations. Mathematical Programming Computation, 10.1007/s12532-017-0127-0

B. M. Poling, J. M. Prausnitz, J. P. O'Connell (2001). The properties of gases and liquids. Mc-Graw Hill, 5th ed.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50039-2

Study on the Measurement of Air Resistance of Plate Fin Heat Exchanger in an Open Environment

Jian Wang[a*], Tong Chen[a], Jingqing Wang[a], Mingming Che[b], Shu Jiangzhou[a]

[a]*College of Automation, Hangzhou Dianzi University, Xiasha Higher Education Zone, Hangzhou 310018, China*
[b]*Hangzhou Oxygen Plant Group Co., LTD., No.592, North Zhongshan Road, Hangzhou 310014, China*
wj@hdu.edu.cn

Abstract

A test system for the air resistance of plate fin heat exchanger (PFHE) in an open environment was developed. A combination of three nozzles is used to measure the air volume flowrate through the PFHE. The combination of three nozzles keeps the pressure difference between their two ends less than 800Pa, therefore a high precision measurement is implemented. The influence of temperature & humidity on the density of wet air in the open environment is discussed, and the variety of Reynolds number with the change of humidity is analysed. A compensation formula is present to cope with the fluctuation of temperature and humidity. An expert fuzzy-PID controller is developed to control the air volume flowrate with a faster speed and higher precision. The results show that the performance of the proposed measurement system is satisfied.

Keywords: open environment, plate fin heat exchanger, high precision measurement, air resistance, fuzzy-PID control.

1. Introduction

Because of the advantages of compact structure, high heat transfer efficiency, lightweight, the plate fin heat exchanger (PFHE) has been wildly used in the fields such as air separation, petrochemical industry, refrigeration, liquefaction, and so on (M. Liu, 2007). The most significant features of the PFHE is its air resistances under the certain airflows on the standard working condition. A high precision test of its air resistance is the most necessary step for PFHE before it leaving the factory (NB/T47006, 2009). However, most of the PFHEs are too large to be put into the standard closed test room. It is a challenge to make a high precision test for those large PFHEs in an open environment (X. Li, 2011). This paper focuses on how to make a high precision test of air resistance of PFHE in an open environment. There are two mainly parts, one is how to get a high precision measurement of air volume flowrate, and another is how to make a high precision, quickly control of air volume.

2. The test system for PFHE

The structure of test system (air volume flowrate: 1000~6000m^3/h) is shown in Figure 1. A standard measuring box with three nozzles with different throat sizes, 110mm, 150mm, 189mm, respectively, is used to measure the air volume flowrate, and an induced draft fan is used to produce the required air volume by changing the speed of

motor with a frequency conversion. When the air volume flowrate is adjusted to the design value, the pressure difference between the import and export of the PFHE will be its air resistance.

A fast speed control is necessary for the test system, so that the system can step into the steady state quickly. Meanwhile, a high precision measurement of air volume flowrate is required to as the feedback information to the control system.

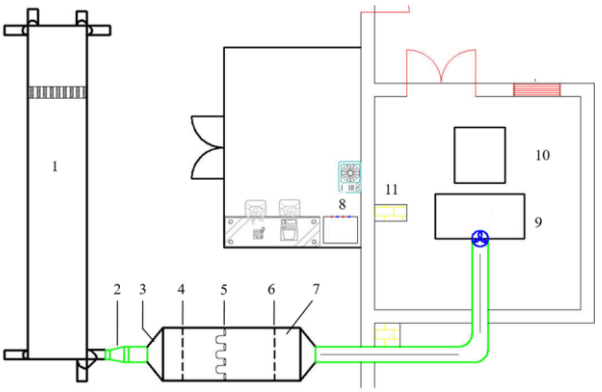


Figure 1. Structure of test system of air resistance of PFHE.
1—PFHE; 2—Adjustable pipeline; 3—Conical diffuser; 4，6—Uniform flow plate; 5—Nozzles; 7—Standard air volume measuring box; 8—Control system; 9—Induced draft fan; 10—Frequency conversion motor; 11—Frequency converter.

3. High precision measurement of air volume flowrate

The general expression for the volume flowrate, Q_v, through the measuring box is given by Eq.(1) (ISO 5167-3, 2003). By measuring the pressure difference Δp between the inlet and outlet of the nozzles, volume flowrate is given.

$$Q_v = \frac{C}{\sqrt{1-\beta^4}} \varepsilon \frac{\pi}{4} d^2 \sqrt{\frac{2\Delta p}{\rho_T}} \tag{1}$$

Where, β is the radio of the throat diameter of nozzles d to the internal diameter D of the measuring pipe upstream of the primary device; C is the discharge coefficient; ρ_T is the upstream air density; ε is the expansibility factor, when $\Delta p/P_\alpha \leq 0.4$ (P_α is the pressure of upstream air), $\varepsilon \approx 1$.

When more than one nozzles are used to measure the volume flowrate, the total volume flowrate Q_t is the sum of all nozzles, as Eq.(2).

$$Q_t = \sum Q_i \tag{2}$$

The discharge coefficient, C, is given by Eq.(3) (ISO 5167-3, 2003).

$$C = 0.9900 - 0.2262\beta^{4.1} - (0.00175\beta^2 - 0.0033\beta^{4.15})\left(\frac{10^6}{Re_D}\right)^{1.15} \tag{3}$$

Where, Re_D is pipe Reynolds number, is given by Eq.(4).

$$Re_D = \frac{vd}{\gamma} \tag{4}$$

Where, v is the velocity of air, and γ is the kinematic viscosity of fluid.

The air density ρ_T and discharge coefficient C change when the environment changing, especially when the fluid is not a single component.

4. High precision control of air volume

The high precision control of air volume flowrate is another essential technology in this test system, which is required not only high precision close to the set point, but also high speed to get the steady state. An expert fuzzy-PID method is presented to meet the requirements.

The expert system refers to that many experiments have been done to get the relationship between several common set values of air volume and the frequency when the test system was set up. Then, we can fit the relationship between the air volume and the frequency by using these data. When adjusting the air volume flowrate to a certain value, the fitted frequency getting by the experiment is used to get the initial value of air volume. Therefore, an approximate value of the set point is obtained very quickly, instead of getting the set value slowly from zero by using a standard PID. Then, a fuzzy PID algorithm is used to get the high precision control of air volume under the changing working condition. The fuzzy method is used to obtain the parameters of PID algorithm dynamically in a changing working condition, especially in different seasons in Hangzhou. By this way, a quick, high precision control method of air volume is realized.

5. Simulation results and discussion

This paper mainly discuss the condition of wet air. The temperature, humidity and components of wet air are the influence factors of wet air on the air volume flowrate.

5.1. Effect of Temperature and Humidity on Air Volume Flowrate

The density of wet air is given by Eq.(5).

$$\rho_\alpha = \frac{1000\ (P_\alpha\text{-}0.378P_V)}{287\Theta} = 3.48\frac{P_\alpha}{\Theta}\left(1 - 0.378\frac{\psi P_b}{P_\alpha}\right) \tag{5}$$

Where, Θ is the absolute temperature of wet air; P_V is the partial pressure of steam in the wet air, $P_V = \psi \cdot P_b$, ψ is relative humidity of wet air; and P_b is the saturated vapor pressure; P_α is the air pressure under test conditions.

In the open environment, the wet air, which flows through the PFHE, is not on the standard condition (0 ℃, 1 atm, ψ = 0%). So, when calculating the air volume flowrate, the effect of temperature, pressure and humidity should be taken into account.

However, the pressure is changed just a little tiny from 1 atm when in a natural open environment (less than 0.5 %), so we could ignore it and focus on the other two aspects: temperature and humidity. Figure 2 and Figure 3 give the effects of temperature and relative humidity on the air volume flowrate respectively: the higher the temperature and humidity, the greater the error. In the typical working condition in the rain season in South China, when the temperature up to 45 ℃ from 0 ℃ (standard condition), the error will grow up to 7.5 %.

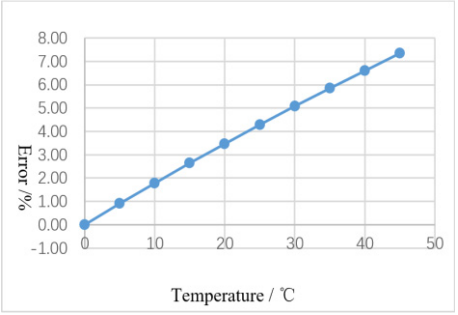


Figure 2. Effect of temperature on the error

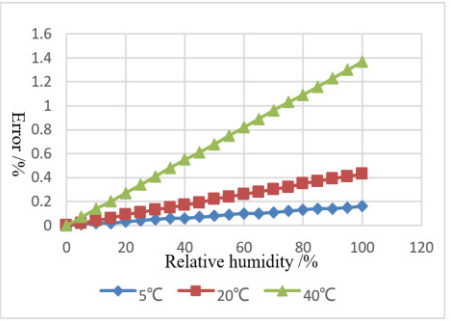


Figure 3. Effect of relative humidity on the error

Where, the effect of temperature and humidity on the error is different: the former is higher than the latter. Figure 3 is the effect of relative humidity at 5 ℃, 20 ℃, and 40 ℃ respectively, which are mostly the temperatures of winter, spring/fall, and summer in the South China. Those two figures show that the temperature has higher effect on error than the humidity, so the temperature correction is necessary when measuring the volume flowrate, but the humidity is not.

5.2. Effect of Air components on Air Volume Flowrate

When calculating the volume flowrate by the Eq.(1), the pipe Reynolds number, Re_D, is needed. But the air components effect the dynamic viscosity, μ, and the kinematic viscosity, γ, of the wet air, shown by Eq.(6) and Eq.(7) respectively.

$$\mu = \frac{\sum_{i=1}^{n} \alpha_i M_i^{1/2} \mu_i}{\sum_{i=1}^{n} \alpha_i M_i^{1/2}} \tag{6}$$

$$\gamma = \frac{\mu}{\rho} \tag{7}$$

Where, α_i is the volume percentage of gas component i in the mixed gas; M_i is the molecular weight of gas component i; μ_i is the dynamic viscosity of gas component i.

In this paper, the fluid is wet air, which is mixed with steam and dry air. The molecular weight of dry air and steam are 28.959 and 18.016 respectively. The percentage of dry air and steam are effected by the relative humidity of wet air, shown as Eq.(8) and Eq.(9) respectively.

$$\alpha_1 = \frac{V_1}{V_1+V_2}\times 100\% = \frac{P-\psi P_b}{P}\times 100\% \tag{8}$$

$$\alpha_2 = \frac{V_2}{V_1+V_2}\times 100\% = \frac{\psi P_b}{P}\times 100\% \tag{9}$$

Changing of component in wet air makes the volume flowrate different from the standard working condition.

For example, there we use the data from an abominable working condition obtained from the test site in Lin'an, Zhejiang Province, South China on Jul. 9, 2016: relative humidity ψ = 90 %, ambient temperature t = 28 ℃, ambient atmospheric pressure p = 100,600 Pa.

By calculating the Q_α by Eq.(1)~(4) under the real working condition, and compared with the standard working condition Q_s the measuring error is shown by Eq.(10).

$$\varepsilon rror = \left(\frac{Q_\alpha - Q_s}{Q_s}\right)\times 100\% = \left(\frac{5.470.90-5428.47}{5428.47}\right)\times 100\% = 0.78\% \tag{10}$$

It is known that the affection of component is far less than the temperature and relative humidity, so it can be ignored in some non-strictly required test.

5.3. High precision control of Air Volume Flowrate

An expert fuzzy-PID is used to get the fast high precision control. Here, we use the transfer function of controlled object in this test system in Lin'an, Hangzhou, which is described by Eq.(11).

$$G_o(s) = \frac{2.9386s^2 + 0.5877}{s^3 + 0.4326s^2 + 0.0665s + 0.0047} \tag{11}$$

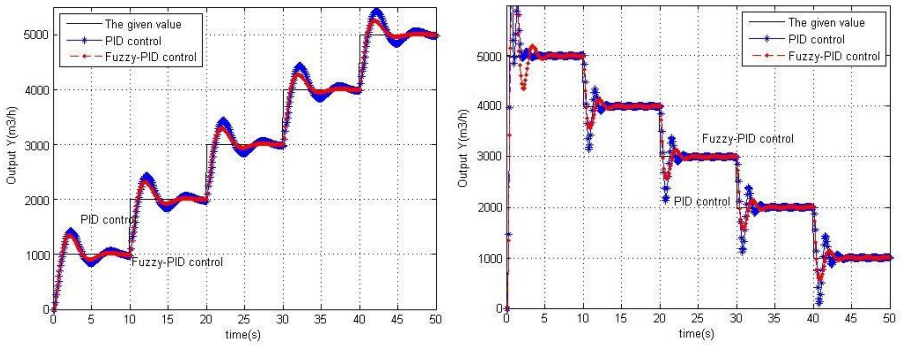


Figure 4: Comparison between fuzzy PID and traditional PID control

Two kinds of adjustment, one goes up from low value to some certain set points and another goes down from high value to the set points, to validate the algorithm (shown by Figure 4).

The fuzzy-PID adjusts the PID parameters automatically when the condition is changing by using fuzzy law. Comparing with a traditional PID ($K_p(0) = 0.07$, $K_i(0) = 0.05$, $K_d(0) = 0.12$, self-tuning by Ziegler-Nichols law and not changing), the fuzzy-PID has a faster turning period and less overshoot. Fuzzy-PID is easier to get the steadiness than traditional PID, especially when the setting value is changed.

6. Conclusions

When test the air resistance of PFHE in an open environment, a fast high precision control is necessary. A high precision measurement of air volume flowrate is the basis of the control. This paper discusses the factors that effect on the high precision measurement, and the compensation formulas are given. An expert fuzzy-PID method is presented to make an advanced adjustment process, which has a shorter regulation time and less overshoot than the traditional PID.

Conflict of interest: The authors declared that they have no conflicts of interest to this work.

Acknowledgments: This work was supported by the Natural Science Foundation of China (61374142) and the Natural Science Foundation of Zhejiang (LY16F030006).

References

ISO 5167-3, 2003, Measurement of fluid flow by means of pressure differential devices inserted in circular cross-section conduits running full-Part 3: Nozzles and Venturi nozzles, International Standard

M. Liu, N. Li, and Q. Dong, 2007, Research progress of plate fin heat exchanger, Chemical Equipment and Pipeline, 44, 6, 9-12.

NB/T47006, 2009, Aluminum plate fin heat exchanger, 2009, Industry Standard of China

X. Li, S. Jiangzhou, A. Jiang, X. Wu, and J. Wang, 2011, Development of test and controlling system on the large plate-fin heat exchanger gas flow resistance characteristics, Proc. of 30th CCC, 6, 22-24

L. Kong, 2014, Engineering Fluid Mechanics(3rd Edition)

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50040-9

Systems Platform as a Synthesis and Decision Tool in Algae Biorefineries

Melina Psycha, Filopoimin Lykokanellos, Antonis C. Kokossis

School of Chemical Engineering, National Technical University of Athens, Zografou Campus, 9, Iroon Polytechniou Str., GR-15780 Athens, Greece

akokossis@mail.ntua.gr

Abstract

The collection and exploitation of emerging data regarding algae systems is a difficult and challenging task, which requires coordination between the different sources. Relations between algae feedstocks, technologies and products are ambiguous at best and the ones that are established need to be documented. The systems platform presented in this paper deals with the challenges and sets a benchmark for a wide reange of products and pathways. Drawing on the platform, relations between the elements of the database are formed on a conceptual level indicating feasible solutions to the respective problem. Ontology engineering is the tool to achieve classification of the elements as well as to form connections between them based on provided properties. In combination with optimization and knowledge engineering, the platform offers high-capacity screening with featuring criteria (techno-economic, environmental etc.), hosts models of different scales, enables bounds on the number of options to screen and delivers the best set of solutions according to each different criterion.

Keywords: systems platform, synthesis, algae, biorefineries, ontologies

1. Introduction

Over the last decades, algae attract plenty of research due to their versatility in terms of products and their flexibility in terms of integration with other types of biorefineries. The plethora of algal strains can provide a multitude of specialty chemicals of pharmaceutical, nutraceutical and cosmeceutical use as well as products that can serve the aquaculture sector. Among them are products of high market value and compounds that cannot be produced by another natural source in a specific form.

As the incentives for the design of algae biorefineries are strengthened, the challenges that need to be overcome become more evident. The specialty products included in algae value chains have short life cycles, volatile markets and equally volatile or even unknown prices. Moreover, the algae value chains are characterized by competitive and alternative chemistries and include chemicals with unknown thermodynamic properties. The lack of industrial applications and the consequent lack of information on industrial scale only add to the challenges of the incentive. The majority of the data derive from lab-scale or pilot facilities, which on the one hand provides useful information but on the other hand, it is difficult to obtain the information due to the multiple sources. The scattered data can be experimental or even empirical and their collection and exploitation is a difficult and challenging task, which requires coordination between the different sources.

The imperative need for a centralized repository in the field of algae biorefineries is evident. Such ventures have been implemented for the case of lignocellulosic (Siougkrou and Kokossis, 2016) and waste (Barla et al., 2016) biomass systems. However, the unique nature of algae systems requires a holistic approach to the problem. The systems platform presented in this paper addresses the challenging task of collecting and exploiting emerging data in algae systems. The platform exploits Information and Communication Technology (ICT) Infrastructures to provide users with various online services in regard to the synthesis of algae biorefineries.

2. Multi-objective methodology

2.1. Ontology engineering

Ontologies are considered a formal way to describe taxonomies and classification networks, essentially defining the structure of knowledge for various domains: the nouns representing classes of objects and the verbs representing relations between the objects. Ontologies have been successfully used for engineering purposes (Brandt et al., 2008; Cecelja et al., 2015), especially because of their dynamic characteristic, which allows for any data to be updated without redesigning the database itself. The platform exploits semantics and ontology engineering to express the system flows, process components and their relationships and finally to represent the synthesis options. An ontology which formulates the value chain of an algae biorefinery has been designed using Protégé, a free, open-source ontology editor and framework for building intelligent systems. Figure 1 illustrates the basic classes and subclasses of the constructed ontology; the key classes are feedstock, technologies, products and intermediates.

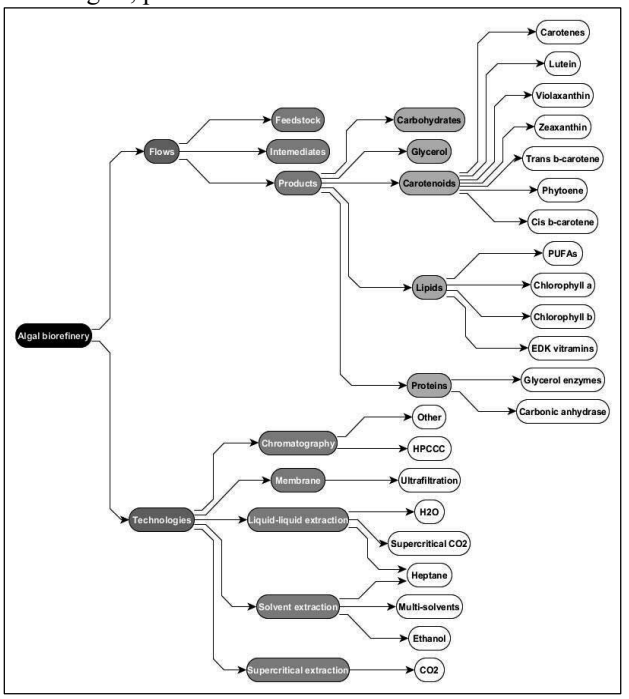


Figure 1: Platform ontology classification (colored background: classes and sub-classes; no background: individuals)

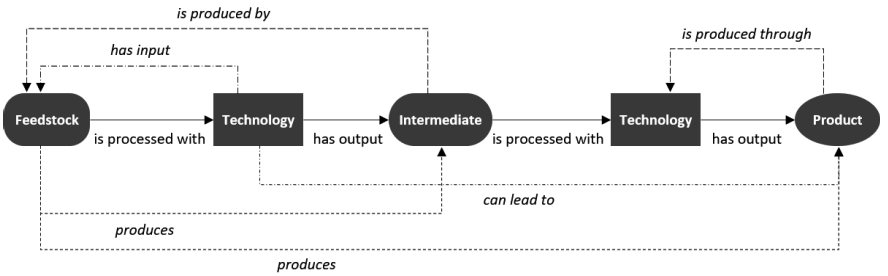


Figure 2: Object properties for a generic synthesis pathway

The ontology main components include the basic or "ground level" objects or instances, which represent the *individuals*, the sets, collections, concepts or types of objects, which are the *classes*, and the aspects, properties, features, characteristics, or parameters that objects (and classes) can have, which are the *attributes*. The attributes which relate individuals to individuals are called object properties and the ones which relate individuals to literals are categorised as datatype properties. The object properties defined for the representation of a typical synthesis pathway are depicted in Figure 2. For each individual only has_input and has_output properties are explicity used to represent a relation. All the other statements are implicity constructed through the first-order logic rules and axioms provided by the OWL 2 language.

2.2. Knowledge engineering and optimization

The platform repository includes all data and information regarding the microalgal value chain (feedstock, intermediates and products), the involved processes for the extraction or conversion of compounds to products and the relevant economic parameters for the final evaluation of processing paths (Psycha and Kokossis, 2017). The models incorporated into the repository come from different sources (literature, experimental, customized models) and as a result are of different types and scales. Specifically, they can range from a detailed equation-based model to a conceptual black box with basic input-output data. The same case applies to physical property models; thus, it is of high importance for the repository to be able to accommodate all available data. Specific information is filled and stored in the database, which is the same for every model regardless of its complexity. In that way, all models provide the required use, which can vary depending on the detail of the respective model.

The concept of the platform is to include all available information without focusing on specific results. On that note, all possible options are incorporated into the background modelling; however, these options are not exhaustive, since some combinations may be encouraged or prohibited due to practical limitations. The background modelling follows the guidelines of the synthesis model developed in previous work (Psycha and Kokossis, 2016) according to which all paths are screened simultaneously and evaluated to yield optimal solutions. With the use of mathematical techniques (integer cuts), cluster of solutions can be retrieved to the user's convenience.

3. Systems platform

3.1. Architecture

The systems platform has been developed upon Yii framework, a high-performance PHP framework for creating Web 2.0 applications. Yii is considered to be very fast and secure featuring the Model-View-Controller (MVC) software design pattern. Thus, the platform architecture is MVC based and it consists of two separated layers, the back-end and the front-end. The former contains two types of data collections and the system controllers. The first type data structures are stored into a database using the relational model and managed by the MySQL database management system. The second ones lie on a RDF triple store powered by Ontotext GraphDBFree, a highly-efficient graph database used as a semantic repository for the platform ontology. The front-end provides the user interface and its functionalities in conjunction with the back-end. It is implemented using the HTML, CSS and JavaScript languages and two powerful components, Bootstrap and yFiles for HTML. yFiles for HTML is a JavaScript diagramming for analyzing, drawing and arranging graphs. yfiles enables the graphical visualization of the synthesis pathways.

3.2. Services

The systems platform offers services related to synthesis as well as graphical representation of possible options (superstructure). The user is able to select between the formulation of value chain and/or the formulation of processing paths as well as define the starting point of the search (feedstock, technologies, products). Moreover, the platform dynamically builds and visualizes the processing pathways using modelling components and displays the paths by utilizing the constructed process designs and integrated schemes. The model screens for optimal solutions and the paths are highlighted on the superstructure scheme thus facilitating the user. The corresponding output data are calculated and economic information on each solution path are available. In addition, charts are outlined including costs breakdown and product/technology percentage frequencies.

The dynamic nature of the platform supports the addition of data and methods; thus, it is constantly being updated as new information and research emerge. The goal is also to activate user-driven entries for prices and other parameters in order to produce customized solutions according to provided data and deal with uncertainties (uncertainty analysis).

4. Benchmark implementation

The platform offers the option of building the value chain of the selected feedstock, which includes potential fractions and products from the specified alga without the processing technologies. Presently, three algae are incorporated (Dunaliella, Haematococcus and Nannochloropsis) and more will be added as the database grows. Figure 3 illustrates the value chain formulation of Dunaliella and the respective superstructure (processing paths formulation), which includes the suitable technologies to recover the desired products. All the blocks of the superstructure correspond to the individuals listed in Figure 1. For the formulation of the processing paths, the user has the ability to select any combination of technologies and/or product types and if the corresponding connection exists in the ontology, it will appear in the demonstration area of the platform. For the purpose of this paper, only the feedstock is selected, thus all possible options appear.

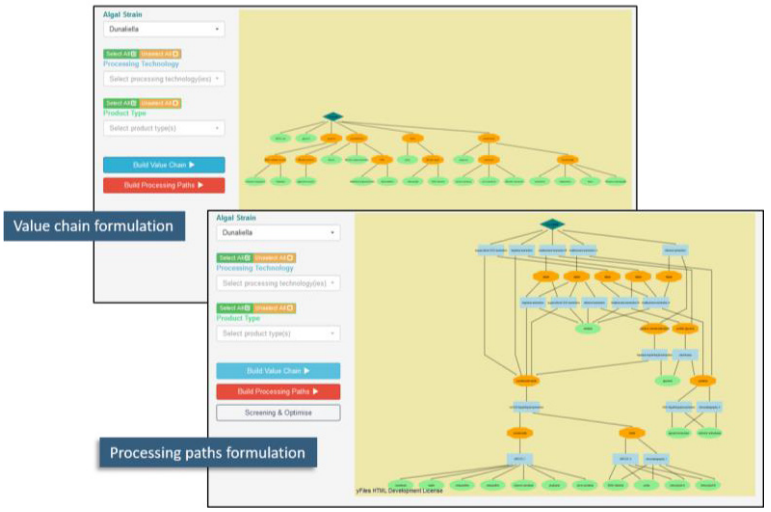


Figure 3: Platforms representation of value chains and processing paths

The key aspect of the platform is that it works as a screening and optimization tool to be used ahead of design and provide designing options according to the users input. As shown in Figure 4, the best ten solution paths are produced (in terms of profitability) in a decreasing order, which allows the user to better comprehend the problem. The analytics section of the platform includes further information and analysis on the results for each produced processing path as well as for the entire group of solutions (capital and operating costs, revenue, product and process frequency). One can select a solution and observe the graphical representation of each path, as depicted in Figure 4. Specifically, each solution path is highlighted to indicate the selected technologies, intermediates and products.

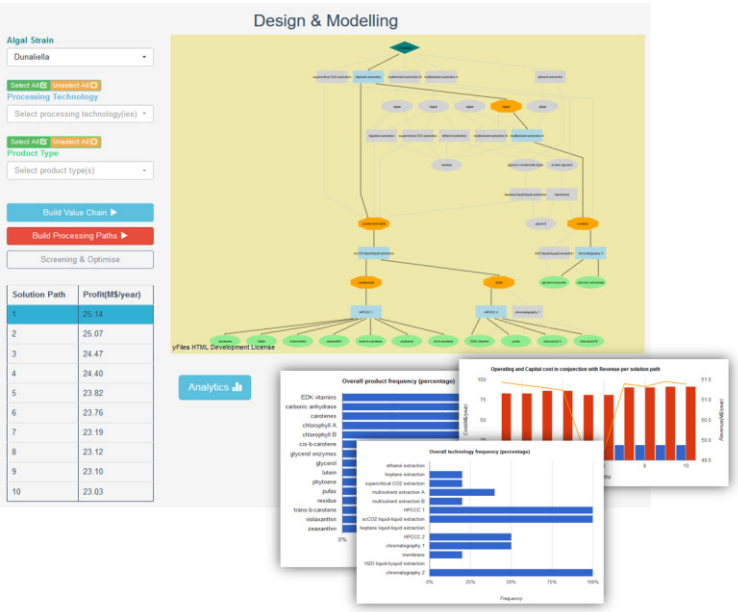


Figure 4: Screening, optimization and analytics of the platform

5. Conclusions

The systems platform for algae systems constitutes a powerful modelling tool, which dynamically builds and graphically represents value chains and processing paths from algal strains. Moreover, high-throughput screening and optimization are featured, and multiple solution paths are produced highlighting the flexibility of the platform as well as the importance of previewing opportunities ahead of detailed designs. The respective numeric results are calculated (profit, capital and operating cost, revenue) and are graphed in column charts. Furthermore, the product and technology frequencies are estimated taking into consideration all solution paths. The incorporation of other strains and the integration between various value chains is a promising venture and can yield significant results in the future.

Acknowledgements

Collaborative project D-Factory supported by the 7th Framework Programme (FP7) for research and development is acknowledged for input data and financial support.

References

E. Siougkrou, A. Kokossis, 2016, Development of Semantically-enabled Community Hubs in Biorefineries and Biorenewables, Computer Aided Chemical Engineering, 38, 2013-2018

F. Barla, F. Lykokanellos, A.C. Kokossis, 2016, Discovering Valorisation Paths in Waste Biorefineries Using an Ontology Engineering Approach, Computer Aided Chemical Engineering, 38, 2079-2084

F. Cecelja, N. Trokanas, T. Raafat, M. Yu, 2015, Semantic Algorithm for Industrial Symbiosis Network Synthesis, Computers and Chemical Engineering, 83, 248-266

M. Psycha, A.C. Kokossis, 2016, Synthesis and Optimization of Microalgae Biorefineries, Computer Aided Chemical Engineering, 38, 325-330

M. Psycha, A.C. Kokossis, 2017, Techno-economic Evaluation of an Integrated Microalga Biorefinery Targeting the Co-production of Specialty Chemicals, Computer Aided Chemical Engineering, 40, 1981-1986

S.C. Brandt, J. Morbach, M. Miatidis, M. Theissen, M. Jaeke, and W. Marquardt, 2008, An Ontology-based Approach to Knowledge Management in Design Processes, Computers and Chemical Engineering, 32, 320-342

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50041-0

Superstructure Optimization of Oleochemical Processes with Surrogate Models

Mark Jones[a,b*], Hector Forero-Hernandez[a,b], Alexandr Zubov[b], Bent Sarup[a], Gürkan Sin[b]

[a]*Edible Oil System Business Unit, Alfa Laval Copenhagen A/S, 2860 Søborg, Denmark*
[b]*Process and Systems Engineering Center (PROSYS), Technical University of Denmark, Building 229, 2800 Lyngby, Denmark*
mark.jones@alfalaval.com

Abstract

In this work we present a framework to generate surrogate models from rigorous process models embedded in a modelling environment or a process simulator. These unit operations (i.e. process flowsheet subsystems) are treated as black-box models to generate data for fitting and deriving the surrogate. Further, the methodology includes the formulation of a superstructure optimization problem and solving it to identify the optimal process flowsheet structure and point of operation from the possible alternatives. The superstructure optimization incorporates selection and interconnection of each surrogate and multi-objective optimization in respect to total annual cost and environmental impact. In this paper we highlight the surrogate building step of the methodology with a rigorous counter-current spray column model and assess the performance of different surrogate modelling methods.

Keywords: Polynomial Chaos Expansion, Gaussian Process Regression, Optimization, Superstructure Generation

1. Introduction

Surrogate models are popular to substitute black-box models which are either computationally expensive to evaluate, have noisy output behaviour or don't supply gradients (Quirante et al., 2015). Surrogates can also be beneficial to perform e.g. sensitivity analysis or optimization through Monte Carlo simulation (Crestaux et al, 2009). Superstructure optimization is another field where surrogates can be embedded in the problem formulation to perform rigorous calculations. In this work multivariate adaptive regression splines (Friedman, 1993), polynomial chaos expansion (PCE) (Feinberg and Langtangen, 2015) and Gaussian process regression (GPR) (Rasmussen and Williams, 2006) are applied to gain insights how these different surrogate modelling techniques perform in the context of oleochemical process modelling.

2. Overview of Surrogate Modelling Methods

2.1 Multivariate Regression Splines

We assume:

$$y_j = f_j(x_1, \ldots, x_p) + \epsilon_j \quad (1)$$

as the output to input relation of the system under research where f is the approximating function and ε the error. A training data set is fit to obtain $\widehat{f_j}$ to each function f_j with the expansion of basis functions:

$$\hat{f}(x_1, \dots, x_p) = \sum_{m=0}^{M} a_m B_m(x_1, \dots, x_p) \tag{2}$$

Least-squares minimization gives the coefficients a_m weighting the basis functions which take the form of a hinge function $h(x_i)$. The linear combination of a constant, a linear function and several hinge functions is iteratively calculated with a two-step forward/backward algorithm. In the first forward step the surrogate model is being built by adding hinge functions while maximizing the goodness of fit until a user-defined level of complexity has been reached. The second pruning and pass step will compare subsets of the potential hinge functions by means of a true cross-validation score (GCV) to prevent overfitting. The subset with the lowest GCV is chosen.

2.2 Polynomial Chaos Expansion (PCE)

In case of PCE a sum of weighted multivariate polynomials represents the surrogate to be built from input-output data:

$$\hat{f}(x) = \sum_{n \in N^M} c_n \Psi(x) \tag{3}$$

To retrieve the Fourier coefficients c_n a linear regression is performed where the multivariate Legendre polynomials $\psi(x)$ are fitted to the output generated in the simulation step with uniformly distributed samples within the boundaries of the input parameters x_i.

2.3 Gaussian Process Regression

Gaussian Processes can be described as a linear combination of polynomials and basis functions where we introduce a mean function $\hat{\mu}$ for the polynomial and a covariance function (also known as kernel) $R(x_i, x_j)$ for the basis part to describe a normal distribution over a set of functions:

$$\hat{f}(x^*) = \hat{\mu}(x^*) + r^T R^{-1}(y - 1\hat{\mu}) \tag{4}$$

The predictive mean $\hat{\mu}$, variance and log marginal likelihood (LML) are obtained through matrix inversion using Cholesky factorization as described by Rasmussen and Williams (2006) with Algorithm 2.1.

3. Methodology for Surrogate-based Superstructure Optimization

The application of surrogate modelling is subject to the variation of the input parameters of the model. The steps of the methodology are presented as follows:

1. *Sampling of the Design Space*: The independent variables (i.e. design degrees of freedom) of the process under research are identified and the boundaries specified. Different experimental designs exist to cover the domain of variation such as Monte Carlo samples obtained with random generators, Latin hypercube

samples or quasi-random (low-discrepancy) sequences, e.g. Sobol, Hammersley or Halton sequences. A hypercube is generated and used for the next step to perform Monte Carlo simulation. Due to optimization (step 6) the bounds of the independent variables will get iteratively updated to ensure feasible solutions.

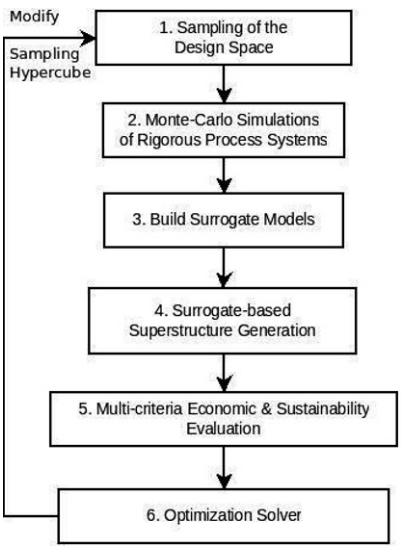


Figure 1: Methodology for superstructure generation with surrogate models

2. *Monte Carlo Simulation of Rigorous Process Systems*: The model is evaluated for each sample and the matrix of observations (model outputs) is stored for the surrogate modelling step.

3. *Build Surrogate Models*: To build a surrogate model the sampling hypercube and matrix of observations obtained from step 1 and 2 are used to apply methods such as multivariate regression splines, polynomial chaos expansion or Gaussian process regression as described in section 2 of this work.

4. *Surrogate-based Superstructure Generation*: A superstructure formulation of the interconnected surrogates and all possible combinations of process structures with operating set points are generated through enumeration or transforming a disjunctive program into a mixed integer non-linear program (MINLP).

5. *Multi-criteria Economic & Sustainability Evaluation*: The different process structures are evaluated in terms of capital, operating and total annual costs and total environmental impact is assessed to take the sustainability of a process into account.

6. *Optimization Solver*: In this step the optimization problem is solved where the economic and environmental objective function is formulated as the minimization of total annual cost and environmental impact subject to boundary conditions. To ensure that we obtain a feasible solution the bounds of the independent variables get updated every time a solution is found which violates the specified bounds defined in the previous iteration step.

4. Rigorous Counter-Current Spray Column Model

This work concentrates on the oleochemical domain where in the following case the chemicals taking part in the mass transfer and reaction phenomena are: triglycerides, fatty acids, glycerol and water.
The counter-current spray column model serves as the rigorous model to be substituted by a surrogate. The hydrolysis reaction in the column produces 3 moles fatty acids and 1 mole glycerol from 1 mole triglyceride and 3 moles water. The rigorous model comprises

a finite volume discretization of the column with a two-phase system made up by a continuous (oil) and dispersed (aqueous) phase. The water and the glycerol components in the dispersed and in the continuous phase form both a liquid-liquid equilibrium between the phases which dictates the theoretical extent of the mass transfer. The model includes back-mixing (axial dispersion) and is based on the works by Spencer et al. (1981) and Ricker et al. (1981). The base case uses the experimental data given by Jeffreys et al. (1961) in the experimental run number 6.
The design space of the model is summarized in Table 1. The same boundaries are defined to generate identical hypercubes for all surrogate modelling methods. The sample size of the hypercube is 500 and populated with Sobol sequences. The model is solved with the input hypercube and the output (overall conversion) is stored for the surrogate modelling step.

Table 1: Design parameters with design space boundaries for the counter-current spray column model

Parameter	Unit	Lower Bound	Upper Bound
Feed flow rate	kmol/h	2000	3500
Solvent to feed ratio	-	1.4	1.9
Temperature	K	473.15	533.15

5. Comparison between Surrogate Modelling Methods

The sampling and observation data was split half wise for training and testing. The surrogate models were built with the train set and the predictions were then compared with the test set. The coefficient of determination and the mean square error are given in Table 2 to compare the different surrogate modelling methods as well as the accuracy of mapping the rigorous (black-box) model to a surrogate model.

Table 2: Coefficient of determination (R^2) and mean square error (MSE) for the predictive performance of the surrogate models derived from a rigorous spray column model (black-box) with a sample size of 500

Surrogate Modelling Method	R^2	MSE
Multivariate Regression Splines	96.2	0.1e-03
Polynomial Chaos Expansion	99.7	0.55e-05
Gaussian Process Regression	99.9	0.40e-05

With x_1, x_2 and x_3 being the feed flowrate, solvent to feed ratio and temperature, the model described with multivariate regression splines is:

$$\hat{f}(x_1, x_2, x_3) = 0.1318 + 0.0016 * h(x_3 - 510.767) - 3.3521 * 10^{-8} * h(x_1 - 2.4716 * 10^6) + 6.5607 * 10^{-8} * h(2.4716 * 10^6 - x_1) - 0.0529 * h(x_2) + 0.0021 * h(x_3 - 487.095) - 0.0010 * h(487.095 - x_3) \quad (5)$$

The surrogate obtained through polynomial chaos expansion is as follows:

$$\hat{f}(x_1, x_2, x_3) = 4.3580 - 0.0230 * x_3 + 0.5354 * x_2 + 3.1109 * 10^{-5} * x_3^2 - 0.0015 * x_2 * x_3 + 4.0402 * 10^{-7} * x_1 - 1.1506 * 10^{-9} * x_1 * x_3 + 0.0418 * x_2^2 + 2.8782 * 10^{-8} * x_1 * x_2 + 1.5982 * 10^{-14} * x_1^2 \quad (6)$$

In respect to the Gaussian process regression surrogate, the covariance function was defined as a Matérn kernel:

$$R = k(r) = \left(1 + \frac{\sqrt{3}r}{l}\right) exp\left(-\frac{\sqrt{3}r}{l}\right) \tag{7}$$

Figure 2 shows how the cross-validation score, in this case R^2, behaves as a function of the number of samples used for training the model. With a training set of more than 15 samples the fitted model will give predictions with R^2 greater 95 %.

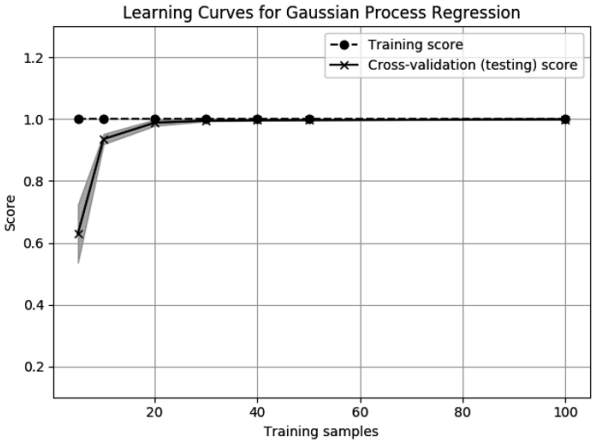


Figure 2: Comparison between the behaviour of training score and testing score as a function of the used training samples

6. Superstructure Optimization

The intention of building surrogate models from rigorous models embedded in different modelling environments and process simulators is to use the surrogates for generating and evaluating a superstructure for the most efficient design in respect to raw material and product specifications. Figure 3 shows the oleochemical process domain under research in this project. We presented in this publication the surrogate building procedure of a rigorous spray column model written in Fortran. The other rigorous models for a continuous stirred tank reactor (Fortran, Matlab), flash based glycerol purification (Pro/II), distillation column (Python) and solvent crystallization (Python) undergo the same procedure as described in the previous chapter. A superstructure can then be formulated with the models obtained with multivariate regression splines, polynomial chaos expansion or Gaussian process regression. The program will then choose between either a CSTR or spray column for the hydrolysis step, distillation or solvent crystallization for the separation of fatty acids and the number of flash purification units to recover higher graded glycerol based on economic feasibility and product specifications. The possible process structures can be generated through an enumeration approach where every possible configuration is evaluated by means of total annual cost and environmental impact. Another approach would be to formulate a disjunctive program with the surrogates and perform Big-M or convex hull transformation to obtain a MINLP to solve for.

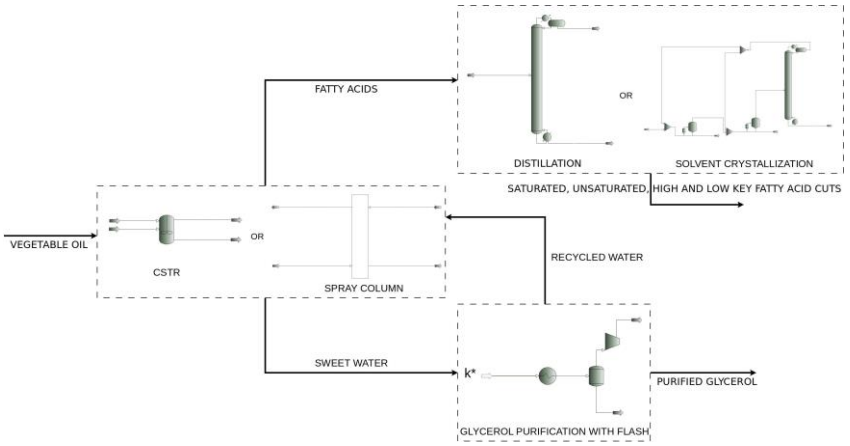


Figure 3: Schematic representation of possible process steps for the oleochemical process domain under research

7. Conclusions

We showed how superstructure optimization can be applied with rigorous models by using surrogate modelling techniques and highlighted the generation of these surrogates. Promising results have been obtained by Henao and Maravealis (2011) who used neural networks to obtain the surrogates. We, in contrast, applied multivariate regression splines, polynomial chaos expansion and Gaussian process regression to compare the different methods and evaluate in future work how the surrogates perform for superstructure optimization. The beauty of multivariate regression splines and polynomial chaos expansion are the simple models received and which are easily introduced in a disjunctive programming superstructure formulation whereas Gaussian process regression has shown to give the best predictive results compared to the two other methods and the surrogates can be used in the enumeration approach to evaluate all possible process configurations.

References

T. Crestaux, O. Le Maître and J. Martinez, “Polynomial Chaos Expansion for Sensitivity Analysis”, Reliability Engineering and System Safety, 94, 1161-1172 (2009)

J. Feinberg and H.P. Langtangen, “Chaospy: An open source tool for designing methods of uncertainty quantification”, Journal of Computational Science, 11, 46-57 (2015)

J. H. Friedman, “Fast MARS”, Technical Report No. 110, Laboratory for Computational Statistics, Department of Statistics, Stanford University (1993)

C. A. Henao and C. T. Maravealis, “Surrogate-Based Process Synthesis”, Computer Aided Chemical Engineering, 28, 1129-1134 (2010)

N. Quirante, J. Javaloyes, R. Ruiz-Fermenia and J.A. Caballero, “Optimization of Chemical Processes Using Surrogate Models Based on a Kriging Interpolation”, Computer Aided Chemical Engineering, 37, 179-184 (2015)

C. E. Rasmussen and C. K. I. Williams, “Gaussian Processes for Machine Learning”, MIT Press (2006)

N. L. Ricker, F. Nakashio and C.J. King, “An Efficient General Approach for Computation of Countercurrent Separation Processes with Axial Dispersion”, AIChE Journal, 27, 2, 277-284 (1981)

J. L. Spencer, L. Steiner and S. Hartland, “Model-Based Analysis of Data from Countercurrent Liquid-Liquid Extraction Processes”, AIChE Journal, 27, 6, 1008-1016 (1981)

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50042-2

Sequential Design of Experiments to Maximize Learning from Carbon Capture Pilot Plant Testing

Frits Byron Soepyan[a], Christine M. Anderson-Cook[b], Joshua C. Morgan[a], Charles H. Tong[c], Debangsu Bhattacharyya[d], Benjamin P. Omell[a*], Michael S. Matuszewski[a], K. Sham Bhat[b], Miguel A. Zamarripa[a], John C. Eslick[a], Joel D. Kress[b], James R. Gattiker[b], Christopher S. Russell[b], Brenda Ng[c], Jeremy C. Ou[c], David C. Miller[a]

[a]*National Energy Technology Laboratory, 626 Cochrans Mill Rd., Pittsburgh, PA 15236, USA*
[b]*Los Alamos National Laboratory, P.O. Box 1663, Los Alamos, NM 87545, USA*
[c]*Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA 94550, USA*
[d]*West Virginia University, Department of Chemical and Biomedical Engineering, Morgantown, WV 26506, USA*
Benjamin.Omell@netl.doe.gov

Abstract

Pilot plant test campaigns can be expensive and time-consuming. Therefore, it is of interest to maximize the amount of learning and the efficiency of the test campaign given the limited number of experiments that can be conducted. This work investigates the use of sequential design of experiments (SDOE) to overcome these challenges by demonstrating its usefulness for a recent solvent-based CO_2 capture plant test campaign. Unlike traditional design of experiments methods, SDOE regularly uses information from ongoing experiments to determine the optimum locations in the design space for subsequent runs within the same experiment. However, there are challenges that need to be addressed, including reducing the high computational burden to efficiently update the model, and the need to incorporate the methodology into a computational tool. We address these challenges by applying SDOE in combination with a software tool, the Framework for Optimization, Quantification of Uncertainty and Surrogates (FOQUS) (Miller et al., 2014a, 2016, 2017). The results of applying SDOE on a pilot plant test campaign for CO_2 capture suggests that relative to traditional design of experiments methods, SDOE can more effectively reduce the uncertainty of the model, thus decreasing technical risk. Future work includes integrating SDOE into FOQUS and using SDOE to support additional large-scale pilot plant test campaigns.

Keywords: Bayesian method; computational tool; space-filling design; minimizing uncertainty

1. Introduction

Pilot plant test campaigns can be expensive and time-consuming, limiting the number of test conditions that can be evaluated. Thus, a prudent choice of design schemes is needed to maximize the amount of knowledge gained from pilot plant experiments. Standard approaches to designing experiments are based on the properties of the input space, without leveraging the error in prediction of the output space (Politis et al., 2017). The

general procedure for many designed experiments follows a static approach: generate the complete set of input combinations, use them to perform the experiment, analyze the experimental results, and terminate the test campaign without adjustment. With this static approach, there is a possibility of under- or over-sampling, or placing runs in locations that provide minimal new information. Since pilot scale testing is expensive, test runs should be maximally informative. Therefore, a sequential design of experiments (SDOE) for pilot scale testing is proposed. This technique uses a combination of exploration (collecting data throughout the design space, especially in poorly-represented areas characterized by high uncertainty) and exploitation (sampling in the most advantageous regions of the known design space) (Garud et al., 2017). Intermediate results are used to determine the location of new design points by applying optimization strategies.

While SDOE can provide test guidance to improve learning, there are challenges to enable its effective use, including managing the computational burden, and incorporating the methodology into a computational tool. These challenges are exacerbated by the fact that accounting for variable uncertainty leads to large and computationally expensive models (Weaver et al., 2016). Thus, a computational tool that combines optimization, uncertainty quantification, and surrogate model development, organized into a Bayesian design of experiments framework, can enable a practical and efficient means to implement SDOE to maximize learning from experimental tests at both the bench and pilot scales.

We addressed these challenges by first using SDOE in conjunction with the Framework for Optimization, Quantification of Uncertainty and Surrogates (FOQUS) (Miller et al., 2014a, 2016, 2017), a software tool developed as part of the Carbon Capture Simulation Initiative (CCSI) (Miller et al., 2014b). FOQUS can interface with commercial process simulators to enable uncertainty quantification, optimization, and the development of surrogate models. Its ability to manage parallel execution of process simulations enables the efficient generation of reduced-order models that are essential for SDOE. SDOE uses a Bayesian method that leverages new data and process models to determine an optimal set of test conditions. This procedure can reduce the total number of experiments required to explore the input/output space of the process model, while improving the model's accuracy. Thus, the approach identifies the experiments and conditions that will provide the most benefit to increased understanding of the process, thereby reducing technical risk. The methodology has the potential to speed up deployment of new technologies at a reduced cost by improving test campaigns.

We implemented pilot scale SDOE for a solvent-based CO_2 capture plant at the National Carbon Capture Center (NCCC), where the surrogate model predicts the CO_2 capture percentage of the plant given a set of input variables. The results demonstrate reduced uncertainty compared with traditional design of experiments methods. This paper is organized as follows. Section 2 outlines the methodology for SDOE. Section 3 describes a case study used to demonstrate the methodology. Section 4 presents and discusses the results of implementing SDOE on the case study. Finally, Section 5 provides the conclusions and future direction of this work.

2. Methodology

The process for SDOE is as follows:

1.) Identify the criteria for optimization. It is possible to consider more than one criterion, and the relative importance of each criterion can be changed as the experiment progresses. Common criteria include: refining the region of interest,

reducing the uncertainty in the estimation of model parameters, reducing the uncertainty in the model's predictions for new observations in the design space, reducing the error in the model's predictions, or optimizing the value of an output response.

2.) Develop a working model of the process based on the underlying science to estimate the value of the criterion/criteria defined in Step 1.
3.) Identify the input variables or factors to be varied during the experiment, define their ranges, and identify constraints or regions in the input space that are not of interest or are intractable.
4.) Identify input combinations of the variables from Step 3 to be considered for the experimental runs during the sequential design.
5.) Develop a working model of the process that can be updated with new data. Sometimes, this model is the same as in Step 2. This model will quantify changes in the criterion/criteria from Step 1 using results from previous experiments.
6.) Develop a plan for N number of sequential batches, with the size of N determined by a realistic time frame to complete the experiments. Each batch consists of a unique subset of the candidate input combinations from Step 4. For each batch:
    a. Run each candidate input combination through the working model.
    b. Measure the responses (i.e., outputs).
    c. Update the working model from Step 5.
    d. Optimize the location of new experimental runs with respect to the criterion/criteria from Step 1.
7.) Execute the plan developed in Step 6.
    a. Identify the initial batch of experimental runs (developed in Step 6).
    b. Run this batch of experiments, update the working model with new experimental data, and use this model to select the next batch of runs.
    c. Iterate Steps 7a and 7b for the entire budget of runs.

3. Case Study

In this section, we describe the implementation of SDOE for a test campaign at NCCC. The objective is to determine the optimal input combinations for improving our surrogate model's predictions of the plant's CO_2 capture percentage (as a means to quantify the predictive capability of the model) throughout a range of different operating conditions. The surrogate model was used to compute the CO_2 capture percentage, because it is less computationally expensive than using the process simulator Aspen Plus® and offers an adequate compromise between computation time and accuracy.

NCCC, located in Wilsonville, AL, USA, is a test facility that uses flue gas provided by Plant Gaston's 880 MW coal-fired power plant operated by Southern Company (Brown et al., 2017). One of the post-combustion CO_2 removal systems used at NCCC is a solvent test unit consisting of an absorber and a stripper. For this case study, the solvent is a non-proprietary aqueous solution of monoethanolamine (MEA) (Bumb et al., 2017) that is commonly used as a baseline for comparing more advanced solvents. MEA has a fast reaction rate with CO_2 and a chemical structure that allows for high water solubility and alkalinity (necessary for CO_2 capture) (Morgan et al., 2017).

Table 1 summarizes the five available configurations of the CO_2 absorber column of the MEA system. Because of limitations on the amount of data that can be collected, we focus on the C3i configuration. The following variables were investigated: flue gas flowrate (1,000 to 3,000 kg/h), CO_2 mass fraction in the flue gas (0.125 to 0.175), lean solvent

loading (0.1 to 0.3 mol CO_2/mol MEA), and lean solvent flowrate (3,000 to 13,000 kg/h). Some combinations of the factors were not reasonable, as they would not produce meaningful results. Therefore, the region of interest was defined as an irregular region inside the space parametrized by the ranges above.

Table 1. The five configurations of the CO_2 absorber column of the MEA system.

Notation	Number of Beds	Intercooling	Number of Available Experimental Data Points	Number of New Experimental Runs
C3i	3	Yes	15	24
C3	3	No	1	4
C2i	2	Yes	2	4
C2	2	No	2	4
C1	1	No	3	8

Time and resource constraints allowed for a total of 44 new experimental runs. Given the available experimental data, the first stage in determining the location of new input combinations uses space-filling designs for each configuration. The procedure involves exploration of the region of interest (CO_2 capture percentage of 50 % to 95 %) and reduction of the uncertainty in the surrogate model's prediction. This process is conditional on the locations in the design space that has already been explored with historical data. The second stage uses SDOE to minimize the estimated width of the confidence interval of the CO_2 capture percentage. It incorporates the results from the first stage to determine the location of the next experiment. For the space-filling design, the minimax criterion (Masoudi et al., 2017) was used. This objective seeks to minimize the maximum distance between any location in the design space of interest and where data have already been obtained. Given the limited number of experiments that can be run, this condition ensures that new input combinations are located far from existing experimental conditions, which allows for a more effective coverage of the input space.

4. Results and Discussions

Figure 1a shows a comparison of the surrogate model's prediction uncertainty before and after Bayesian update of the surrogate model's parameter distributions (denoted as "Before Update" and "After Update") for the C3i configuration. The uncertainty is quantified using the width of the 95 % confidence interval (CI), where a smaller value denotes smaller uncertainty. The objective is to reduce the uncertainty of the detailed models from Aspen Plus (i.e., the hydrodynamics and mass transfer models) by reducing the size of the model's parameter space to better understand the probability of observing a given CO_2 capture percentage value within a certain range. Figure 1a shows a reduction in the uncertainty of the surrogate model's output, which was calculated to range from 22.2 % to 70.2 %. This plot highlights the benefits of additional data and sequentially updating the surrogate model with new experimental results. In Figure 1a, the first fourteen points (1 to 14) represent the width of the 95 % confidence interval of the surrogate model for the new data obtained using the space-filling design (denoted as "Stage 1"), and the last three points (15 to 17) represent the width of the 95 % confidence interval of the surrogate model after the second phase of SDOE (denoted as "Stage 2").

The performance of the surrogate model's predictions after updating the model with data from SDOE is shown in Figure 1b, where the standard deviation of the percentage error is 7.70 %. Figure 1b shows that most of the historical data (denoted as "Historical") are clustered close to the 100 % CO_2 capture percentage region, whereas the new data

obtained using the space-filling design (denoted as "Stage 1") provide better coverage of the desired range of CO_2 capture percentage. The second stage data (denoted as "Stage 2") are collected to further reduce the width of the confidence interval. They are located in the 90 % to 100 % CO_2 capture percentage region because of the large estimated uncertainty (as shown in Figure 1a) of the surrogate model's predictions in this region. More points in this region would be maximally advantageous for refining the prediction.

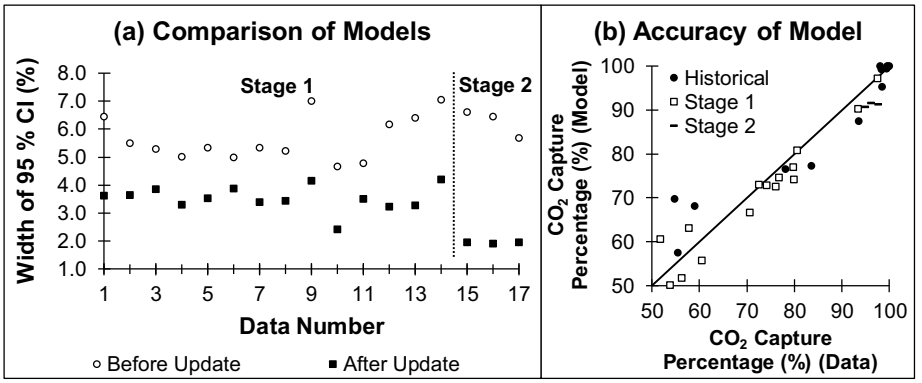


Figure 1. Performance of surrogate model: (a) Reduction of uncertainty, (b) Parity plot of CO_2 capture percentage in the absorber (solid line represents parity line).

5. Conclusions and Future Direction

SDOE was applied to the NCCC test campaign and used to update the test plan as new experimental data were obtained. The results show a reduction in the uncertainty of our surrogate model's predictions of the CO_2 capture percentage. SDOE provides a flexible method that can be adapted to multiple study objectives, and make optimal use of experimental resources. Potential improvements to SDOE include developing algorithms capable of generating accurate surrogate models with fewer data points for cases where the evaluation of the detailed model (e.g., the prediction from Aspen Plus) is computationally expensive, and developing algorithms for switching between the detailed model and the surrogate model to improve the Bayesian inference. Future work includes integrating SDOE into FOQUS, and using SDOE to support large-scale pilot tests at Technology Centre Mongstad (TCM), a post-combustion carbon capture facility located in Mongstad, Norway that can operate with a flue gas flow rate of 60,000 Sm^3/h (Bumb et al., 2017).

6. Acknowledgements and Disclaimer

The authors acknowledge the research funding provided by the Carbon Capture Simulation for Industry Impact ($CCSI^2$), a collaboration among national laboratories, industry and academic institutions for developing, demonstrating, and deploying computational tools for carbon capture. This research was supported in part by an appointment to the National Energy Technology Laboratory Research Participation Program, sponsored by the U.S. Department of Energy and administered by the Oak Ridge Institute for Science and Education.

any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

References

N. Brown, G. Heller, G. Staab, T. Silverman, R. Kupfer, R. Brown, A. Brown, 2017, Novel advanced solvent-based carbon capture pilot demonstration at the National Carbon Capture Center, Energy Procedia, 114, 1075–1086.

P. Bumb, A. Patkar, R. Mather, R. Kumar, J. Hall, F. Morton, J. Anthony, 2017, Field demonstration of advanced CDRMax solvent at the US-DOE's National Carbon Capture Centre and the CO2 Technology Centre Mongstad DA, Norway, Energy Procedia, 114, 1087–1099.

S.S. Garud, I.A. Karimi, M. Kraft, 2017, Design of computer experiments: A review, Computers and Chemical Engineering, 106, 71–95.

E. Masoudi, H. Holling, W.K. Wong, 2017, Application of imperialist competitive algorithm to find minimax and standardized maximin optimal designs, Computational Statistics and Data Analysis, 113, 330–345.

D.C. Miller, D. Agarwal, D. Bhattacharyya, J. Boverhof, Y.-W. Cheah, Y. Chen, J. Eslick, J. Leek, J. Ma, P. Mahapatra, B. Ng, N.V. Sahinidis, C. Tong, S.E. Zitney, 2016, Innovative computational tools and models for the design, optimization and control of carbon capture processes, Computer Aided Chemical Engineering, 38, 2391–2396.

D.C. Miller, D. Agarwal, D. Bhattacharyya, J. Boverhof, Y. Chen, J. Eslick, J. Leek, J. Ma, P. Mahapatra, B. Ng, N.V. Sahinidis, C. Tong, S.E. Zitney, 2017, Innovative computational tools and models for the design, optimization and control of carbon capture processes, Process Systems and Materials for CO2 Capture: Modelling, Design, Control and Integration, A.I. Papadopoulos, P. Seferlis (eds.), John Wiley & Sons Ltd, 311–342.

D.C. Miller, B. Ng, J. Eslick, C. Tong, Y. Chen, 2014a, Advanced computational tools for optimization and uncertainty quantification of carbon capture processes, The 8th International Conference on Foundations of Computer-Aided Process Design, 13–17 July 2014, Cle Elum, WA, USA.

D.C. Miller, M. Syamlal, D.S. Mebane, C. Storlie, D. Bhattacharyya, N.V. Sahinidis, D. Agarwal, C. Tong, S.E. Zitney, A. Sarkar, X. Sun, S. Sundaresan, E. Ryan, D. Engel, C. Dale, 2014b, Carbon Capture Simulation Initiative: A case study in multiscale modeling and new challenges, Chemical and Biomolecular Engineering, 5, 301–323.

J.C. Morgan, A.S. Chinen, B. Omell, D. Bhattacharyya, C. Tong, D.C. Miller, 2017, Thermodynamic modeling and uncertainty quantification of CO_2-loaded aqueous MEA solutions, Chemical Engineering Science, 168, 309–324.

S.N. Politis, P. Colombo, G. Colombo, D.M. Rekkas, 2017, Design of experiments (DoE) in pharmaceutical development, Drug Development and Industrial Pharmacy, 43, 6, 889–901.

B.P. Weaver, B.J. Williams, C.M. Anderson-Cook, D.M. Higdon, 2016, Computational enhancements to Bayesian design of experiments using Gaussian processes, Bayesian Analysis, 11, 191–213.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50043-4

Integrated Optimal Design Proposal Based on a Multiscale Approach: Validation and Experimental Adjustment of Properties

Cristhian Tinjacá, Juan Pablo Gallo-Molina, Óscar Álvarez, Jorge Gómez*

Product and Process Design Group, Chemical Engineering Department
Universidad de Los Andes, Bogotá, Colombia
jorgomez@uniandes.edu.co

Abstract

The chemical industry has evolved in recent years and has driven the research in the manufacture of high added-value products that satisfies specific requirements of consumers. The design of these products requires methodologies that integrate the multiscale approach, manufacturing process, consumer considerations and market parameters. In this work, a computational framework that allows integrated optimal design for Oil-in-Water emulsions was used. Predicted properties by the mathematical models were compared with the experimental behavior of these colloidal systems. The results concerning to computational and experimental design gave valuable information about the robustness of the optimal model and allows an understanding of phenomena in the microscopic structure of emulsions, which will helps in the improvement of computational design modules.

Keywords: Computational framework, Integrated optimal design, O/W emulsion, Experimental adjustment

1. Introduction

Oil-in-Water (O/W) emulsions are widely used in a variety of consumer products, such as moisturizing creams, sunscreens and cosmetics. These systems are thermodynamically unstable colloids and their performance as products is determined by the complex set of interrelationships among molecular, microscopic and macroscopic variables (Pradilla et al, 2015). In this work, the properties predicted by an optimal product design model were compared with experimental data of O/W emulsions. This computational tool was developed with a focus on moisturizing creams and the objective of producing two optimal cream formulations (Torres et al, 2016): one that allows to obtain the maximum profit and another that assures the highest preference based on physical characteristics perceived by the consumer (Bagajewicz et al, 2011).

The optimal product design model was formulated with a multiscale approach encompassing three levels. At the molecular scale, the properties of the surfactant or surfactants mixture were predicted through molecular interactions and structure (Mattei et al, 2014). The microscopic level was reflected in droplet diameter calculations. Finally, apparent viscosity was selected as a macroscopic variable, which was calculated from microscopic and molecular responses (Smith and Ierapepritou, 2010). This multiscale design was implemented in Pyomo, which allows the integration of a modular representation for product structure and properties, consumer preferences and

economic performance and it was using Bonmin solver for programming problem in order to reduce the computational time according to Torres et al (2016). It was possible to obtain cream formulations, properties and basic manufacturing conditions by defining impeller diameter and type into de computational framework.

2. Methodology

The optimal formulations were taken as a starting point for the experimental design and some operating conditions that were not considered in the property models, such as mixing time, incorporation flow rate and tank diameter, were well-defined in the laboratory. Although the computational tool provides the emulsifier mass fraction, it does not consider the cost of surfactant molecules and, consequently, this formulation contains a large mass of emulsifiers. As a result, these fractions were adjusted in the experimental design. Once these conditions were fixed (see Table 1), the moisturizing cream design was performed in order to compare droplet diameter and apparent viscosity predicted with experimental information of these microscopic and macroscopic variables.

Table 1. Design parameters for formulation

Design parameters		
Emulsifier mass fraction	[%]	4.00
Emulsifiers	[−]	Tween 20 Span 80
D_{Imp}/D_{Tank}	[−]	0.75
Impeller type	[−]	Vertical Flat blade
Mixing Time	[s]	900
Incorporation flow rate	$\left[\frac{m^3}{s}\right]$	$5x10^{-7}$

The emulsions were prepared from two optimal formulation cases: maximum consumer preference and maximum profit. With the surfactant adjustment, the resultant formulations had a 20% dispersed phase concentration in both cases. Considering that emulsion properties depends on every aspect related to molecular interactions, microscopic structure and operating conditions in the manufacturing process, the product design began only with water and mineral oil/petrolatum at a first stage in order to obtain experimental measures for an early approach with the computational framework and, subsequently, follow step-by-step the predicted formulation at subsequent stages of this research. This methodology allows a better understanding of the manufacturing process of emulsions, and consequently, it is possible to evaluate robustness, scope and every consideration in mathematical models.

3. Mathematical models and experimental measurements

The results of optimal formulations and experimental measures of droplet diameter distribution, mean drop diameter d_{32} and viscosity as a function of shear rate are reported in Table 2

Table 2. Formulation of preference and profit case obtained with computational tool (models).

Compound		Formulation case	
		Preference	**Profit**
Continuous phase	[%]	73.53	71.95
-Water	[%]	94.10	95.00
-PPG	[%]	4.54	0.00
-PEG	[%]	0.00	4.86
-Urea	[%]	1.36	0.14
Dispersed phase	[%]	20.03	10.04
-Mineral oil	[%]	50.00	0.00
-Almond oil	[%]	50.00	0.00
-Petrolatum	[%]	0.00	99.50
-Coconut oil	[%]	0.00	0.50
Emulsifiers	[%]	4.00	4.00
HLB	[−]	12.03	13.65
Thickener	[%]	0.00	3.00
Agitation velocity	[rad/s]	104,72	104,72
Results			
Droplet Diameter	[µm]	19.11	25.76
Apparent viscosity	[Pa*s]	0.385	0.394

For drop diameter, the model proposed by Tcholakova et al (2011) was used. This approach depends on density of continuous phase ρ_C, viscosity of dispersed phase μ_D, interfacial tension σ_{ow} and other factors that quantify the hydraulic and differential pressure contributions. This model also takes into account the turbulent energy dissipation given by agitation velocity N_s, impeller diameter D_{Im} and parameter b_1 related to impeller type.

$$d_{32} = A_1\left(1 + \frac{A_2 * \mu_D * \epsilon^{1/3} * d_{32}^{2/3}}{\sigma_{ow}}\right)^{3/5} \sigma_{ow}^{3/5} * \rho_C^{-\frac{3}{5}} * \epsilon^{-\frac{2}{5}} \quad (1)$$

$$\epsilon = b_1 * N_s^3 D_{Im}^2 \quad (2)$$

The model used for viscosity prediction in diluted emulsions was developed by Oldroyd (1953) and complemented by Pal (2000). In this case, the emulsion viscosity depends on the viscosity of continuous phase μ_C, viscosity ratio of both phases κ, volumetric fraction of dispersed phase ϕ and capillary number N_{Ca} that includes shear rate $\dot{\gamma}$ and mean drop diameter d_{32}.

$$\mu_e = \mu_c\left[1 + \frac{5\kappa + 2}{2(\kappa + 1)}\phi + \frac{(5+2)^2}{10(\kappa+1)^2}\phi^2\right]\left[\frac{1 + \lambda_1\lambda_2 N_{CA}^2}{1 + \lambda_1^2 N_{CA}^2}\right] \quad (3)$$

$$\lambda_1 = \frac{(19\kappa + 16)(2\kappa + 3)}{40(\kappa + 1)}\left[1 + \frac{19\kappa + 16}{5(\kappa+1)(2\kappa+3)}\phi\right] \quad (4)$$

$$\lambda_2 = \frac{(19\kappa + 16)(2\kappa + 3)}{40(\kappa + 1)}\left[1 - \frac{3(19\kappa + 16)}{10(\kappa + 1)(2\kappa + 3)}\phi\right] \quad (5)$$

$$N_{Ca} = \frac{\mu_{cd_g}\dot{\gamma}}{2\sigma_{ow}} \quad (6)$$

The contribution of a rheology modifier was described with the Huggins equation with specific parameters for the thickener used in experimental design. Where $[\eta_i]$ represents the relative increase in viscosity, $[\eta]$ represents the intrinsic viscosity, C_{th} is the thickener concentration and k_H is the Huggins slope.

$$[\eta_i] = [\eta]C_{th} + k_H[\eta^2]C_{th}^2 \quad (7)$$

$$[\eta] = 45.8 * \exp(-0.0005\dot{\gamma}) \quad (8)$$

$$k_H = 0.0047 * \dot{\gamma} - 0.0005 \quad (9)$$

For experimental measures of drop diameter and viscosity, a particle size analyzer and a rotational rheometer were used respectively. Considering the shear rate of 100 s^{-1} defined by Wibowo and Ng (2001) for moisturizing creams and pastes viscosity measurements were performed in a range of $[90,120]\ s^{-1}$. As mentioned before, O/W emulsions with just mineral oil and water were initially prepared. The resulting measures of this basic formulation are reported on Table 3. Table 4 reports results for experimental measures of profit and preference approach that includes a second profit formulation in which was eliminated the thickener with the purpose to identify it direct effects in the emulsion structure.

Table 3. Basic formulation results from experiments

Basic Formulation		
Compound		**Value**
Continuous phase	[%]	80.00
-Water	[%]	100
Dispersed phase	[%]	20.00
-Mineral oil	[%]	100
Emulsifiers	[−]	4.00
Thickener	[%]	0.00
Agitation velocity	[rad/s]	104,72
Results		
Droplet Diameter	[μm]	17.60
Apparent viscosity	[Pa*s]	0.0017

Table 4. Formulation and experimental results for preference and profit case approach

Compound		Approach		
		Preference	**Profit**	**2nd Profit**
Continuous phase	[%]	80.00	80.00	80.00
-Water	[%]	94.10	95.00	95.00
-PPG	[%]	4.54	0.00	0.00
-PEG	[%]	0.00	4.86	4.86
-Urea	[%]	1.36	0.14	0.14
Dispersed phase	[%]	20.00	20.00	20.00
-Mineral oil	[%]	50.00	0.00	0.00
-Almond oil	[%]	50.00	0.00	0.00
-Petrolatum	[%]	0.00	99.50	99.50
-Coconut oil	[%]	0.00	0.50	0.50
Emulsifiers	[%]	4.00	4.00	4.00
Thickener	[%]	0.00	3.00	0.00
Agitation velocity	[rad/s]	104,72	104,72	104,72
Results				
Droplet Diameter	[μm]	19.70	0.20	23.30
Apparent viscosity	[Pa*s]	0.00289	0.645	0.00271

4. Optimal model vs Experimental information

As it can be seen in Table 4, the measured properties of the basic formulation were lower in comparison to the predictions, for the reason that these models include the contribution to viscosity of additional compounds. The results of the optimal cases were significantly closer to mathematical models for one of two estimated properties, depending on the case evaluated. For the preference approach, the comparison for mean droplet diameter denoting a difference of 3% related to calculated diameter. However, in the case of viscosity, a divergence between experimental and predicted properties was noted. This was attributed to the effect of the thickener. The profit approach includes a high concentration of thickener, which produces high viscosities in the emulsions. Although the values of viscosity were not quite close in comparison to the model, both values stayed in the same order of magnitude. Nevertheless, predicted droplet diameter was significantly smaller from the computational model. In order to determine which phenomena drive this behavior, it is necessary to analyze the microscopic behavior of these emulsified systems.

An ample space between droplets exists in diluted emulsions. However, flocculation and coalescence mechanisms tend to produce larger drops. When a thickener is used, its molecules occupy the space between particles, preventing coalescence and favoring smaller drop sizes (Akiyama et al, 2005). In the computational tool, this effect was not included. For this reason, a second profit approach formulation was proposed without thickener. As show in Table 4, the drop diameter obtained implies a difference of 9%.

5. Conclusions and future work

The comparison between the experimental and computational data gave valuable information about the robustness degree of the optimal model. The mean diameter calculated in two formulations, without the effects of thickener, predicted satisfactorily the microscopically behavior of emulsion evidenced in experimental execution; hence the mathematical model for microstructure of emulsion was appropriated in the design

of computational framework. On the other hand, the viscosity model that includes the contribution of thickener in the properties calculation, did not predict entirely the structure of emulsion due to the fact that this model not consider the effects of rheology modifier in coalescence of dispersed drops.

As previously exposed, it was possible to conclude that the computational framework was a useful tool for emulsion design and gives accurate information about drop diameter that keeps relation with stability of these colloidal systems. Also, the experimental design performed in this work had an important role in the validation of predicted properties, it allows to understand phenomena that not consider the mathematical model and as a result of this is possible to overcome some limitation in the computational tool by designing a specific module that includes the thickener effects in coalescence of drops for specific concentrations defined in a new experimental design. So is possible to integrate the microscopically and macroscopically effects of rheology modifier in the emulsion structure, and consequently, providing an accurate prediction of properties that have an influence on operating conditions, consumer preferences and economic yield.

References

D. Pradilla, W. Vargas and O. Alvarez, 2015, "The application of a multi-scale approach to the manufacture of concentrated and highly concentrated emulsion", Chemical Eng. Res. Design, vol. 95, 162-172.

J. Torres, O. Alvarez and J. Gomez 2016, "Diseño computacional múltiescalar integrado de productos cosméticos estructurados", Universidad de los Andes.

M. Bagajewicz, S. Hill, A. Robben, L. Heyde, M. Sanders, E. Sposato, C. Baade, S. Manora and J. Coradin,2011, "Product Design in Price-Competitive Markets: A case study of a Skin Moisturizing Lotion", AIChE Journal, vol. 57, no. 1, 160 – 177.

M. Mattei, G. Kontogeorgis and R. Gani, 2014, "A comprehensive framework for surfactant selection and design for emulsion based chemical product design", Fluid Phase Equilibria, vol. 362, 288-299.

B. Smith and M. Ierapepritou,2010," Integrative Chemical Product Design Strategies: Reflecting Industry Trends and Challenge", Computers and Chemical Engineering, vol. 34, 857-865.

S. Tcholakova, I. Lesov, K. Golemanov, N. Denkov, S. Judat, R. Engel y T. Danner,2011, "Efficient Emulsfication of Viscous Oils at High Drop Volume Fraction", Langmuir, vol. 27, 14783-14796.

J. Oldroyd, 1953, "The Elastic and Viscous Properties of Emulsions and Suspensions", Proceedings of the Royal Society of London, vol. 218, 122-132

R. Pal, 2001, "Evaluation of theoretical viscosity models for concentrated emulsions at low capillary numbers", Chemical Engineering Journal, vol. I, nº 81, 15-21.

C. Wibowo and K. Ng, 2001, "Product-Oriented Process Synthesis and Development: Creams and Pastes", Process Systems Engineering, vol. 47, nº 12.

E. Akiyama, A. Kashimoto, K. Fukuda, H. Hotta, T. Suzuki and T. Kitsuki, 2005, "Thickening properties and emulsification mechanisms of new derivatives of polysaccharides in aqueous solution", Journal of Colloid and Interface Science, vol. 282, 448-457.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50044-6

Catalytic Production of Gamma-Valerolactone from Two Different Feedstocks

Juyeon Kim[a*], Jaewon Byun[a], Yuchan Ahn[a] and Jeehoon Han[a,b]

[a]*School of Semiconductor and Chemical Engineering, Chonbuk National University, Jeonju 54896, Korea*
[b]*School of Chemical Engineering, Chonbuk National University, Jeonju 54896, Korea*
jujuy2621@gmail.com

Abstract

Large-scale processes for catalytic production of gamma-valerolactone (GVL) from two different feedstocks (levulinic acid; LA and ethyl levulinate; EL) are presented based on experimental results. In LA-based process, cellulose and hemicellulose of corn stover are converted to LA using dilute sulfuric acid catalyst and subsequently converted to GVL over $RuSn_4$/C catalyst in the present of GVL solvent. Instead of LA, EL converted to GVL on Pt-ZSM 35 catalyst using ethanol solvent obtained as a by-product in EL-based process. The simulation study including process design, heat integration, and economic analysis is performed to show the economic feasibility for the EL-based process compared with LA-based process. The minimum selling price of EL-based GVL ($ 1.64/kg) is higher than LA-based GVL ($ 1.16/kg). The sensitivity analysis shows that if EL-based process uses lower feedstock cost and energy requirements, it will be comparable with the LA-based process.

Keywords: Gamma-valerolactone; Ethyl levulinate; Levulinic acid; Heterogeneous catalytic hydrogenation; Techno-economic analysis

1. Introduction

Gamma-valerolactone (GVL) derived from lignocellulosic biomass can be used as a value-added platform chemical for the production of biofuels, fuel additives, and polymers. Ethyl levulinate (EL) or levulinic acid (LA) can be converted to GVL through the heterogeneous catalytic hydrogenation. GVL is mainly produced by LA, obtained from the hydrolysis of lignocellulosic biomass (cellulose and hemicellulose) (Alonso et al., 2013). Recently, several conversion studies of EL or LA into GVL by the heterogeneous catalytic hydrogenation were reported (Quiroz et al., 2016; Li et al., 2017). We present two catalytic conversion strategies for the production of GVL from different feedstocks. In the EL-based process, EL can be converted to GVL on Pt-ZSM 35 catalyst using ethanol solvent obtained as a by-product, while in the LA-based process, LA over $RUSn_4$/C catalyst can be converted to GVL using solvent GVL (Alonso et al., 2013; Chen et al., 2016). Both processes have three common advantages as follows: (1) The use of ethanol and GVL as reaction solvent enables to reduce its makeup of reaction solvent; (2) Heterogeneous catalysts can be effectively separated and recycled; (3) Both processes show 99% high yield for GVL production. Here, compared to LA-based process, EL-based process has a higher feedstock cost but a lower separation cost because of the lower boiling point of the solvent.

In this work, integrated processes are developed combining separation systems with reaction systems by using ASPEN Plus process simulator based on the same processing

rate of the feedstock (2,000 dry tonnes of biomass per day) as well as the same economic assumptions and parameters. To reduce the total energy requirement of the process, we design a heat exchanger network by using the ASPEN Energy Analyzer. Finally, we conduct techno-economic analysis and comparison of minimum selling price (MSP) between both processes shows which process is an economically competitive approach.

2. Technical overview

A EL-based process is presented for the catalytic production of GVL from biomass-derived EL based on the experimental studies, and is compared with biomass derived LA-based GVL production.

Biomass derived LA-based GVL production consists of two catalytic conversion subsystems: (i) biomass deconstruction and (ii) GVL production (Figure 1). In the biomass deconstruction, cellulose and hemicellulose can be converted to LA and furfural (FF) with 61 % and 56 % molar yields at 443 K and 16 bar, containing dilute sulfuric acid (SA) catalyst in a GVL–water solution (Alonso et al., 2013). In the GVL production, FF and LA can be converted to GVL though three catalytic tubular reactors connected in series. First, FF can be converted to furfuryl alcohol (FFA) with a 90 % molar yield at 373 K and 35 bar in the first reactor using Pt_3Sn/SiO_2 catalyst. Next, FFA subsequently converted to LA with a 70 % molar yield in the second reactor using Amberlyst 70 catalyst. Finally, the LA can be converted to GVL with a 99 % molar yield at 493 K and 35 bar on a RuSn (1:4)/C catalyst.

Chen et al. (Chen et al., 2016) reported that EL is hydrogenated to GVL with a high yields (99 %) using commercial zeolite supported Pt (Figure 2). They investigated the influence of four reaction conditions (catalyst, temperature, pressure and solvent) in GVL yields. First, 1 % Pt/ZSM-35 was the most efficient catalyst among other catalysts (Pt/MCM, Pt/SAPO, Pt/HY, Pt/USY and Pt/MOR). EL is hydrogenated with different range of temperature (180-220 ℃), and the highest yield of GVL is achieved at 200 ℃. As increment of H_2 pressure (10-60 bar), the H_2 concentration on the surface of the catalyst that accelerates the reaction rate is increased. Finally, the study confirmed that the selectivity of GVL was influenced by solvent. Consequently, Pt/ZSM-35 was found to show the highest activity as a results of the catalyst screening, under optimized conditions (200 ℃, 60 bar, ethanol solvent) and 100 % conversion of EL with a high GVL selectivity (99 %) were achieved. Additionally, the EtOH is produced as a byproduct.

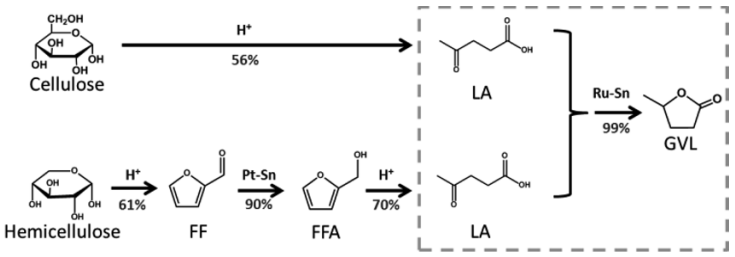


Figure 1. Reaction scheme of the LA-based GVL production from lignocellulosic biomass.

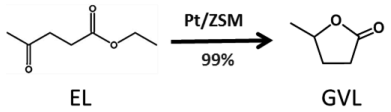


Figure 2. Reaction scheme of the EL-based GVL production.

3. Techno-economic results and discussion

Two large-scale processes for catalytic production of GVL from two different feedstocks (LA and EL) were designed based on the technologies outlined and the experimental data described in the previous section by using ASPEN Plus process simulator. In the ASPEN simulation, standard method was used for the convergence. The base capacity of both GVL production processes was set to 2,000 tonnes per day (t/d) of dry corn stover.

3.1. LA based process: simulation and energy analysis

The LA-based process (Han, 2016) consists of five subsystems: biomass deconstruction, GVL production, GVL recovery, wastewater treatment (WWT), and storage (Figure 3). A mixture of shredded corn stover in GVL–water solution (4:1) containing 0.1 M SA is supplied to a biomass deconstruction reactor (R-1 in Figure 3) for coproduction of LA and FF from cellulose (41 %) and hemicellulose (24 %) of corn stover. For the subsequent catalytic conversion of LA and FF, SA is removed by precipitation using $Ca(OH)_2$ (S-1 in Figure 3). In the GVL production subsystem, LA and FF are simultaneously converted to GVL (R-2 in Figure 3). The resulting mixture containing GVL, water, biomass residues (BR), and CO_2 is sent to GVL recovery subsystem to effectively separate GVL and water from CO_2 and BR (S-2 and S-3 in Figure 3). The water and wet BR are treated in WWT subsystem. Most of GVL used as a solvent is recycled back to the biomass deconstruction subsystem, and the remaining GVL is fed to distillation column (D-1 Figure 3) for producing a high purity (99.9 wt%) of product. Finally, the GVL (375 t/d) is obtained at the bottom of D-1.

Next, we performed an energy analysis of the proposed process. To produce GVL, the process requires 186 MW of heating, 196 MW of cooling, and 1 MW of electricity. We conduct a heat integration to minimize energy requirements with heat exchanger network (HEN) between hot and cold process streams by using the ASPEN Energy Analyzer. The heat integration result shows considerable energy reduction in the heating requirements of the process (99 MW). In the biomass deconstruction and GVL production subsystems, most of the heat requirement in their cold process streams could be transferred from hot process streams in GVL recovery subsystem.

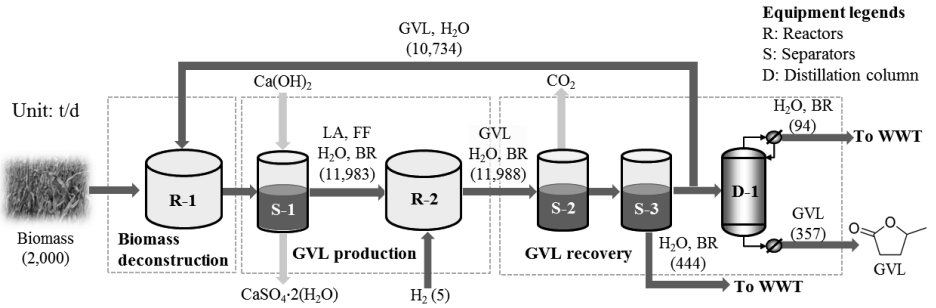


Figure 3. Simplified process flow diagram of LA-based GVL production process.

The energy efficiency which is the ratio of the energy output (GVL) to energy input (biomass feedstock, H_2 and primary energy) is increased from 18.1 % (before heat integration) to 23.2% (after heat integration).

3.2. EL based process: simulation and energy analysis

The EL-based process consists of three main subsystems, as shown in Figure 4: GVL production, GVL recovery, and storage. We used the same cellulose and hemicellulose-to-EL mass yields (54 %) in the previous study (Chen et al., 2016). Based on this yield and feedstock compositions given in LA-based process, the amount of EL required for EL-based process was estimated to 702 t/d. Feed solution (1.2 wt% of EL in EtOH solvent) and H_2 are supplied to the catalytic reactor (R-1 in Figure 4) in GVL production subsystem. EL is converted to GVL through the hydrogenation at 60 bar and 200 ℃ with Pt-ZSM catalyst. The resulting mixture containing dilute GVL (1.3 wt%) is sent to GVL recovery subsystem to effectively separate GVL from ethanol solvent. High purity of GVL (99.9 wt%, 490 t/d) is obtained at the bottom of distillation column (D-1 in Figure 4), while high purity of EtOH (99.9 wt%, 225 t/d) is obtained at the top of D-1. The most of EtOH is recycled back to the GVL production and the remaining EtOH (2%) is sold as byproduct.

The process requires 249 MW of heating, 254 MW of cooling, and 2.8 MW of electricity. Considerable amount of heat (162 MW) are required in GVL production subsystem to raise temperature of conversion reactor. To reduced energy requirements of the process, we also performed heat integration. But, the result of heat integration shows only 1% of total energy reduction, resulting still large amount of heat energy requirement in the GVL production system. There are a few chances to heat transfer between hot process streams and cold process steam in the other subsystems because most of the temperature of cold process stream is higher than them of hot process streams. In this study, EL-based process has a portion of process flows to produce GVL from the biomass compared to LA-based process. The limited number of process streams reduces the optimization possibility of HEN which can reduce heat energy requirements to be minimized. The energy efficiency of EL-based process is increased from 25.3 % (before heat integration) to 25.5 % (after heat integration).

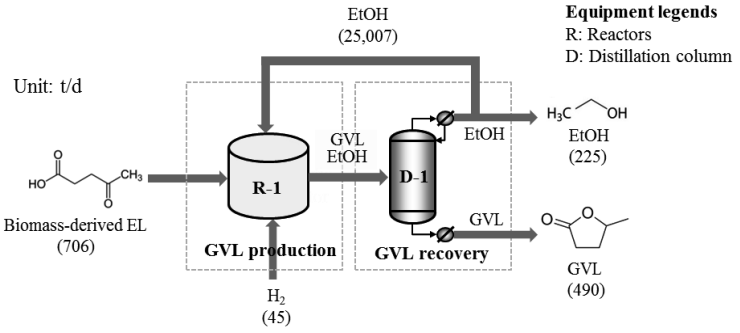


Figure 4. Simplified process flow diagram of EL-based GVL production process.

3.3. Techno-economic comparison of two processes

In this section, we conducted techno-economic comparison of two processes to present possibility that EL-based GVL production process could be alternative to LA-based

process. Based on the simulation results, capital and operating costs were calculated by using ASPEN Economic Analyzer. We performed an economic analysis by using a discounted cash flow to determine a MSP of GVL that makes the net present value equal to zero. The MSP (\$1.64 per kg; $/kg_{GVL}$) of GVL for EL-based process was higher than that of LA-based process ($1.16/kg_{GVL}$). This is because the high feedstock and utility costs are required in EL-based strategy. In case of feedstock cost, EL price is much higher than LA derived from biomass (Hayes, 2013; Han et al., 2014). Although the production of GVL from EL-based process (490 t/d) was higher than LA-based process (375 t/d), total cost of EL-based process is much higher than that of LA-based process. Utility cost of EL-based process (\$ 77 million per year; \$ million/y) is 97 % higher than LA-based process (\$ 39 million/y) (Table 1). The EL-based process only included GVL production and recovery subsystems, while the LA-based process included various subsystems to produce GVL from biomass. Process integration with the process that can produce EL from biomass at low cost and supply heating energy to EL-based process can decrease the MSP of GVL. We performed sensitivity analysis to present the impact of process integration on the MSP of GVL. The EL price could be reduced by improving economics of the EL production process, and the utility costs could be reduced by heat integration with other process using waste heat.

Table 1. Annualized capital cost and operating cost of EL-based GVL production process (\$ million / y).

	ACC	OC
Subsystem		
GVL production	16.9	254.4
GVL & ETOH recovery	4.0	2.8
Storage	1.6	1.1
Utilities(HEN)	6.2	77.0
Total ACC/OC	28.7	335.3
TC	364.0	

ACC : Annualized capital cost OC : Operating cost

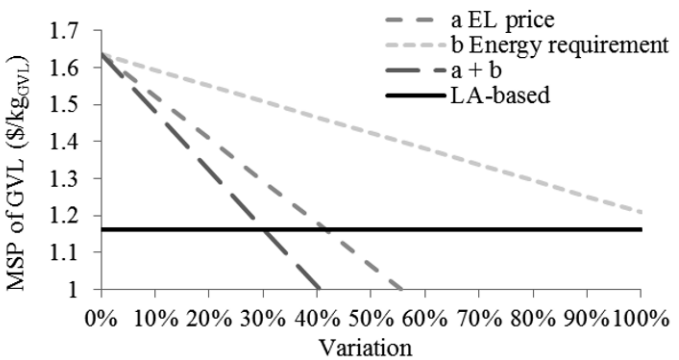


Figure 5. Sensitivity analysis of the MSP of GVL with respect to major cost drivers.

The feedstock cost is the most significant factor affecting the MSP of GVL. When the EL price is decreased by 42 %, the MSP of GVL is decreased to that of LA-based strategy (\$ 1.16/kg_{GVL}). Even if the energy requirements are decreased to zero, the MSP of GVL is still higher than that of LA-based process. When the EL price and energy requirements are simultaneously decreased by more than 30 %, the MSP of GVL can be lower than that of LA-based process (Figure 5).

4. Conclusion

Two processes for catalytic production of GVL from EL and biomass derived-LA were designed. Techno-economic comparison presented a replaceability of feedstock with the results that EL-based GVL production process could be alternative to LA-based process. The EL-based process had higher MSP (\$1.64/kg_{GVL}) because the feedstock and energy costs of EL-based process were higher despite larger production of GVL from EL-based process (490 t/d) compared to LA-based process (375 t/d, \$1.16/kg_{GVL}). The MSP for the EL-based process could be lower than that of LA-based process when feedstock and energy costs decrease by 30%.

References

Alonso, D. M., Wettstein, S. G., & Dumesic, J. A. (2013). Gamma-valerolactone, a sustainable platform molecule derived from lignocellulosic biomass. Green Chemistry, 15, 584-595.

Alonso, D. M., Wettstein, S. G., Mellmer, M. A., Gurbuz, E. I., & Dumesic, J. A. (2013). Integrated conversion of hemicellulose and cellulose from lignocellulosic biomass. Energy & Environmental Science, 6, 76-80.

Chen, C.-B., Chen, M.-Y., Zada, B., Ma, Y.-J., Yan, L., Xu, Q., Li, W.-z., Guo, Q.-X., & Fu, Y. (2016). Effective conversion of biomass-derived ethyl levulinate into γ-valerolactone over commercial zeolite supported Pt catalysts. RSC Advances, 6, 112477-112485.

Han, J. (2016). A bio-based 'green'process for catalytic adipic acid production from lignocellulosic biomass using cellulose and hemicellulose derived γ-valerolactone. Energy Conversion and Management, 129, 75-80.

Han, J., Sen, S. M., Alonso, D. M., Dumesic, J. A., & Maravelias, C. T. (2014). A strategy for the simultaneous catalytic conversion of hemicellulose and cellulose from lignocellulosic biomass to liquid transportation fuels. Green Chemistry, 16, 653-661.

Hayes, D. (2013). Report on Optimal Use of DIBANET Feedstocks and Technologies. In.

Li, C., Xu, G., Zhai, Y., Liu, X., Ma, Y., & Zhang, Y. (2017). Hydrogenation of biomass-derived ethyl levulinate into γ-valerolactone by activated carbon supported bimetallic Ni and Fe catalysts. Fuel, 203, 23-31.

Quiroz, J., Mai, E. F., & da Silva, V. T. (2016). Synthesis of nanostructured molybdenum carbide as catalyst for the hydrogenation of levulinic acid to γ-Valerolactone. Topics in Catalysis, 59, 148-158.

Yang, T., Li, H., He, J., Liu, Y., Zhao, W., Wang, Z., Ji, X., & Yang, S. (2017). Porous Ti/Zr Microspheres for Efficient Transfer Hydrogenation of Biobased Ethyl Levulinate to γ-Valerolactone. ACS Omega, 2, 1047-1054.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50045-8

Multilayer and Multiobjective Design Platform for Drug Product Manufacturing Processes of Biopharmaceuticals

Haruku Shirahata[a*], Shant Dakessian[a,b], Masahiko Hirao[a], Hirokazu Sugiyama[a]

[a]*Department of Chemical System Engineering, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, 113-8656, Tokyo, Japan*
[b]*Department of Chemistry and Applied Biosciences, ETH Zurich, Vladimir-Prelog-Weg 1-5/10, 8093 Zurich, Switzerland*
h-shirahata@pse.t.u-tokyo.ac.jp

Abstract

A platform is presented that enables systematic alternative generation and multiobjetive evaluation for designing biopharmaceutical drug product manufacturing processes. In the platform, the layers of product, unit operation, equipment technology, and operation are defined to assist generation of process alternatives as the combination of information at each layer. Multiple indicators are defined to evaluate economy, environment, supply, and product quality, which are aggregated to total score using weighting factors. A case study demonstrated that the developed platform can facilitate comprehensive process synthesis and multiobjective evaluation considering the novel disposable equipment technology for biopharmaceutical manufacturing known as single-use technology.

Keywords: Pharmaceuticals, Single-use technology, Process alternative generation, Multiobjective evaluation, Decision-making

1. Introduction

Pharmaceutical manufacturing is shifting from large-scale single-product production to small-scale and multiproduct production. In biopharmaceutical manufacturing, continuous technology or single-use technology (SUT) are developed as the new equipment technologies (Konstantinov and Cooney, 2015, Levine et al, 2012). SUT uses pre-sterilized and disposable resin-made equipment, which can substantially reduce changeover operations at the shop floor, e.g., cleaning and sterilization. The conventional technology to SUT is known as multi-use technology (MUT) that uses stainless steel equipment and requires cleaning and sterilization at every batch. The choice of these technologies can affect the entire process regarding various aspects.

Many authors investigated the process design considering SUT in drug substance and drug product manufacturing, e.g., Walther et al. (2015) and Shirahata et al. (2017a), respectively. In these previous works, the number of the considered process alternatives were limited due to the reliance on heuristics. Generation of process alternatives has been investigated for decades for commodity chemical manufacturing (Kirkwood et al., 1988), however, less attention has been paid to pharmaceutical manufacturing. In this work, we focus on biopharmaceutical drug product manufacturing, i.e., sterile filling of compounded drug solution into glass containers, and present a design platform for generating and evaluating process alternatives with considering novel equipment technologies such as SUT.

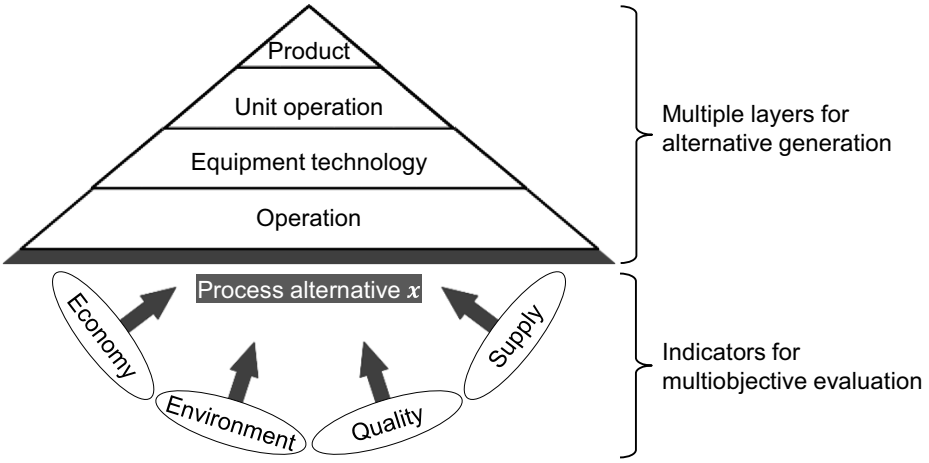


Figure 1. Hierarchical structure of the developed platform

2. Developed platform

Figure 1 shows the hierarchical structure of the developed platform with multiple layers for generating alternatives, and with the multiple indicators for evaluating alternatives.

2.1. Multiple layers for alternative generation

Four layers are defined to consider numerous candidate alternatives in a hierarchical manner. The first and the second layer, "product" and "unit operation", consider information on the product and the combinations of unit operations, respectively. The third layer, "equipment technology", is defined to incorporate novel equipment technologies, e.g., SUT, with considering materials, layout, and sizes of the equipment. The fourth layer, "operation", considers manufacturing conditions and scheduling. At each layer of "product", "unit operation", "equipment technology", and "operation", row vectors, $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, and $\boldsymbol{d}$, are defined, respectively, as shown in Eqs.(1–4), which contain the elements corresponding to the layer.

$$\boldsymbol{a} = (\text{Production lifetime, Annual production amount, Dorsage form, } \dots) \quad (1)$$

$$\boldsymbol{b} = (\text{Sterilization method, Presence of freeze dry, } \dots) \quad (2)$$

$$\boldsymbol{c} = (\text{Equipment material, Layout, Shape, } \dots) \quad (3)$$

$$\boldsymbol{d} = (\text{Production sequence, Filling speed, Temperature, } \dots) \quad (4)$$

A process alternative $\boldsymbol{x}$ is defined by specifying the elements of the vectors $\boldsymbol{a}$ to $\boldsymbol{d}$ as values or discrete choices (Eq. (5)). The symbol $\oplus$ indicates an operation of the direct sum.

$$\boldsymbol{x} = \boldsymbol{a} \oplus \boldsymbol{b} \oplus \boldsymbol{c} \oplus \boldsymbol{d} \quad (5)$$

2.2. Indicators for multiobjective evaluation

Multiobjective evaluation of the generated alternative $\boldsymbol{x}$ is then performed as shown in Eq.(6) by evaluation indicator y_{obj} with evaluation model f_{obj} for each design objective *obj*. For example, net present value (NPV) is used as y_{obj} (*obj* = Economy) as shown in Eq.(7). Here, C_0 [USD], s [–], n_{tax} [–], t [–], i [–], n [–], C [USD], and m [–] stand for investment cost, salvage rate, taxation lifetime, taxation rate, interest rate, production lifetime, monthly expense, and number of month per year, respectively.

$$y_{obj} = f_{obj}(\boldsymbol{x}) \qquad \text{with } obj = \{\text{Economy, Environment, Supply, Quality}\} \tag{6}$$

$$\begin{aligned}\text{NPV(cost)}(\boldsymbol{x}) = C_0(\boldsymbol{x}) - \frac{(C_0(\boldsymbol{x}) - sC_0(\boldsymbol{x}))}{n_{\text{tax}}}\, t \sum_{j=1}^{n_{\text{tax}}} \frac{1}{(1+i)^j} - \frac{sC_0(\boldsymbol{x})}{(1+i)^n} \\ + mC(\boldsymbol{x}) \sum_{j=1}^{mn} \frac{1}{\left(1 + {}^{i}/_{m}\right)^j}\end{aligned} \tag{7}$$

Life cycle CO_2 emissions, impact of time delay in product supply, and impact of interaction between the product and the equipment material are used for *obj* = Environment, Supply, and Quality, respectively, with the models f_{obj} from Shirahata et al. (2017b). The general design problem is formulated as shown in Eq.(8), where total score T [–] is obtained by aggregating four y_{obj} for supporting the decision-maker by providing a single objective value. The parameters w_{obj} [–] and $\alpha^{interaction}$ [–] stand for the weighting factors for *obj* and for *interaction*, respectively. The suffix *interaction* [–] stands for the two types of the interaction between the drug solution and the equipment materials, which are resin-made (for SUT) or stainless steel (for MUT).

$$\begin{aligned}\min_{\boldsymbol{x}} T(\boldsymbol{x}) = w_{\text{Economy}} \frac{y_{\text{Economy}}(\boldsymbol{x})}{\max_{\boldsymbol{x}} y_{\text{Economy}}(\boldsymbol{x})} + w_{\text{Environment}} \frac{y_{\text{Environment}}(\boldsymbol{x})}{\max_{\boldsymbol{x}} y_{\text{Environment}}(\boldsymbol{x})} \\ + w_{\text{Supply}} \frac{y_{\text{Supply}}(\boldsymbol{x})}{\max_{\boldsymbol{x}} y_{\text{Supply}}(\boldsymbol{x})} \\ + w_{\text{Quality}} \sum_{interaction} \alpha^{interaction} \frac{y_{\text{Quality}}^{interaction}(\boldsymbol{x})}{\max_{\boldsymbol{x}} y_{\text{Quality}}^{interaction}(\boldsymbol{x})}\end{aligned} \tag{8}$$

s. t. $\quad 0 \le \alpha \le 1, 0 \le w \le 1, \sum \alpha = 1, \sum w = 1$

3. Case study

A case study was conducted in which the developed platform was applied to the design of a commercial production process of biopharmaceutical drug manufacturing.

3.1. Alternative generation at product, unit operation, and operation layers

At the "product" layer, two options for annual production amount, i.e., 10,000 or 250,000 L/year, and three options for production lifetime, i.e., 1, 10 or 20 year, were considered. At the "unit operation" layer, no option was considered but nine unit operations were set according to the industrial standard, which are partly shown in Figure 2. The "operation" layer considered two options for production sequence, i.e., alternating or campaign production. As for the remaining elements, typical values and choices were specified based on industrial expert information. The alternatives generated at these three layers are hereafter referred to as "production situation", i.e., in total twelve production situations (I–XII) are considered further in this work.

The specified elements for two production situations I and XII are shown in Eqs.(9) and (10) as the representative examples. The production situation I and XII describe a "short-

term, small-scale and multiproduct" production, and a "long-term, large-scale and single-product" production, respectively.

$$\boldsymbol{x}^{\mathbf{I}} = (\text{Production lifetime, Annual production amount, ..., Production sequence, ...}) \quad (9)$$

$$= (1\ \text{year}, 10{,}000\ \text{L/year}, \ldots, \text{Alternating production}, \ldots)$$

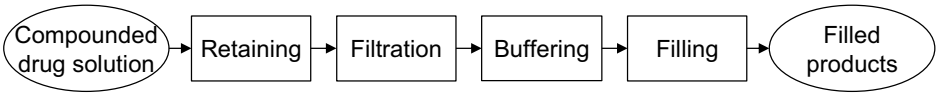
$$\boldsymbol{x}^{\mathbf{XII}} = (20\ \text{year}, 250{,}000\ \text{L/year}, \ldots, \text{Campaign production}, \ldots) \quad (10)$$

3.2. Alternative generation at equipment technology layer

The "equipment technology" layer considered the choice of resin-made and stainless steel equipment for each of the nine unit operations, except for the operation to send drug solution for filling glass containers. This operation can only be performed with resin-made tube because peristaltic pump is used for filling biopharmaceutical drug products. After considering all combinations, in total $2^8 = 256$ process alternatives were generated per product situation. Among all the alternatives, the one consisting only of the stainless steel options was named as "full-MUT"; the one with all resin-made equipment was named as "full-SUT"; the remaining 254 alternatives consisting of both materials were denoted as "hybrid (HYB)". Figure 2 shows four representative alternatives; the full-MUT, two HYBs, and the full-SUT, to which the IDs of 1, 128, 130, and 256 were assigned, respectively. For example, HYB-128 has a stainless steel mixing tank and resin-made equipment for the remaining, which is recognized as a typical hybrid process industrially. HYB-130 is introduced for a comparative purpose.

3.3. Multiobjective evaluation

The generated alternatives were evaluated under each production situation regarding economy, environment, supply, and quality, which were further aggregated to T. Figure 3 shows the evaluation results for all $\boldsymbol{x}$ in production situation XII using equal values for w and α, i.e., $w = 0.25$ and $\alpha = 0.5$. The objective function T is to be minimized, i.e., $\boldsymbol{x}^{\mathbf{XII}}_{\mathbf{130}}$ was the worst, and $\boldsymbol{x}^{\mathbf{XII}}_{\mathbf{256}}$ was the best. The alternative $\boldsymbol{x}^{\mathbf{XII}}_{\mathbf{130}}$ used resin-made mixing tank and filling needles, and stainless steel equipment for the remaining equipment. This combination made the manufacturing footprint the largest, which made the environmental evaluation the worst due to large energy consumption required for the heating, ventilation, and air conditioning (HVAC). In drug product manufacturing, HVAC is the technology to maintain the cleanliness of the production area, and known to be cost- and energy-intensive (Müller et al., 2014).



Equipment	Mixing tank	Filter housing	Buffer tank	Filling needle	Process alternative ID
Equipment material*	0	0	0	0	1 (Full-MUT)
	0	1	1	1	128 (HYB)
	1	0	0	1	130 (HYB)
	1	1	1	1	256 (Full-SUT)

*{0, 1}={Stainless-steel, Resin}

Figure 2. Process alternatives with combination of equipment material

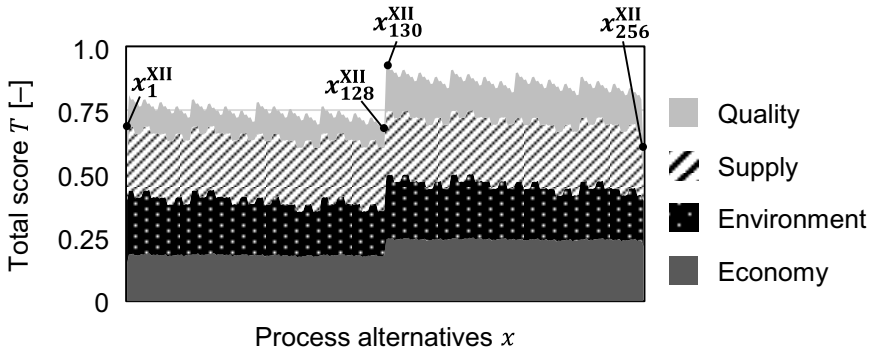


Figure 3. Total score T of 256 process alternatives under production situation XII (long-term, large-scale and single-product production) using equal values for the weighting factors w and α

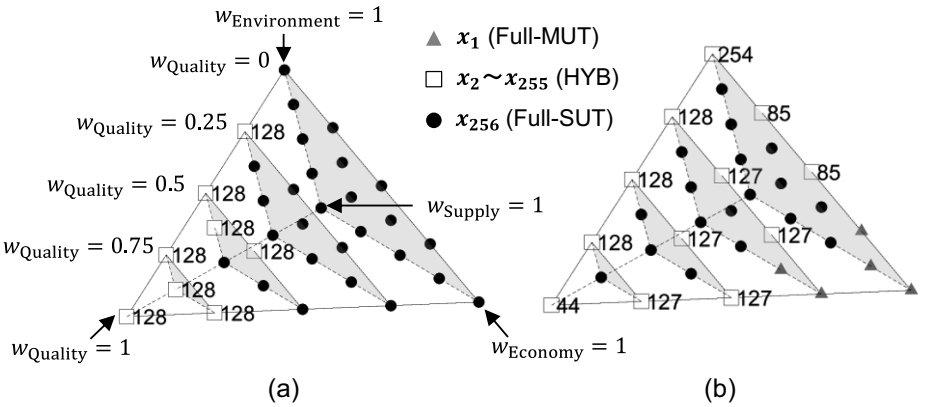


Figure 4. Sensitivity of weighting factor w to total score T at (a) production situation I (short-term, small-scale and multiproduct production) and (b) production situation XII (long-term, large-scale and single-product production)

3.4. Sensitivity analysis of weighting factor w

Sensitivity analysis on the factor w was performed, in order to investigate the impact of the priority of the design objectives on T. The tetrahedron representation in Figure 4 visualizes the sensitivity of the four weighting factors at one time (Shirahata et al., 2017b). The inner space corresponds to the possible range of w, where each vertex represents w = 1; at each point the alternative $\boldsymbol{x}$ with the smallest T is displayed. Figure 4 shows the result for production situations I (Figure 4(a)) and XII (Figure 4(b)), where full-SUT was commonly the best at the gravity point, i.e., w = 0.25. As for the area around the vertex with w_{Economy} = 1, full-SUT and full-MUT were the best in the case of (a) and (b), respectively. This disagreement is because full-MUT had large investment cost in (a), and full-SUT had small investment cost but large long-term operating expenses in (b). The area around the vertex of $w_{\text{Environment}}$ = 1 displayed the alternatives with the small footprint; the area around the vertex w_{Quality} = 1 displayed the hybrid alternatives that had lower impacts of interactions between the equipment material and the drug product than full-SUT or -MUT. Full-SUT was the best at the area around the vertex of w_{Supply} = 1 because

full-SUT showed longer but more importantly less frequent delay of product supply than full-MUT.

4. Conclusions and outlook

We presented a platform, which enabled systematic alternative generation and multiobjective evaluation for designing biopharmaceuticals drug product manufacturing processes. Multiple layers of the platform categorized the information on product, unit operation, equipment technology, and operation, which assisted hierarchical and comprehensive generation of alternatives. Especially, the "equipment technology" layer enabled to consider the novel equipment technology SUT by including the element of equipment material. The platform was demonstrated in a case study where 256 alternatives for the equipment technology under 12 production situations were evaluated considering economy, environment, supply, and quality. Under the production situations of a short-term, small-scale and multiproduct production, and a long-term, large-scale and single-product production, the full-SUT alternative resulted the best when the priorities of all indicators were the same. The sensitivity analysis on the weighting factors indicated that the best alternative could differ depending on the priority, i.e., when making decisions, the "mind-set" must be appropriately reflected in the weighting factors. The way of the systematic reflection is the future research subject; also we plan to extend the study boundary to cover drug substance manufacturing processes, and to include continuous technology as an option to be generated in the platform.

Acknowledgement

The authors thank the industrial experts of the International Society of Pharmaceutical Engineering (ISPE) Japan. Financial support by Grant-in-Aid for Young Scientists (A) No. 17H04964 from Japan Society for the Promotion of Science is acknowledged.

References

R. L. Kirkwood, M. H. Locke, J. M. Douglas, 1988. A prototype expert system for synthesizing chemical process flowsheets, *Comput. Chem. Eng.*, 12(4), 329–343.

K. B. Konstantinov, C. L. Cooney, 2015. White paper on continuous bioprocessing May 20–21, 2014 Continuous Manufacturing Symposium, *J. Pharm. Sci.*, 104, 813–820.

H. L. Levine, J. E. Lilja, R. Stock, H. Hummel, S. D. Jones, 2012. Efficient, flexible facilities for the 21st century, *Bioprocess Int.*, 10(11)s, 20–30.

G. Müller, H. Sugiyama, S. Stocker, R. Schmidt, 2014. Reducing energy consumption in pharmaceutical production processes: framework and case study. *J. Pharm. Innov.*, 9(3), 212–26.

H. Shirahata, M. Hirao, H. Sugiyama, 2017a. Decision-support method for the choice between single-use and multi-use technologies in sterile drug product manufacturing, *J. Pharm. Innov.*, 12(1), 1–13.

H. Shirahata, M. Hirao, H. Sugiyama, 2017b. Risk evaluation models for the design of parenterals manufacturing processes, *Comput. Aided Chem. Eng.*, 40, 2791–2796.

J. Walther, R. Godawat, C. Hwang, Y. Abe, A. Sinclair, K. B. Konstantinov, 2015. The business impact of an integrated continuous biomanufacturing platform for recombinant protein production, *J. Biotechnol.*, 213, 3–12.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50046-X

Computational fluid dynamics model on a compact-type steam methane reformer for highly-efficient hydrogen production from natural gas

Son Ich Ngo[a], Dan Duc Nguyen[a], Young-il Lim[a,*], Woohyun Kim[b], Dongju Seo[b], and Wang-Lai Yoon[b]

[a]Center of Sustainable Process Engineering (CoSPE), Department of Chemical Engineering, Hankyong National University, Gyeonggi-do, Anseong-si, Jungang-ro 327, 17579, Korea
[b]Hydrogen & Fuel Cell Research Department, Korea Institute of Energy Research (KIER), 71-2, Jang-dong, Yuseong-gu, Daejeon, 305-343, Korea
limyi@hknu.ac.kr

Abstract
A computational fluid dynamics (CFD) model for compact-type steam methane reformer (SMR) was proposed for highly-efficient hydrogen production from natural gas. The compact-type SMR included two domains: the combustor at the center for supplying the reaction heat demand and sleeve-shape reactor with a catalyst for steam methane reforming reactions. The $k-\varepsilon$ realizable turbulence model and discrete ordinates (DO) method of radiative transfer equation (RTE) were employed to the CFD model. A catalyst volume-based reactions mechanism from literatures was adopted for steam methane reforming reactions at the sleeve-shape porous catalyst zone. The CFD model results were validated with our experiment data and other studies on SMR in terms of the temperature profile, producer gas compositions, production rate, heating efficiency, and wall heat flux. Pressure, temperature, flow stream lines, producer gas compositions, and reaction rates were then analyzed. The present compact type SMR reactor showed a flatter heat flux profile than conventional reactor designs in literatures.
Keywords: Steam methane reformer; Combustion; Computational fluid dynamics; Radiation; Compact-type reactor design.

1. Introduction
The steam methane reforming (SMR) as a hydrocarbon fuel processing technology accounts for a significant proportion of worldwide hydrogen production (Kumar et al., 2016; Tran et al., 2017). The SMR process is an overall endothermic process in which the methane (natural gas) reacts with super-heated steam in the presence of a nickel-based catalyst to produce hydrogen, carbon dioxide, and carbon monoxide (Tran et al., 2017; Xu and Froment, 1989). The SMR unit has two main modules including a combustor and reactor. The combustor (or furnace) produces heat demand by reacting the methane with oxygen in the air, whereas the reactor consumes heat to produce the hydrogen from steam and natural gas (Seo et al., 2006). There are several configurations of the combustor such as top-fired, bottom-fired, side-fired, and terrace wall-fired reformers (Lao et al., 2016; Tran et al., 2017).
With the dramatic increase of computational power, computational fluid dynamics (CFD) has become a powerful tool for design and hydrodynamics evaluation of chemical processes (Ngo et al., 2015). CFD technology has been successfully applied to the

simulation of industrial furnaces (Han et al., 2007), SMR tube reactors (Behnam et al., 2012), and newly designed SMR reactors (Ngo et al., 2015). CFD models in combination with an equilibrium reaction kinetics (Xu and Froment, 1989) predicted the SMR performance for various reactor scales (Lao et al., 2016). Unfortunately, there is no rigorous three-dimensional (3D) CFD study on a compact SMR reactor, considering the effect of the sleeve between the combustor and reactor on heat transfer.

In this study, a 3D CFD model is developed to investigate hydrodynamics, heat transfer and producer gas compositions of a compact SMR reactor with an H_2 production rate of 2.5 Nm^3/hr (6 kg/day). The effect of a sleeve protecting the catalytic reactor against the hot combustion gas on heat transfer is rigorously examined. The realizable $k-\varepsilon$ turbulence model for gas mixtures and the radiative transfer equation (RTE) by a discrete ordinates (DO) method are incorporated into the CFD model. An equilibrium SMR reaction kinetic model (Xu and Froment, 1989) is employed to the catalyst-bed zone which was assumed as porous media. The CFD results are validated with experimental data in terms of axial temperatures, producer gas compositions, and thermal efficiency. The CFD model will provide a tool for improving the design and the thermal efficiency of the compact SMR reactor with the sleeve.

2. Compact steam methane reforming (SMR) process

A compact SMR process with a height of 460 mm and an outer diameter of 114 mm is shown in Figure 1. The compact SMR reactor is composed of two primary domains: (a) the fluid domain consists of the combustor and reactor, and (b) the solid domain covers walls between fluids. The heat is generated by combustion of CH_4 with air at the center, then transferred through the sleeve gap. The sleeve gap size is 3 mm. It is assumed that the combustion gas enters at 1,100 ºC and 1 atm. The feed gas (steam and NG) is injected by 16 nozzles through a porous heat carrier zone for pre-heating the mixture.

Material properties of solids (wall, heat carrier and catalyst), gas components, and fluid mixtures are listed in Table 1. The refractory wall has a low thermal conductivity. The metallic heat carrier with a high porosity (γ) of 0.95 shows a high density and conductivity. A Ni-based catalyst supported on Al_2O_3 was used in this experiment, which ensures a low-pressure drop in the tubular reactor (ReforMax, 2017). A high thermal conductivity of 235 W/m/K at an actual operating temperature of about 700 ºC was used for the catalyst, which is higher than typical catalysts (Palma et al., 2012). The incompressible ideal gas was used for the density of the gas mixture. The ideal gas mixing-law was applied for the specific heat capacity, thermal conductivity, and viscosity.

3. Modelling and simulation

A steady-state 3D simulation of the compact SMR process was performed using a commercial CFD code, ANSYS Fluent R17 (ANSYS Inc., USA). The gas phase was assumed to be incompressible. The CFD model included a realizable $k-\varepsilon$ turbulence model, discrete ordinates (DO) radiation method and a species transport model with SMR reactions.

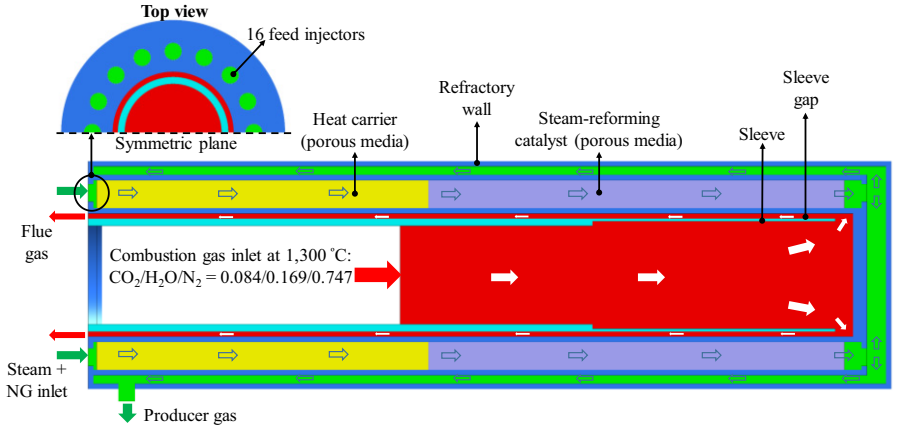


Figure 1. Geometry and mesh structure of compact SMR process.

Table 1. Material properties.

Name	Density (ρ, kg/m^3)	Specific heat capacity (c_p, J/kg/K)	Thermal conductivity (k, W/m/K)
Wall	7,874	449	42
Heat carrier	8,920	384.4	348
Catalyst	2,719	990	235
Mixture	Incompressible ideal gas	Ideal-gas mixing law	

Table 2. Boundary conditions.

Boundary name	**Type**	**Value**		
		Velocity (m/s)	Molar ratio	Temperate (°C)
Combustor inlet	Mass flow inlet	4.25	$CO_2/H_2O/N_2$ = 0.084/0.169/0.747	1300
SMR inlet	Mass flow inlet	3.24	CH_4/H_2O = 0.25/0.75	300
		Absolute pressure (Pa)		
Combustor outlet	Pressure outlet	2.026×10^5		
SMR outlet	Pressure outlet	2.026×10^5		
		Heat transfer coefficient (W/m^2/K)		Free stream temperature (°C)
Outer wall	Convection	9		300

An equilibrium chemical reaction model (R_n) of SMR on nickel-based catalyst pellets proposed by Xu and Froment (Xu and Froment, 1989), which was derived from the Langmuir-Hinshelwood mechanism, is used in this study. SMR boundary conditions at the inlet and outlet are listed in Table 2.

4. Results and discussions

The CFD model was solved using a finite volume-based solver, ANSYS Fluent R17 (ANSYS Inc., USA). The calculation was performed on a workstation with 24 cores (2.7

GHz CPU) and 128 GB RAM. The mesh independence test was first presented. Then the effectiveness factor function was introduced, and the CFD results were validated with experimental data. Finally, the performance criteria were analyzed for the reformer.

The producer gas composition (y_i, dry basis), H_2 production rate (F_P, $\mathrm{Nm^3/h}$), and lower heating value (LHV) efficiency (η_{LHV}) are compared for the CFD model, our experiment, pure kinetic model, and literature data of industrial-scale SMR processes. The wet-based producer gas compositions along the normalized catalyst length from the present CFD model are compared to the data from a 1D tube-segment mathematical model in an industrial-scale SMR process (Latham, 2008) in Fig. 2a. The gas compositions at the end of the normalized reactor length are almost the same for the two results. Fig. 2b shows the wall heat fluxes through sleeve wall ($q_{w,S}$, $\mathrm{kW/m^2}$) and catalyst outer wall ($q_{w,R}$, $\mathrm{kW/m^2}$) with respect to the normalized length of the reactor. The q_w has a symmetric shape along the reactor length due to a counter-current heat exchange through the sleeve. Whereas, the $q_{w,S}$ of Lao *et al*. (2016) decreases monotonically because of a co-current flow in a top-fired reactor. This SMR reactor with the sleeve leads to a flatter heat flux than the top-fired one. The mean value of $q_{w,S}$ of this study (=55 $\mathrm{kW/m^2}$) is lower than that (=75 $\mathrm{kW/m^2}$) of the industrial-scale SMR process (Lao et al., 2016). The flatter heat flux has a more advantage in operating and fixed costs (Ricca et al., 2017) and a lower risk in hot-spot phenomena (Palma et al., 2012).

Higher thermal efficiency was observed with a more uniform temperature distribution of the reformer. The sleeve used in this study helps to uniformly distribute the temperature of the wall of the reformer (or heat flux) over the axial direction, as mentioned earlier. The uniformity of heat flux was used to evaluate how much the heat flux at the inner wall of the catalyst reactor ($U_{q,R}$) and sleeve wall ($U_{q,S}$) are distributed evenly.

$$U_{q,R} = 1 - \frac{\int \left| q_{w,R} - \overline{q}_{w,R} \right| dA}{2\overline{q}_{w,R} A_{R,i}}, \quad \overline{q}_{w,R} = \frac{\int q_{w,R} dA}{A_{R,i}} \tag{1}$$

$$U_{q,S} = 1 - \frac{\int \left| q_{w,S} - \overline{q}_{w,S} \right| dA}{2\overline{q}_{w,S} A_{S,i}}, \quad \overline{q}_{w,S} = \frac{\int q_{w,S} dA}{A_{S,i}} \tag{2}$$

Where A is area, q is heat flux through reactor inner wall ($q_{w,R}$) and sleeve inner wall ($q_{w,S}$) for each computational face cell i.

Table 3 summaries the $U_{q,R}$ and $U_{q,S}$ for present CFD result and literature. Both of $U_{q,R}$ and $U_{q,S}$ are higher than a conventional top-fired SMR reactor (Lao *et al*., 2016). The CFD result proves the function of sleeve gap on flattening the heat flux in present SMR reactor design. The amount of heat flux through the inner wall of the catalyst reactor is around 3.4 times higher than the heat flux from the combustor to the inner wall of the sleeve because of the convection flow inside the sleeve gap. However, the direct heat transfers from combustor to sleeve gap through sleeve wall at the top part of sleeve is higher due to high temperature of the inlet, therefore the total heat flux through sleeve gap is smoothly distributed from the top to the bottom of reactor. As the result, the heat flux through reactor in presence of sleeve wall is more uniformly distributed than the conventional top-fired reactor design.

Table 3. Heat transfer coefficient and temperature uniformity of annulus SMR reactor

	$U_{q,R}$ (this study)	$U_{q,S}$ (this study)	$U_{q,R}$ (Lao *et al*., 2016)
Values	0.89	0.72	0.69

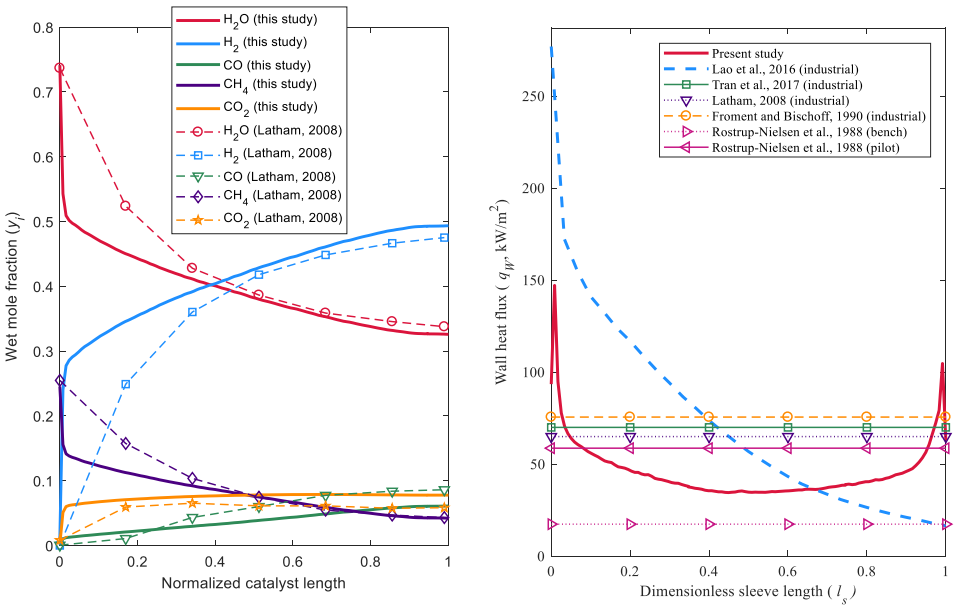
Characteristics of flow streamlines are illustrated in Fig. 3a. Figs. 3b and 3c show contours of absolute pressure and temperature on the symmetric plane. In the pressure contour, the pressure drops in the combustor and reactor are around 360 and 654 Pa, respectively. The pressure drop in the combustor is mainly caused by the sleeve gap, whereas that in the reactor is caused by the porous heat carrier and catalyst beds.

5. Conclusions

A three-dimensional computational fluid dynamics (CFD) model was developed for a compact type steam methane reforming (SMR) reactor to evaluate wall heat flux and to investigate the effect of a sleeve between the combustor and reactor on heat flux. The $k-\varepsilon$ realizable turbulence model and discrete ordinates (DO) method of radiative transfer equation were employed to the CFD model. The CFD results such as temperature profile, producer gas composition, production rate, heating efficiency, and wall heat flux were validated with experiment data and other studies in the literatures. Pressure, temperature, flow stream lines, and producer gas compositions inside the reactor were analysed. The compact-type SMR reactor showed flatter heat flux, compared to conventional top-fired SMR reactors.

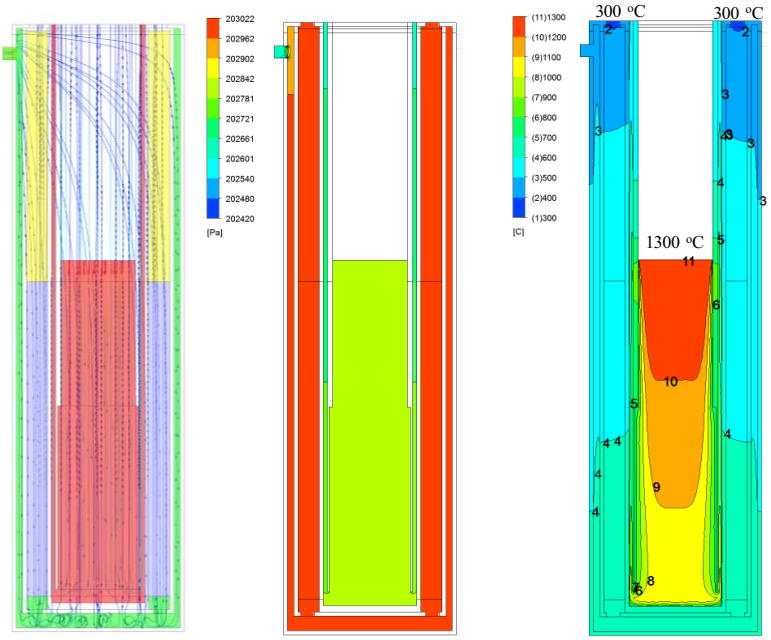
Acknowledgements

This work was conducted under the framework of the research and development program of the Korea Institute of Energy Research (B7-2424).



(a) Producer gas composition (b) Heat flux

Figure 2. Comparison of (a) producer gas compositions (y_i) and (b) heat flux between this study and literature.



(a) Flow stream lines (b) Absolute pressure (c) Temperature

Figure 3. Flow streamlines, pressure and temperature profiles.

References

Behnam, M., Dixon, A.G., Wright, P.M., Nijemeisland, M., Stitt, E.H., 2012. Comparison of CFD simulations to experiment under methane steam reforming reacting conditions. Chem. Eng. J. 207, 690–700.

Han, Y.L., Xiao, R., Zhang, M.Y., 2007. Combustion and Pyrolysis Reactions in a Naphtha

Kumar, A., Baldea, M., Edgar, T.F., 2016. Real-time optimization of an industrial steam-methane reformer under distributed sensing. Control Eng. Pract. 54, 140–153.

Lao, L., Aguirre, A., Tran, A., Wu, Z., Durand, H., Christofides, P.D., 2016. CFD modeling and control of a steam methane reforming reactor. Chem. Eng. Sci. 148, 78–92.

Latham, D., 2008. Mathematical modelling of an industrial steam methane reformer. Dep. Chem. Eng. Queen's University.

Ngo, S.I., Lim, Y. Il, Song, B.H., Lee, U. Do, Lee, J.W., Song, J.H., 2015. Effects of fluidization velocity on solid stack volume in a bubbling fluidized-bed with nozzle-type distributor. Powder Technol. 275, 188–198.

Palma, V., Ricca, A., Ciambelli, P., 2012. Monolith and foam catalysts performances in ATR of liquid and gaseous fuels. Chem. Eng. J. 207, 577–586.

ReforMax, C., 2017. Clariant ReforMax LDP Plus: a new generation of reforming catalysts with ultra-low pressure drop. Focus Catal. 2017, 4. doi:10.1016/J.FOCAT.2017.05.020

Ricca, A., Palma, V., Martino, M., Meloni, E., 2017. Innovative catalyst design for methane steam reforming intensification. Fuel 198, 175–182.

Seo, Y.-S., Seo, D.-J., Seo, Y.-T., Yoon, W.-L., 2006. Investigation of the characteristics of a compact steam reformer integrated with a water-gas shift reactor. J. Power Sources 161,

Tran, A., Aguirre, A., Durand, H., Crose, M., Christofides, P.D., 2017. CFD modeling of a industrial-scale steam methane reforming furnace. Chem. Eng. Sci. 171, 576–598.

Xu, J., Froment, G.F., 1989. Methane steam reforming, methanation and water-gas shift: I. Intrinsic kinetics. AIChE J. 35, 88–96.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50047-1

Computational fluid dynamics of gas-liquid bubble column with hydrocracking reactions

Bay Van Tran[a], Son Ich Ngo[a], Young-il Lim[a,*], Woohyun Kim[b], Kang-Seok Go[b], and Nam-Sun Nho[b]

[a] *CoSPE, Department of Chemical Engineering, Hankyong National University, Gyeonggi-do, Anseong-si, Jungang-ro 327, 17579 Republic of Korea*
[b] *Center for Convergent Chemical Process (CCP), 141 Gajeong-ro, Yuseong-gu, Daejeon, 34114 Republic of Korea*
limyi@hknu.ac.kr

Abstract

A gas-liquid Eulerian computational fluid dynamics (CFD) model coupled with a population balance equation (PBE) was presented to investigate hydrodynamics of a bubble column with hydrocracking reactions in H_2-vacuum residue oil system (H_2-VR) in terms of pressure drop, gas holdup, mean bubble size, and bubble surface area. The column had a height 1.9 m and 0.05 m in inner diameter. The column was initially filled with a vacuum residue (VR) and vacuum gas oil (VGO) mixture of a height of 1.9 m. The VR/VGO mixture with 80% VR and 20% VGO by weight, was injected at a velocity of 0.271 mm/s. The H_2 was injected through the gas distributor at a superficial gas velocity of 6.4 mm/s. The bubble column was operated at 160 bar, 430°C. For the H_2-VR system at the high pressure and temperature, the pressure drop, the total gas holdup, the mean bubble size and the specific bubble surface area were estimated to be 5,984 Pa/m, 8%, 3.2 mm and 150 m^2/m^3, respectively, using the CFD-PBE model. The compositions of the components were also calculated. This CFD study provides a potential to design the bubble column and to optimize the operating conditions.

Keywords: Bubble column, Computational fluid dynamics (CFD), Population balance equations (PBE), Specific surface area.

1. Introduction

Bubble column reactors are widely used in chemical, petrochemical, biochemical, and metallurgical industry for gas-liquid and gas-liquid-solid contacting or chemical reactions (Bhole et al., 2008; Liang et al., 2016). Because of the large interfacial area available, bubble columns offer excellent heat- and mass-transfer characteristics (Monahan et al., 2005).

The gas holdup and the bubble size are two important parameters in the bubble column design. The BSD determines gas holdup and gas-liquid interfacial area, which is of importance for mass transfer between the phases (Sharaf et al., 2016). The gas holdup is influenced by pressure, liquid and gas velocities, viscosity, surface tension, and liquid and gas densities.

Computational fluid dynamics (CFD) has emerged as a promising tool for the simulation of hydrodynamics in bubble columns (Bhole et al., 2008; Liang et al., 2016). Two approaches including the Euler-Lagrange (E-L) approach and the Euler-Euler (E-E) approach are primarily used (Liang et al., 2016; Ngo et al., 2015; Pham et al., 2015).

Since the E-L model is normally limited to a relatively small number of particles or bubbles because of computational expense (Bayraktar et al., 2013; Liang et al., 2016), the E-E approach is typically preferred for modeling the bubble column (Bhole et al., 2008).

The population balance equation (PBE) coupled with the flow equations is a promising approach to simulate gas-liquid dispersed flows (Bayraktar et al., 2013). The coupled model takes into account bubble breakup and coalescence, describing the gas holdup and the BSD (Liang et al., 2016).

Fluid dynamic studies on multiphase systems at high pressure and temperature remain scanty. Few researchers have addressed the effect of pressure on hydrodynamics of the bubble column, using the multiphase Eulerian CFD model with PBE.

This study investigated hydrodynamics of a bubble column reactor with chemical reaction at 160 bar and 430°C, using a multiphase Eulerian CFD model with PBE. Firstly, the gas–liquid Eulerian CFD model coupled with a turbulence and PBE is developed for a pilot-scale pressurized bubble column. The reaction model is also integrated into the CFD model. Finally, the hydrodynamic parameters and compositions of components are estimated in a H_2/vacuum residue system.

2. Geometry, physical properties and boundary condition

A pilot-scale bubble column with a height of 1.9 m and an inner diameter of 0.05 m was constructed to perform catalytic hydrocracking of heavy at 160 bar and 430°C. A perforated plate was used as the gas distributor.

2.1. Geometry and meshing of CFD calculation domain

The two dimensional (2D) symmetric geometry and the mesh structure of the bubble column are depicted in Fig. 1. In this study, two inlet boundary conditions were considered. One is the premixed inlet, where a premixed mixture of both gas and liquid is injected to the bottom of the bubble column. The other is the individual inlet, where gas and liquid enter the bubble column uniformly and individually. For the outlet boundary condition, a virtual outlet with an extended column is used to minimize the effect of the reverse flow at the exit on the composition of the liquid phase.

2.2. Physical properties and boundary conditions

The physical properties such as molecular weight (MW), density (ρ), viscosity (μ) and surface tension (σ) between the gas and liquid phases affect the bubble behavior (Bhole et al., 2008). Table 1 shows the physical properties of each component in the liquid and gas phases. All the species except H_2 are pseudo-components classified by the boiling point range. The boundary conditions of the CFD model are shown in Table 2.

3. Gas–liquid Eulerian CFD model couple with population balance model

The fluid flow was assumed to be incompressible and isothermal, and the gas was considered as a dispersed phase. A gas–liquid Eulerian CFD model was developed for the bubble column. The governing equation included the continuity equation for total mass conservation and the momentum equation with drag and non-drag forces. The modified Tomiyama model with a drag force correction factor was used (Liang et al., 2016). The k-ω turbulence equation, the population balance equation (PBE) and the reaction model were integrated into the CFD model.

Table 1. Physical properties of gas and liquid at 160 bar and 430°C.

Component	Phase	MW(kg/kmol)	ρ (kg/m^3)	μ (Pa·s)	σ (N/m)
Vacuum residue	Liquid	629.8	816.3	1.21×10^{-3}	20.23×10^{-3}
Vacuum gas oil	Liquid	377.9	685.9	2.07×10^{-4}	9.41×10^{-3}
Distillate	Liquid	195.6	577.5	5.30×10^{-5}	1.45×10^{-3}
Naphtha	Liquid	120.8	419.7	4.30×10^{-5}	5.40×10^{-5}
Gas	Liquid	30.0	160.0	1.50×10^{-7}	-
	Gas	30.0	86.0	2.19×10^{-5}	-
Hydrogen	Liquid	2.0	74.0	3.80×10^{-9}	-
	Gas	2.0	5.2	1.60×10^{-5}	-

Table 2. Boundary conditions of CFD model.

	Boundary conditions
Gas & liquid inlets	Incompressible gas, mass flow inlet. Gas load (H_2): U_G = 6.4 mm/s, P = 160 bar, T = 430°C. Liquid load (VR/VGO): U_L = 0.271 mm/s, P = 160 bar, T = 430°C.
Gas outlet	Degassing.
Liquid outlet	Pressure outlet.

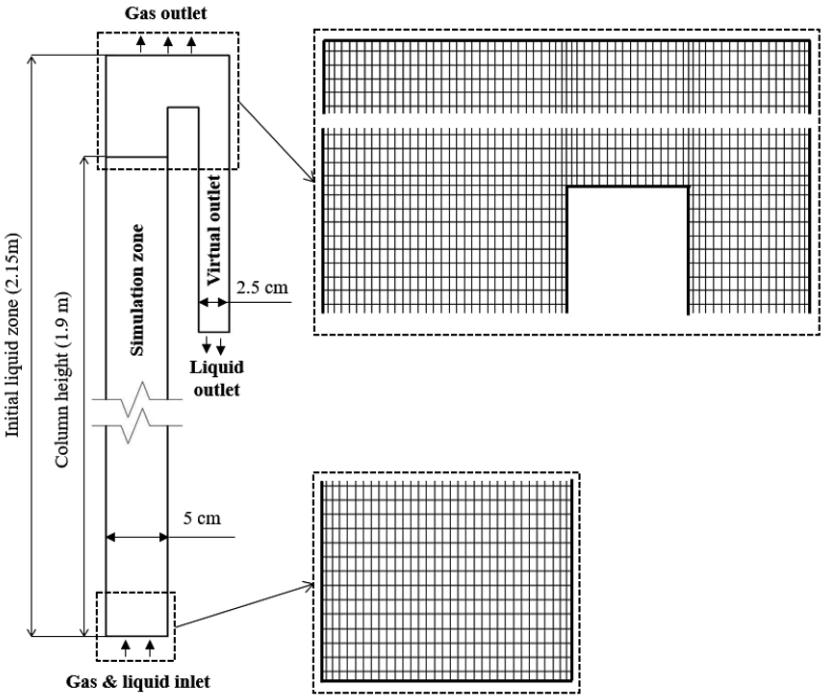


Fig. 1. Geometry and meshing of CFD calculation domain

4. CFD simulation results

The CFD-PBE model was first solved on different mesh numbers to examine their effect on hydrodynamics in the bubble column. Then, CFD results of the inlet boundary

conditions (individual and premixed inlets) were compared. Finally, the hydrodynamics and the composition of species after 80 minutes were estimated in the H_2/vacuum residue system.

4.1. Mesh independent test without chemical reaction

The effect of the mesh number is shown in Fig. 2 in terms of the specific pressure drop ($\Delta P/L$), the total gas holdup ($\overline{\alpha}_G$), the Sauter diameter (d_{32}) and the specific surface area (a_s) at 160 bar and 430°C. The specific pressure drop and gas holdup are almost the same in the case of medium and fine meshes. The Sauter diameter and specific surface area change with the number of cells. However, almost the same BSD shape is observed for medium and fine meshes (Fig. 3a). Thus, in this study, the medium mesh (30,000 cells) was selected to reduce the computational cost.

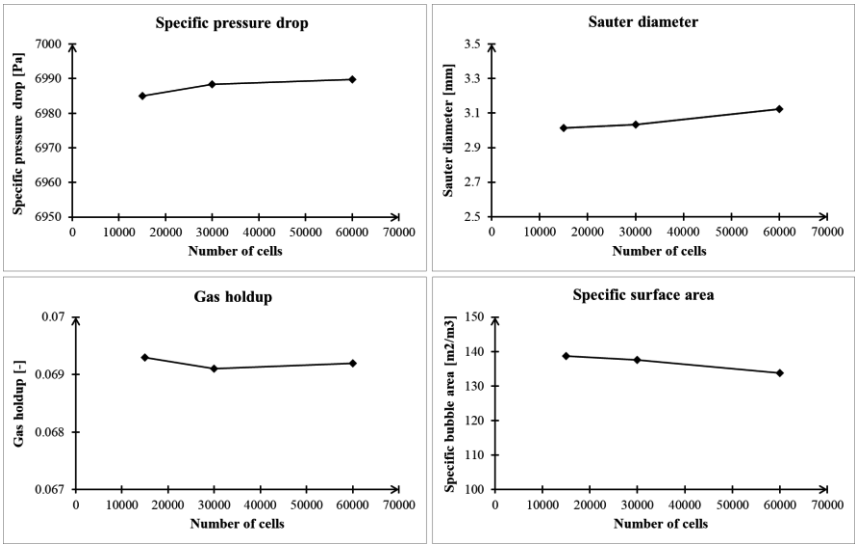


Fig. 2. Mesh independent test on coarse, medium and fine meshes with premixed inlet and virtual outlet with $N_b = 20$, $0.1 \leq d_b \leq 30$ mm at 160 bar and 430°C, $U_G = 6.4$ mm/s, $U_L = 0.271$ mm/s (time-averaged value).

4.2. Comparison of individual and premixed inlets without chemical reaction

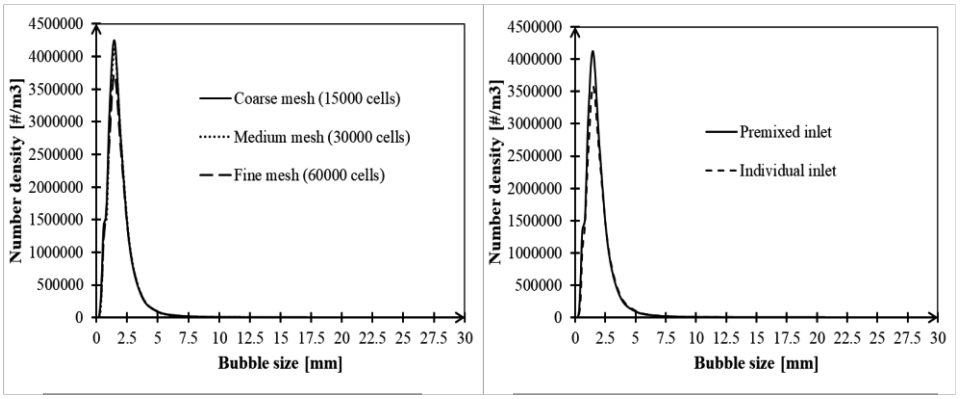
The effect the inlet boundary condition on hydrodynamics of the bubble column was investigated. Fig 3b compares the BSD between the two cases. The BSD of individual inlet is a little narrower than that of the premixed inlet. Table 3 compares the hydrodynamic parameters of the two cases. The difference of the parameters does not exceed 5%. Thus, to simplify the geometry and reduce the calculation time, the premixed inlet was selected in this study.

4.3. Hydrodynamics and product composition after reaction

The gas holdup reaches a stable state after about 0.4 h, as shown in Fig. 4a. Fig. 4b shows the BSD at t = 80 min after the chemical reactions take place. The hydrodynamics at t = 80 min are reported in Table 4.

Table 3. Hydrodynamic parameters of individual and premixed inlets with N_b = 20, 0.1 ≤ d_b ≤ 30 mm at 160 bar and 430°C, U_G = 6.4 mm/s, U_L = 0.271 mm/s (time-averaged value).

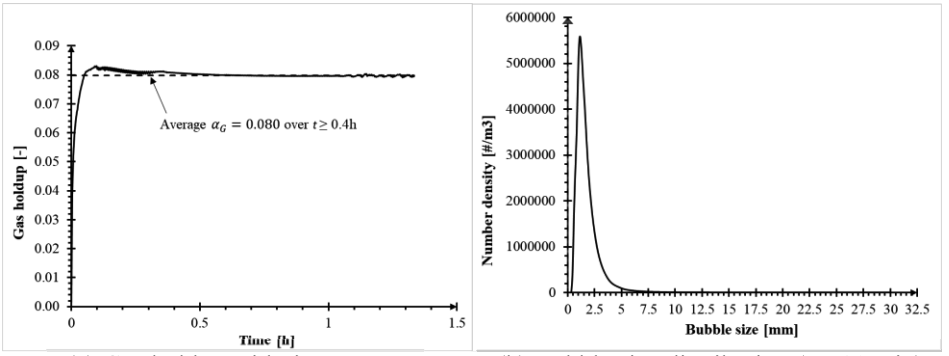
	$\Delta P/\Delta L$ (Pa/m)	$\bar{\alpha}_G$	d_{32}	a_s (m^2/m^3)	Calculation time (h)
Premixed inlet	6,988	0.069	3.0	138	5
Individual inlet	6,821	0.068	3.1	132	6
Difference (%)	-2.4	-1.9	2.4	-4.2	20



(a) BSD on coarse, medium, fine meshes of premixed inlet

(b) Comparison of BSD between individual and premixed inlets.

Fig. 3. BSD on coarse, medium, fine meshes of premixed inlet and comparison of BSD between individual inlet and premixed inlet with N_b = 20, 0.1 ≤ d_b ≤ 30 mm at 160 bar and 430°C, U_G = 6.4 mm/s, U_L = 0.271 mm/s (time-average value).



(a) Gas holdup with time

(b) Bubble size distribution (t = 80 min)

Fig. 4. Evolution of gas holdup with time and bubble size distribution (BSD).

Table 4. Hydrodynamic parameters of bubble column reactor with N_b = 20, 0.1 ≤ d_b ≤ 30 mm at 160 bar and 430°C, U_G = 6.4 mm/s, U_L = 0.271 mm/s, t = 80 min.

$\Delta P/\Delta L$ (Pa/m)	$\bar{\alpha}_G$	d_{32}	a_s (m^2/m^3)
5,984	0.080	3.2	150

After 80 min of reaction, the concentrations of vacuum residue, vacuum gas oil, distillate, naphtha and gas are about 52%, 39%, 3.4%, 4.7% and 0.9% by weight, respectively. All the species except gas approach monotonically a steady-state with time. Gas increases rapidly at the initial moment and decreases gradually toward a stable state. Fig. 5 shows the concentration of species at the exit of the reactor column with time.

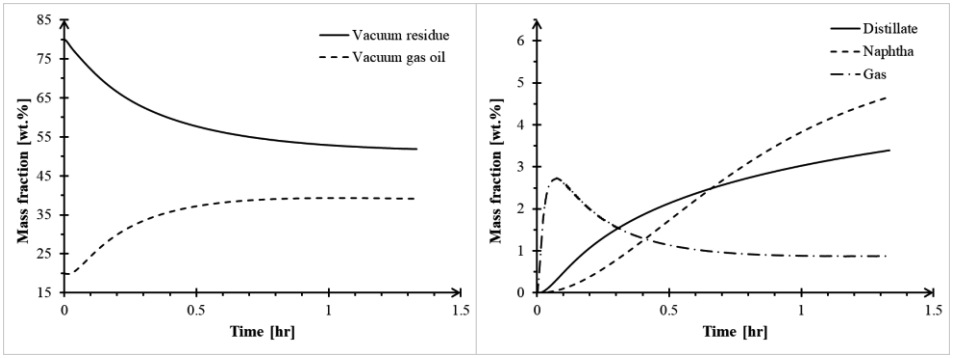


Fig. 5. Concentration of species with time at bubble column exit.

5. Conclusions

For a H_2/heavy oil system, the gas–liquid Eulerian CFD model coupled with a k-ω turbulence equation and a population balance equation (PBE) was developed to investigate hydrodynamics and the product composition of a bubble column reactor. The hydrodynamics were not sensitive to the gas-liquid mixing condition at the inlet. Thanks to the virtual outlet with an extended column, the reverse flow was avoided and the computational time was reduced. Future work should focus on the validation of the product composition with experimental data.

References

Bayraktar, E., Mierka, O., & Turek, S., 2013. Numerical Solutions of Population Balance Equations within Liquid/Gas-Liquid Flow Simulations. Chemie Ingenieur Technik, (85), 1137-1145.

Bhole, M. R., Joshi, J. B., & Ramkrishna, D., 2008. CFD simulation of bubble columns incorporating population balance modeling. Chemical Engineering Science, (63), 2267-2282.

Liang, X.-F., Pan, H., Su, Y.-H., & Luo, Z.-H., 2016. CFD-PBM approach with modified drag model for the gas–liquid flow in a bubble column. Chemical Engineering Research and Design, (112), 88-102.

Monahan, S. M., Vitankar, V. S., & Fox, R. O., 2005. CFD predictions for flow-regime transitions in bubble columns. AIChE Journal, (51), 1897-1923.

Ngo, S. I., Lim, Y.-I., Song, B.-H., Lee, U.-D., Lee, J.-W., & Song, J.-H., 2015. Effects of fluidization velocity on solid stack volume in a bubbling fluidized-bed with nozzle-type distributor. Powder Technology, (275), 188-198.

Pham, D. A., Lim, Y.-I., Jee, H., Ahn, E., & Jung, Y., 2015. Effect of ship tilting and motion on amine absorber with structured-packing for CO2 removal from natural gas. AIChE Journal, (61), 4412-4425.

Sharaf, S., Zednikova, M., Ruzicka, M. C., & Azzopardi, B. J., 2016. Global and local hydrodynamics of bubble columns – Effect of gas distributor. Chemical Engineering Journal, (288), 489-504.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50048-3

Effect of ship motion on separation efficiency in crude oil separator with coalescer

Thuy T. Le[a], Young-Il Lim[a,*], Chi-Kyun Park[b], Byung-Don Lee[b], Byung-Gook Kim[b], and Dong-Ha Lim[c]

[a]*CoSPE, Department of Chemical Engineering, Hankyong National University, 327 Jungang-ro, Anseong, Gyonggi-do, 17579 Korea*
[b]*Jeonjin EnTech Co., LTD, 38 Gadal 1-ro, Gangseo-gu, Busan, 46729 Korea*
[c]*KITECH Busan, 60 Gwahaksandan 1-ro, Busan, 46742 Korea*
limyi@hknu.ac.kr.

Abstract

Three-phase (air-oil-water) crude oil is separated using the gravity separator with internal structures such as coalescer, demister, and weir plate. The objective of this study is to develop a multiphase liquid/liquid/gas computational fluid dynamics (CFD) model and to predict hydrodynamics and separation performance under the ship motion. An Eulerian-Eulerian CFD model combined with the k-ε turbulence model and rotation matrix was used. The air-phase was considered continuous, while oil and water phases were assumed as mono-dispersed secondary-phases with an average diameter of 100 μm. The momentum equation included drag force, lift force and resistance of porous media. The exit pressures of oil and water were examined, which played an important role in determining the liquid level of oil and water in the separator. It was found that the exit pressures of water and oil were at 6.3 kPa and 5.1 kPa, respectively, which showed an effective separation of water and oil. The mean residence time for water and oil obtained from the CFD simulation was around 15 minutes. The vertical velocities of oil and water in the gravity direction along flow direction were analyzed at several liquid levels. The predicted oil separation efficiency (99.85%) were in line with experimental data (99.96 %). The CFD results also showed under the pitching motion. Three ship motions influenced significantly the separation performance in the separator.

Keywords: Separation efficiency, Liquid outlet pressure, Residence time, Three ship motions, Computational fluid dynamics (CFD)

1. Introduction

The separation processes of gas, oil and water are of importance in the petroleum industry. Currently, there are several available separation methods including filtration, membrane separation, gravity or centrifugal settling, chemical demulsification, pH adjustment, heat treatment, and electrostatic demulsification (Noik et al.; Roques-Carmes et al., 2014; Šećerov Sokolović et al., 2009; Zhao & Li, 2011; Zolfaghari et al., 2016). The gravity separation processes based on density differences of immiscible phases are often used in industry. However, the gravity settling approach requires a big equipment size to reach effective separation, which is not practical and economic, especially, not convenient with offshore conditions. Hence, the three-phase gravity separators are equipped with a different type of internals to enhance droplet coalescence and to optimize their length (Kharoua et al., 2013).

Recently, the experiments and CFD simulations have been used to study the effect of design and operating parameters on the separation efficiency. Speth et al. (2002) presented coalescence of secondary dispersions in fiber beds. Hernandez-Perez and Abdulkadir (2010) indicated a strong dependency of phase separation on the flow velocity and the droplet size. Sokolovic et al. (2010) investigated the effect of the coalescer geometry on steady-state bed coalescence by experiment. The effect of changing the position and size of baffles and modifying the thickness of liquid levels on separation performance were considered in Pourahmadi Laleh et al. (2013). The performance of internal structures on separation efficiency were investigated by Kharoua et al. (2012)

However, few researchers have addressed the effect of exit pressure of oil and water outlets. Moreover, limited studies examined the performance of separator under the off-shore condition. This study proposes a computational fluid dynamic (CFD) model to predict internal flow patterns and the effects of exit pressure and ship motions on separation efficiency.

2. Oil separation with coalescer

A three-phase separator includes some internal structures such as perforated plate, coalescer and demister. There are one inlet and three outlets. **Table 1** shows physical properties of three phase and the mass flow rates at the inlet. The operating pressure and temperature are set to 1 bar and 20°C, respectively.

Table 1 Physical properties of air, water and oil and inlet mass flow rate.

Component	Phase	Mass flow rate (kg/s)	ρ (kg/m^3)	μ (Pa.s)	σ (N/m) with air
Air	Gas	0.0444	1.2	$1.84\text{x}10^{-5}$	-
Water	Liquid	0.1067	998	$1.00\text{x}10^{-3}$	$72.8\text{x}10^{-3}$
Oil[a]	Liquid	0.0711	831	$7.21\text{x}10^{-3}$	$27.2\text{x}10^{-3}$

[a]Oil: synthetic oil

The geometry of the separator is shown in **Fig. 1**. The horizontal separator is 2.66 m long with a diameter of 0.78 m. Since the geometry is symmetric, a 3D symmetric geometry is used as the CFD domain. The polyhedral mesh is constructed and the number of cells is about 495,000.

3. CFD model

A 3D porous media CFD model was developed for the oil separator. The fluid flow was assumed to be incompressible, isothermal and Newtonian fluid. The unsteady turbulent multiphase flow was solved in a commercial CFD code, ANSYS Fluent 17.0. In the Eulerian-Eulerian model, the air-phase was considered continuous, while water and oil assume as mono-dispersed secondary phases. The two secondary phases were

represented as an average diameter of 100 μm. In this study, the momentum equations included drag force, lift force and resistance of porous media.

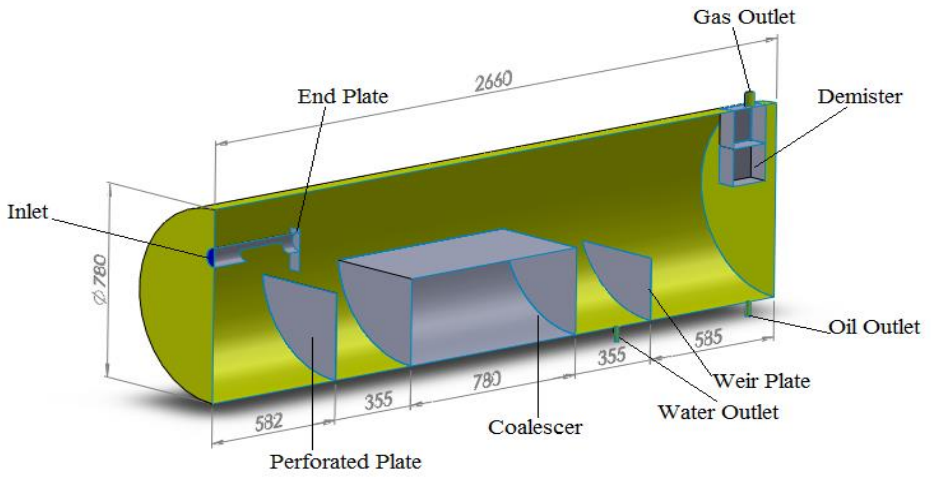


Fig. 1. Geometry of air-oil-water separator.

The phase-coupled SIMPLE method was chosen for pressure-velocity coupling. The first-order upwind and QUICK schemes were used for the spatial discretization of the momentum and volume fraction, respectively. The maximum number of iterations was limited to 50 at each time step, and the convergence tolerance was 1×10^{-3}. The calculation time was about 7 days with a flow time of 40 minutes.

The inlet and outlet boundary conditions were set to the velocity inlet (mass flow inlet) and the pressure outlet, respectively. The pressure at the gas outlet was atmospheric, while those at the liquid outlets were monitored to maintain the liquid levels and to ensure an overall phase mass balance. At the beginning of simulation, the separator is filled with gas (air) and liquid (water).

4. Results and discussions

The effect of exit pressure in water outlet and oil outlet on the liquid level were investigated. The oil separation efficiency predicted by CFD was compared with experiment data. Next, the dynamic behavior of mass flow rate in oil and water outlet under pitching motion was shown. Finally, the oil separation performance between vertical standing and pitching motion was compared.

4.1. Effect of exit pressure on separation efficiency

Table 2 shows the effect of exit pressure on oil separation efficiency. The oil outlet pressure was kept constant (P_O = 4.2 kPa), and the water outlet pressure was changed from 5.6 to 6.4 kPa. As the pressure at the water outlet increases, the level of the oil layer in the water side decreases, and the oil level in the oil side increases. Thus, increasing the water exit pressure decreases the oil separation efficiency. When P_o changes from 4.2 to 5.4 kPa at P_W=5.6 kPa, the oil content in the outlet water increases and the separation

efficiency decreases. Therefore, it is necessary to choose a proper range of outlet pressures of water and oil.

Table 2. Effect of exit pressures to separation efficiency

			P_O = 4.2 kPa			P_W = 5.6 kPa		
Pressure in outlet (kPa)			P_W = 5.6	P_W = 6.0	P_W = 6.4	P_O = 4.2	P_O = 4.7	P_O = 5.4
Volume fraction (%)	Outlet gas	water	6.0E-07	6.3E-07	8.8E-07	6.0E-07	9.2E-07	8.1E-07
		oil	4.4E-06	2.1E-06	2.9E-06	4.4E-06	8.6E-07	7.8E-07
		air	99.99	99.99	99.99	99.99	99.99	99.99
	Outlet oil	water	0.17	0.55	1.83	0.17	0.12	0.11
		oil	99.83	99.45	98.17	99.83	99.88	99.88
		air	4.7E-05	6.6E-05	2.6E-04	4.7E-05	6.4E-05	8.5E-05
	Outlet water	water	99.95	99.96	99.94	99.95	99.06	71.10
		oil	0.044	0.035	0.057	0.044	0.94	28.90
		air	1.5E-06	4.2E-07	2.5E-09	1.5E-06	4.9E-05	1.2E-04
Liquid level (cm)	Water side	Water	12.1	16.2	18.6	12.1	7.3	6.0
		Oil	12.9	8.8	6.4	12.9	17.7	19.0
	Oil side	Oil	4.2	5.9	11.5	4.2	11.5	19.9

In this study, the suitable water outlet pressure and oil outlet pressure were found to be P_W = 6.3 kPa and P_O = 5.1 kPa for effective separation. **Table 3** shows the separation efficiency obtained from the simulation and experimental data. The oil separation efficiency obtained from the CFD simulation agrees well with experimental data.

Table 3. Comparison of CFD results with experimental data.

		CFD results	Exp. data	Error (%)
Outlet gas	air	99.99	-	-
Outlet water	water	99.96	-	-
Outlet oil	oil	99.85	99.96	0.11

4.2. Effect of ship motion on separation efficiency

To investigate the effect of the pitching motion, the exit pressures in water and oil outlets kept constant at 6.3 kPa and 5.1 kPa, respectively. For the case of 3° pitching motion, there are some backflows (positive values), which is not realistic. The fluctuation amplitude of the mass flow rate is bigger than that of 1.5° in both the water and oil outlets, as shown in **Fig. 2**.

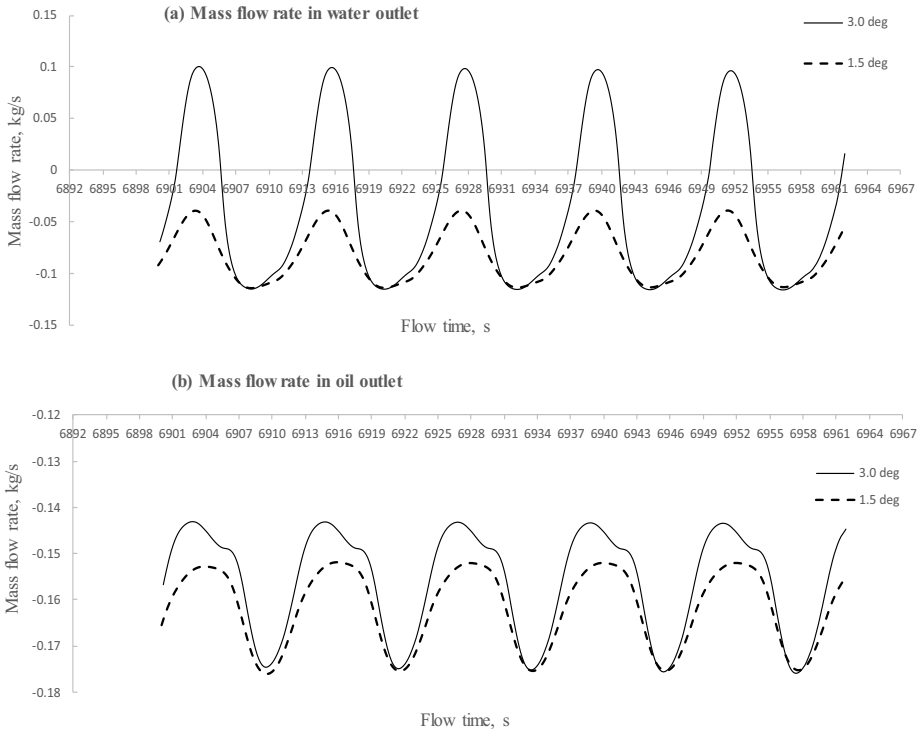


Fig. 2. Dynamic behavior of mass flow rate in water and oil outlets.

Table 4 summerizes the separation efficiency of vertical standing and pitching motion cases. The pitching motion decreases significantly the separation efficiency by 13.3% for the 1.5° and 17.0 % for 3° pitching motion. The separation efficiencies under pitching motion are low because of serious fluctuation caused by spurious backflows from the water outlet. These spurious backflows would be resolved by dynamic pressure control of the water and oil outlets.

Table 4. Comparison of separation efficiency for vertical standing and pitching motion.

	Vertical standing	Pitching motion									
Angular (degree)	0	1.5					3				
Time	0	0s	3s	6s	9s	mean	0s	3s	6s	9s	mean
Outlet gas	99.99	99.99	99.99	99.99	99.99	99.99	99.99	99.99	99.99	99.99	99.99
Outlet water	99.96	99.68	99.69	99.98	99.73	99.77	99.96	99.50	98.35	99.24	99.27
Outlet oil	99.87	87.26	86.52	85.70	86.88	86.59	83.56	83.38	81.66	82.79	82.85

5. Conclusions

A 3D CFD model for an air-water-oil separator with the perforated plate, coalescer and demister was built to identify hydrodynamics inside the three phase separator and to predict the oil separation efficiency under offshore operation.

The effect of exit pressures in water and oil outlets on the separation efficiency was investigated. It was found that the exit pressures at 6.3 kPa and 5.1 kPa for water and oil, respectively, were suitable for maintaining a liquid level of 25 cm with a good separation efficiency. The predicted oil separation efficiency (99.85%) agreed well with experimental data (99.96 %). The pitching motion decreased significantly on the separation efficiency by 13.3% for 1.5° and 17.0% for 3° pitching motion.

The CFD model proposed in this study did not include the population balance model (PBM) that accounts for the droplet size distribution, coalescence, and breakup phenomena of the secondary phases.

Acknowledgements

This research was supported by the Ministry of Trade, Industry & Energy(MOTIE), Korea Institute for Advancement of Technology(KIAT) through the Encouragement Program for The Industries of Economic Cooperation Region.

References

Hernandez-Perez, V., & Abdulkadir, M., The Effect of Mixture Velocity and Droplet Diameter on Oil-water Separator using Computational Fluid Dynamics (CFD), World Academy of Science, Engineering and Technology, 4, 2010.

Kharoua, N., Khezzar, L., & Saadawi, H., Using CFD to Model the Performance of Retrofit Production Separators in Abu Dhabi. In Abu Dhabi International Petroleum Exhhibition & Conference. Abu Dhabi, UA: Society of Petroleum Engineers, 2012.

Kharoua, N., Khezzar, L., & Saadawi, H., CFD Modeling of a Horizontal Three-Phase Separator: A Population Balance Approach, American Journal of Fluid Dynamics, 3, 101-118, 2013.

Noik, C., Chen, J., & Dalmazzone, C. S. H., Electrostatic Demulsification on Crude Oil: A State-of-the-Art Review. In: Society of Petroleum Engineers.

Pourahmadi Laleh, A., Svrcek, W. Y., & Monnery, W., Computational Fluid Dynamics-Based Study of an Oilfield Separator--Part II: An Optimum Design, 2013.

Roques-Carmes, T., Monnier, H., Portha, J.-F., Marchal, P., & Falk, L., Influence of the plate-type continuous micro-separator dimensions on the efficiency of demulsification of oil-in-water emulsion, Chemical Engineering Research and Design, 92, 2758-2769, 2014.

Šećerov Sokolović, R., Sokolović, S., & Šević, S., Oily water treatment using a new steady-state fiber-bed coalescer, Journal of hazardous materials, 162, 410-415, 2009.

Sokolovic, R. M., Govedarica, D. D., & Sokolovic, D. S., Separation of oil-in-water emulsion using two coalescers of different geometry, Journal of hazardous materials, 175, 1001-1006, 2010.

Speth, H., Pfennig, A., Chatterjee, M., & Franken, H., Coalescence of secondary dispersions in fiber beds, Separation and Purification Technology, 29, 113-119, 2002.

Zhao, H., & Li, G., Application of Fibrous Coalescer in the Treatment of Oily Wastewater, Procedia Environmental Sciences, 10, 158-162, 2011.

Zolfaghari, R., Fakhru'l-Razi, A., Abdullah, L. C., Elnashaie, S. S. E. H., & Pendashteh, A., Demulsification techniques of water-in-oil and oil-in-water emulsions in petroleum industry, Separation and Purification Technology, 170, 377-407, 2016.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50049-5

Mathematical Modeling of a Moving-Bed Reactor for Chemical Looping Combustion of Methane

Anca Ostace[a], Andrew Lee[b], Chinedu O. Okoli[c], Anthony P. Burgard[c], David C. Miller[c], Debangsu Bhattacharyya[a*]

[a]*Department of Chemical and Biomedical Engineering, West Virginia University, 401 Evansdale Drive, Morgantown, WV 26506, USA*
[b]*KeyLogic Systems, National Energy Technology Laboratory, 626 Cochrans Mill Road, Pittsburgh, PA 15236, USA*
[b]*National Energy Technology Laboratory, 626 Cochrans Mill Road, Pittsburgh, PA 15236, USA*
Debangsu.Bhattacharyya@mail.wvu.edu

Abstract

A flexible first principles mathematical model of a counter-current moving-bed reactor has been developed and implemented in Pyomo as part of the Institute for the Design of Advanced Energy Systems (IDAES) process systems engineering framework. The model consists of a set of differential and algebraic equations obtained by applying the mass, energy, and pressure balances. The model is applied to the simulation of the reducer reactor for the chemical looping combustion of methane using an iron-based oxygen carrier. The effect of several system and operating parameters on the methane and oxygen carrier conversions are investigated to help assess the performance of the moving-bed reducer. The results show that nearly full methane conversion (95.75 %) can be attained in the counter-current moving-bed regime using relatively low oxygen carrier/methane mass flow ratios.

Keywords: chemical looping combustion, moving-bed reactor model, IDAES, Pyomo

1. Introduction

Chemical looping combustion (CLC) has emerged as a new high-efficiency, advanced energy technology that has the potential of outperforming conventional technologies for fossil fuel combustion with CO_2 capture. In CLC, oxygen is transferred from air to a fuel via oxidation-reduction cycling of a solid oxygen carrier (OC) between two reactors. In the reducer, the OC reacts with the fuel, generating CO_2 and water. In the oxidizer, the reduced OC is regenerated with air, producing a hot O_2-depleted air stream that can be used to produce steam for power production. Thus, CLC provides a means for oxy-combustion without an expensive air separation unit.

To date, many different fuels (syngas, natural gas, biogas, coal, and biomass) have been successfully combusted at small to intermediate scales using nickel-, iron-, copper-, manganese-, or calcium-based OCs, in different types of gas-solid contactors operating in circulating fluidized bed, bubbling fluidized bed, and moving-bed (MB) regimes. Given that the performance of a CLC process is determined by the properties and performance of the particles, the system configuration, and the reactor design, choosing the optimal combination of these factors for a specific process becomes very important for achieving a highly efficient, cost-effective process. Computational tools that rely on advanced multi-scale models provide a systematic approach to rapidly synthesize and

optimize processes, accelerating the development of new technologies like CLC to maturity (Miller et al., 2014). Important requirements of mathematical models that can be used to optimize complex, integrated processes while supporting scale up are that they should be flexible and accurate for the parameter and variable space under investigation (Miller et al., 2017) and have exact first and second derivatives. To facilitate the development of custom models that are not available in commercial simulators and enable advanced process systems engineering (PSE) capabilities that support optimization, control and uncertainty quantification, the U.S. Department of Energy (DOE) initiated the Institute for the Design of Advanced Energy Systems (IDAES). IDAES is developing a next generation PSE framework to enable the rapid modeling, design, and optimization of new, innovative advanced energy systems such as CLC. In this paper, we present the development of a detailed, flexible, hierarchical, first principles model of a MB reducer within the IDAES PSE framework. The model is implemented in Pyomo (Hart et al., 2017), a Python-based open-source software package. Due to its hierarchical, modular structure, the model can be applied to the simulation of different gas-solid processes in a straightforward manner, given sufficient reaction information. To demonstrate the capabilities of the model, the MB reducer is applied to the combustion of methane (CH_4) using an iron-based OC. Additionally, a series of case studies are presented that evaluate the effect of several design and operating parameters on the conversion of both the OC and CH_4. Ultimately, this model will be used to support optimization of integrated CLC systems.

2. Model formulation

Fresh OC (hematite, 45 % Fe_2O_3 on an alumina inert support) enters at the top of the MB reducer and counter-currently interacts with a CH_4 gas stream fed at the bottom. A CO_2 and H_2O gas stream leaves at the top of the reducer, while spent OC (magnetite) is retrieved at the bottom. The H_2O vapor can be condensed from the reducer exhaust gas, leaving a concentrated CO_2 stream ready for compression and utilization or storage. The spent OC is circulated to the oxidizer, where it is reconstituted to its oxidized state using air and then reintroduced to the reducer. The reduction reaction of iron oxide Fe_2O_3 with CH_4 is given by the endothermic reaction:

$$CH_{4(g)} + 12Fe_2O_{3(s)} \rightarrow CO_{2(g)} + 2H_2O_{(g)} + 8Fe_3O_{4(s)} \quad (1)$$

The overall performance of the process is determined by reduction reaction (1), because oxidation is fast and thermodynamically favored (Li et al., 2010). Therefore, studying and modeling the reducer reactor is essential for the design of an efficient CLC process. To date, several MB adsorber models have been proposed for carbon capture (Kim et al. 2013, 2016, and Mondino et al. 2017); however, MB reducer models are scarce. While Tong et al. (2013) and Kathe et al. (2016) have presented models based on idealized reactors in series, no previous studies have reported a first principles mathematical model of a MB reducer for CLC suitable for equation-oriented flowsheet optimization. In the following, the MB reducer model is presented.

The model is a 1-D, steady-state, non-isothermal, gas-solid counter-current MB reactor model, based on the following assumptions:

1. Radial concentration and temperature gradients within the bed are negligible.
2. Axial dispersion and thermal conduction along the axial direction are negligible.
3. The solids flow with uniform velocity and the reactor bed voidage is constant.
4. No channeling occurs.

Table 1. Mass and energy balance equations of the MB reducer model.

$0 = -\frac{d(u_g C_i)}{dz} - (1-\varepsilon)\rho_s \frac{\upsilon_i}{\upsilon_{Fe_2O_3}} r_{gen}$	(2)
$0 = \frac{d(u_s q_j)}{dz} - (1-\varepsilon)\rho_s MW_{s,j} \frac{\upsilon_j}{\upsilon_{Fe_2O_3}} r_{gen}$	(3)
$0 = -\rho_g C_{p,g} u_g \frac{dT_g}{dz} - (1-\varepsilon) h_f \frac{6}{d_p}(T_g - T_s) - h_w \frac{4}{D_r}(T_g - T_w)$	(4)
$0 = \rho_s C_{p,s} u_s \frac{dT_s}{dz} + (1-\varepsilon) h_f \frac{6}{d_p}(T_g - T_s) + (1-\varepsilon)\rho_s(-\Delta H_{rxn}) r_{gen}$	(5)
$0 = a_{w,in} h_w (T_g - T_w) - a_{w,out} U (T_w - T_\infty)$	(6)

5. The ideal gas law holds in the gas phase.
6. The pressure balance is given by the Ergun equation, using the slip velocity between the gas and the solid phases.

The mass and energy balance equations of the mathematical model are presented in Table 1. Eq. (2) and (3) are the gas and solid component mass balance, Eq. (4), (5) and (6) are the gas phase, solid phase and reactor wall energy balance equations. In addition to the balance equations, the MB model includes the Ergun equation and additional correlations used to compute gas and solid properties, heat transfer coefficients, etc.

In the balance equations, $C_i(z)$ (mol/m^3) is the gas phase concentration of component i, $q_j(z)$ (kg/m^3) the mass concentration of the solid component j, $T_g(z)$, $T_s(z)$, $T_w(z)$ and T_∞ (K) are the gas, solid, reactor wall and ambient temperatures, $u_g(z)$ and u_s (m/s) are the superficial gas and solid velocities, υ_i is the stoichiometric coefficient of component i, $MW_{s,j}$ (kg/mol) the molecular weight of the solid component j, $\Delta H_{rxn}(z)$ (kJ/mol) the heat of reaction, $h_f(z)$ and $h_w(z)$ (kW/m^2/K) are heat transfer coefficients between the gas and particles and gas and reactor wall, $U(z)$ (kW/m^2/K) is the overall heat transfer coefficient, $C_{p,g}(z)$ and $C_{p,s}(z)$ (kJ/kg/K) are the gas and solid heat capacities, $\rho_g(z)$ and ρ_s (kg/m^3) the gas and particle densities, and $a_{w,in}$ and $a_{w,out}$ (m^{-1}) are internal and external specific heat transfer areas.

For modeling the reaction kinetics, it is assumed that diffusion through the particle grain is the dominating resistance (Abad et al., 2007), resulting in the reaction rate expression given by Eq. (7). It is also assumed that CH_4 reforming is negligible. The kinetic parameters are from Abad et al. (2007), as shown in Table 2.

$$r_{gen} = \frac{x_{Fe_2O_3} a_{vol}}{MW_{s,Fe_2O_3}} \frac{3bkC_{CH_4}^n}{(\rho_m r_g)} (1 - X_s)^{2/3} \tag{7}$$

Eq. (7) is written on Fe_2O_3 basis, where r_{gen} (mol/kg$_{cat}$/s) is the reaction rate, $x_{Fe_2O_3}$ the mass fraction of Fe_2O_3, a_{vol} the available reaction volume per volume of OC, X_s the fractional conversion of Fe_2O_3, b the reaction stoichiometric coefficient of Fe_2O_3, and $k = k_0 \exp\left(-E/(RT_g)\right)$ the kinetic rate constant.

The MB reducer model is implemented in Pyomo, where the differential equations with appropriate boundary conditions are discretized on a non-uniform grid using the shifted Gauss-Radau roots orthogonal collocation method over 20 finite elements with 3 collocation points within each finite element, using the automatic discretization

Table 2. Kinetic model parameters, and MB reactor and particle parameters.

Kinetic model parameters		MB reactor and particle parameters	
Grain radius, r_g (m)	2.6×10^{-7}	Particle diameter, d_p (m)	1.5×10^{-3}
Molar density, ρ_m (mol/m^3)	32.8×10^3	Particle density, ρ_s (kg/m^3)	3251.75
n (-)	1.3	Reactor diameter, D_r (m)	6.5
k_0 (m$^{1.9}$/mol$^{0.3}$/s)	0.0008	Reactor height, H (m)	5
Activation energy, E (kJ/mol)	49	Reactor bed voidage, ε (-)	0.4

capability of pyomo.dae (Hart et al., 2017), and solved using IPOPT (Wächter and Biegler, 2005). An initialization scheme has been developed and implemented that sequentially activates the initially deactivated mass and energy balance equation terms and uses the solution obtained at the end of each step as the initial guess for the solution of the following step. The reaction model and properties of the gas and solid phases are computed in separate property modeling blocks, thus keeping the MB model general and easily applicable to other gas-solid CLC processes. Consistent with the IDAES modeling conventions, the model consists of a collection of modeling components that can be conveniently deactivated or activated as needed. This grants flexibility to the model and enables its straightforward application to the simulation of, for example, an isothermal model by simply deactivating the energy balance equations, or of a cold flow process (i.e., no reaction) by deactivating the terms corresponding to the reaction in both the mass and energy balances, or of an adiabatic process by deactivating the reactor wall energy balance Eq. (6) and the third term in Eq. (4). In the following section, results from the CH_4 CLC process obtained from adiabatic simulations are presented.

3. Results and discussion

To demonstrate the capabilities of the model, a conceptual 100 MW_{th} MB-based CLC reducer is simulated. For the given processing capacity of the fuel $C_{plant} = 100$ MW_{th}, the feed flow rate of CH_4 F_{CH_4}(mol/s) is calculated as $F_{CH_4} = C_{plant}/H_{CH_4}$ (Li et al., 2010), where $H_{CH_4} = 889$ kJ/mol HHV is the higher heating value of CH_4. Then, the inlet flow rate of the gas mixture is calculated to be 245.495 mol/s with the following composition: $y_{CH_4} = 0.458$, $y_{CO_2} = 0.499$, and $y_{H_2O} = 0.043$. The inlet molar fractions, inlet pressure (1.37 bara) and temperatures are referenced from NETL's CLC reference steam generation system (Keairns et al., 2014) and the Ohio State University's CLC technology (Kathe et al., 2016). The reactor dimensions are calculated such that the superficial gas velocity u_g does not exceed the minimum fluidization velocity of the OC particles at any point in the reactor. It is worth noting that the gas velocity can increase up to three-fold in the reducer. The reducer design parameter values and OC particle characteristics used in this study are listed in Table 2. These are not optimized values, as optimization of the CLC system is outside of the scope of this study.

The reducer performance under different gas and solid flow conditions is analyzed by quantifying their effect on the CH_4 and OC conversions. The operating parameters that are varied are the stoichiometric OC/fuel ratio, the gas and the solids inlet temperatures. The values of the different conditions considered are presented in Table 3, together with the steady state conversion of CH_4 and Fe_2O_3. The stoichiometric OC/fuel ratio for this system is 266.15, i.e. 266.15 kg of OC (of which 45% is Fe_2O_3) are needed for the combustion of 1 kg of CH_4, which gives an Fe_2O_3/CH_4 mass flow ratio of $\phi = 119.77$.

Table 3. Case study operating conditions and resulting CH_4 and Fe_2O_3 to Fe_3O_4 conversions.

Case	T_g (°C)	T_s (°C)	$M_{OC,in}$ (kg/s)	ϕ (-)	CH_4 conversion (%)	Fe_2O_3 conversion (%)
1	600	910	479.01	119.77	95.75	95.75
2	600	910	239.51	59.88	49.49	98.99
3	600	910	718.52	179.65	99.79	66.53
4	300	910	479.01	119.77	95.75	95.75
5	600	600	479.01	119.77	79.21	79.21

Case 1 tests the performance of the MB reducer under stoichiometric feed conditions. High conversions are attained: both CH_4 and Fe_2O_3 conversions are 95.75 %. Next, the Fe_2O_3/CH_4 mass flow ratio is decreased to $\phi = 59.88$ by decreasing the inlet OC flow to $M_{OC,in} = 239.54$ kg/s (half of the OC flow for Case 1). Operation at an OC/fuel ratio that is 50 % below the stoichiometric requirement leads to 98.99 % conversion of Fe_2O_3, while leaving 50.51 % uncombusted CH_4. For Case 3, the Fe_2O_3/CH_4 mass flow ratio is increased to ϕ =179.65, i.e. 50 % above the stoichiometric OC/fuel requirement. The excess amount of OC leads to near total (99.79 %) combustion of CH_4 with 33.47 % unconverted Fe_2O_3. In Case 4, the effect of a lower inlet gas temperature is investigated, with all other inlet and operating conditions being identical to the conditions of Case 1. Decreasing the inlet gas temperature has a negligible effect on the reactor performance, even though the decrease in the inlet gas temperature is significant (300 °C). This is because thermal equilibrium in the reducer is reached almost instantaneously, given the extremely rapid heat transfer between the gas and solid phases, and the much larger thermal mass of the solids. In Case 5, the inlet temperature of the solids is decreased by 300 °C, keeping all other conditions identical to the conditions of Case 1. The conversions of CH_4 and Fe_2O_3 drop by 16.54 percentage points compared to Case 1. The sensitivity of the Fe_2O_3 conversion to the investigated system conditions is shown in Figure 1 (a) along the normalized reactor bed length. The CH_4 concentration profiles depicted in Figure 1 (b) are consistent with the trends observed in the Fe_2O_3 conversion profiles. Finally, Figure 1 (c) shows the solid temperature profiles along the reactor bed for each of the studied cases.

The current version of the MB model can be used for the design and optimization of CLC reducers. Model validation, testing, and further development are ongoing, to improve the MB model's accuracy and establish its reliability and robustness under a wide range of applications (e.g. optimization and process flowsheeting).

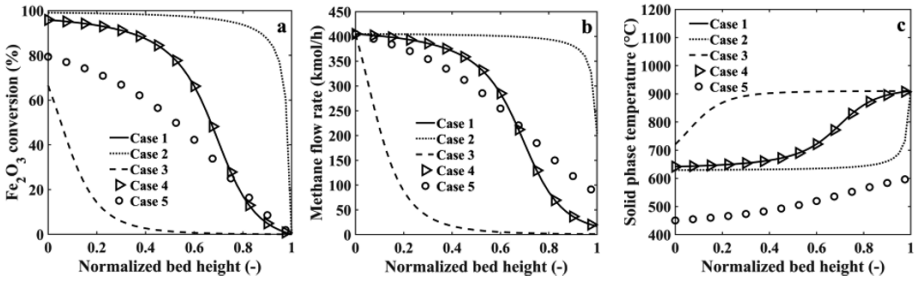


Figure 1. (a) Fe_2O_3 conversion, (b) CH_4 molar flow rate, and (c) solid temperature profiles along the normalized length of the reactor bed (gas inlet: $z = 0$, solids inlet: $z = 1$).

4. Conclusions

The development of a detailed first principles model of a gas-solid counter-current MB reducer has been presented and applied to the simulation of the CLC of methane using an iron-based oxygen carrier. The model is flexible and modular, enabling its straightforward adaptation for the simulation of different gas-solid processes. It is also suited for the design and optimization of CLC processes. The reducer has been designed to simulate a conceptual 100 MW_{th} MB-based CLC system, and the conversions of methane and oxygen carrier have been investigated under different simulation conditions. The conversions have been found to be determined by the oxygen carrier/fuel flow ratio and the solids inlet temperature. The gas inlet temperature has a negligible effect on the reducer performance. The main advantage of operating in the MB regime is the capability of attaining very high methane conversion (95.75 %) at solids circulation rates corresponding to the oxygen carrier/fuel stoichiometric ratio.

Acknowledgements and Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. A. Ostace and D. Bhattacharyya would like to acknowledge the financial assistance provided by the National Energy Technology Laboratory through the Mission Execution and Strategic Analysis Contract DE-FE00025912, as part of the U.S. DOE IDAES initiative.

References

A. Abad, J. Adanez, F. Garcia-Labiano, L. F. de Diego, P. Gayan, and J. Celaya, 2007, Chem Eng Sci, vol. 62, pp. 533-549.

W. E. Hart, C. D. Laird, J. P. Watson, D. L. Woodruff, G. A. Hackebeil, B. L. Nicholson, and J. D. Siirola, 2017, Pyomo -Optimization Modeling in Python, Second Edition, vol. 67, Springer.

M. V. Kathe, A. Empfield, J. Na, E. Blair, and L.-S. Fan, 2016, Appl Energ, vol. 165, pp.183-201.

D. Keairns, R. Newby, V. Shah, M. Turner, M. Woods, A.Zoelle, B. A. Hamilton, S. Ostheim, S. Carpenter, G. Richards, D. Straub, and R. Breault, 2014, Industrial Carbon Management Initiative: Chemical Looping Combustion Techno-Economic Study, U. S. DOE, NETL.

H. Kim, D. C. Miller, S. Modekurti, B. Omell, D. Bhattacharyya, and S. E. Zitney, 2016, AIChE J, vol. 62, no. 11, pp. 3899-3914.

K. Kim, Y. Son, W. B. Lee, and K. S. Lee, 2013, Int J GreenH Gas Con, vol. 17, pp. 13-24.

F. Li, F. Wang, D. Sridhar, H. R. Kim, L. G. Valazques-Vargas, and L.-S. Fan, 2010, Chemical Looping Systems for Fossil Energy Conversions, Ch. 3, John Wiley & Sons, Inc., pp. 143-214.

D. C. Miller, D. Agarwal, D. Bhattacharyya, J. Boverhof, Y. Chen, J. Eslick, J. Leek, J. Ma, P. Mahapatra, B. Ng, N. V. Sahinidis, C. Tong, and S. E. Zitney, 2017, Process Systems and Materials for CO_2 Capture, Ch. 12, John Wiley & Sons, Ltd., pp. 311-342.

D. C. Miller, B. Ng, J. Eslick, C. Tong, and Y. Chen, 2014, Comput-Aided Chem En, Elsevier, 8th International Conference on Foundations of Computer-Aided Process Design – FOCAPD.

G. Mondino, C. Grande, and R. Blom, 2017, Energies, vol. 10, p. 745.

A. Tong, L. Zeng, M. V. Kathe, D. Sridhar, and L.-S. Fan, 2013, Energ Fuel, vol.27, p .4119.

A. Wächter and L. T. Biegler, 2006, Math Program, vol. 160, no. 1, pp. 25-57.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50050-1

Computational fluid dynamics (CFD) modelling and optimum gap size of a compact steam methane reforming (SMR) reactor

Dan Duc Nguyen[a], Son Ich Ngo[a], Young-Il Lim[a,*], Woohyun Kim[b], Dongju Seo[b], and Wang Lai Yoon[b]

[a]*Center of Sustainable Process Engineering (CoSPE), Department of Chemical Engineering, Hankyong National University, Gyonggi-do Anseong-si Jungangno 327, 17579 Korea*
[b]*Hydrogen Laboratory, KIER, 152 Gajeong-ro, Yuseong-gu, Daejeon, 34129 Korea*
limyi@hknu.ac.kr

Abstract

A compact steam methane reforming (SMR) reactor for H_2 production from natural gas was investigated using a steady-state three-dimensional (3D) computational fluid dynamics (CFD) model with heat transfer and chemical reactions. The SMR reactor consisted of two main individual streams including combustion gas for reaction heat supply, and natural gas feed into a porous catalyst bed. The realizable k–epsilon turbulence model with enhanced wall treatment, the discrete ordinates (DO) radiation model, and the volumetric reactions in the species transport model were employed into the CFD model. A reaction kinetics model for SMR of Xu and Froment (1989) was applied to porous media as the catalyst bed. CFD model results for the temperature profile and producer gas composition were validated with experimental data. Optimum sleeve gap sizes between the burner and catalytic reaction zone for various velocities were determined, maximizing the H_2 production rate. The present CFD model showed a potential for the development of new design of the SMR reactor.

Keywords: Hydrogen production; Steam methane reforming (SMR); Computational fluid dynamics (CFD); Chemical reaction rate model; Optimum gap size.

1. Introduction

Hydrogen is one of the most important materials for many petroleum and chemical processes in manufacturing industry. In petroleum refineries, hydrogen is produced for many applications including the hydrotreating and hydrocracking processes. Additionally, in the field of energy, hydrogen is used to run fuel cells which convert hydrogen to electricity (Olivieri & Vegliò, 2008). One of the industrially established methods to effectively produce H_2 is steam methane reforming (SMR) (Chen et al., 2011). In this process, natural gas and steam are converted into hydrogen, carbon dioxide and carbon monoxide in the presence of a nickel-based catalyst under high pressure and high temperature (Olivieri & Vegliò, 2008; Zamaniyan et al., 2010).

The earliest model for SMR reactions was developed in the 1960s. The reaction model was applied to the design, optimization, and monitoring of SMR reactions (McGreavy, 1969). Currently, an SMR reaction kinetic model on nickel-alumina catalyst (Xu & Froment, 1989), which was derived from Langmuir-Hinshelwood mechanism, was often

accepted in academia as well as in industry (Lao et al., 2016; Lee et al., 2013; Seo et al., 2006; Tran et al., 2017).

Computational fluid dynamics (CFD) is a powerful tool for the identification of the intrinsic nature of chemical processes. Specifically, many CFD researches have realized successfully physical and chemical phenomena inside SMR tube reactors in microscopic-scale, bench-scale as well as industrial-scale (Dixon, 2014; Seo et al., 2006; Tran et al., 2017). However, little attention is paid to an annulus SMR reactor with a narrow sleeve.

In this study, a CFD model is developed for a compact SMR reactor involving two main zones: combustion and catalyst zones. The CFD model includes the turbulence equation, radiative heat transfer model, and chemical reaction kinetics for combustion and SMR reactions. The CFD results are validated by comparing with available data in our experiment and literatures. Finally, an optimum gap size of the sleeve showing the highest thermal efficiency is found.

2. SMR reactor

2.1. Geometry of SMR reactor

The geometry of the SMR reactor is illustrated in Fig. 1a The SMR reactor is divided into two main domains: the fluid and solid domains.

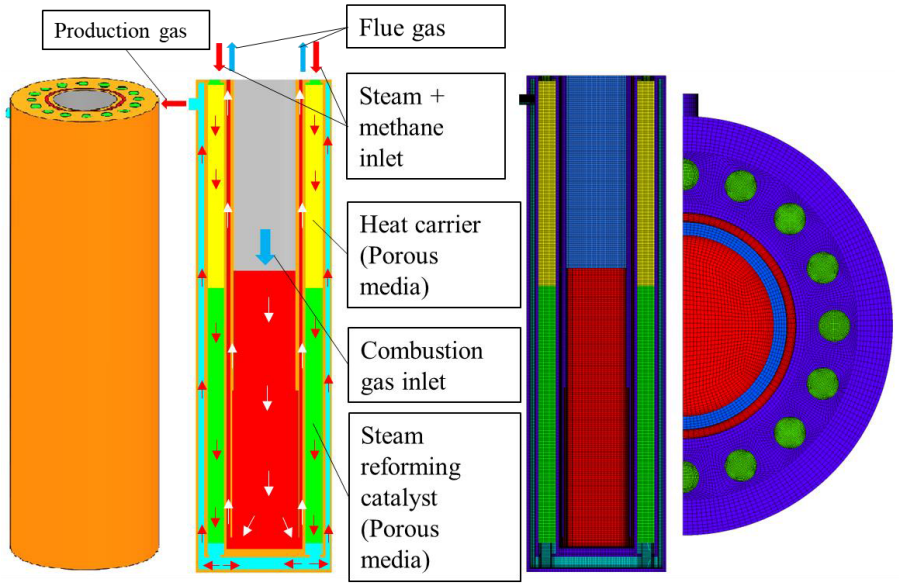


Fig. 1. Schematic diagram of SMR reactor and hexahedral mesh structure.

2.2. Meshing of CFD domain

The 3D symmetric CFD domain is shown in Fig. 1b, which was constructed in a meshing software, ICEM-CFD. The number of hexahedral cells is approximately 2.6 million with a minimum orthogonal quality of 0.52 and a maximum aspect ratio of 29.7.

3. CFD model with chemical reactions

3.1. CH_4 combustion model

The SMR reaction is a highly endothermic process. The heat required for the reforming reaction is supplied by the combustion gas via a gas burner. The air–methane mixture with an oxygen molar ratio to methane of 2.58 is injected into the combustion chamber and is completely oxidized to generate carbon dioxide, water and a large amount of thermal energy required for the SMR reaction. A simplified global kinetic model for the combustion of methane is adopted to reduce the computational time. A single-step combustion mechanism is given as follows:

$$CH_4 + 2O_2 \Rightarrow CO_2 + 2H_2O \quad (1)$$

The premixed combustion is resolved by using the eddy-dissipation (ED) model as the turbulence and chemistry interaction in the volumetric reaction.

3.2. SMR reaction model

The SMR reaction on nickel–based catalyst is expressed by the following reversible equations:

$$\begin{aligned} &CH_4 + H_2O \leftrightarrow CO + 3H_2 && \Delta H_{\circ 298} = 206\ \text{kJmol}^{-1} \\ &CO + H_2O \leftrightarrow CO_2 + H_2 && \Delta H_{\circ 298} = -41.1\ \text{kJmol}^{-1} \\ &CH_4 + 2H_2O \leftrightarrow CO_2 + 4H_2 && \Delta H_{\circ 298} = 164.9\ \text{kJmol}^{-1} \end{aligned} \quad (2)$$

The SMR reaction kinetics model is integrated into the CFD model by using user–defined functions.

3.3. Performance criteria of SMR reactor

The sleeve gap sizes of 2, 3, 4, 5, and 7 mm were tested at three velocities of 6, 7.5, and 9 m/s. The thermal heat efficiency (η) based on lower heating value (*LHV*) was used to find an optimal gap size.

$$\eta = \frac{F_{product,H2} \times LHV_{H2}}{(F_{CH4,burner} + F_{CH4,feed}) \times LHV_{CH4}} \times 100 \quad (3)$$

where F (Nm^3/h) is the volumetric flow rate.

4. Results and discussions

The CFD model with a gap size of 3 mm was solved at 7.5 m/s. The temperature and concentration profiles obtained from the CFD model were compared to experimental data and values published in literatures. The sleeve gap size maximizing the thermal efficiency was found at each feed velocity.

4.1. Validation of CFD model

Fig. 2 shows the temperature along the catalytic reactor length. The temperature profile from the CFD model with combustion reaction is higher than that without combustion reaction. The CFD results of temperature show a similar tendency to the experimental ones.

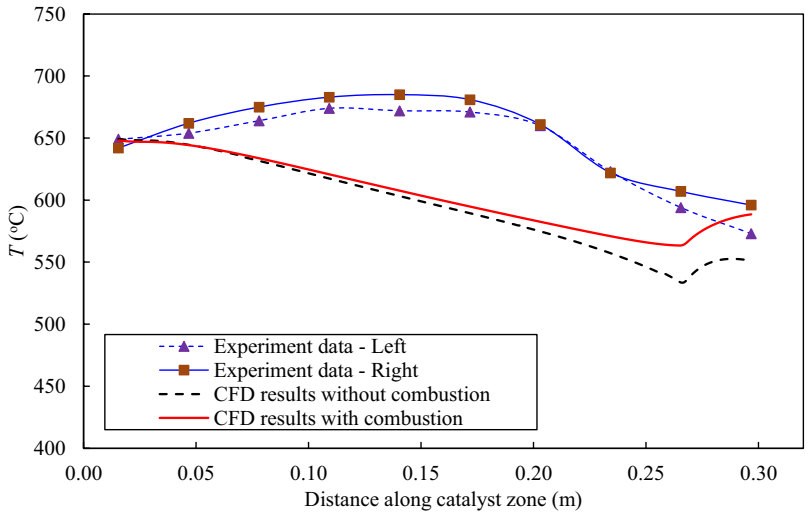


Fig. 2. Comparison of temperatures obtained from experimental data and CFD results with and without combustion reaction.

Fig. 3 depicts mole fractions of the producer gas along the normalized catalyst length. Two literature data are compared to the CFD results. The CFD results have a good agreement with the literature data. Table 1 reports the mole composition, H_2 production rate and *LHV* heat efficiency for this CFD study, experimental data, and the two literature data. The CFD results are in good agreement with the others.

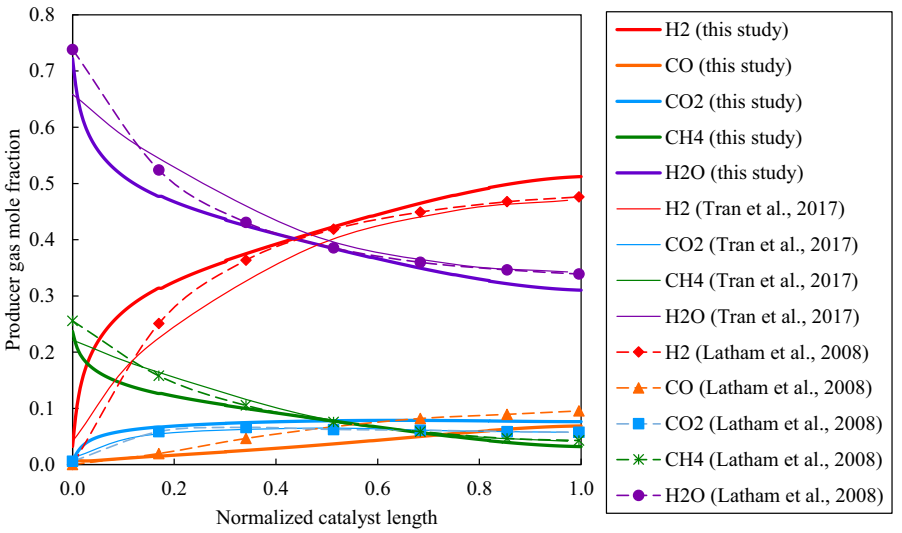


Fig. 3. Comparison of product gas composition along catalyst length with literature

Table 1. CFD model validation on producer gas composition

Parameter	CFD result w/o combustion	CFD result with combustion	Exp. Data 1	Exp. Data 2	(Xu & Froment, 1989)	(Latham, 2008)
H_2	0.7318	0.7266	0.727	0.7395	0.7206	0.71
CO	0.0902	0.0308	0.0665	0.08	0.1089	0.13
CH_4	0.0627	0.0841	0.0845	0.0615	0.0704	0.07
CO_2	0.1153	0.1585	0.122	0.119	0.1001	0.087
H_2 production rate (Nm^3/h)	2.64	2.57	2.73	2.56	2.51	2×10^5
LHV efficiency (%)	49.33	48.05	48	50.1	46.8	

4.2. Optimum sleeve gap size

Fig. 4 shows the thermal efficiency according to the sleeve gap size. The maximum heat efficiency is observed at each gas velocity inside the sleeve gap. As the velocity increases from 6 to 9 m/s, the optimum gap size decreases from 5 to 3 mm.

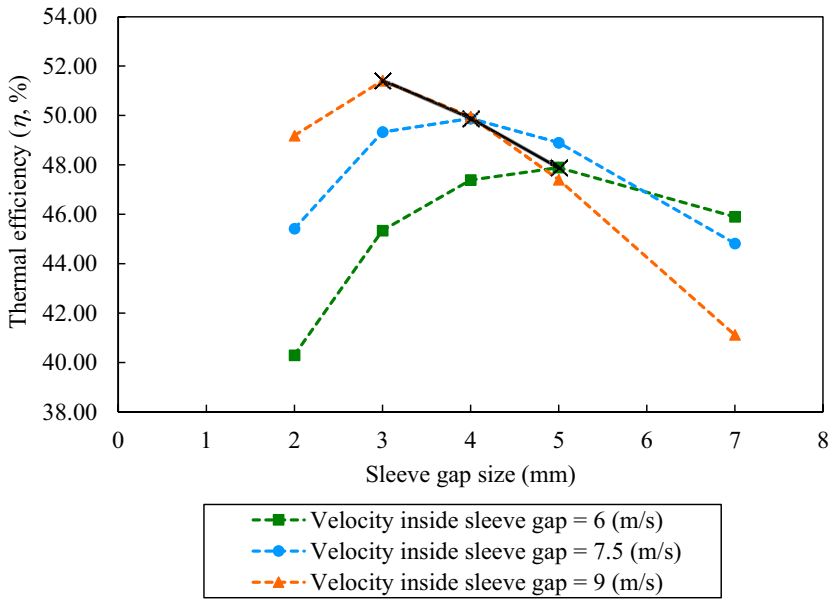


Fig. 4. Thermal efficiency with respect to sleeve gap size for each gas velocity inside sleeve gap.

5. Conclusions

A three-dimensional (3D) computational fluid dynamics (CFD) model was constructed for a steam methane reforming (SMR) reactor with catalyst in the reaction zone and porous media in the heat carrier zone. The CFD simulation including heat transfer, radiation and reaction models was performed at 2 bar and about 700 ºC. The coarse, medium and fine meshes were built for the mesh independent test and the medium mesh was selected for further study.

The CFD results agreed well with experimental data in terms of temperature and gas compositions. The optimum sleeve gap size was found at each velocity of the sleeve gap, maximizing the thermal efficiency.

Acknowledgements

This work was conducted under the framework of the research and development program of the Korea Institute of Energy Research (B7-2424).

References

Chen, D., Lødeng, R., Svendsen, H., & Holmen, A., 2011. Hierarchical Multiscale Modeling of Methane Steam Reforming Reactions. Industrial & Engineering Chemistry Research, (50), 2600-2612.

Dixon, A. G., 2014. CFD study of effect of inclination angle on transport and reaction in hollow cylinder catalysts. Chemical Engineering Research and Design, (92), 1279-1295.

Lao, L., Aguirre, A., Tran, A., Wu, Z., Durand, H., & Christofides, P. D., 2016. CFD modeling and control of a steam methane reforming reactor. Chemical Engineering Science, (148), 78-92.

Latham, D., 2008. Mathematical Modeling of an Industrial Steam Methane Reformer (Masters Thesis). Queen's University.

Lee, J. S., Seo, J., Kim, H. Y., Chung, J. T., & Yoon, S. S., 2013. Effects of combustion parameters on reforming performance of a steam–methane reformer. Fuel, (111), 461-471.

McGreavy, C. N., M. W., 1969. Development of a mathematical model of a steam methane reformer. In: Proceedings of the Conference on the Industrial Applications of Dynamic Modelling), 16–18.

Olivieri, A., & Vegliò, F., 2008. Process simulation of natural gas steam reforming: Fuel distribution optimisation in the furnace. Fuel Processing Technology, (89), 622-632.

Seo, Y.-S., Seo, D.-J., Seo, Y.-T., & Yoon, W.-L., 2006. Investigation of the characteristics of a compact steam reformer integrated with a water-gas shift reactor. Journal of Power Sources, (161), 1208-1216.

Tran, A., Aguirre, A., Durand, H., Crose, M., & Christofides, P. D., 2017. CFD modeling of a industrial-scale steam methane reforming furnace. Chemical Engineering Science, (171), 576-598.

Xu, J., & Froment, G. F., 1989. Methane steam reforming, methanation and water-gas shift: I. Intrinsic kinetics. AIChE Journal, (35), 88-96.

Zamaniyan, A., Behroozsarand, A., & Ebrahimi, H., 2010. Modeling and simulation of large scale hydrogen production. Journal of Natural Gas Science and Engineering, (2), 293-301.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50051-3

Integrated Bio-refinery Utilizing Brown Macroalgae: Process Design, Simulation and Techno-economical Assessment

Boris Brigljević[a], Peyman Fasahati[b], J. Jay Liu[a*]

[a]*Department of Chemical Engineering, Pukyong National University, 365 Sinseon-ro, 48547, Busan, Korea*
[b]*Department of Chemical and Biological Engineering, University of Wisconsin–Madison, WI 53706-1691, Madison, USA*
jayliu@pknu.ac.kr

Abstract

Biofuel production using carbohydrate rich brown seaweed is being intensely researched in the last few years as a viable replacement for 1st and 2nd generation biomass. Process design and simulation is generally used when evaluating technological and market capabilities of large scale, novel, fuel production processes. This study, for the first time, presents process design solution for a biorefinery, an integrated anaerobic digestion, and PHP (Pyrolysis, Heat and Power) process. The process utilizes 3rd generation biomass (marine macroalgae) and yields wide range of products: mixed alcohols, ethanol, biodiesel equivalent, biogas, and electricity. Technoeconomical assessment of the biorefinery is performed and compared to single conversion pathways (biochemical and thermochemical). It is highly expected that the synergistic effect of this integrated process should yield a more economically and environmentally viable biofuel production, compared to stand-alone pathways.

Keywords: Biofuel, Seaweed, Process simulation, Process integration

1. Introduction

The strain between energy supply and demand, is increasing, as the finite nature of fossil fuels is becoming ever more evident. Consequently, innovation and research in every part of the energy sector is increasing rapidly. Biofuels present a green alternative to crude oil derivatives, due to their inherit renewable nature and possibility of carbon neutrality. Since the biofuel boom of 2005[1] advocates against biofuels argue that the global usage of food crops for biofuel production will inevitably affect global food prices. Hence, research has shifted toward food crop residues such as corn stover[2], rice husk[3], seed shell[4], etc., as well as woody biomass and forest residues[5]. Biofuels derived from those materials are defined as second generation biofuels[6]. With valid management and supply chains second generation biofuels do present advantage over food crops. However, indisputably, second generation biofuels are still terrestrially derived, which means they do require land in some point of their lifecycle. Utilization of aquatic biomass, or third generation biofuels[7,8] present a number of advantages over terrestrial biomass. First and most obvious advantage is that marine biomass (microalgae and macroalgae) does not require land for cultivation. Secondly, they have inherent faster growth rates and better photosynthetic efficiency[9]. Finally, their physiology does not require lignin nor cellulose as support materials[10–12], which both present challenges in biofuel

production from woody biomass. On the other hand inherently high moisture, mineral content and its seasonal variation[13,14] presents new set of challenges for potential biofuel production processes. This means that well established conversion processes for biofuel production need to be modified and evaluated if this relatively new type of feedstock is to be effectively utilized. Hence, algae cultivation studies[15], as well as, standalone conversion processes which utilize marine macroalgae as feedstocks have been evaluated. Biochemical pathways such as anaerobic digestion[16], bioethanol and mixed alcohol processes[17,18] have been investigated. Thermochemical pathways such as gasification[19], pyrolysis[20] and liquefaction[21] have also been investigated on industrial scale. This work however, presents a conceptual study of an industrial integrated process combining biochemical and thermochemical pathway and which utilizes brown macroalgae as a feedstock. General aim of the study is to define beneficial economic and possibly environmental implications of combining two pathways into a seaweed based biorefinery.

2. Materials and methods

2.1. Biochemical Conversion Pathway

The biochemical conversion process simulation model, design and economics are based on published work by Fasahati et al.[22]. Main products of the process were ethanol and mixed alcohols, produced via hydrogenation of volatile fatty acids (VFAs). The VFAs were obtained by inhibitor controlled anaerobic digestion of brown seaweed *Saccharina japonica* as a feedstock. The process was designed and simulated in Aspen Plus on an industrial scale of 500,000 tonnes of dried seaweed (18 % wt. moisture) per year. Major product process streams and pricing is outlined in table 2. Combined mixed alcohol yield of the process compared to bone dry seaweed weight was 29 % wt., of which 75 % wt. was ethanol after separation. The process is highly water intensive and consumes 532 T/h. Solid residue of anaerobic digestion or digestate is produced at 41.5 T/h (35% wt. moisture) and at this process design is marketed as crop fertilizer. The total installed cost of the process was determined to be 85.9 mil USD, and the calculated minimum product selling prices (MPSPs) of ethanol and mixed alcohols were 0.589 USD/L, and 1.143 USD/L respectively.

2.2. Thermochemical Conversion Pathway

The thermochemical conversion process of algae to diesel equivalent fuel (Figure 1) was simulated in Aspen Plus at an industrial scale of 320,000 dry tonnes per year. Major product process streams and pricing is outlined in table 2 Experimental data for the simulation model was based on published works by Choi et al. [23,24] for fast pyrolysis of brown algae *Saccharina japonica* in a fluidized bed reactor. As received feed with water content of 35 wt. % and mineral content of 26 wt. % is pressed and dried to a water content of 5 wt. %. The seaweed is pyrolyzed in a fluidized bed reactor, after which pyrolysis vapors are separated from char in a cyclone and sent to phase separation section. Here, pyrolysis vapors are liquified, and organic pyrolysis liquid (biocrude) is separated from aqueous phase and non-condensable gasses. Biocrude is upgraded to a diesel equivalent fuel via catalytic hydrogenation. Required hydrogen is partly purchased (45 wt. %) while the rest can be obtained by recycling the gasses from hydrogenation section via pressure swing absorption (PSA). The yield of diesel upgraded fuel compared to dry seaweed (absent mineral content) was 14.5 wt. %. As a thermochemical process it is highly heat intensive and consumes 9.2 T/h of natural gas and produces 29.1 T/h of

aqueous phase. Net produced power in the heat and power section was 1.5 MW. Calculated total installed cost of the process was 56.9 mil USD and a MPSP of 1.257 USD/L was determined.

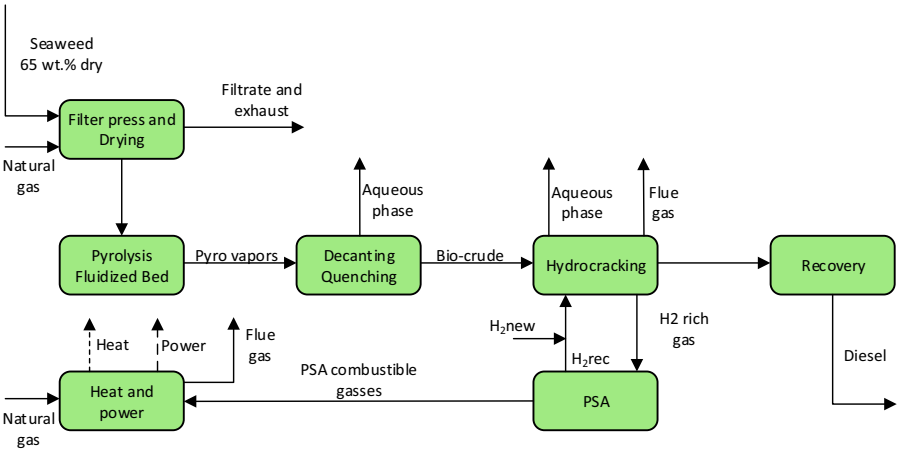


Figure 1. Block flow diagram of the fast pyrolysis route

2.3. Combined Process Key Integration Points

Integration design concept of biochemical and thermochemical pathways is illustrated in figure 2. The two pathways are integrated in three key points. Primarily, anaerobic digestate will now be utilized as a feedstock for fast pyrolysis process. This would accomplish two objectives, production of higher value-added product, as well as reducing the mineral content in the pyrolysis feedstock due to water washing during the fermentation. Secondly, both pathways require the same type of product upgrading, as well as hydrogen recycling, which presents the potential of lowering the total capital investment of combined process. Thirdly, utilization of produced aqueous phases in the pyrolysis section would reduce the water consumption of the biochemical conversion. Finally, the Heat and power production section of the pyrolysis section can provide most of the heat and power required for the combined process.

Table 1. Economics summary of standalone pathways and integrated process

USD / 2016	Thermal-pathway	Bio-pathway	Integrated
Inside battery limit (ISBL)	53,300,000	78,330,000	119,279,000
Total installed cost	56,880,000	85,926,000	122,858,000
Total direct costs	66,207,000	99,634,000	143,732,000
Total indirect costs	39,724,000	59,780,000	86,239,000
Fixed capital investment	105,932,000	159,414,000	229,972,000
Total capital investment	114,426,000	172,084,503	243,092,000

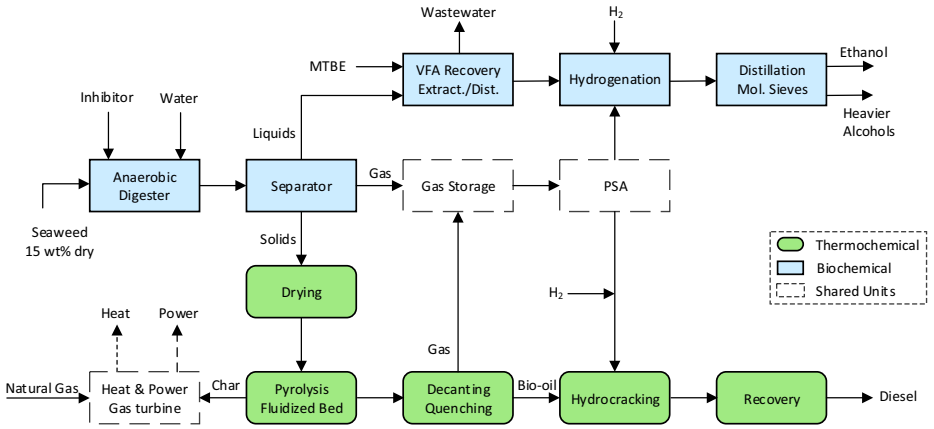


Figure 2. Block flow diagram of integrated process

Table 2. Summary of major streams and pricing of bio and thermal pathway

Bio	Flow kg/h	Price USD/kg	Thermal	Flow kg/h	Price USD/kg
Raw materials					
Seaweed (dry weight)	50800	0.07	Digestate	27050	0.0
Iodoform	10.2	5.68	H2	250	1.8
MTBE	420	1.18	Nat. Gas	9227	0.2
Catalyst	0.12	55.82	Water	351000	2.00E-04
H2	680	1.80			
Water	532400	2.00E-04			
Cooling chems	1.7	0.36			
Waste products					
Effluent	403200		flue gas	25927	
Marketable products					
Digestate	27050	/	Diesel eqv.	3600	MPSP
Mixed Alcohols	6325	MPSP	Power	1.5 MW	
Ethanol	8420	MPSP			

3. Results and discussion

3.1. Technoeconomical Assessment of Integrated Process vs. Standalone Pathways

Economics summary of standalone pathways and integrated process are outlined in table 1. As it can be observed, on considered industrial scale, integrated process shows a reduction in capital investment by 15.2 % compared to combined two standalone pathways of the same scale. This reduction is accredited to shared equipment of hydrogen

recycling and upgrading sections, as well as single heat and power production section (Figure 3). Total utilities savings of integrated process was calculated at 6.33 mil USD per year of which the majority is due to the hydrogen savings (85 %), heat and power savings (14 %) and water usage savings (1%). Even though in sheer mass consumption per year water was significantly reduced compared to hydrogen, greater price of the latter affected a bigger ration in utilities savings.

3.2. Minimum Product Selling Price

Due to reduction in capital investment cost and utilities savings MPSPs of main products ethanol, mixed alcohols and diesel equivalent fuel were reduced by 5.6 % compared to product prices of standalone pathways, assuming equal percentage reduction of all product prices. If MPSP of diesel equivalent fuel remained equal to that of standalone pyrolysis process in the integrated process, the MPSPs of alcohols can be reduced by 7.5 %. Likewise, if MPSPs of alcohols remained fixed, the MPSP of diesel equivalent fuel can be reduced by as much as 21.9 % in the integrated process.

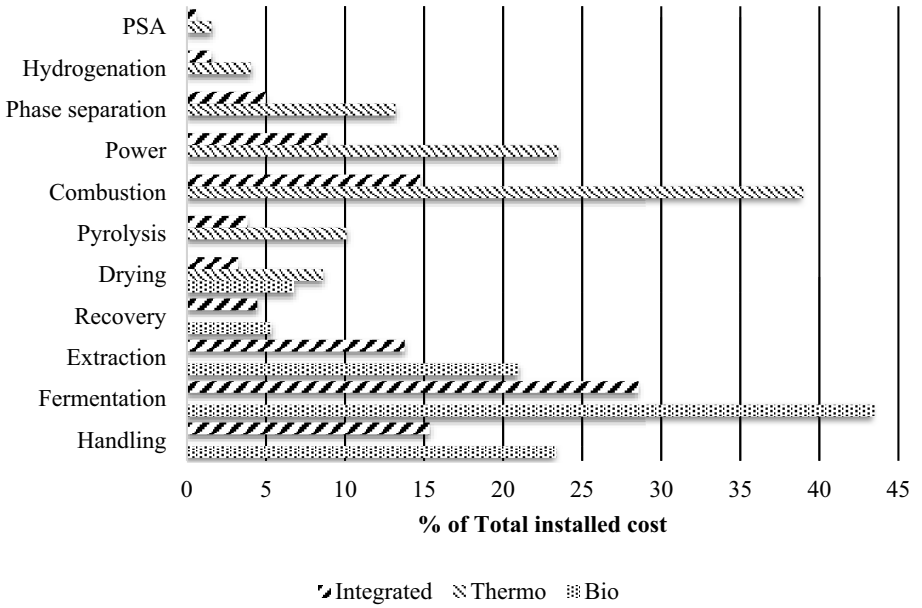


Figure 3. Capital cost breakdowns by process sections of standalone pathways and integrated process

4. Conclusions

In conclusion it was shown that seaweed based biorefinery consisting of mixed alcohol production and diesel fuel via biochemical and thermochemical pathway respectively has performed more favorably economically compared to standalone pathways of the same scale. This was shown through reduced total capital investment and annual utilities savings. More favorable economics also affected the reduction of minimum product selling prices of 5.6 % for all products equally or 7.5 % to 21.9 % when individual products are viewed.

References

[1] M. Giampietro, K. Mayumi, 2009, The biofuel delusion: The fallacy of large scale agro-biofuels production, Routledge, .

[2] N. Murali, S. Fernandez, B.K. Ahring, 2017, Fermentation of wet-exploded corn stover for the production of volatile fatty acids, Bioresource Technology, 227, 197–204.

[3] M. Ebrahimi, O.B. Villaflores, E.E. Ordono, A.R. Caparanga, 2017, Effects of acidified aqueous glycerol and glycerol carbonate pretreatment of rice husk on the enzymatic digestibility, structural characteristics, and bioethanol production, Bioresource Technology, 228, 264–271.

[4] N. Pragya, N. Sharma, B. Gowda, 2017, Biofuel from oil-rich tree seeds: Net energy ratio, emissions saving and other environmental impacts associated with agroforestry practices in Hassan district of Karnataka, India, Journal of Cleaner Production, 164, 905–917.

[5] X. Zhang, A.O. Oyedun, A. Kumar, D. Oestreich, U. Arnold, J. Sauer, 2016, An optimized process design for oxymethylene ether production from woody-biomass-derived syngas, Biomass and Bioenergy, 90, 7–14.

[6] Q. Nguyen, J. Bowyer, J. Howe, S. Bratkovich, H. Groot, E. Pepke, K. Fernholz, 2017, Global Prduction of Second Generation Biofuels: Trends and Influences, Dovetail Partners Inc,.

[7] S.A. Jambo, R. Abdulla, S.H. Mohd Azhar, H. Marbawi, J.A. Gansau, P. Ravindra, 2016, A review on third generation bioethanol feedstock, Renewable and Sustainable Energy Reviews, 65, 756–769.

[8] M. Ghadiryanfar, K.A. Rosentrater, A. Keyhani, M. Omid, 2016, A review of macroalgae production, with potential applications in biofuels and bioenergy, Renewable and Sustainable Energy Reviews, 54, 473–481.

[9] K. Gao, K.R. McKinley, 1994, Use of macroalgae for marine biomass production and CO2 remediation: a review, Journal of Applied Phycology, 6, 45–60.

[10] G. Roesijadi, S.B. Jones, L.J. Snowden-Swan, Y. Zhu, 2010, Macroalgae as a biomass feedstock: a preliminary analysis, PNNL 19944, Richland, .

[11] J. Milledge, B. Smith, P. Dyer, P. Harvey, 2014, Macroalgae-Derived Biofuel: A Review of Methods of Energy Extraction from Seaweed Biomass, Energies, 7, 7194–7222.

[12] A. Ross, J. Jones, M. Kubacki, T. Bridgeman, 2008, Classification of macroalgae as fuel and its thermochemical behaviour, Bioresource Technology, 99, 6494–6504.

[13] J.M.M. Adams, A.B. Ross, K. Anastasakis, E.M. Hodgson, J.A. Gallagher, J.M. Jones, I.S. Donnison, 2011, Seasonal variation in the chemical composition of the bioenergy feedstock Laminaria digitata for thermochemical conversion, Bioresource Technology, 102, 226–234.

[14] J.S. Rowbotham, P.W. Dyer, H.C. Greenwell, M.K. Theodorou, 2012, Thermochemical processing of macroalgae: a late bloomer in the development of third-generation biofuels?, Biofuels, 3, 441–461.

[15] R. Davis, J. Markham, C. Kinchin, N. Grundl, E.C.D. Tan, D. Humbird, 2016, Process Design and Economics for the Production of Algal Biomass: Algal Biomass Production in Open Pond Systems and Processing Through Dewatering for Downstream Conversion, National Renewable Energy Laboratory, 128.

[16] P. Fasahati, C.M. Saffron, H.C. Woo, J.J. Liu, 2017, Potential of brown algae for sustainable electricity production through anaerobic digestion, Energy Conversion and Management, 135, 297–307.

[17] P. Fasahati, J.J. Liu, 2016, Application of MixAlco® processes for mixed alcohol production from brown algae: Economic, energy, and carbon footprint assessments, Fuel Processing Technology, 144, 262–273.

[18] P. Fasahati, H.C. Woo, J.J. Liu, 2015, Industrial-scale bioethanol production from brown algae: Effects of pretreatment processes on plant economics, Applied Energy, 139, 175–187.

[19] C.O. Okoli, T.A. Adams, B. Brigljević, J.J. Liu, 2016, Design and economic analysis of a macroalgae-to-butanol process via a thermochemical route, Energy Conversion and Management, 123,.

[20] B. Brigljevic, J. Liu, 2017, Process design and simulation of industrial scale biofuel production via pyrolysis of saccharina japonica, in: 2017 6th Int. Symp. Adv. Control Ind. Process., : pp. 493–498.

[21] A. Giaconia, G. Caputo, A. Ienna, D. Mazzei, B. Schiavo, O. Scialdone, A. Galia, 2017, Biorefinery process for hydrothermal liquefaction of microalgae powered by a concentrating solar plant: A conceptual study, Applied Energy, 208, 1139–1149.

[22] P. Fasahati, J.J. Liu, 2015, Impact of volatile fatty acid recovery on economics of ethanol production from brown algae via mixed alcohol synthesis, Chemical Engineering Research and Design, 98, 107–122.

[23] H.V. Ly, S.S. Kim, H.C. Woo, J.H. Choi, D.J. Suh, J. Kim, 2015, Fast pyrolysis of macroalga Saccharina japonica in a bubbling fluidized-bed reactor for bio-oil production, Energy, 93, 1436–1446.

[24] J.H. Choi, S.S. Kim, H.V. Ly, J. Kim, H.C. Woo, 2017, Effects of water-washing Saccharina japonica on fast pyrolysis in a bubbling fluidized-bed reactor, Biomass and Bioenergy, 98, 112–123.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50052-5

Process modeling for steam biomass gasification in a dual fluidized bed gasifier

Quang-Vu Bach, Hye-Ri Gye and Chul-Jin Lee*

School of Chemical Engineering and Materials Science, Chung-Ang University, Seoul 06980, Republic of Korea
cjlee@cau.ac.kr

Abstract

Steam biomass gasification has been receiving a lot of attentions as a green process to produce hydrogen rich fuels. In this process, biomass is converted into product gas, which consists of mainly H_2, CO, CH_4, and other combustible gases, at elevated temperatures and under presence of air and steam as oxidizing agents. The product gas can be used directly as a gaseous fuel or upgraded to syngas (i.e., H_2 + CO) for production of other liquid biofuels and chemicals. Fluidized bed reactor has been preferable for biomass gasification due to excellence gas-solid contact as well as heat and mass transfer. Conventional fluidized bed gasifier involves air and steam in a single reactor, which normally yields product gas consisting of nitrogen at a high concentration. Consequently, heating value of the product gas is dramatically reduces. In contract, a dual fluidized bed concept can produce nitrogen-free product gas, which is advantageous for process up-scaling.

In this study, a biomass gasification process with a dual fluidized bed technology is constructed in Aspen Plus software. Model validation is carried out using available experimental data. The model can provide in details the performance of the gasification process in terms of product gas distribution and heating value as well as the process efficiency.

Keywords: Biomass gasification; Dual fluidized bed gasifier, Process modeling.

1. Introduction

Biomass gasification has been receiving a lot of attentions as a green and clean process to produce hydrogen rich fuels with minimal environmental impacts (Farzad et al., 2016). In this process, biomass is converted into product gas, which consists of mainly H_2, CO, CH_4, and other combustible gases, at high temperatures and under a controlled oxidative atmosphere. The product gas can be used directly as a gaseous fuel (Cuéllar-Franca & Azapagic, 2015; Streimikiene, 2010) or upgraded to syngas for production of other liquid biofuels and chemicals (Alghurabie et al., 2013; Takeshita, 2011).

Among the gasifiers, fluidized bed reactor has been preferable for biomass gasification due to excellence gas-solid contact as well as heat and mass transfer. Conventional fluidized bed gasifier involves air and steam in a single reactor, which normally yields product gas consisting of nitrogen at a high concentration. Consequently, heating value of the product gas is dramatically reduces. A dual fluidized bed (DFB) reactor (Kern et al., 2013) can overcome this problem of steam biomass gasification by employing two separate fluidized bed reactors: a combustor and a gasifier. In the combustor, part of char is burnt with air to heat bed material (e.g., sand). The hot bed material is then

transferred to the gasifier, where steam is introduced as gasifying agent, for char gasification.

Although some studies have been conducted experimentally to understand the effect of the process parameters on the performance of DFB gasifier (Corella et al., 2007; Kern et al., 2013), current knowledge on the DFB technology is still limited due to limited experimental points. Process simulation can help increase the number of investigated points and provide more information about the technology.

In this study, a biomass gasification process with a dual fluidized bed technology is constructed in Aspen Plus software. Model validation is carried out using available experimental data. The model can provide in details the performance of the gasification process in terms of product gas distribution and heating value as well as the process efficiency.

2. Process simulation

2.1. Model description

The constructed gasification model in Aspen Plus is presented in Figure 1. The stream BIOMASS containing dried biomass is fed to the reactor PYRO, which simulates the decomposition of biomass at moderate temperature (i.e., pyrolysis) and converts the non-conventional component into conventional components. The block CHARSEP is used to split a portion of char, which is burnt at the COMBUST reactor to obtain the heat Q-COMB. The splitting ratio is set to a value that the heat from the char combustion can balance the heat required for the char gasification and decomposition. Other components after splitting char are fed to the NS reactor to convert nitrogen and sulfur in the biomass into NH_3 and H_2S, respectively. Thereafter, the two gases are separated from the main stream by the block NSSEP and mixed to the product gas later. Main gasification reactions occur at the GASIFIER block, where steam is introduced at 150 °C. After gasification, the stream OUTGASIF goes through the block H2OSEP to remove water from the product gas. Then the dry gas (stream DRY-GAS) is mixed with the stream NS-GAS2 containing NH_3 and H_2S in the MIX-GAS before cooled to room temperature by the HX-CG. The heater HX-NS is used to increase the temperature of the stream NS-GAS1 equal to that of the stream DRY-GAS.

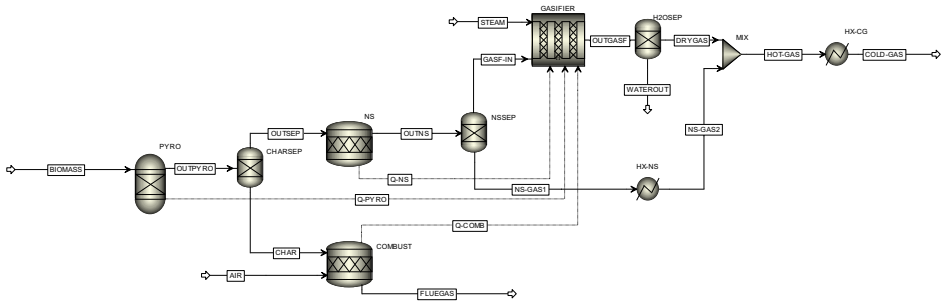


Figure 1. Biomass gasification model in Aspen Plus.

2.2. Assumptions

In order to simplify the process simulation, some common assumptions are made:

- The stream class used in the model is MIXCISLD, in which biomass are considered as non-conventional components.
- Char is 100 % carbon, and ash is an inert component.
- The properties method is Peng-Robinson with Boston-Mathias function (PR-BM), which is suitable for high temperature like gasification.
- All calculations are in steady-state.
- The system operates at atmospheric pressure and all pressure drops are neglected.
- Tar formation is neglected.
- Air consists of 79 % nitrogen and 21 % oxygen on molar basis.

2.3. Materials

In this study, Norway spruce from (Tapasvi et al., 2012) are chosen as feedstock for the gasification simulation. On the other hand hardwood chips, blended pellets and sewage pellets from an experimental DFB gasification study (Schmid et al., 2012) are chosen for model validation. The fuel properties of these materials are list in Table 1.

Table 1. Fuel proprieties of various biomass materials (on dry basis).

Feedstock	Proximate analysis			Ultimate analysis				
	Ash	VM	FC	C	H	N	S	O
Norway Spruce	0.23	86.34	13.43	49.94	6.34	0.07	0.05	43.38
Hardwood chips	1.0	84.0	15.0	48.8	5.9	0.15	0.05	44.1
Blended pellets	1.6	82.4	16.0	49.9	5.7	0.25	0.05	42.5
Sewage sludge pellets	41.5	55.4	3.1	29.7	3.7	3.90	1.00	20.2

3. Results and discussion

3.1. Model validation

Experimental data from a DFB gasification study (Schmid et al., 2012) was adopted for validation of the model in this work. The extracted data from the simulation were compared with the experimental values in the previous work. The product gas compositions from the experiments and simulations are presented in Figure 2, which shows good agreements between the modeled and experimental data. The validation indicates that the constructed model is appropriate for a variation of feedstock and could be employed for deeper investigations of different biomass gasification.

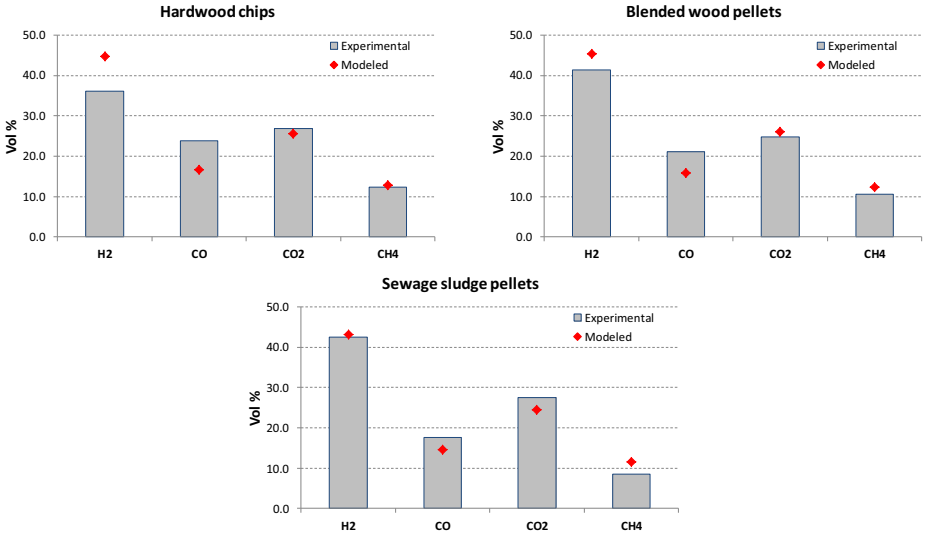


Figure 2. Validation of product gas composition from different feedstocks.

3.2. Effects of gasification temperature

Effects of temperature on the spruce gasification are presented in Figure 3. Figure 3A shows the composition of product gas at different temperature while Figure 3B demonstrates the LHV (lower heating value) of cold gas and the CGE (cold gas efficiency) with varying the temperature. It can be from Figure 3A that both the H_2 and CH_4 contents decrease slightly while CO content increases when temperature rises from 700 to 900 °C. On the other hand, CO_2 content shows a decrease with increasing temperature. In addition, the contributions of NH_3 and H_2S in the product gas are insignificant due to low nitrogen and sulfur contents in the biomass. Figure 3B reveals that the LHV of cold gas increases but the CGE decreases with increasing the gasification temperature. The increase of cold LHV with temperature is due to the reduction of CO_2 and the increase of CO. In addition, the decrease of CGE with temperature can be explained by that more char is split in order to balance for higher heat demand at the GASIFIER, thus less char is gasified and less gas yield is obtained at higher temperature.

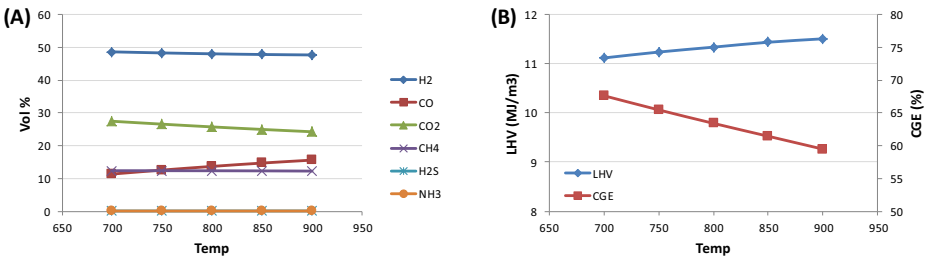


Figure 3. Effects of temperature in spruce gasification.

3.3. Effect of steam-to-biomass ratio

The changes in product gas composition with varying SBR (steam-to-biomass ratio) are presented in Figure 4A, which shows increasing trends for both the H_2 and CO_2 contents and decreasing trend for both the CO and CH_4 contents while increasing the SBR. In details, the contents of H_2 and CO_2 are increased respectively from 45.6 and 24.7 vol% to 50.2 and 26.5 vol%; whereas the contents of CO and CH_4 are decreased respectively from 15.8 and 13.7 vol% to 12.1 and 11.2 vol%. Due to the change in the composition, the LHV of cold gas decreases (from 11.8 to 10.9 MJ/m^3) but the CGE increases (from 62.2 to 64.2 %) with increasing the SBR, as shown in Figure 4B.

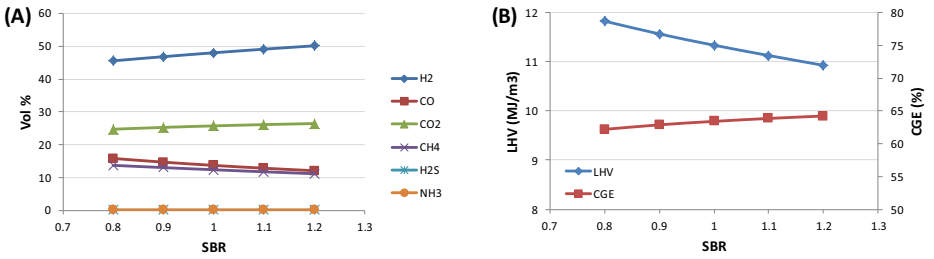


Figure 4. Effects of SBR in spruce gasification.

4. Conclusions

A model for biomass steam gasification with the DFB technology has been successfully constructed in the commercial Aspen Plus simulator, and validated using available experimental data. Simulation results from the model show that the H_2 and CO_2 contents in the product gas decreases with increasing temperature and decreasing SRB, while the CO and CH_4 contents show opposite trends. In addition, the cold gas LHV is enhanced with increasing gasification temperature and decreasing SBR. The CGE is, however, reduced when gasification temperature increases and SBR decreases.

Nomenclature

Abbreviation	Description
CGE	Cold gas efficiency
DFB	Dual fluidized bed
LHV	Lower heating value
SBR	Steam-to-biomass ratio

References

Alghurabie, I.K., Hasan, B.O., Jackson, B., Kosminski, A., Ashman, P.J. 2013. Fluidized bed gasification of Kingston coal and marine microalgae in a spouted bed reactor. *Chemical Engineering Research and Design*, **91**(9), 1614-1624.

Corella, J., Toledo, J.M., Molina, G. 2007. A Review on Dual Fluidized-Bed Biomass Gasifiers. *Industrial & Engineering Chemistry Research*, **46**(21), 6831-6839.

Cuéllar-Franca, R.M., Azapagic, A. 2015. Carbon capture, storage and utilisation technologies: A critical analysis and comparison of their life cycle environmental impacts. *Journal of CO2 Utilization*, **9**, 82-102.

Farzad, S., Mandegari, M.A., Görgens, J.F. 2016. A critical review on biomass gasification, co-gasification, and their environmental assessments. *Biofuel Research Journal*, **3**(4), 483-495.

Kern, S., Pfeifer, C., Hofbauer, H. 2013. Gasification of wood in a dual fluidized bed gasifier: Influence of fuel feeding on process performance. *Chemical Engineering Science*, **90**, 284-298.

Schmid, J.C., Wolfesberger, U., Koppatz, S., Pfeifer, C., Hofbauer, H. 2012. Variation of feedstock in a dual fluidized bed steam gasifier—influence on product gas, tar content, and composition. *Environmental Progress & Sustainable Energy*, **31**(2), 205-215.

Streimikiene, D. 2010. Comparative assessment of future power generation technologies based on carbon price development. *Renewable and Sustainable Energy Reviews*, **14**(4), 1283-1292.

Takeshita, T. 2011. Competitiveness, role, and impact of microalgal biodiesel in the global energy future. *Applied Energy*, **88**(10), 3481-3491.

Tapasvi, D., Khalil, R., Skreiberg, Ø., Tran, K.-Q., Grønli, M. 2012. Torrefaction of Norwegian Birch and Spruce: An Experimental Study Using Macro-TGA. *Energy & Fuels*, **26**(8), 5232-5240.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50053-7

A Systematic Methodology for Property Model-Based Chemical Substitution from Chemical-based Products

Spardha Jhamb[a*], Xiaodong Liang[a], Rafiqul Gani[a,b], Georgios M. Kontogeorgis[a]

[a]Department of Chemical and Biochemical Engineering, Technical University of Denmark, DK-2800, Kgs. Lyngby, Denmark
[b]PSE for SPEED, Skyttemosen 6, DK-3450 Allerod, Denmark
spajha@kt.dtu.dk

Abstract

In this paper, a general model-based methodology for chemical substitution, which takes into account different problem definitions depending on the objective for substitution, is presented. The associated property models and modeling tools are also described. The application of the methodology is shown through an example on substitution of a chemical from a chemical-based product. It is about finding a substitute for a non-biodegradable surfactant used in products from the cosmetics and personal care sector. Amino acid based surfactants are found to be particularly viable substitute candidates. However, this example requires the development of a set of new group contribution-based models for a number of useful properties of amino acids, which are also presented. Besides this, several other known substitution problems are solved using the developed methodology.

Keywords: Chemical Substitution, Model-based Methodology, Property Models.

1. Introduction

Chemical-based products, which could be structured product formulations, single molecule products or blends of molecules have brought substantial benefits for humans and have been a significant part of the world economy. Life may be difficult to imagine without these structured products. But like every coin has two sides, some chemicals constituting these products can be detrimental to us and the environment. This is primarily due to the hazardous environment-related properties that some of these chemicals possess. Besides, there are many more chemicals which have not been evaluated due to lack of resources for experiments and lack of rigorous estimation methods (Hukkerikar *et al.*, 2012). Hence, there is an urgent need to identify substitute chemicals that will not be dangerous to the environment, toxic to human health and harmful for the eco-systems.

The concerns about the environment and human well-being has given rise to the REACH regulation implemented by the European Chemical Agency (Echa.europa.eu, 2007), which compels companies operating in Europe to stop the use of hazardous substances and replace them with environmentally benign chemicals. Therefore, there is a need to develop a systematic, model-based methodology that can help to find substitutes to existing chemicals in order to improve the environmental impact, while still delivering the same or improved product functionality.

In this work, the objective of developing a chemical substitution methodology is to quickly and reliably identify promising candidates through model-based techniques and

only then proceed to conduct experiments in order to verify and evaluate their performance and applicability. In this way, the experimental resources are used for verification rather than the time-consuming, trial-and-error search. Besides, while seeking alternative substitutes for the undesirable chemicals, the trial and error based approach will have a very large search space. This is avoided by a reverse design approach (Gani, 2004) wherein, predictive property models are made use of. Here, alternative chemicals are found by matching the desired properties of the original product and avoiding the undesired properties of the 'compound to-be substituted'. The developed methodology can be used in systematically identifying, comparing and selecting safer and environmentally compatible alternatives to the chemicals of concern; and finally designing safe chemical product formulations with the same or better product performance.

2. Methodology for Property Model-based Chemical Substitution

2.1. Components of Methodology Development

The model-based methodology development first requires the following three components that need to be very comprehensive in order to solve a wide range of chemical substitution problems:

2.1.1. Environmental Health and Safety (EH&S) Property Criteria to select 'Objective for Substitution'

The EH&S properties enable us to determine whether the 'compound to-be substituted' is an environmental hazard or a human health hazard and what specific danger it has, if it is continued to be used. The criteria that aid in selecting the objective for substitution have been defined by making use of the 'substance classification and labelling' requirements under REACH regulations (Fisk, 2004).

2.1.2. Property Model Library

The property model library (Figure 1) consists of pure component, mixture and product performance property models for a variety of compounds. The product performance models are further classified according to the class of products i.e. blends of molecules, homogeneous liquid phase formulations, emulsions and devices.

2.1.3. Experimental Property Database

The experimental property database available at KT-Consortium contains 40 pure component properties of different types (primary, secondary and functional) and 9 mixture properties of more than 13,000 compounds classified into nine main categories according to molecular structure: normal fluid, polar associating, polar non-associating, multifunctional (with respect to groups), water, polymer, electrolyte, steroid and amino acid (Marrero and Gani, 2004). This database would be required either when the property prediction models are not available or when a developed property model needs to be verified.

2.2. A Model-based Methodology

Based on the regulatory needs along with the regulatory lists like Hazardous Substances Data Bank (HSDB) and REACH Substances of Very High Concern (SVHC) list, the 'compound to-be substituted' can be recognized, when the list of ingredients and their chemical composition in the product are available. In order to find a substitute for this compound, one can follow the Tasks 1 - 7 as outlined in Figure 2 and described below.

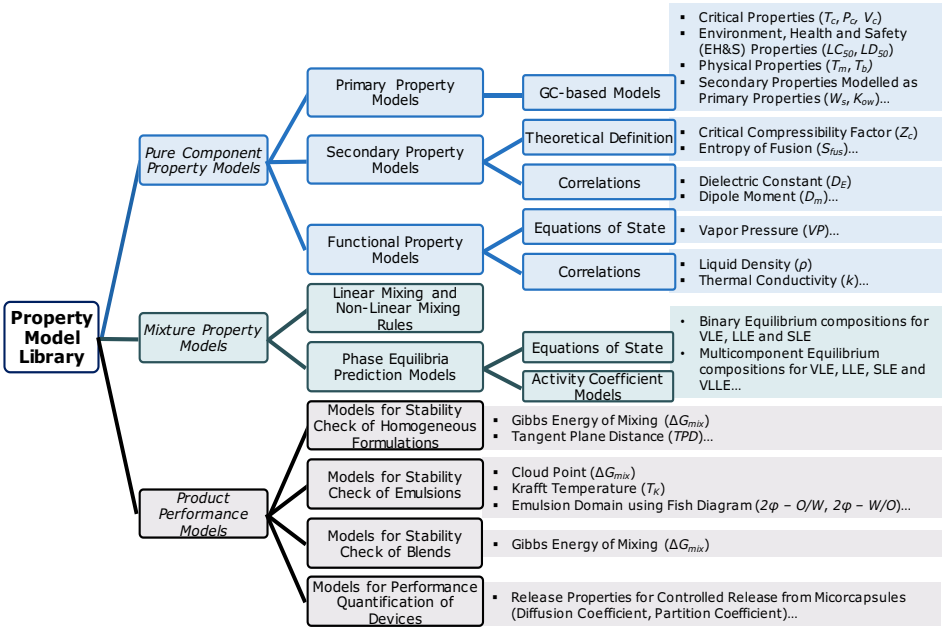


Figure 1: Classification of Properties in the Property Model Library

Task 1. Identification of Problem Type: Firstly, the function of the 'compound to-be substituted' in the original product and the 'objective for substitution' based on the 'EH&S property criteria' are identified. This sets the type of substitution problem that we have i.e. whether the compound to-be substituted is toxic to the aquatic environment, persistent in the environment, toxic on ingestion etc.

Task 2. Problem Definition: Next, the problem is mathematically formulated wherein, the desired and undesired needs of the substitute are recognized and converted to target pure component and mixture properties.

Task 3. Constraint Selection: Constraints on pure components and mixture properties are selected such that the performance and functional properties are as close as possible to the 'compound to-be substituted' while the EH&S properties are as far as possible.

Task 4. Identification / Generation of Substitute Candidates

Sub-Task 4A. Generation of Substitute Candidates: The substitute candidates are generated by Computer-Aided Molecular Design (CAMD), which makes use of property models.

Sub-Task 4B. Database Search: Alternatively, the substitute candidates can be identified by means of a database search after expanding the experimental database.

Task 5. Substitute Compound Verification: In-case it was chosen to generate substitute candidates by CAMD in Sub-Task 4A, but if some of the property models have known uncertainties, then these properties can be checked at the verification stage using an experimental property database or rigorous pure component property models.

Task 6. Final Product Verification: After the list of verified substitute candidates is obtained, it is required that it is tested to check it fits in the original mixture. This can be done by first screening the candidates based on mixture properties calculated through linear mixing rules, then by rigorous thermodynamic and property models and finally by checking the experimental mixture property database and/or conducting experiments. If at this stage, the verification fails, then the problem formulation needs to be changed,

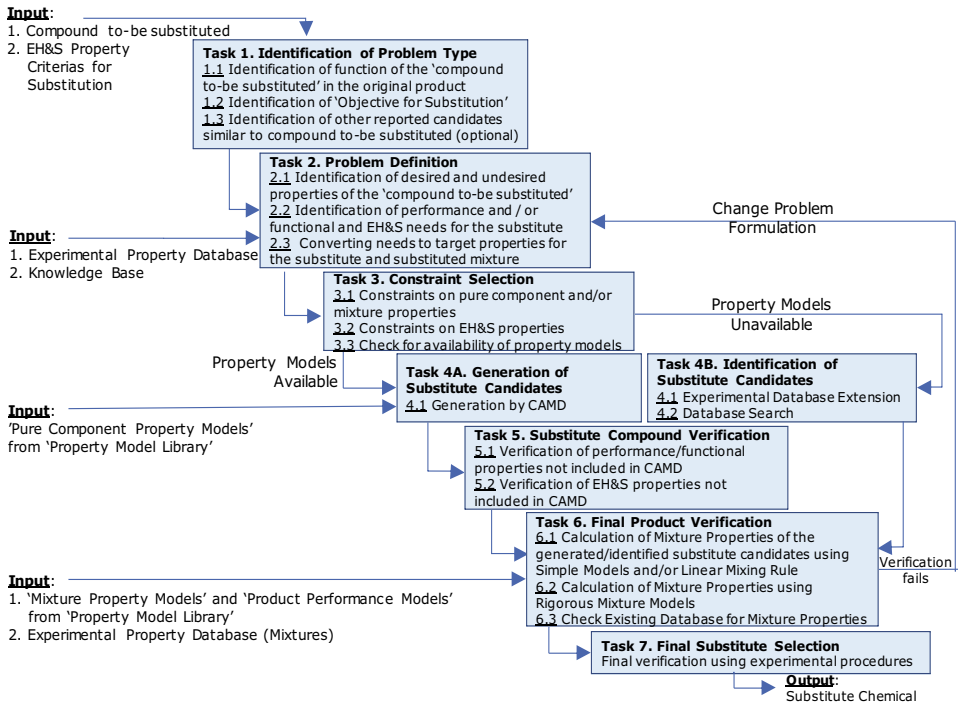


Figure 2: Methodology for Model-based Chemical Substitution

that is, some of the properties need to be compromised for other important ones or the constraints need to be relaxed.

Task 7. Final Substitute Selection: Finally, the substitute can be selected after all its properties have been checked through experimental procedures.

It should be noted that in case a needed GC property model is not available, the method cannot be applied. However, if sufficient data can be found for the selected target property, a GC model can quickly be developed through the property modelling tool-box in ProCAPD (Kalakul *et al.*, 2017) by regressing the group contribution parameters.

3. Practical Examples Employing the Developed Methodology

3.1. An Example: Substitution of Cetrimonium Bromide form a Hair Conditioner

Cetrimonium bromide is a conventional mono-quaternary ammonium cationic surfactant, used in the hair conditioner, 'Rejuvenol Keratin after Treatment Conditioner' (Whatsinproducts.com, 2001). However, it is persistent in the environment and toxic to freshwater crustaceans (Carlroth.com. 2006). In order to find a substitute for this surfactant, the desired and undesired needs are identified. These are then converted to target pure component and mixture properties as shown in Table 1 by making use of the knowledge base.

Two amino acid-based surfactants (Pérez *et al.*, 2002) are found to be viable substitutes with a biodegradation rate of greater than 60% and low toxicity to the aquatic environment. However, due to lack of GC-based models for EH&S properties and the surface-active properties for 'amino acids and their derivatives', the substitute candidate identification is carried out by means of a 'database search' (Sub-Task 4B).

Table 1: Needs, Target Properties and Constraints for Example

Need	Target Property	Target Property Constraints
Ability to form micelles / Solubilize dirt and oils	Critical Micelle Concentration (*CMC*)	$CMC < 0.02$ mol L^{-1}
Ability to lower surface tension	Surface Tension at CMC (σ_{CMC})	$\sigma_{CMC} < 40$ mN m^{-1}
Soluble in Water	Water solubility (log W_s)	$log\ W_s > 1.5$ log(mg L^{-1})
Non-toxic to aquatic environment	Immobilization Concentration (IC_{50})	$-log\ IC_{50} < 4.5$ log(mol m^{-3})
Biodegradable	Biodegradation Rate	> 60% in 28 days

A comparison of the properties of these two identified substitutes with cetrimonium bromide is shown in Table 2.

Table 2: Comparison of Target Properties of Cetrimonium Bromide with Substitute Candidates

Compound	CMC (mol L^{-1})	σ_{CMC} (mN m^{-1})	Log W_s log(mg L^{-1})	-Log IC_{50} log(mol m^{-3})	Biodegradation Rate (%)
Cetrimonium Bromide	0.0009	38.0	4.56	6.88	44
CAM (N_α-acyl-arginine methyl ester, n=8*)	0.0160	40.0	1.53	4.11	95
LKM (N_ε-acyl lysine methyl ester, n=10*)	0.0055	31.0	2.16	4.40	90

*The surfactant is the acyl derivative of an amino acid ester where *n* denotes the alkyl chain length

In order to solve the problem by making use of property models, the missing property models are first required to be developed. Therefore, GC-based models for the prediction of three physicochemical properties of amino acids and their derivatives (normal melting point, T_m, aqueous water solubility, W_s, octanol water partition coefficient, K_{ow}) are developed using the Marrero and Gani group contribution method (Jhamb *et al*., 2018). The performance of these models, determined using the statistical indicators namely, R-square and average absolute error, are summarized in Table 3.

Once GC-based models for the prediction of all the pure component properties necessary to solve this problem (Table 1) are available in the property model library, CAMD can be used to generate substitute candidates with the tool, ProCAMD in ICAS (Gani *et al.*, 1997). As a first input to the CAMD algorithm, the building blocks of amino acids can be provided as they are inherently known to be environmentally compatible and safe molecules.

Table 3: Model Performance Statistics of Developed GC-based Property Models for Amino acids

Property	Number of datapoints	R-square	Average Absolute Error	Maximum Average Absolute Error
T_m	239	0.94	10.81 *K*	42.82 *K*
Log W_s	211	0.94	0.19 *log(mg L^{-1})*	0.97 *log(mg L^{-1})*
Log K_{ow}	335	0.99	0.16	1.18

3.2. Other Examples Solved using the Developed Methodology

Besides the example discussed above, three others have been successfully solved using the developed methodology. Two out of these are well-known problems (ethylene glycol substitution from an engine coolant; methylene chloride substitution from a paint and epoxy remover), while the third one is to substitute an aromatic, toxic solvent used for the dissolution of UHMW-PE in its gel spinning process. It is to be noted however, that the final selection of the substitute for all the solved examples, does need experimental verification (Task 7).

4. Conclusion and Future Work

A systematic model-based methodology for chemical substitution has been developed and has been applied to substitute environmentally hazardous and/or unsafe chemicals in liquid phase product formulations from various industrial sectors. The challenges and opportunities for chemical substitution in devices, solid state products, polymers and plastics etc. still need to be evaluated. While some of the missing property models have been developed and validated, for instance, the group contribution models for predicting the physico-chemical properties of the amino acid class of compounds (Jhamb *et al*., 2018), the development of many more property models is necessary to expand the scope, significance and applicability of model-based methodologies like this.

References

Atsdr.cdc.gov. 2010. *Toxicological Profile for Ethylene Glycol - U.S. Department for Health and Human Services*. [online] Available at: https://www.atsdr.cdc.gov/toxprofiles/tp96.pdf [Accessed 26 Nov. 2017].

Carlroth.com. 2006. Safety Data Sheet - Cetyltrimethylammonium bromide. [online] Available at: https://www.carlroth.com/downloads/sdb/en/9/SDB_9161_IE_EN.pdf [Accessed 26 Nov. 2017].

Echa.europa.eu. (2007). *REACH Legislation - ECHA*. [online] Available at: https://echa.europa.eu/regulations/reach/legislation [Accessed 26 Nov. 2017].

Fisk, P., 2004. Classification and labelling under REACH. In: *Chemical Risk Assessment: A Manual for REACH*. John Wiley & Sons, Ltd, 317 – 339. Available at: http://onlinelibrary.wiley.com/book/10.1002/9781118683989 [Accessed 26 Nov. 2017].

Gani R., Hytoft G., Jaksland C., Jensen A.K., 1997. Computers Chemical Engineering 21(10), 1135–1146

Gani, R., 2004. *Computers and Chemical Engineering* 28(12) 2441-2457.

Hukkerikar, A.S., Kalakul, S., Sarup, B., Young, D.M., Sin, G., Gani, R., 2012. *Journal of Chemical Information and Modeling* 52(11) 2823−2839

Jhamb, S, Liang, X, Gani, R, Hukkerikar, A.S, 2018. *Chemical Engineering Science* 175 148-161

Kalakul, S, Cignitti, S, Zhang, L, Gani, R, 2017, VPPD-Lab: The Chemical Product Simulator, Computer Aided Chemical Engineering, 39, 61-94

Marrero, J. and Gani, R., 2004., *Computer Aided Chemical Engineering*, 19, 45 - 57.

Pérez, L., García, M.T., Ribosa, I., Vinardell, M.P., Manresa, A., Infante M.R., 2002. *Environmental Toxicology and Chemistry* 21(6) 1279-1285.

Whatsinproducts.com, 2001. *CPID*. [online] Available at: https://www.whatsinproducts.com/chemicals/index/1 (accessed 26 Nov. 2017).

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50054-9

Introducing the Concept of Floating Pumps in the Synthesis of Multipurpose Batch Plants

Shaun Engelbrecht[a], Thokozani Majozi[ab*]

[a]*University of the Witwatersrand, Johannesburg, 2017, South Africa*
[b]*NRF/DST Chair in Sustainable Process Engineering, Johannesburg, 2017, South Africa*
thokozani.majozi@wits.ac.za

Abstract

Presented in this manuscript is a novel technique for the exploitation of a unique feature of multipurpose batch plants, so called floating pumps. A MILP mathematical formulation is proposed to determine the optimal number and allocation of pumps within the batch plant while simultaneously synthesizing the plant and optimizing the batch production schedule with variable material transfer times. It was shown by way of an example that it is possible to decrease the number of pumps required for operation from seven to two. By using fewer pieces of equipment which are optimally allocated, equipment degradation due to nonuse can be reduced. It is, therefore, possible to reduce capital and maintenance costs by implementing the formulation while still achieving or improving on existing throughput.

Keywords: Pump, Batch, Synthesis, Multipurpose, Optimization

1. Introduction

Batch processes are commonly used to produce high value products in low volumes. Multipurpose batch plants afford great flexibility by producing a number of different products using the same equipment. This flexibility allows for improved equipment utilization while adapting to rapid market fluctuations. In the selection of batch equipment, the processing requirements and production material properties dictate the design. Most batch chemical plants, such as in the food, pharmaceutical and agrochemical industries, predominantly contain liquids that need to be transferred between processing units using pumps.

Batch plant optimization and scheduling have been studied since the 1970s and gained considerable attention in the field of mathematical modelling during the 1990s. The use of mathematical modelling has made it possible to determine the optimal processing time and material allotted to processing units in order to manage a batch plant efficiently. Sparrow et al. (1975) presented an initial formulation to address the general scheduling problem of multiproduct batch plants by looking at heuristic and branch and bound techniques. A large number of scheduling models based on various insights followed. Notably, within these scheduling models, mainly the major processing equipment of the batch plant, such as the reactors, separators, heaters and storage, were mathematically scheduled, yet focus has not been given to the optimal synthesis or scheduling of pumping equipment within multipurpose batch plants.

Loonkar and Robinson (1970) noted that the capital cost of a batch plant is also dependent on the rating of pumping equipment, heat exchanger duty and the size of the processing vessels. Hegyhati and Friedler (2011) noted that materials transfers are generally simplified to fixed durations when scheduling plant production, regardless of the batch size and costs associated with the transfers. Lee et al. (2015) considered fixed material transfer times in their work in order to allow for heat integration during the transfer, however, each transfer time was a fixed parameter which was independent of the amount of material transferred.

Barbosa-Póvoa and Macchietto (1994) stated that piping connections must be present between vessels in order to facilitate transfer of material and presented a general MILP model that allowed for transfers only if a relevant connection was available. Hegyháti et al. (2009) noted that many models overlooked the subtle transfer infeasibility that arose when various operational philosophies were employed. Specifically noted was that mathematical models inadvertently allowed material to be transferred to and from a processing unit at the same time, so called cross-transfer.

2. Motivation

In this manuscript, the concept of floating pumps is introduced. These floating pumps are not necessarily dedicated to particular units or sections in the plant, as shown in Figure 1(a) but can be reassigned to different processing units and sections as required. By using floating pumps, flexibility in the batch plant is increased. It is evident from Figure 1(b) that the proposed allocation of pumps would result in increased pumping equipment utilization, while reducing the total number of pumps required. This is favorable since prolonged disuse of pumps may increase maintenance costs due to clogging or blockages and structural degradation. Notably, this can be achieved while preserving or even improving the production throughput of cases where no flexible pump allocation is taken into consideration. This synthesis method can be applied for plant design where it is desirable to limit the number of equipment units required, in order to save on capital investment and optimally use available space. It can also be applied in retrofit design in order to eliminate unnecessary processing and pumping equipment in a plant, thereby saving on maintenance costs.

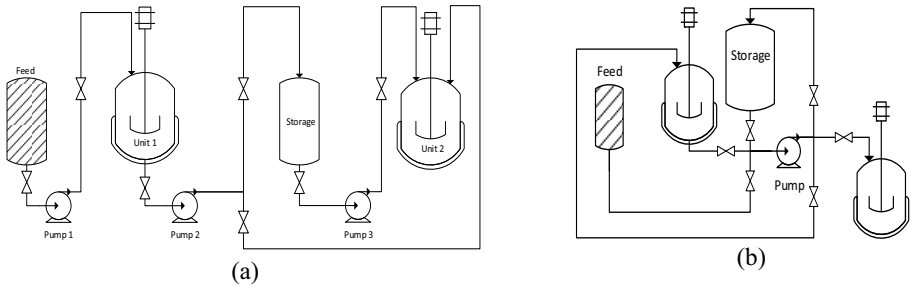


Figure 1 (a) Conventional pump allocation and (b) proposed pump allocation

3. Key Constraints

The key constraints that pertain to pump allocation are shown below. These constraints may be adapted into scheduling models. The scheduling model used in this manuscript is reformulated from the models presented by Seid and Majozi (2013) and Lee et al. (2015) to allow for pumping allocation and variable transfer times. In the formulation, v represents any vessel, j a processing unit, v_v a storage vessel, ζ a pump, v_s a source vessel, v_d a destination vessel, s a state, $s_{in,j}$ a task and p a time point.

Constraints (1) and (2) state that pumping (μ) may only occur if inlet (i) and outlet (o) piping connections are available.

$$\mu(v', v, \zeta, p) \leq i(\zeta, v') \tag{1}$$

$$\mu(v', v, \zeta, p) \leq o(v, \zeta) \tag{2}$$

Constraints (3) and (4) state that a pump may only pump to a single unit or from a single unit at a given time point, respectively. This may only occur if the pump exists.

$$\mu(j, v_s, \zeta, p) + \sum_{j'} \mu(j', v_s', \zeta, p) \leq e(\zeta) \tag{3}$$

$$\mu(v_d, j, \zeta, p) + \sum_{j'} \mu(v'_d, j', \zeta, p) \leq e(\zeta) \tag{4}$$

Constraint (5) states that a pump must be active between a receiving storage vessel and a unit in order to transfer material.

$$m^{in}(s, j, p) \leq V^U(v_v) \sum_{\zeta} \mu(v_v, j, \zeta, p) \tag{5}$$

Similarly, Constraint (6) states that a pump must be active between a receiving unit and a storage vessel to transfer material.

$$m^{out}(s, j, p) \leq V^U(v_v) \sum_{\zeta} \mu(j, v_v, \zeta, p) \tag{6}$$

Additionally, Constraint (7) states that a pump must be active between two different units to transfer material.

$$m_t(s, j, j', p) \leq V^U(j) \sum_{\zeta} \mu(j, j', \zeta, p) \tag{7}$$

Constraints (8) – (10) state that the transfer time is dependent on the respective material transfer and the rating of the pump used.

$$R(\zeta) m_t(s, j, j', p) - M\big(1 - \mu(j, j', \zeta, p)\big) \leq t_t(s, j, j', \zeta, p) \leq R(\zeta) m_t(s, j, j', p) + M\big(1 - \mu(j, j', \zeta, p)\big) \tag{8}$$

$$R(\zeta) m^{in}(s, j, p) - M\big(1 - \mu(v, j, \zeta, p)\big) \leq t_t(s, v, j, \zeta, p) \leq R(\zeta) m^{in}(s, j, p) + M\big(1 - \mu(v, j, \zeta, p)\big) \tag{9}$$

$$R(\zeta) m^{out}(s, j, p) - M\big(1 - \mu(j, v, \zeta, p)\big) \leq t_t(s, j, v, \zeta, p) \leq R(\zeta) m^{out}(s, j, p) + M\big(1 - \mu(j, v, \zeta, p)\big) \tag{10}$$

Constraint (11) states that there is no transfer time required to transfer material from a unit to itself.

$$t_t(s, j, j, \zeta, p) = 0 \tag{11}$$

Constraints (12) and (13) assign slots for pumping to the scheduling formulation and states that sufficient time for pumping into a task and out of a task must be given, respectively.

$$t^{in}_{dur}\big(s_{in,j}, p\big) \geq \sum_{v_S} t_t(s^{sc}, j, v_S, \zeta, p) \tag{12}$$

$$t^{out}_{dur}\big(s_{in,j}, p\big) \geq \sum_{v_v} t_t(s^{sp}, v_v, j, \zeta, p) \tag{13}$$

4. Illustrative Example

An illustrate example is presented based on the scenario given by Lin and Floudas (2001), which involves a multipurpose batch plant capable of performing 4 distinct tasks. Three major processing units are available to perform the processing tasks, along with a storage vessel that is available to store state 4 (S4). The recipe representation for the scenario is given in Figure 3. Data pertaining to the equipment may be found in Table 1 and economic data in Table 2, as adapted from Lin and Floudas (2001). The data pertaining to pumping rates for the various available pumps are given in Table 3. The illustrative pumping rates were calculated such that no pump will be active for more than five minutes when the pumps are placed conventionally with the plant operating at full capacity. A time horizon of 12 hours was considered.

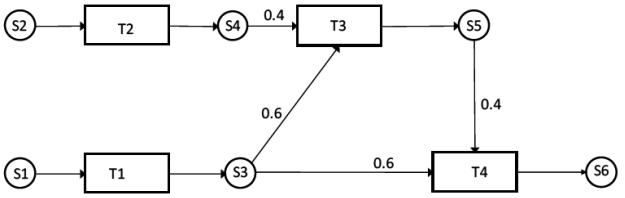


Figure 2 STN Representation for Illustrative Example

Figure 3 shows the typical flowsheet for the illustrative example and depicts the pumps which would conventionally be used to transfer material between different areas within the plant.

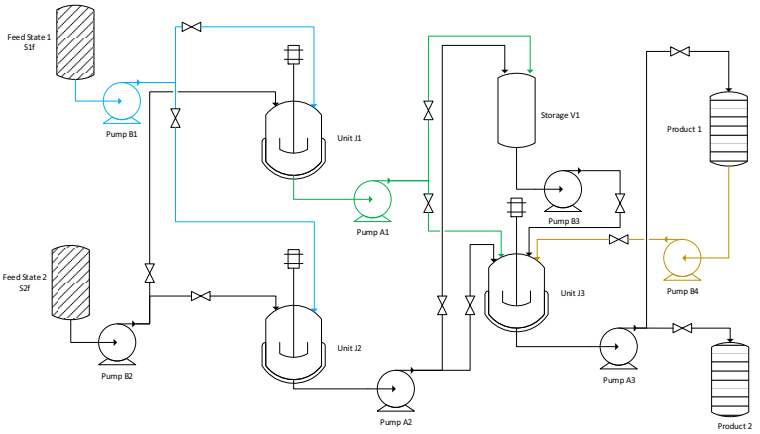


Figure 3 Conventional pump allocation and plant flowsheet

Table 1. Equipment data for the first illustrative example

Unit / Storage Vessel	Capacity	Suitability (Task/Storage)	Task Duration (Hours)	Cost (Cost Units)
Unit J1	50-150 mu	T1 T2	2 2	100
Unit J2	50-150 mu	T1 T2	2 2	150
Unit J3	50-200 mu	T3 T4	4 2	120
Vessel 4	100 mu	S4	-	15

Table 2. Economic data for the first illustrative example

Material/Equipment	Price (Cost units)
S1, S2	0.1 /mu
S5 (Product 1)	0.4 /mu
S6 (Product 2)	0.6 /mu
Pump base cost	10
Pump rating cost	0.001
Piping cost per connection	5

Table 3. Pump ratings for the first illustrative example

Pump	Pump Rating (seconds / unit material)	Pump	Pump Rating (seconds / unit material)
A1	2.00	**B1**	2.00
A2	2.00	**B2**	2.00
A3	1.50	**B3**	1.50
		B4	1.50

The MILP problem was solved using the CPLEX solver. The optimal pumping schedule may be found in Figure 4 and the optimal processing schedule in Figure 5. Figure 6 shows the resulting optimal plant pumping structure needed to cater for the example. By flexibly allocating pumps, the plant required two pumps to transfer material throughout the batch plant whereas seven pumps were required when flexible pump allocation was not considered. The synthesis results also show that a storage vessel is not required for S4. The problem was solved using four time points, yielding an objective value of 1.27×10^6 cost units in a CPU solution time of 2.45 seconds.

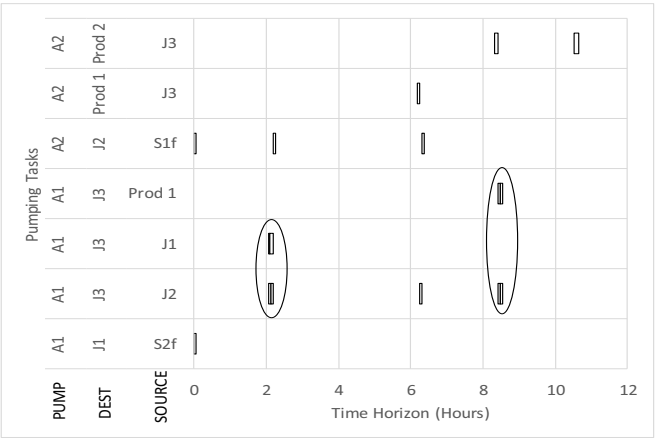


Figure 4 (a) Optimal pumping scheduling using floating pumps

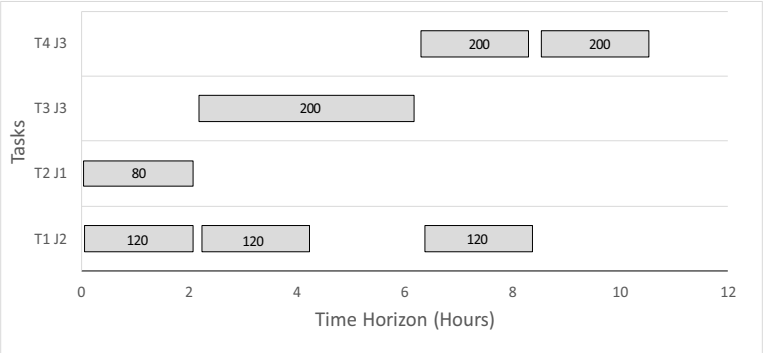


Figure 5 Optimal processing schedule using floating pumps

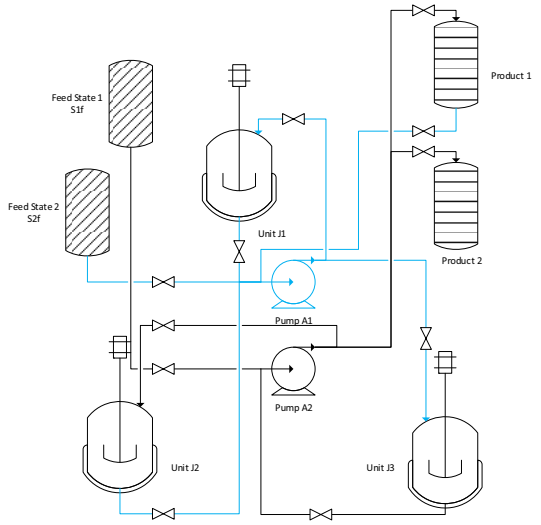


Figure 6 Optimal floating pump allocation and connections

5. Conclusions

A MILP mathematical formulation is introduced which is capable of synthesizing multipurpose batch chemical plants while flexibly allocating pumps to transfer different materials to units during the time horizon of interest. Previously, no consideration has been given to the pumping equipment in the synthesis and scheduling of batch plants. The illustrative example showed it is possible to decrease the total number of pumps required from seven to two pumps. Worthy of mention is that product integrity issues would have to be taken into consideration whenever the concept of roving pumps is introduced in a practical process. It is also evident that the roving pumps result in combinatorially complex operations that would ultimately require advanced control mechanisms in order to avoid cross contamination. These aspects will be the subject of future investigation.

References

Barbosa-Póvoa, A.P., Macchietto, S., 1994. Detailed design of multipurpose batch plants. Comput. Chem. Eng. 18, 1013–1042.

Hegyhati, M., Friedler, F., 2011. The role of energy consumption in batch process scheduling, in: 21st European Symposium on Computer Aided Process Engineering - ESCAPE 21. pp. 1979–1983.

Hegyháti, M., Majozi, T., Holczinger, T., Friedler, F., 2009. Practical infeasibility of cross-transfer in batch plants with complex recipes: S-graph vs MILP methods. Chem. Eng. Sci. 64, 605–610.

Lee, J., Seid, E.R., Majozi, T., 2015. Heat integration of intermittently available continuous streams in multipurpose batch plants. Comput. Chem. Eng. 74, 100–114.

Lin, X., Floudas, C., 2001. Design, synthesis and scheduling of multipurpose batch plants via an effective continuous-time formulation. Comput. Chem. Eng. 25, 665–674.

Loonkar, Y.R., Robinson, J.D., 1970. Minimization of Capital Investment for Batch Processes Calculation of Optimum Equipment Sizes. Ind. Eng. Chem. Process Des. Dev. 9, 625–629.

Seid, E.R., Majozi, T., 2013. Design and synthesis of multipurpose batch plants using a robust scheduling platform. Ind. Eng. Chem. Res. 52, 16301–16313.

Sparrow, R.E., Forder, G.J., Rippin, D.W.T., 1975. The choice of equipment sizes for multiproduct batch plants. Heuristics vs. branch and bound. Ind. Eng. Chem. Process Des. Devel. 14, 197–203.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50055-0

Conceptual Process Design and Economic Analysis of Oxidative Coupling of Methane

Hamid Reza Godini[a], Mohammadreza Azadi[a], Mohammadali Khadivi[a], Abolfazl Gharibi[a], Seyed Mahdi Jazayeri[a], Daniel Salerno[a], Alberto Penteado[a], Babak Mokhtarani[b], Alvaro Orjuela[c], Tim Karsten[a], Günter Wozny[a], Jens-Uwe Repke[a]

[a]*Chair of Process Dynamics and Operation, Berlin Institute of Technology, Berlin, Germany*
[b] *Chemistry and Chemical Engineering Research Centre of Iran,Tehran, Iran*
[c]*Department of Chemical and Environmental Engineering, Universidad Nacional de Colombia, Bogotá D.C., Colombia*
hamid.r.godini@tu-berlin.de

Abstract

Addressing the need for comprehensive analysis of the potentials and characteristics of ethylene production via Oxidative Coupling of Methane (OCM) process was the main motivation of performing the current study. In this context, technical, environmental and economic characteristics of alternative OCM process design structures including the integrated OCM and ethane dehydrogenation process for the industrial plants with the annual capacity of one million t ethylene production were analyzed and compared by predicting their performances using Aspen-Plus simulation and Aspen Economic Process Analyzer. The performance of the reactor section, carbon dioxide separation section and adsorption section in this simulation were mapped and validated using the observed performances of the OCM miniplant scale experimental facility constructed in TU Berlin.

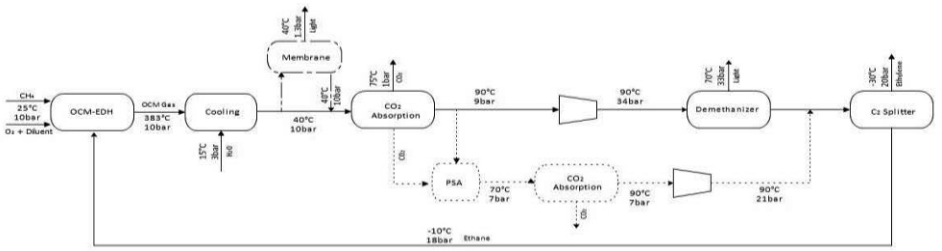
It was found that the operating cost in all cases is the main source of cost, so that even the one-year operating cost is estimated to be 3-4 times of the fixed cost. In the operating cost, the raw material cost stands for the major part of the cost to be around 750 million Euro annually. Most of the utility costs is needed for the energy-intense cryogenic distillation, which is avoided in the demethanizer section when adsorption unit is utilized. However, one percent loss of ethylene and significant increase on the fixed-cost are the main disadvantages of using adsorption technology. Yet, this process structure provides the fastest return of investments around 9 y based on the considered costs and the assumptions in this study. Using selective carbon dioxide membrane separation technology was also investigated and showed a marginal contribution specially by considering the expected 2% loss of ethylene in this case. If better membrane technologies for CO_2 separation or preferably for ethylene separation becomes available, it can be beneficial in the whole economy of the process and in this case the best observed total energy consumed can be improved beyond current 40-50 GJ/t ethylene. Beside trying to reduce the total cost, the amount of the generated CO_2 and the total energy consumed associated with production of one t of ethylene, potential of using bio-based materials for instance for producing the required adsorbents and also utilizing the whole OCM process for bio-based methane feedstocks were also taken into consideration. The technical challenges regarding the OCM catalyst and reactor operation should be addressed prior to any attempt for industrial scale operation of this process.

Keywords: OCM process, Economic evaluation, Ethylene adsorption, Process scale analysis.

1. Introduction

Ethane and naphtha cracking are the main stream mature technologies currently used for ethylene production in the scale of 160 million t/y as the third largest chemicals produced worldwide. Achieving around 60 percent conversion and 80 percent selectivity towards ethylene are the promising performance indicators of these processes. As an attempt for diversifying the feedstocks and replacing a cheaper and more available feedstock for producing ethylene from methane, Oxidative Coupling of Methane (OCM) process has been investigated for more than 35 years. OCM technology specifically is recommended for utilizing the remote and waste methane sources in the places where ethane or naphtha are not easily available in the quantity needed for the current and future capacity of the crackers. During the last four decades, different aspects of the OCM process such as catalysts' materials and characteristics, reactor concepts, process structures etc. have been investigated and their performances have been improved. However, the OCM process has not been so far implemented industrially mainly due to the fact that globally observing, ethylene production via cracking is preferred both technically and economically. To mention the technical reasons, it can be highlighted that a) the previous reported catalysts has not been enough selective towards ethylene and more importantly has not been enough stable, b) operation of the exothermic high temperature OCM reactor in large scale is a huge challenge, c) separation of the OCM products and unreacted reactants requires an energy/cost-efficient removal/purification system. Although, the challenges regarding OCM economic perspective and its catalytic performance remains, the progresses achieved in these areas in the last decade specially based on the extensive experiences achieved by our research group (Cluster of Excellence "Unifying Concepts in Catalysis" UniCat, www.unicat.tu-berlin.de) are the major motivations for re-evaluating the OCM process in this study. In this paper, the observed performances of the catalyst-reactor section, carbon dioxide removal and ethylene separation in the UniCat miniplant experimental facility are set as the base for the simulation-economic study of four alternative OCM plants targeting one million t/y ethylene production. The specifications and the assumptions associated with the performances of these sections will be reviewed here while the specifications of the experimental setup will be described later in this paper for each unit operation.

Figure 1 shows the schematic representation of the OCM process in which the unit operations and their interactions have been demonstrated.



— – — first alternative for case 3 (Hybrid membrane-absorption for CO_2 removal)
…… second alternative for case 4 (Adsorber)

Figure 1. Schematic representation of the OCM covering four process structures

1.1. Catalyst-Reactor Section

Among the selected tested OCM catalysts, the benchmark Mn-Na_2WO_4/SiO_2 catalyst has shown a promising performance in terms of the selectivity and stability, therefore it has been widely recommended for this process (Yildiz*, et al.* 2016). This catalyst has been

synthesized and tested in different-scales reactors in the UniCat miniplant (Sadjadi*, et al.* 2015, Godini*, et al.* 2014) resulted in one of the best so far observed OCM catalysts performances in membrane reactor (Godini*, et al.* 2014). It worth emphasizing that such reactor performance has been achieved under the set of operating conditions which are also practically relevant regarding the industrial-scale operation of this process. For instance, in the reactor section, the utilized level of the methane-to-oxygen feed ratio enabled securing the minimum level of methane conversion and ethylene yield respectively at 40% and 60% believed roughly to be the minimum values of these factors associated with potential economic OCM process in industrial scale. Moreover, the level of inert gas dilution, which will impose a significant cost in the downstream units to be separated, should be kept as low as possible. The observed reactor's product compositions are reported in Table 1. The utilized feed stream composed of methane, oxygen and nitrogen respectively to be 64%, 32% and 4%. The feed was introduced to the pre-heating section of the reactor with 600 ˚C and the products have been cooled down to 650 ˚C before utilizing further heat integration.

Table 1. OCM reactor performance indicators of 50% conversion and 23% C_2-yield; product composition observed experimentally and mapped in the performed model-based simulation for 4% N_2-diluted feed of methane-to-oxygen ratio 2

CH_4	H_2O	N_2	CO	CO_2	C_2H_4	C_2H_6	C_3H_6	C_3H_8	H_2
0.31	0.28	0.04	0.04	0.13	0.05	0.02	0.01	0.002	0.12

The experimentally observed performance of the OCM and EDH reactors were mapped in the Aspen Plus simulation to simulate the outlet compositions of the reactors and conducting the techno-economic analysis of the process scenarios using Aspen Economic Analyzer. The required amount of catalysts were calculated based on the productivities of these catalyst which later was resulted in determining the dimension of the reactor by considering the density of those catalyst and effective volume of the shell-tube reactors.

1-150 gr catalysts have been tested in the UniCat OCM reactors with the reactor diameter of 7-55 mm. In the cases of integrated reactors, the design of the reactor section has been considered as an integrated form of OCM and an endothermic ethane dehydrogenation (EDH) reaction system. The performance of the EDH section has been also considered to be around 56% ethane conversion with approximately 50% ethylene yield.

The assumptions related to this reactor section is that the catalyst is stable for 5 y, heat-integration is manageable, and the operating pressure is 10 bar. The reason for the last one is that in this case, carbon dioxide removal in the next section takes place efficiently under this pressure and any need for compressing the gas from the reactor section into the CO_2-removal section or other downstream units imposes a huge operating cost due to the need for special corrosive resistance material and relatively high compression duty there. Moreover, supplying methane-rich and oxygen-rich stream of 10 bar from the previous units is not complicated. In any case, the performance of the OCM reactor should be evaluated in the context of the whole OCM process.

1.2. Carbon Dioxide Removal Section

As mentioned earlier, since more than 40 percent of the converted methane appears in the form of carbon oxides, it is important to remove it from the reactor outlet gas stream as quickly and as efficiently as possible. This is achieved via different techniques. For instance, in the UniCat miniplant facility, the performance of the amine absorption unit and the polymer membrane module for this task were experimentally investigated.

1.2.1. Amine Absorption

Wide range of operating conditions and amine materials have been tested for carbon dioxide removal from the OCM reactor outlet stream. It was experimentally shown that Methyl diethanolamine (MDEA) solution under 10 bar pressure in the miniplant-scale absorption column with 40 mm diameter and 5 m length, under 30 ˚C, provides one of the best removal performances in terms of lowest energy consumption and ethylene lost. It was decided to keep the ethylene lost below 5% as ethylene lost more than that value is obviously is economically unjustified.

In the simulation, MDEA solution of 10% MDEA+ 88% H_2O + 2% Piperazin was used.

In HYSYS, thermodynamic equations method recommended for CO_2 Absorption "Acid gas" was used in an absorber-column with 48 stages under 10 bar pressure. For the stripping, a recoiled absorber with 15 stages under 1.5 bar has been used.

1.2.2. Membrane

Using polymer membrane especially cellulose acetate, polyimides or poly-(ethylene oxide) active layer have been tested in the UniCat miniplant-scale Flat sheet envelope type of membrane module (provided by Helmholtz-Zentrum Geesthacht) for CO_2 removal task for this application. It was shown that the hybrid separation via combination of membranes and absorption provides the best separation performance in terms of minimizing the required energy and ethylene lost. Further information in this regard are available elsewhere (Brinkmann, *et al.* 2015). A one-dimensional solution-diffusion model was developed to represent the performance of such membrane modules in the industrial scale techno-economic analysis.

1.3. Ethylene Separation Section

The limited range of methane conversion below 50% per pass and consequently the huge amount of the remaining unreacted methane which is needed to be separated from the rest of gaseous species is a major source of deficiency in the OCM process as this is usually done using the energy-intense cryogenic distillation.

1.3.1. Separating unreacted methane and other light gases via Demethanizer

Right after the carbon dioxide removal section, by providing around 33.5 bar pressure and condenser temperature of -111.9 ˚C, separation performance of this system for the industrial scale system was simulated using RadFrac column with 30 stages, Reflux ratio of 1.5 and distillate to feed ratio of 0.83 in a 8.5 m diameter, 30 m tall column which traces the methane in the C_2-rich stream by minimum ethylene loss of 0.27 %.

1.3.2. C_2-Splitter

Similar procedure was applied for separating ethane from ethylene in a 5.4 m diameter 18 m height RadFrac-column with 66 stages under 20 bar pressure and around -29 ˚C operating temperature with Reflux ratio of 5.3 and distillate to feed ratio of 0.648 to secure ethylene purity of 99.5 %.

1.3.3. Adsorption

As an alternative in the downstream of the OCM process, a pressure/temperature swing adsorber PSA/TSA system with 60 mm diameter and one-meter length was tested experimentally in the UniCat miniplant to separate the C_2-products after the reactor as quick and efficiently as possible. Improved Zeolite 13X was used as an efficient adsorbent for such task showing a significant adsorption capacity. Similar to the other unit operations, the experimentally observed performance of the adsorption unit in the miniplant scale was mapped for the large scale OCM plant and all operating aspects associated to it, such as considering 5 columns from which one is always in adsorption

cycle and three are in desorption cycle and one is used as by-pass for backup, were considered in the techno economic analysis and CO_2 was used as strip gas.

The selected four process scenarios and the results of their techno-economic analysis are listed in Table 2. Case 1 is the reference OCM process. Case 2 has an EDH reactor. Case 3 is similar to case 2 but utilizes a hybrid membrane absorption system. Case 4 is similar to case 2, but instead of demethanizer uses a PSA/TSA system for ethylene separation.

Table 2. Selected OCM process cases and their performance-cost items

case #	OCM Reactor	integrated potential	CO_2 removal		C_2 separation		C_2 splitter	P.O. Period [y]
		EDH	Amin absorption	Polymer membrane	Demethanizer	Adsorption		
case 1	44 [€/t C_2H_4]	-	21 [€/t C_2H_4]	-	127[€/t C_2H_4]	-	3.9 [€/t C_2H_4]	-
case 2	51 [€/t C_2H_4]		20 [€/t C_2H_4]	-	100[€/t C_2H_4]	-	5.9 [€/t C_2H_4]	9.52
case 3	51 [€/t C_2H_4]		17 [€/t C_2H_4]		100[€/t C_2H_4]	-	6.3 [€/t C_2H_4]	9.96
case 4	54 [€/t C_2H_4]		35 [€/t C_2H_4]	-	-	11 [€/t C_2H_4]	6.6 [€/t C_2H_4]	8.95

The reported costs here include the utility cost per t of ethylene in which the equipment cost has been depreciated and included (€/t C_2H_4).

The details of the costs associated with each task and section of these case-processes have been also shown in Table 2. As seen here, it is clear that by increasing the ethylene production capacity by integrating the EDH reactor, the separation costs per t of ethylene is reduced. It is also seen that the major part of the cost is associated with separation of the unreacted methane and light gases, when demethanizer is used. Using adsorption, demethanizer can be even completely avoided. Therefore, the operating cost for C_2-separation can be reduced significantly. It should be mentioned however, that these are the operating costs and in order to have a base for comparing the economy of these cases, fixed costs and pay-out period of these processes should be also taken into analysis. The results of the performed techno-economic analysis of these scenarios are reported in Table 3. For such analysis, the cost of raw materials, main and side products, utilities etc. have been collected for the same local and have been doubled checked with the industrial sources. For instance, one t of methane, ethylene, ethane and one Kcal generated HP-steam were respectively considered to cost 110 €, 900 €, 250 € and 6.45E-06 €.

Table 3. Results of the performed techno-economic analysis on the OCM process cases

	Case 1	**Case 2**	**Case 3**	**Case 4**
Ethylene Lost [%]	0.8	0.3	2.1	1.0
Operating Cost [€]	994,295,520	1,031,630,000	1,025,270,000	899,646,000
Utility Cost [€]	168,363,160	196,424,000	191,280,000	75,921,500
Raw Material Cost [€]	730,328,360	734,983,000	734,948,000	742,601,000
Fixed Cost	265,754,000	349,605,000	364,794,000	388,877,000
Total energy per t of Ethylene [GJ/t C_2H_4]	53.95	46.87	43.03	53.50
Total energy cost per t of Ethylene [€/t C_2H_4]	170.33	156.70	156.10	115.95
Total CO_2 Produced per t of Ethylene [t CO_2/t C_2H_4]	3.82	3.03	3.06	3.31

Table 3. Results of the performed techno-economic analysis on the OCM process cases (continued)

Utility cost per t of Ethylene with equipment cost depreciated included [€/t C_2H_4]				
Section of reaction	44.08	50.87	51.18	53.97
Section of CO_2 separation	21.10	20.23	16.75	34.79
Section of Demethanizer	126.92	100.22	100.28	0.00
section of C_2 splitter	3.89	5.89	6.33	6.56
Section of adsorption	0.00	0.00	0.00	11.05

2. Conclusions

The performance of the OCM reactor, carbon dioxide removal and C_2-separation sections have been analyzed in a miniplant scale experimental facility and their results have been mapped in the large scale OCM plants representing four different process-structures.

Recycling ethane back to the OCM-EDH reactor section securing significant ethane dehydrogenation and heat integration has a significant positive impact on the economy of the OCM process. Moreover, using a low-pressure adsorption unit enables one to even replace the cryogenic distillation used for demethanizer and its related required compression and cooling duties, which stand for the major portion of operating cost of the whole OCM process. Fixed-cost in this case will be increased however. Similarly, using polymer membrane reduces the required energy for CO_2 separation which is not a significant part of the whole OCM process cost though. All these conclusions are only valid when the remaining operating challenges for the OCM catalyst and reactor are addressed as assumed in this study.

Acknowledgement

The authors acknowledge the financial support from the Cluster of Excellence “Unifying Concepts in Catalysis” coordinated by the Technische Universität Berlin and funded by the German Research Foundation -Deutsche Forschungsgemeinschaft-.

References

M. Yildiz, Y. Aksu, U. Simon, T. Otremba, K. Kailasam, C. Göbel, F. Girgsdies, O. Görke, F. Rosowski, A. Thomas, R. Schomäcker, S. Arndt, Silica material variation for MnxOy-Na2WO4/SiO2, Appl. Catal. A, 525 (2016), 168-179.

S. Sadjadi, S. Jaso, H. R. Godini, S. Arndt, M. Wollgarten, R. Blume, O. Görke, R. Schomäcker, G. Wozny and U. Simon, Feasibility study of the Mn–Na2WO4/SiO2 catalytic system for the oxidative coupling of methane in a fluidized-bed reactor Catalysis, Catalysis Science and Technology, 5 (2015), 942–952.

H. R. Godini, A. Gili, O. Görke, U. Simon, K. Hou, and G. Wozny, Performance analysis of a porous packed bed membrane reactor for Oxidative Coupling of Methane: Structural and operational characteristics, Energy Fuels, 28, 2 (2013), 877–89.

T. Brinkmann, C. Naderipour, P. Jan, J. Wind, W. Thorsten, E. Erik, D. Müller, G. Wozny, B. Hoting, Pilot scale investigations of the removal of carbon dioxide from hydrocarbon gas streams using poly (ethylene oxide)–poly (butylene terephthalate) polyActive™) thin film composite membranes. Journal of Membrane Science, 489 (2015), 237-247.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50056-2

Energy-economic multi-objective modeling framework for simultaneous multistream heat exchangers and process optimization

Yingzong Liang[a], Keat Ping Yeoh[a], Ergys Pahija[a], Pui Ying Lee[a], Chi Wai Hui[a,*]

[a]*Department of Chemical and Biological Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, N.T., Hong Kong*
kehui@ust.hk

Abstract

This paper presents a multi-objective modeling framework for simultaneous process-heat integration optimization with multistream heat exchangers (MHEXs). The trade-off between operating cost and heat integration of the process is considered. The proposed framework comprises heat recovery of MHEXs, thermodynamics, vapor-liquid equilibrium, and other unit operations of process. The MHEXs model takes advantage of the compact formulation of the multi-M approach (Hui, 2014) to improve solution efficiency. The model is formulated as a mixed-integer nonlinear programming (MINLP) model and applied to an industrial air separation unit (ASU) to demonstrate its effectiveness. Multi-objective optimization is performed to investigate the trade-off between heat recovery of MHEXs and energy consumption of air compressors.

Keywords: multistream heat exchanger, heat integration, process optimization, multi-objective optimization

1. Introduction

Multistream heat exchanger (MHEX) is a single unit process for simultaneous heat transfer between more than two streams. MHEXs find applications in a variety of energy intensive processes that includes, but are not limited to, air separation (Pattison and Baldea, 2015), food and pharmaceutical production (Galeazzo et al., 2006), hydrocarbon processing (Shah et al., 2000), hydrogen purification and liquefaction (Christopher and Dimitrios, 2012), liquefied natural gas (LNG) production (Kamath et al., 2012), power generation (Memmott et al., 2017), and refrigeration processes (Thonon, 2008).

Despite wide application, it is often challenging to simultaneously optimize MHEXs within process due to the restriction of temperature difference and occurrence of phase change. Further complicating the issue is the variable stream data, e.g. inlet and outlet temperatures and flow rates. Hasan et al (2009) applied the concept of superstructure (Yeomans and Grossmann, 1999) to simulate MHEX, and formulated an MINLP model to optimize heat recovery of MHEX. However, the complex superstructure is often computationally expensive, which brings greater difficulty to the simultaneous process and MHEX optimization. Kamath et al. (2012) presented a nonlinear programming (NLP) formulation of MHEX based on the concept of Pinch analysis (Kemp, 2011,Liang and Hui, 2016). The major challenge of the NLP formulation is the

nonlinear approximation of *max* function, which leads to errors in energy balance and numerical difficulties.

Therefore, an efficient modeling approach of MHEX is significant for the simultaneous MHEXs and process optimization. Moreover, existing methods rarely performed multi-objective optimization of processes with multiple MHEXs. In practice, however, multiple factors are considered in the design so that process can achieve balanced energy-economic performance. Thus, there is a growing interest to develop a rigorous and computationally efficient modeling framework and investigate the tradeoff between different factors (e.g. heat recovery and operating cost). Heat integration modeling approach (Hui, 2014) and multi-objective optimization method (Hwang and Masud, 1979) can be applied to the simultaneous MHEXs and process optimization.

The objective of this paper is to propose a framework for simultaneous MHEXs and processes optimization that addresses the modeling of MHEX and considers trade-off between operating cost and heat recovery. The proposed framework avoids discontinuous *max* function and nonlinear approximation when modeling the MHEX. The resulting optimization problem is formulated as a multi-objective MINLP model, which simultaneously maximizes heat recovery of MHEXs and minimize operating cost of process. The model subjects to the constraints of energy balance, constraints to enforce temperature difference of heat transfer in MHEXs, and constraints of the thermodynamics and other unit operations. We apply the model to an industrial air separation process with multiple MHEXs to demonstrate its effectiveness.

2. Model formulation

(MP1) summarizes the proposed framework for simultaneous MHEXs and process optimization. The model considers heat recovery of MHEXs, thermodynamics, and other unit operations. **(MP1)** is formulated as a multi-objective MINLP model for the nonlinear thermodynamics constraints and discrete nature of the multi-M formulation. For convenience, parameters are denoted by upper case letter, and variables by lower case letter in this paper.

Heat transfer of an MHEX can be regarded as heat integration of multiple process streams with no heating and cooling utilities. Therefore, the concept of Pinch analysis is applicable to model MHEXs. However, simultaneous MHEXs and process optimization usually requires nonlinear formulation or binary variables, which are computationally expensive. To enhance solution efficiency, we propose a method adapted from Hui's (2014) multi-M formulation that circumvents nonlinear formulation and minimizes the number of binary variables. The model assumes counter-current of hot and cold streams at the MHEXs. Additionally, the heat capacity of a stream is assumed temperature independent in liquid phase, vapor phase, and phase change. The stream data (e.g. inlet and outlet temperatures and flow rate) can be varied or fixed.

(MP1)

$$max\ \varphi(x) = \sum_{i \in H} CP_{i,e} \cdot f_{i,e}\left(t_{i,e}^{in} - t_{i,e}^{out}\right) \tag{1}$$

$$min\ \psi(x) = \sum_{n \in N} cost_n \tag{2}$$

s.t.

$$\sum_{i\in H} CP_{i,e}\cdot f_{i,e}\left(t_{i,e}^{in}-t_{i,e}^{out}\right)=\sum_{i\in C} CP_{i,e}\cdot f_{i,e}\left(t_{i,e}^{out}-t_{i,e}^{in}\right)\quad e\in E \tag{3}$$

$$qsoa_{j,e}\geq qsia_{j,e}\quad j\in I,e\in E \tag{4}$$

$$qsoa_{j,e}=\sum_{i\in H} CP_{i,e}\cdot f_{i,e}\left(tp_{i,j,e}^{in}-tp_{i,j,e}^{out}\right)\quad j\in I,e\in E \tag{5}$$

$$qsia_{j,e}=\sum_{i\in C} CP_{i,e}\cdot f_{i,e}\left(tp_{i,j,e}^{out}-tp_{i,j,e}^{in}\right)\quad j\in I,e\in E \tag{6}$$

$$tp_{i,j,e}^{in}\leq t_{i,e}^{in}+M1_{i,j,e}^{in}\left(1-y_{i,j,e}^{in}\right)\quad i\in I,j\in I,e\in E \tag{7}$$

$$t_{i,e}^{in}\leq tp_{i,j,e}^{in}\quad i\in I,j\in I,e\in IE \tag{8}$$

$$tp_{i,j,e}^{in}\leq t_{j,e}^{p}+M2_{i,j,e}^{in}\cdot y_{i,j,e}^{in}\quad i\in H,j\in I,e\in E \tag{9}$$

$$t_{j,e}^{p}\leq tp_{i,j,e}^{in}\quad i\in H,j\in I,e\in E \tag{10}$$

$$tp_{i,j,e}^{in}\leq t_{j,e}^{p}-\Delta T_{e}^{min}+M2_{i,j,e}^{in}\cdot y_{i,j,e}^{in}\quad i\in C,j\in I,e\in E \tag{11}$$

$$t_{j}^{p}-\Delta T_{e}^{min}\leq tp_{i,j,e}^{in}\quad i\in C,j\in I,e\in E \tag{12}$$

$$tp_{i,j,e}^{out}\geq t_{i,e}^{out}\quad i\in I,j\in I,e\in E \tag{13}$$

$$tp_{i,j,e}^{out}\geq t_{j,e}^{p}\quad i\in H,j\in I,e\in E \tag{14}$$

$$tp_{i,j,e}^{out}\geq t_{j,e}^{p}-\Delta T_{e}^{min}\quad i\in C,j\in I,e\in E \tag{15}$$

$$t_{j,e}^{p}=t_{i,e}^{in}\quad i\in H,j\in I,e\in E \tag{16}$$

$$t_{j,e}^{p}-\Delta T_{e}^{min}=t_{i,e}^{in}\quad i\in C,j\in I,e\in E \tag{17}$$

$$g(x,w)=0 \tag{18}$$

$$h(x,w)\leq 0 \tag{19}$$

H and C are the sets of hot streams and cold streams, respectively. I is a set of process streams ($I=H\cup C$). E is a set of MHEXs. $f_{i,e}$, $CP_{i,e}$, $t_{i,e}^{in}$ and $t_{i,e}^{out}$ are molar flow rate, heat capacity, inlet temperature, and outlet temperature of process streams, respectively. $tp_{i,j,e}^{in}$ and $tp_{i,j,e}^{out}$ respectively are pseudo inlet and outlet temperatures of stream i at MHEX e regarding pinch candidate j. $qsoa_{j,e}$ is enthalpy of hot streams above pinch candidate j, and $qsia_{j,e}$ is enthalpy of cold streams above pinch candidate j. $t_{j,e}^{p}$ is pinch candidate. $y_{i,j,e}^{in}$ is binary variables to determine the value of pseudo inlet temperatures. $M1_{i,j,e}^{in}$ and $M2_{i,j,e}^{in}$ are sufficiently large numbers. ΔT_{e}^{min} is the minimum temperature difference for heat transfer.

Eq. (1) is the objective function to maximize overall heat recovery of MHEXs. Eq. (2) is the objective function to minimize operating cost of process. Eq. (3) represents energy balance of MHEXs, in which heat recovery obeys the conservation of energy. Eqs. (4)-(17) guarantee feasible heat transfer between hot and cold streams with temperature differences greater than or equal to ΔT_{e}^{min}. Eq. (4) ensures heat surplus above all pinch candidates. Eqs. (5) and (6) calculate the enthalpy of hot and cold streams above $t_{j,e}^{p}$ using the pseudo temperatures, which are determined by the multi-M constraints (Eqs. (7)-(15)). Eqs. (16) and (17) calculates pinch candidate $t_{j,e}^{p}$. Eqs. (18) and (19) denote the constraints of thermodynamics and unit operations (e.g. compressors and valves).

Here we use the calculation of pseudo temperatures of hot streams (Eqs. (7)-(10), (13), and (14)) to illustrate the multi-M formulation. Pseudo inlet temperature $tp_{i,j,e}^{in}$ is calculated via Eqs. (7)-(10). For $t_{i,e}^{in}\geq t_{j,e}^{p}$, $y_{i,j,e}^{in}$ must be equal to 1, and hence Eq. (7) is

reduced to $tp_{i,j,e}^{in} \leq t_{i,e}^{in}$ whilst Eq. (9) is relaxed into $tp_{i,j,e}^{in} \leq t_{i,e}^{in} + M1_{i,j,e}^{in}$. Together with Eq. (8), there is $tp_{i,j,e}^{in} = t_{i,e}^{in}$. Similarly, for $t_{i,e}^{in} \leq t_{j,e}^{p}$, we have $tp_{i,j,e}^{in} = t_{j,e}^{p}$. Pseudo outlet temperature $tp_{i,j,e}^{out}$ is governed by Eqs. (13) and (14) without using binary variable. The constraints ensure the hot streams' enthalpy above $t_{j,e}^{p}$ is not overestimated. Distinct from previous methods, the multi-M formulation avoids nonlinear formulation and reduces the number of binary variables, which are crucial to improve efficiency of an MINLP model.

3. Illustrative example

We use an ASU process to demonstrate the proposed modeling framework. Figure 1 illustrates a simplified flowsheet of the MHEXs in an industrial ASU. The process involves 2 MHEXs, 3 hot streams, 4 cold streams, 2 water coolers, 2 air compressors, 1 expansion turbine, and 1 expansion valve. Cooling of the ASU is delivered by the compressors, and is recovered by the MHEXs. The thermodynamics of the process is modeled based on SRK method. Process streams data are listed in Table 1. Composition of the process streams are presented in Table 2. The operating conditions of the units are specified in Table 3. The minimum temperature difference ΔT_e^{min} is equal to 3 K. Multi-objective optimization is achieved by ε-constraint method (Hwang and Masud, 1979). The model is implemented on GAMS 27.4 and solved by BARON.

The optimal solutions of the example are listed in Table 4. 2 points of interest, namely the maximum heat recovery solution and minimum operating cost solution, are presented. The results suggest that increasing outlet temperature and flow rate of H2 can increase heat recovery. However, the increase flow rate of H2 results in greater energy consumption of the compressors, which leads to higher operating cost.

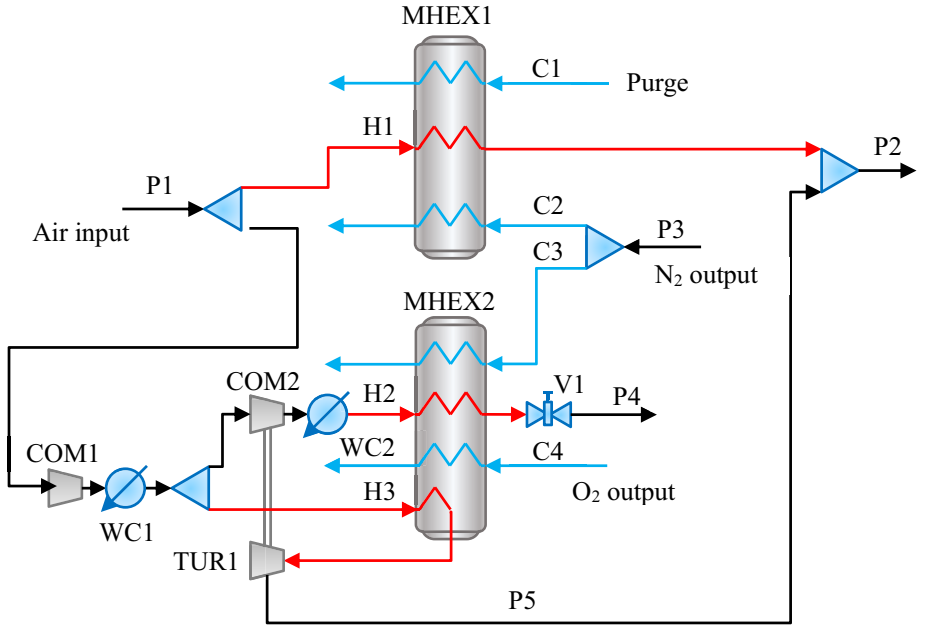


Figure 1 Simplified MHEXs of an industrial ASU.

Table 1 Data of process streams.

Stream	$t_{i,e}^{in}$ (K)	$t_{i,e}^{out}$ (K)	$f_{i,e}$ (mol/s)	$CP_{i,e}$ (J/mol·K)	$p_{i,e}$ (kPa)
H1 (g-m)	284.2	98.15-102.2	≤800.0	-	610.0
H2 (g-l)	313.2	98.15-102.2	≤800.0	-	6350
H3 (g-g)	313.2	-	≤800.0	30.94	4000
C1 (g-g)	97.70	-	40.00	30.36	590.0
C2 (g-g)	89.15	-	≤620.0	29.38	130.0
C3 (g-g)	89.15	-	≤620.0	29.38	130.0
C4 (l-g)	99.99	-	140.0	-	8500
P1 (g)	-	284.2	800.0	-	610.0
P2 (m)	-	98.15-102.2	≤800.0	-	610.0
P3 (g)	-	89.15	620.0	-	130.0
P4 (m)	-	98.15-102.2	≤800.0	-	610.0
P5 (m)	-	98.15-102.2	≤800.0	-	610.0

Table 2 composition of the process streams (mol %).

	P1	C1	P3	C4
N_2	78.85	65.27	96.00	1.00
O_2	21.15	34.73	4.000	1.000

Table 3 Operating condition of the units.

Unit	Outlet pressure (kPa)	Unit	Outlet temperature (K)
COM1	4000	WC1	313.2
COM2	6350	WC2	313.2
TUR1	610.0		
V1	610.0		

Table 4 Optimal solutions of the ASU example.

	Max heat recovery	Min operating cost
Heat recovery (kW)	5276	5170
Operating cost	6813	6372
H1 flow rate (mol/s)	331.3	325.1
H2 flow rate (mol/s)	191.7	185.3
H3 flow rate (mol/s)	277.1	289.6
H1 outlet temperature (K)	100.5	100.5
H2 outlet temperature (K)	97.20	96.70
H3 outlet temperature (K)	142.8	152.2
C2 flow rate (mol/s)	365.9	352.9
C3 flow rate (mol/s)	245.1	267.1
C1 outlet temperature (K)	251.2	253.5
C2 outlet temperature (K)	251.2	253.5
C3 outlet temperature (K)	310.1	298.5
C4 outlet temperature (K)	310.1	298.5

4. Conclusions

A modeling framework was developed for simultaneous MHEXs and process optimization. In this model, MHEXs were modeled effectively using the multi-M formulation (Hui, 2014) that allowed variable stream data, avoided nonlinear approximation, and reduced the number of binary variables. Moreover, the proposed framework can be integrated with rigorous thermodynamics method, which enables the framework to be compatible with various processes. The proposed framework was applied to a realistic ASU design problem with multiple MHEXs. The results showed that the framework successfully handled complex process optimization problem. Eventually the framework will be extended to the entire ASU flowsheet including distillation columns, which is able to consider trade-offs between compressor power consumption, oxygen purity, heat recovery, and investment of equipment.

Acknowledgement

The authors appreciate the financial support from the Hong Kong Research Grants Council General Research Fund (16211117), Guangzhou Science and Technology Project (201704030136), and the Studentship from the Energy Concentration program of the School of Engineering at the Hong Kong University of Science and Technology.

References

Christopher K, Dimitrios R. A review on exergy comparison of hydrogen production methods from renewable energy sources. Energy & Environmental Science 2012;5:6640-51.

Galeazzo FC, Miura RY, Gut JA, Tadini CC. Experimental and numerical heat transfer in a plate heat exchanger. Chemical Engineering Science 2006;61:7133-8.

Hasan M, Karimi I, Alfadala H, Grootjans H. Operational modeling of multistream heat exchangers with phase changes. AIChE J 2009;55:150-71.

Hui C. Optimization of heat integration with variable stream data and non-linear process constraints. Comput Chem Eng 2014;65:81-8.

Hwang CL, Masud ASM. Multiple Objective Decision Making, Methods and Applications: A State-of-the-art Survey. In Collab. with SR Paidy and K. Yoon. : Springer-verlag; 1979.

Kamath RS, Biegler LT, Grossmann IE. Modeling multistream heat exchangers with and without phase changes for simultaneous optimization and heat integration. AIChE J 2012;58:190-204.

Kemp IC. Pinch analysis and process integration: a user guide on process integration for the efficient use of energy. : Butterworth-Heinemann; 2011.

Liang Y, Hui C. A shortcut model for energy efficient water network synthesis. Appl Therm Eng 2016;96:88-91.

Memmott MJ, Wilding PR, Petrovic B. An optimized power conversion system concept of the integral, inherently-safe light water reactor. Ann Nucl Energy 2017;100:42-52.

Pattison RC, Baldea M. Multistream heat exchangers: Equation- oriented modeling and flowsheet optimization. AIChE J 2015;61:1856-66.

Shah R, Thonon B, Benforado D. Opportunities for heat exchanger applications in environmental systems. Appl Therm Eng 2000;20:631-50.

Thonon B. A review of hydrocarbon two-phase heat transfer in compact heat exchangers and enhanced geometries. Int J Refrig 2008;31:633-42.

Yeomans H, Grossmann IE. A systematic modeling framework of superstructure optimization in process synthesis. Comput Chem Eng 1999;23:709-31.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50057-4

An approach to optimal design of pressure-swing distillation for separating azeotropic ternary mixtures

Xu Huang[a], Yiqing Luo[a], Xigang Yuan[a,b*],

[a] *School of Chemical Engineering and Technology, Tianjin University, 300072, Tianjin, China*
[b] *State Key Laboratory of Chemical Engineering, Tianjin University, 300072, Tianjin, China*
yuanxg@tju.edu.cn

Abstract

The separation of ternary mixture with azeotropes by distillation into three high-purity products is difficult due to the existence of distillation boundaries. However, combinations of pressure swing distillation (PSD) configurations and recycle streams can provide alternatives for feasible pathways to the separation of the mixtures. In this work, a two-step approach is proposed to the optimal design of pressure swing distillation system to separate azeotropic ternary mixtures. An example of separating a ternary mixture with azeotropes of C2H5OH/C4H8O-01/C4H8O-02 is used for illustrating and validating the proposed method.

Keywords: ternary mixture; binary azeotrope; PSD; optimization;

1. Introduction

Several methods in industry are used to separate mixtures with azeotropic, including extractive distillation, azeotropic distillation, pressure swing distillation and relatively new membrane separation and so on (Luyben, 2012 and Soto, M, 2011, Xia et al. 2012). Application of extractive distillation or azeotropic distillation in industry is relatively mature. PSD is first proposed by Lewis (Lewis,1928), however, the pressure swing distillation from then has been mainly applied for separations of binary azeotropes where the composition of azeotrope is sensitive to pressure change, such as 2-butanone (MEK)-cyclohexane. Studying the separation of binary mixture (ethanol-water) separation, Knapp (Knapp,1992) mentioned that when the two components are in different distillation zones and the azeotropes are insensitive to pressure changes, a distillation boundary line, one end of which is sensitive to the pressure change can be formed by addition of a new component. However, such studies have been mainly focused in on binary azeotropic mixtures (Li et al. 2014, Li and Xu, 2017), few workers have addressed PSD for azeotrope containing ternary mixture separations (Yang and Gao, 2010).

In this paper, a approach to the conceptual design of PSD for separating a kind of azeotropic ternary mixture is proposed. If binary minimum-boiling azeotrope(s) is/are formed in a ternary mixture, and the composition of a binary azeotrope is sensitive to pressure, a three-column PSD process combined with two recycle streams can be designed by the method proposed in the following sections. The separation process of C2H5OH/C4H8O-01/C4H8O-02 ternary mixture will be used as an example to illustrate the method.

2. Generation of PSD flowsheet

2.1 Residual curve analysis

The residual curve map (RCM) at 101.3 kPa of C2H5OH/C4H8O-01/C4H8O-02 mixture is shown in figure 1. Three binary minimum azeotropes by each two of the components result in two distillation boundary lines. Three pure components to be separated are divided into three different distillation regions (as shown in Figure 1). As shown in figure 2, the composition of the C2H5OH/C4H8O-01 azeotrope is pressure sensitive. And thus, if the feed composition of the ternary mixture falls in Zone 2, it seems that a right pressure could be found so that the feed composition is on the C-D line, as shown in figure 3, and a distillation column, say column T1, can be used to separate the ternary mixture into the product of C4H8O-02 at the bottom of T1 and the azeotrope of the other two components at the top. However, if the position of feed locates in the remaining two zones, no matter how the pressure changes, no a feasible distillation exists to separate the ternary mixture into a pure component and a binary azeotrope mixture, as shown by the D-E and B-F lines. So, our design method is valid only when the feed composition locates in Zone 2.

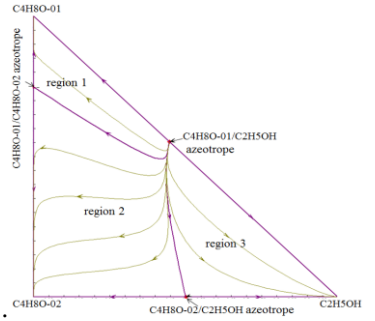


Fig. 1 RCM of the ternary mixture at 101.3 kPa

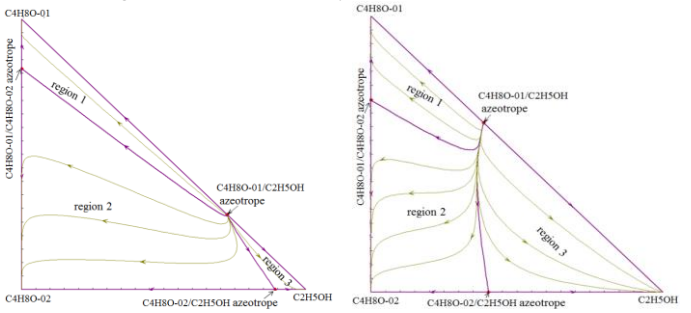


Fig. 2 RCMs of the mixture at different pressures: (a) 600 kPa; (b) 50 kPa

Figure 3 also implies that azeotrope composition at the opposite side of the triangle should be sensitive enough to pressure so that a C-D line can be found for a proper feed composition of the ternary mixture. Inspired by figure 3, we find that the top product can be introduced into a column, let say T2 that operates in a lower pressure so that the bottom product can be pure component B and at the top we can get an azeotrope with richer component A. This azeotrope can then be introduced to a column T3 with higher pressure to produce in the bottom the pure A and the top the azeotrope richer with B, which can be recycled to column T2. As a result, the ternary mixture is separated into three pure components by the three-column distillation system.

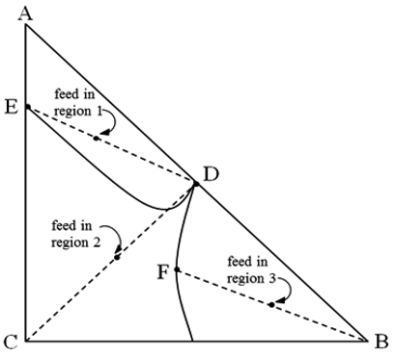


Fig. 3 Possible distillation separations for three kinds of feed compositions

2.2. Synthesis of three-column PSD process

Suppose that the feed flowrate is 10 kmol·h^{-1} with a temperature of 25 ˚C and a pressure of 101.3 kPa. The composition of feed and the product requirements are given in Table 1. It is found in our calculation that the cost of the distillation by column T1 depends on its pressure, and 600 kPa is found having the lowest cost by calculating its TAC and comparing with other four cases having different pressures of T1 column. The corresponding residue curve combination chart is shown in Figure 4. In this case, C2H5OH product (B2) should be added to the feed so that FEED moves to FEED1 that locates on the B1-D1 line as shown in Figure 4. The process chart of the three-column PSD obtained from Figure 4 is shown in Figure 5. The operating pressures for the three columns are 600 kPa, 50 kPa and 101.3 kPa, respectively. The feed stream needs to be pressurized to 600 kPa before mixed with part of C2H5OH product from T2.

Table 1 The composition of raw materials and product specifications

Components	Composition /%(mol)	Product specifications /%(mol)
C2H5OH	33	99
C4H8O-01	20	99
C4H8O-02	47	99

In this case, the corresponding distillation boundaries become the blue curves in Figure 4. The mixed feed stream FEED1 passes through T1 column, C4H8O-02 product is obtained at the bottom, i.e. B1 in Figure 4, and the C2H5OH/C4H8O-01 azeotrope (D1) is obtained at the top. After decompression to 50kpa, the stream of composition of D1 enters T2, the distillation boundaries after decompression becomes green line with in Figure 4. And D1 and C2H5OH product are in the same distillation zone, after separation in T2, the C2H5OH product is obtained at the bottom, i.e. B2 in Figure 4; Part of the product recycles back to form the feed to T1, and the rest is recovered as C2H5OH product. The top product of T2 of composition of D2 is introduced into T3 and separated at pressure 101.3 kPa into C2H5OH/C4H8O-01 azeotrope of composition D3 at the top of T3 and C4H8O-01 product at the bottom, and now, the corresponding distillation boundaries are red lines; then the D3 stream is depressurized to 50 kPa and then is recycled to column T2, to form the feed composition of FEED2 to column T2. Finally, the ternary mixture is effectively separated by three-column PSD process.

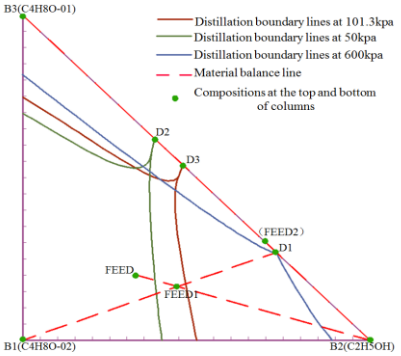


Fig. 4 Distillation boundaries at different operating pressures

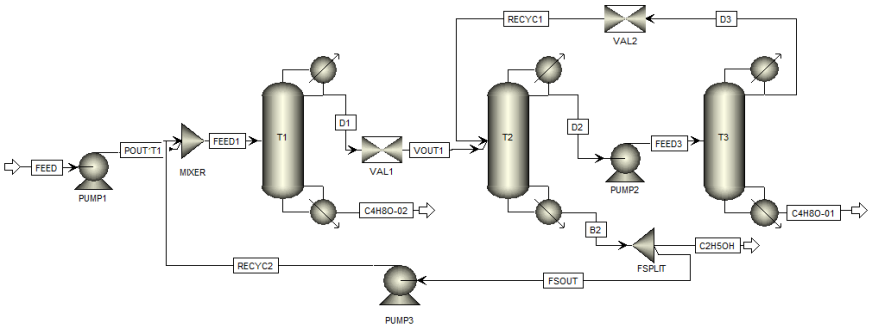


Fig. 5 PSD configuration for the ternary mixture separation

2.3 Selection of column pressures

The position of the original feed in the residue curve is changed by varying pressure of T1column and recycle flow rates of C2H5OH product, as shown in Figure 4 and then column T1 can separate C4H8O-02 product and C2H5OH/C4H8O-01 azeotrope. Figure 6 gives the corresponding distillation boundaries at 600 and 700 kPa respectively. And thus, the corresponding composition of the feed to T1 is FEED1 and FEED1*, respectively. In this way, six different pressures (P1=570, 600, 700, 800, 900 and 1000 kPa) for column T1 are tested to find the best one. It is found that when P1 is 600 kPa, the corresponding temperature of T1' bottom, 419K, attended the maximum temperature that can be provided by the medium-pressure steam used as the hot utility. To use less expensive cooling water as cold utility, P2 (pressure of T2) takes 50 kPa; the corresponding temperature of C2H5OH/C4H8O-01 azeotrope is 326 K. P3 (pressure of T3) is 101.3 kPa, meeting the conditions that the difference between azeotropic compositions at the top of T2 and T3 is greater than 5mol%.

3. Optimization results and discussion

In this paper, the total annual cost (TAC) as a criterion is estimated using the method suggested by Douglas (Douglas, 1988). In this section, column T1 at 600 kPa is optimized as an example to illustrate the optimization.

3.1 Optimization of the flow rate of RECYC2

Figure 7 gives the changes in purity of products at the top and bottom of column T1 with the recycle stream. It shows that when the flowrate of RECYC2 is 2.2 kmol•h^{-1}, the purity of C4H8O-02 product at the bottom of T1 reaches the maximum value.

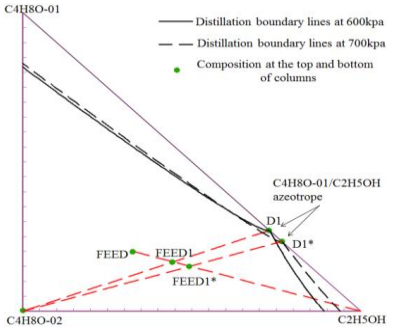


Fig. 6 RCMs of C2H5OH/C4H8O-01/C4H8O-02 at 600kpa and 700kpa

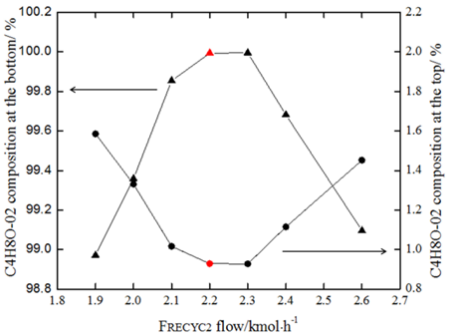


Fig. 7 The dependence of the T1 column product purities on the flow rate of RECYC2.

3.2 Optimization of T1 column

Figure 8 shows the relationship between the number of stages (NS1) in T1 and the reflux ratio (RR1). The value corresponding to where the number of column plates varies most significantly with the slope of reflux ratio curve is selected as the optimum number of theoretical stages, and the corresponding reflux ratio is taken as the optimum. As shown in Figure 8, NS1 is 97 and the corresponding RR1 is 28.6. As shown in Figure 9, the feed stage position (FS1) of T1 is 45.

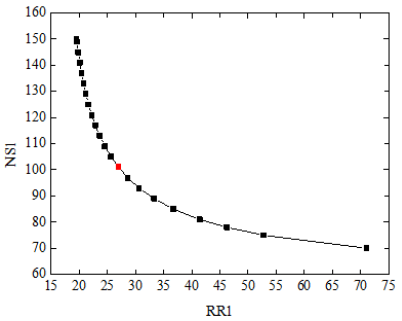


Fig. 8 Relationship of NS1and RR1 in T1column

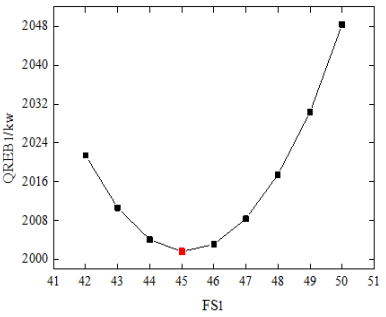


Fig. 9 The dependence of the reboiler heat duty on feed stage of T1 column

3.3 Optimization of columns T2 and T3

A method of sequential iterative optimization for T2 and T3 column is adopted in this paper. This method is widely used in many studies of optimization of PSD. Results of five case studies at different pressures of T1 are collected in Table 2. The heat duty of T1 is much greater than the sum of the heat duties of the other two columns, therefore, the difficulty in separating the mixture lies in the separation of C4H8O-02. It is found that when P1 is 600 kPa, the total annual cost (TAC) is 1025.2×103/$·y^{-1}, the lowest cost value, 6.27% lower than that when P1 is 800 kPa.

Table 2 Result of case studies at different pressures of column T1

Variables	**Case1**	**Case2**	**Case3**	**Case4**	**Case5**
P1/kPa	600	700	800	900	1000
NS1/ NS2/ NS3	97/45/50	102/43/50	95/55/67	104/54/60	95/55/55
FS1	45	79	71	80	70
FSD1/FSRECYC1	19/13	30/27	24/19	22/15	32/23
FS3	25	39	26	24	35

RR1/ RR2/ RR3	28.6/1/1.5	25.6/1.7/1.5	21.2/1.2/2	18.7/1.2/1.5	15.5/1.2/1.5
FRECYC1/kmol • h^{-1}	2.2	3.3	4.5	6.0	8.8
QREB1/kW	2002.1	2052.5	1944.6	1989. 8	2089.7
QREB2+QREB3/kW	650.9	641.2	660.2	629.4	576.0
TAC×10^3/$ • y^{-1}	1025.2	1094.0	1089.5	1102.7	1104.4

4. Conclusions

Using the residue curves under different pressures, a method to synthesis a PSD system separating a kind of ternary mixtures with azeotropes is proposed. Separation of mixture C2H5OH/C4H8O-01/C4H8O-02 is used to illustrate the method. If the feed composition falls in a distillation area containing C4H8O-02 product, separation of three components can be achieved. A three-column PSD process with two recycle streams has been proposed to have achieved effective separation of the mixture. Furthermore, the operating pressure of T1 not only determines the economy of the column itself, but also determines the flowrate of the RECYC2, thereby affecting the economy of the overall process. By choosing five cases with different pressures of T1, this paper has studied the influence of P1 on the economy, with the optimized calculation results shown in Table 2.

Acknowledgements

Support by National Natural Science Foundation of China (Grant 91434204) is acknowledged.

References

Luyben, W. L. (2012) Pressure-swing distillation for minimum- and maximum-boiling homogeneous azeotropes. Ind. Eng Chem. Res. 51: 10881−10886.

Soto, M.; Tlacuahuac, F. (2011) Modeling, simulation and control of an internally heat integrated pressureswing distillation process for bioethanol separation.Computers and Chemical Eng., 35: 1532–1546.

Lewis, W. K.(1928) Dehydratkg'Alcohol and the Like.[P] US. 1676,700

Knapp J P, Doherty M F. (1992) A new pressure-swing-distillation process for separating homogeneous azeotropic mixtures[J]. Industrial & engineering chemistry research, 31(1): 346-357.

Douglas J M. (1988) Conceptual design of chemical processes[M]. McGraw-Hill Science/Engineering/Math.

Li W, Shi L, Yu B, et al.(2013) New pressure-swing distillation for separating pressure-insensitive maximum boiling azeotrope via introducing a heavy entrainer: design and control[J]. Industrial & Engineering Chemistry Research, 2013, 52(23): 7836-7853.

Li Y, Xu C.(2017) Pressure-Swing Distillation for Separating Pressure-Insensitive Minimum Boiling Azeotrope Methanol/Toluene via Introducing a Light Entrainer: Design and Control[J]. Industrial & Engineering Chemistry Research, 2017, 56(14): 4017-4037.

Xia S, Wei W, Liu G, et al.(2012) Pervaporation properties of polyvinyl alcohol/ceramic composite membrane for separation of ethyl acetate/ethanol/water ternary mixtures[J]. Korean Journal of Chemical Engineering, 2012, 29(2): 228-234.

YANG D M， GAO Xiao-Xin.(2010) Simulation for Separation of Water-Isopropanol-Diiospropylamine by Heat Integration Pressure Swing Distillation, The Chinese Journal of Process Engineering, 2010, 10(2) 327-331

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50058-6

Conceptual design of an efficient Hydrogen production process from Natural Gas using an extension to the "G-H" methodology

Avinash Shankar Rammohan Subramanian[a], Rahul Anantharaman[b], Truls Gundersen[a*]

[a]*Department of Energy and Process Engineering, Norwegian University of Science and Technology (NTNU), NO-7491 Trondheim, Norway*
[b]*SINTEF Energy Research, Sem Sælandsvei 11, NO-7465, Trondheim, Norway*
truls.gundersen@ntnu.no

Abstract

Conceptual design of novel efficient hydrogen production processes is essential to realizing a future clean energy scenario. Traditional approaches to process design involve detailed modelling of unit operations followed by simulation and optimization steps. However, several implicit assumptions may be made in the detailed models and these limit the efficiency of the overall process. In addition, potentially efficient novel technologies may not be considered. In this paper, fundamental thermodynamic principles are used within the framework of the "G-H" methodology (G denotes Gibbs free energy and H denotes enthalpy) for synthesis of an efficient hydrogen production process. The optimal chemical route is determined and termed the "1-step reaction". However, the G-H methodology assumes complete conversion, which may not be optimal or even feasible for all reactions. Thus, this paper extends the G-H methodology to make it more realistic by including conversion as a variable. This methodology is then applied to a case study of flowsheet synthesis that uses separation and recycle to achieve 80% conversion. The flowsheet is simulated using Aspen HYSYS and an exergy analysis is performed. An overall exergetic efficiency of 81.8% is achieved and this represents an ideal target to motivate future technology improvements.

Keywords: Conceptual design, Hydrogen production, Process synthesis.

1. Introduction

Hydrogen is a clean fuel and is thus expected to play an important role as an energy carrier in a future decarbonized energy scenario. However, global hydrogen production currently is dominated by fossil-fuel based processes with significant inefficiencies. Conceptual design of novel efficient processes is essential to realizing a hydrogen-based energy economy. Traditional approaches to process design have focused on detailed modelling and optimization of unit operations (such as reactors, distillation columns and heat exchanger networks) followed by assembling these into a flowsheet. The interaction between the different units is not emphasized implying the overall flowsheet may be inefficient even with optimized individual unit operations. More recently, superstructures are generated of unit operation alternatives, followed by simulation and optimization using mathematical programming techniques. However, the optimal solution is only as good as the superstructure considered (Hildebrandt et al., 2015), and potentially efficient processes may be excluded. This is especially true for novel technologies since

superstructures tend to only include well-understood mature technologies. Furthermore, several implicit assumptions (such as chemical routes, temperature and pressure ranges) may be made in the detailed models of the individual unit operations and thus may limit the efficiency of the overall process.

An alternative approach is to use fundamental thermodynamic principles for process design through targeting and synthesis steps. Patel et al. (2007) developed a methodology based on mass, energy and entropy balance perspectives to target for the optimal chemical route while making as few assumptions as possible. Fox et al. (2013) developed the systematic "G-H" methodology, which uses information about the Gibbs free energy (G) and enthalpy (H) for process synthesis to derive the operating conditions (temperatures, pressures, stream flow rates) that result in highest thermodynamic efficiency. It is assumed that complete conversion can be achieved in all the equilibrium-limited reactions. Hildebrandt et al. (2015) illustrated the value of this methodology using a case study of a coal-to-liquids process. The key principle is to begin with a thermodynamically ideal process target and synthesize flowsheets that approach this target. Then, the unit operations of the flowsheet are mapped onto existing technologies. However, there are usually no existing technologies to achieve the ideal process target. In particular, it may not be optimal or even feasible to achieve complete conversion. Thus, this paper extends the G-H methodology to include conversion as a variable such that the resulting flowsheet is realistically achievable. No restrictions are made on the specific method used to achieve this conversion. The extended G-H methodology is then illustrated with a case study of flowsheet synthesis for a single-step hydrogen production process.

2. Methodology and Application to Hydrogen production process design

2.1. Chemical Route Targeting

The target of the overall process is to convert as much feed to useful product(s) as possible while satisfying material balance constraints. Patel et al. (2007) developed a methodology to determine the optimal chemical route that satisfies this target. The methodology is applied in this work to determine the best chemical route that converts a unit mole of methane to the maximum amount of hydrogen. Water and carbon dioxide are chosen as reacting species (Eq.(1)) and the material balance constraints are presented in Eq.(2).

Solving Eq.(2) gives the reaction presented in Eq.(3), hereafter referred to as the "1-step reaction" because it combines the well-known Steam-Methane Reforming (SMR), Water-Gas Shift (WGS) and steam generation processes. The technological feasibility of the 1-step reaction has been experimentally demonstrated by Malerød-Fjeld et al. (2017). A $BaZrO_3$-based proton-conducting electrolyte deposited as a dense film on a porous Ni composite electrode was used as the catalyst. Lee et al. (2007) also presented a process concept in which the SMR and WGS reactions were carried out in the same unit operation. The 1-step reaction requires heat and work in order to proceed. Thus, heat integration

$$1\mathrm{CH_4(g)} + \nu_1\mathrm{H_2O(l)} + \nu_2\mathrm{CO_2(g)} \rightleftharpoons \nu_3\mathrm{H_2(g)} \tag{1}$$

$$1 + \nu_2 = 0; \quad 4 + 2\nu_1 = 2\nu_3; \quad \nu_1 + 2\nu_2 = 0 \tag{2}$$

$$\mathrm{CH_4(g)} + 2\mathrm{H_2O(l)} \rightleftharpoons \mathrm{CO_2(g)} + 4\mathrm{H_2(g)} \quad (\Delta H_{1-step} = 252.95\,\mathrm{kJ/mol},\ \Delta G_{1-step} = 130.69\,\mathrm{kJ/mol}) \tag{3}$$

$$\begin{aligned} &CH_4(g) + 2O_2(g) \rightleftharpoons CO_2(g) + 2H_2O(g) \\ &(\Delta H_{comb} = -802.35\,\text{kJ/mol}, \Delta G_{comb} = -800.71\,\text{kJ/mol}) \end{aligned} \quad (4)$$

with an exothermic process is necessary. This heat could be provided by methane combustion (Eq.(4)) as is done in standard SMR processes today.

2.2. Thermodynamics of Chemical reactions

The G-H methodology uses the results (Eq.(5)) that a reaction's heat (Q) and work (W) requirements can be derived from the 1st and 2nd laws of thermodynamics. Work can be added by compression, separation or together with heat (Patel et al., 2010). In this paper, the reactor work requirements are supplied together with heat. For the work requirements to be exactly supplied together with heat, the reactor has to be run at a particular temperature called the "Carnot temperature", T_{Carnot} (Patel et al., 2010). Eq.(6) gives the relationship between a reaction's heat and work requirements and its Carnot temperature. Patel et al. (2010) demonstrated that reactors running at their Carnot temperature are reversible and hence are thermodynamically efficient. The Carnot temperatures of the 1-step reaction and the combustion reaction are 616.9 K and 145,866.3 K respectively.

$$Q = \Delta H_{reaction} \quad ; \quad W = \Delta G_{reaction} \quad (5)$$

$$W_{comb} = Q_{comb}\left(1 - \frac{T_0}{T_{Carnot}}\right) \quad (6)$$

The Carnot temperature of the combustion process is not technologically feasible, thus the process has to be run at a lower temperature. This deviation introduces irreversibilities implying that the methane combustion reaction is inherently inefficient. A lower operating temperature of 626.9 K is chosen such that the 1-step reactor can be placed inside the furnace with $\Delta T_{\min}$ of 10 K.

2.3. Integrating equilibrium-limited chemical reactions

In order to synthesize a process flowsheet, only the input and output parameters (or the transfer characteristics that relate the output to input parameters) of the different unit operations are relevant. In an equilibrium reactor, conversion is the relevant transfer function and thus an important variable. Equilibrium-limited reactions running at their Carnot temperatures do not necessarily achieve complete conversion. Approaches used to shift the equilibrium to the right include continual separation and removal of products using membrane reactors (Malerød-Fjeld et al., 2017) or sorption-enhanced reactors (Lee et al., 2007), and separation and recycle of unreacted feed. It is commonly not optimal or even technologically feasible to achieve complete conversion. Thus, the G-H methodology has to be extended to account for incomplete conversion as illustrated next.

For a stoichiometric feed, the change in the number of moles of each component is proportional to the stoichiometric coefficient; the extent of reaction (e) is the proportionality constant. Conversion (c) is the fraction of the stoichiometric feed that reacts. Thus, the extent of the reaction is given by: $e = f * c$. f is termed the "multiple of the stoichiometric feed" since the molar flow rate of each reacting component can be determined by multiplying f with the corresponding stoichiometric coefficient. The rationale for using f is to have a single number from which the feed flow rates can be

$$\begin{pmatrix} \Delta H_{overall} \\ \Delta G_{overall} \end{pmatrix} = \begin{pmatrix} \Delta H_{1-step} \\ \Delta G_{1-step} \end{pmatrix} \left(f_{1-step} * c_{1-step} \right) + \begin{pmatrix} \Delta H_{comb} \\ \Delta G_{comb} \end{pmatrix} \left(e_{comb} \right) \quad (7)$$

determined for flowsheet synthesis. With this definition, the overall process thermodynamics (Eq.(7)) can be determined by adding the heat and work requirements of each reaction scaled by the corresponding extent of reaction (with the substitution $f * c$ made for the equilibrium-limited 1-step reaction).

In the G-H methodology (Fox et al., 2013), design decisions are made to determine the unknowns in Eq.(7) ($\Delta H_{overall}, \Delta G_{overall}, f_{1-step}, c_{1-step}, e_{comb}$). These design decisions include: Requiring no wastage of methane (i.e. $e_{comb} = 1 - f_{1-step} * c_{1-step}$), requiring no loss of excess heat irreversibly to the environment (i.e. $\Delta H_{overall} = 0$), and assuming that a conversion target (c_{1-step}) of 0.8 is technologically feasible. Eq.(7) can be solved to give f_{1-step} of 0.95. The conversion target is a variable that can be set to any value. Determining the optimal conversion target requires insight into the specific conversion enhancement method and is as an important task for future work. The flow rates of all streams are derived from f_{1-step} and presented in Figure 1. $\Delta G_{overall}$ is -1.50 kJ/mol implying that the combined process is feasible. Reactor pressures are set such that all the available work can be extracted (Fox et al., 2013). The net work is the difference between the work supplied by expansion of products from the reactor pressure and the work required for compression of feed to reactor pressure. In this case study, no work is extracted since $\Delta G_{overall}$ is negligible thus the pressures of the two reactors are both 1 bar. The proposed flowsheet was simulated using Aspen HYSYS v9.0. Conversion of 80 % is achieved using separation and recycle of unreacted feed. The unreacted methane from the 1-step process is fed to the combustion process. An exergy analysis is then carried out to evaluate process performance.

3. Results and Discussion

Table 1 shows that the furnace destroys the most exergy as expected. The methane combustion reaction is inherently inefficient because its Carnot temperature is not practically achievable. This implies that future design efforts should focus on lowering the extent of the combustion reaction or eliminating it altogether by finding alternative heat sources. An interesting concept was suggested by Adams and Barton (2011) who proposed using the exothermic coal gasification process as a heat source. Other alternatives arise from novel process intensification technologies. For instance, the Protonic Membrane Reactor presented by Malerød-Fjeld et al. (2017) uses the heat generated from the galvanic operation of the membrane. The endothermic 1-step reaction and the exothermic processes are spatially coupled resulting in microscale heat integration. If the 1-step reactor is co-located with other industrial process systems, heat integration with background hot streams may also be possible. Herein lies one of the advantages of the 1-step reaction compared to the SMR or WGS reactions. The Carnot temperature of the 1-step process (616.9 K) is significantly lower than the Carnot temperatures of the SMR (960.8 K) or the WGS (974.7 K) reactions implying there is a greater chance of finding background process streams for heat integration.

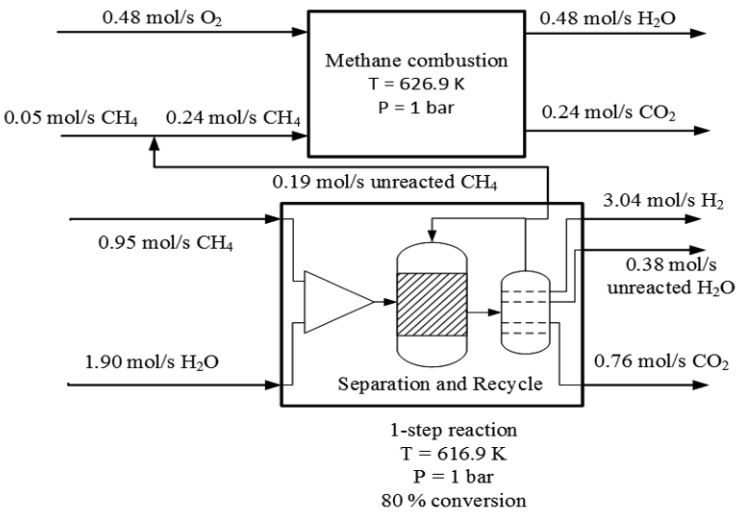


Figure 1: Flowsheet for hydrogen production using the 1-step reaction

Table 1: Exergy analysis of hydrogen production process

Component	Exergy Destruction (%)
Feed mixers	4.1
1-step reactor	6.3
Combustion furnace	65.6
Separation processes	24.0

The separation process, that includes both separation of the product stream into pure components and separation of the unreacted feed for recycle, also destroys a significant amount of exergy. Thus, there is a trade-off between the chosen conversion target and the exergetic efficiency: Larger amounts of exergy are destroyed in the separation process because of higher recycle to achieve a higher conversion. This implies an optimization step is essential to determine the conversion target.

The overall exergetic efficiency ($\frac{\text{Total Useful Exergy Out}}{\text{Total Exergy In}}$) is 81.79%. This represents the upper-bound attainable because of two reasons. First, while the flowsheet in Figure 1 is technologically feasible, it corresponds to the conceptual phase. Detailed modelling of the unit operations inevitably introduces irreversibilities resulting in lower exergetic efficiency (Anantharaman et al., 2013). Second, while the flowsheet represents a thermodynamically optimal process, it may not necessarily be economically optimal. Detailed modelling and optimization with respect to net present value may result in a different design with lower exergetic efficiency.

4. Conclusions

This paper extends the G-H methodology to make it more realistic by including conversion as a variable. This extended methodology is then applied to a case study of flowsheet synthesis for a hydrogen production process that uses separation and recycle as a conversion enhancement method. A conversion target of 0.8 was chosen but this could

be optimized. Determining the optimal conversion target is an important task that requires insight into the specific conversion enhancement technology. Current work by the authors focuses on developing a methodology to estimate the conversion achievable in a membrane reactor at the conceptual stage. The optimal chemical route, termed the 1-step reaction, is determined and validated by existing technologies. The 1-step reaction intensifies the SMR, WGS and the steam generation processes into a single unit operation, thus showing potential to lower capital and operating costs as well as to reduce the number of units. In addition, the 1-step reaction can be run efficiently at a lower temperature than the SMR or WGS reactions, thus increasing the chance of finding suitable background streams for heat integration. The methane combustion reaction destroys the most exergy, thus future design efforts should focus on utilizing alternative heat sources. Current work by the authors focuses on developing polygeneration strategies that include additional feedstocks with exothermic processing steps.

The key principle behind the extended G-H methodology is to begin with a thermodynamically ideal process target and synthesize a feasible flowsheet that approaches this target. Then, the unit operations of the flowsheet are mapped onto existing technologies. If there are no existing technologies, this highlights an improvement potential to motivate the development of novel technologies (Anantharaman et al. 2013). Thus, this approach uses fundamental thermodynamic principles applied at the systems level to influence where to focus technology development efforts rather than the opposite approach of generating a flowsheet based on existing technologies. Therefore, this paper presents a novel and more realistic thermodynamic approach to process synthesis and design.

References

T. A. Adams, P. I. Barton, 2011, Combining coal gasification and natural gas reforming for efficient polygeneration, Fuel Processing Technology 92 (3), 639–655.

R. Anantharaman, K. Jordal, D. Berstad, T. Gundersen, 2013, The role of process synthesis in the systematic design of energy efficient fossil fuel power plants with CO_2 capture, Chemical Engineering Transactions 35.

J. A. Fox, D. Hildebrandt, D. Glasser, B. Patel, 2013, A graphical approach to process synthesis and its application to steam reforming, AIChE Journal 59 (10), 3714–3729.

D. Hildebrandt, D. Glasser, B. Patel, B. C. Sempuga, J. A. Fox, 2015, Making processes work. Computers & Chemical Engineering 81, 22–31.

K. B. Lee, M. G. Beaver, H. S. Caram, S. Sircar, 2007, Novel thermal-swing sorption-enhanced reaction process concept for hydrogen production by low-temperature steam-methane reforming, Industrial & Engineering Chemistry Research 46 (14), 5003–5014.

H. Malerød-Fjeld, D. Clark, I. Yuste-Tirados, R. Zanón, D. Catalán-Martinez, D. Beeaff, S. H. Morejudo, P. K. Vestre, T. Norby, R. Haugsrud, J. M. Serra, C. Kjølseth, 2017, Thermo-electrochemical production of compressed hydrogen from methane with near-zero energy loss. Nature Energy, 1-9.

B. Patel, D. Hildebrandt, D. Glasser, 2010, An overall thermodynamic view of processes: Comparison of fuel producing processes, Chemical Engineering Research and Design 88 (7), 844–860.

B. Patel, D. Hildebrandt, D. Glasser, B. Hausberger, 2007, Synthesis and integration of chemical processes from a mass, energy, and entropy perspective, Industrial & Engineering Chemistry Research 46 (25), 8756–8766.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50059-8

Optimized sustainable molecular and purification process design framework: acetone-butanol-ethanol case study

Jaime D. Ponce-Rocha, Eduardo Sánchez-Ramírez,
Juan G. Segovia-Hernández, Fernando I. Gómez-Castro,
Ricardo Morales-Rodriguez*

Departamento de Ingeniería Química, División de Ciencias Naturales y Exactas, Campus Guanajuato, Universidad de Guanajuato, Noria Alta S/N, Guanajuato, Gto. 36050, México.
ricardo.morales@ugto.mx

Abstract

The design of optimal sustainable processes must be performed relying on systematic steps, which must guide the users to take the best decision in terms of given objectives. This study proposes and implements a framework consisting of organized steps combining methods and tools, using as a case study the separation and purification of the acetone-butanol-ethanol (ABE) mixture obtained from a fermentation broth. A feasible configuration to that end includes the use of liquid-liquid extraction (LLE) columns followed by conventional distillation columns. Thus, the framework includes a molecular design step to tailor-made the extractant agent employed in the LLE unit with the desired properties, the process design of the required equipment for the ABE separation, and finally an optimization step using a hybrid stochastic optimization method, differential evolution with taboo list. The results allowed to reduce the total annual cost by 67 % and the environmental impact by 72 % relying on the eco-indicator.

Keywords: ABE, Eco-Indicator, liquid-liquid extraction, molecular design, heptyl acetate.

1. Introduction

The lignocellulosic butanol has been identified as a potential biofuel since it has showed some advantages compared with bioethanol such as: better blending with gasolines, it has higher energy content, it is less hydrophilic, etc. The production of butanol relying on a fermentation process has been done employing bacteria from the genera Clostridia. The possible advantage of this fermentation is that the bacteria also produce simultaneously acetone and ethanol, which firstly make consider that this could be a multiproduction process (Morales-Rodriguez et al, 2014); however, the simultaneous acetone-butanol-ethanol (ABE) production increases the complexity in the separation of the products, specially, the butanol. This has originated that several studies have aimed to find the best separation and purification configuration, for example, employing conventional distillation, extractive distillation and hybrid processes including liquid-liquid extraction (LLE) and distillation columns (DC) configurations (Kraemer et al., 2010). Errico et al. (2015) proposed 9 alternative hybrid process configurations, which showed an interesting variety of technologies to achieve his separation, for instance, hybrid extraction, thermally coupled alternatives, intensified configurations, etc.

A literature survey allowed identifying the possible process candidates for the ABE separation and purification. Previous works focused on biobutanol purification process have shown some benefits in terms of cost and controllability (Errico et al., 2015). Nevertheless, the first step for the LLE is the identification of the extractant, which is commonly identified and selected employing an experimental trial and error approach; thus, generating some drawbacks for the efficient process design, making it expensive in term of time and economic aspects. Moreover, the process design task not necessarily would provide the optimal process configuration, therefore, an optimization effort is necessary to find the best process design. Thus, the use of a systematic methodology including diverse methods and computer-aided tools could assist to overcome those issues, thereby, the objective of this work is to propose a framework for optimal molecular and sustainable process design to separate and purify the acetone-butanol-ethanol mixture employing a LLE system and conventional DC.

2. Optimized sustainable molecular and process design framework: LLE and DC approach

The framework for optimized molecular and sustainable process design includes 7 steps: 1) the data collection and state of art analysis related with LLE; 2) the molecular design task, aiming to find the potential extracting agents with tailor-made characteristics, employing a computer-aided molecular design (CAMD) (Harper et al., 1999); 3) selection of the extractants based on the cost, market availability and toxicity, and the thermodynamic model suitable for the mixture to separate; 4) evaluation of extractant candidates in a simulation tool (Aspen plus, PRO II, etc); 5) comparison of the evaluated extracting agent with the base case (BC) in order to determine its performance and continue with the design of the subsequent equipment, whether the performance of the candidate is worse than the BC scenario in terms of selected criteria, it is possible to return to step number 3 to select another candidate, otherwise, just continue with the next step; 6) process design employing shortcut methods followed by a sensitivity analysis to refine the process specification, and finally, the selection of configurations with better performance based on product recovery, purity and operating cost; 7) multiobjective optimization employing differential evolution (DE) and taboo list (TL) methods, minimizing the total annual cost (TAC) and the environmental impact (EI) calculated with the Eco-indicator 99; and finally the selection of the best configuration based on a Pareto front.

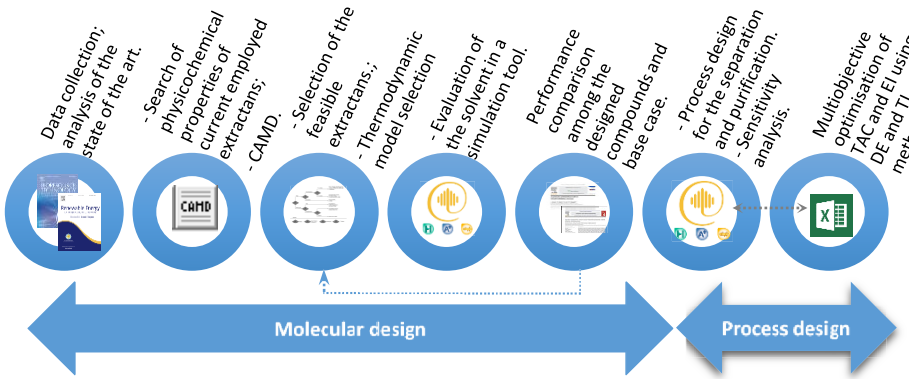


Figure 1. Optimized molecular and sustainable process design framework.

3. Implementation of systematic framework for molecular and sustainable process design: ABE separation from fermentation broth

The framework used as case study the separation of a fermentation broth containing ABE, water, carbon dioxide and hydrogen, since the residual sugars such as glucose and xylose are separated in distillation column before entering to the LLE unit.

3.1. Data collection and state of the art analysis

The implementation of LLE includes the selection of extractant agent. Kraemer et al. (2010) employed 1,3,5-trimetilbenzene (mesitylene). Moreover, Morales-Espinosa et al. (2017) employed hexyl acetate (HEX) as extractant agent having better results than using mesitylene. In any case, there is still necessary to search an environmentally friendlier and economic extractant agent substitute.

3.2. Computer-aided molecular design

The potential extractant agent can be designed relying on CAMD principles. Thus, it is necessary to list the desired characteristics of the reference extractant (mesitylene), such as, partition coefficient, water solubility, vapour pressure, not azeotrope formation with the present compounds, etc. That information is specified in the Pro-CAMD that is available in the Integrated Computer Aided System (ICAS). The Pro-CAMD designed a number of candidates that were reduced by manipulating the lower and upper limit values of the desired characteristics, ending with 846 compounds.

3.3. Selection of the potential candidate to be used as extractant and the thermodynamic model

The discrimination to determine the final list of potential candidates was obtained checking the existence or availability of the compound in the market, and also the commercial prices compared with the reference compound. The heptyl acetate (HEP) as well as the HEX (previously analysed by Morales-Espinosa et al., 2017) were found as the most promising substitutes for the mesitylene. Thus, in this study the HEP was chosen to be evaluated. Since all the involved compounds were known, the NRTL and Hayden O'Connell thermodynamic models were selected to describe the liquid-vapour equilibrium, respectively.

3.4. Evaluation of the potential extractant candidate

The HEP and mesitylene were tested in a LLE unit with the same number of stages (5), process conditions (T = 35 °C and P = 1 atm) and molar-based feeding ratio (1:0.84, mixture and extractant, respectively).

3.5. Comparison of the potential candidate vs base case extractant

The results employing HEP had similar ABE recovery. In addition, the commercial price of mesitylene compared with HEP was 2.62 times higher, thereby, finding the HEP as a better candidate in terms of economics and performance.

3.6. Process design for separation and purification.

The process configuration is shown in Figure 2, which was design relying in heuristic rules, then the equipment design was done using shortcut methods for the DC and a sensitivity analysis to improve the LLE design and the DC. Once, the design is gotten a rigorous simulation is performed using EXTRACT unit for LLE and RADFRAC for the three DC in Aspen plus (AP). The LLE unit (C-101) is employed to remove the water from the fermentation broth and the columns D-101 and D-102 are used to purify the HEP and butanol, respectively. The D-103 unit separates the acetone, hydrogen and carbon dioxide from the ethanol and water traces.

Mass fraction 101	
wA	0.0965
wB	0.1950
wE	0.0316
wCO_2	0.0058
wH_2	0.0001
wH_2O	0.6710
Flow rate (kmol/h)	299.091

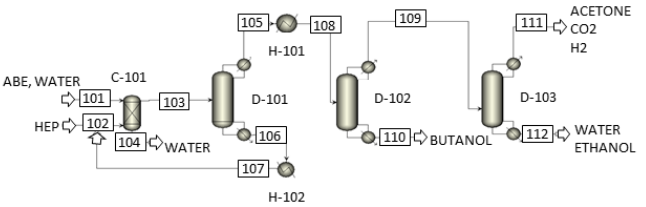


Figure 2. Process configuration for the ABE separation using HEP as extractant agent.

3.7. Multiobjective optimization

3.7.1. Objective Function

The objective function included the simultaneous minimization of the TAC and the eco indicator 99. The minimization of these objectives was subject to the required recoveries and purities in each product stream (see Equation 1).

$$Min\left(TAC, Eco\ 99\right) = f\left(N_{tn}, M_{in}, N_{fn}, R_{rn}, F_{rn}, D_{cn}\right),\ subject\ to\ y_m \geq x_m \tag{1}$$

where N_{tn} are total column stages in LLE column and DCs, M_{in} is the feeding flowrate of extractant to the LLE, N_{fn} is the feed stage of all streams in columns, R_{rn} is the reflux ratio, F_{rn} is the distillate fluxes and D_{cn} is the column diameter, y_m and x_m are vectors of obtained and required mass purities (95%, 99.5% and 83% for ABE) with the mass recoveries (98%) for the m components, respectively. This minimization implied the manipulation of 18 continuous and discrete variables. The TAC calculation was performed as it is shown in Equation 2.

$$TAC = \frac{Capital\ Cost}{Payback\ period} + Operative\ cost \tag{2}$$

On the other hand, the environmental indicator was measured through the eco-indicator 99 based on the methodology of the life cycle analysis (Geodkoop and Spriensma, 2001). In the eco-indicator 99 methodology, 11 impact categories were considered, aggregated into three major damages categories: (1) human health, (2) ecosystem quality, and (3) resources depletion.

3.7.2. Global stochastic optimization strategy

To optimize the process route for biobutanol production, a stochastic optimization method, Differential Evolution with Taboo List (DETL) was used, which have shown being robust to optimize intensified separation systems. Srinivas and Rangaiah (2007) showed that the use of some concepts of the metaheuristic taboo can improve the performance of DE algorithm. The implementation of this optimization approach was made using a hybrid platform where the DETL was coded using Microsoft Excel (ME) and the separation process was rigorously simulated using AP. The vector of decision variables (i.e., the design variables) is sent from ME to AP using DDE (Dynamic Data Exchange) through COM technology. For the optimization of process routes analysed in this study, the following parameters for DETL method were used: 200 individuals, 300 generations, a taboo list of 50% of total individuals, a taboo radius of 0.0000025, 0.80 and 0.6 for crossover and mutation fractions, respectively.

The results of the optimization are illustrated in Figure 3. There are various possible combinations that could be selected as optimal as it is shown inside the oval. For example, the optimal value could be the point (a). However, there are other combinations with lower TAC (point b and c), but higher eco-points. Comparing the eco-points values for point (a) with points (b) and (c), the differences are higher by 0.7%. and 1.21%, respectively; with a difference of 9,609 points/year for point (b) and 16,632 points/year for point (c) compared with the 1,353,369 points/year for point (a). In addition, the TAC was decreased by 5.5% when comparing point (a) and (b). Therefore, the optimal design is the point (b) with TAC = 4,481,469 US$/year and eco indicator = 1,362,978 points/year. The comparison between the optimal scenarios for TAC and eco-points with the base case (BC) scenario, indicated that the base case (point d) was 3 times higher for TAC and 3.6 for the eco-indicator, thereby, the optimization task undoubtedly achieved its objectives. Table 1 shows the optimal (Opt) specifications compared with the BC scenario. The results showed the D-101 and D-102 units are shorter, and the reflux ratio was highly decreased due to the optimization implementation; on the other hand, it was found for D-103 that it is more convenient having a larger unit operation, but whether the reflux ratios are compared, the BC is 11.72 times higher than optimal. For all the distillations columns, the reduction in the reflux ratio results in a reduction of the operation cost because of the heat duty decrease.

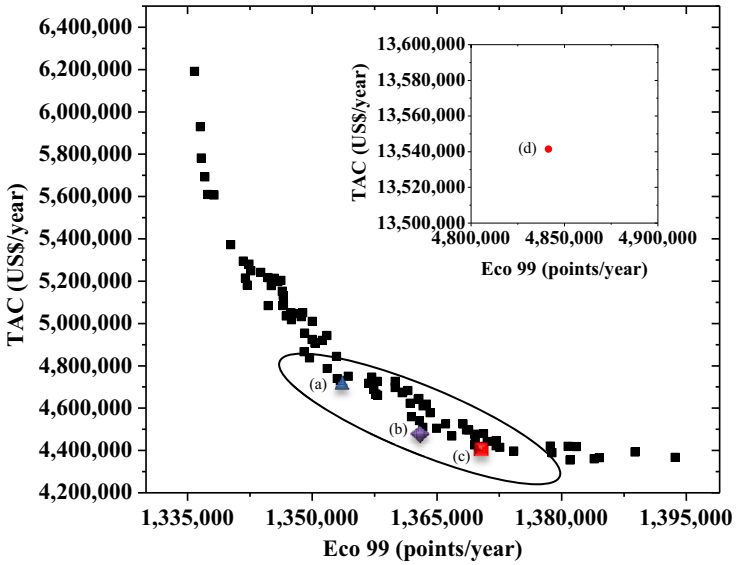


Figure 3. Pareto Chart for determining the optimal process design in terms of economic and environmental aspects.

4. Conclusions

This study proposed a framework for optimized molecular and sustainable purification process design, which was possible due to the combination of different methods and computer-aided tools. The results showed an improved design by reducing the TAC and the eco-indicator 67% and 72%, respectively. The enhancement was achieved due to the appropriate molecular design that provided a cheaper candidate with the similar

performance as some previous extractant agents; in addition with the improvement of the process design by finding the proper unit characteristics relying on the DETL method.

Table 1. Specifications of the unit operations included in the downstream section.

Unit	Base case	Optimal design
C-101	NS=5; P=1 atm; ABE/HEP =1.20; TT= 27.5°C; BT= 34°C	NS=5; P=1 atm; ABE/HEP =1.21; TT= 27.5°C; BT= 34°C
D-101	NS=37; RR=2.6; BR=250 kmol/h; FS=27; P=1 atm; TT=100°C ; BT=192°C	NS=21; RR=0.16; BR=246.7 kmol/h; FS=10; P=1 atm; TT=100°C; BT=192°C
H-101	T=35°C; P= 1 atm	T=35°C; P= 1 atm
D-102	NS=37; RR=7; BR=52.9 kmol/h; FS=2; P=1 atm; TT=78.6°C; BT=115.4°C	NS=28; RR=0.57; BR=52.83 kmol/h; FS=3; P=1 atm; TT=78.6°C; BT=115.4°C
D-103	NS=37; RR=45; DR=29.05 kmol/h; FS=14; P=1 atm; TT=54°C; BT=81°C	NS=61; RR=3.84; DR=29.10 kmol/h; FS=56; P=1 atm; TT=54°C; BT=81°C

NS: number of stage; P: pressure; ABE/HEP ratio in molar base; TT: top temperature; BT: bottom temperature; RR: reflux ratio; BR: bottom rate; FS: Feed stage; T: temperature; DR: Distillate rate.

Acknowledgments

Dr. Morales-Rodriguez acknowledges the partial financial support by the Mexican Bioenergy Innovation Centre, Bioalcohols Cluster (249564) and the Universidad de Guanajuato for the development of this project.

References

M. Errico, E. Sanchez-Ramirez, J.J. Quiroz-Ramírez, J.G. Segovia-Hernandez, B.-G. Rong, 2015, Alternative Hybrid Liquid-Liquid and Distillation Sequences for the Biobutanol Separation. Comput.-Aided Chem. Eng., 37, 1127-1132.

M. Geodkoop, R. Spriensma, 2001, The eco-indicator 99. A damage oriented for life cycle impact assessment. Methodology report and manual for designers, Technical report, PRe´ Consultants, Amersfoort, The Netherlands.

P.M. Harper, R. Gani, P. Kolar, T. Ishikawa, 1999, Computer-Aided molecular Design with combined molecular modelling and group contribution, Fluid Phase Equilib., 158-160, 337-347.

H.-J. Huang, S.Ramaswamy, Y. Liu, 2014, Separation and purification of biobutanol during bioconversion of biomass, Sep. Purif. Technol., 132, 513–540.

K. Kraemer, A. Hardwardt, R. Bronneberg, W. Marquardt, 2010, Separation of butanol from acetone-butanol-ethanol fermentation by a hybrid extraction-distillations process, Comput.-Aided Chem. Eng., 28, 7-12.

N. Morales-Espinosa, E. Sánchez-Ramírez, J.J. Quiroz-Ramírez, J.G. Segovia-Hernández, F.I. Gómez-Castro, R. Morales-Rodriguez, 2017, A Framework for an Optimized Sustainable Product and Process Design: Acetone-Butanol-Ethanol Separation and Purification, Comput.-Aided Chem. Eng., 40, 697-702.

R. Morales-Rodriguez, D. Rodriguez-Gomez, M. Sales-Cruz, J.A. de los Reyes-Heredia, E.S. Pérez-Cisneros, 2014, Model-Based Analysis for Acetone-Butanol-Ethanol Production Process through a Dynamic Simulation, Comput.-Aided Chem. Eng., 33, 133-138.

A. Sanchez, I.Valdez-Vazquez, A. Soto, S. Sánchez, D. Tavarez, 2017, Lignocellulosic n-butanol co-production in an advanced biorefinery using mixed cultures, Biomass Bioenergy, 102, 1-12.

M. Srinivas M., G.P. Rangaiah, 2007, Differential Evolution with Tabu List for Solving Nonlinear and Mixed-Integer Nonlinear Programming Problems, Ind. Eng. Chem. Res., 46, 7126-7135.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50060-4

Simulation of a Dual Mixed Refrigerant LNG Process using a Nonsmooth Framework

Matias Vikse[a*], Harry A. J. Watson[b], Paul I. Barton[b], Truls Gundersen[a]

[a]*Department of Energy and Process Engineering, Norwegian University of Science and Technology (NTNU), Kolbjoern Hejes Vei 1B, NO-7491, Trondheim, Norway*
[b]*Massachusetts Institute of Technology, Process Systems Engineering Laboratory, Cambridge, MA 02139*
matias.vikse@ntnu.no

Abstract

Natural gas liquefaction is an energy intensive process with very small driving forces at cryogenic temperatures. Small temperature differences arise from the excessive exergy destruction that occurs from irreversible heat transfer at low temperatures. As a result, even a small change in driving forces in the low temperature region can propagate into large exergy losses that must be compensated by additional compression power. Along with the significant investments and operating costs associated with these processes, this demands a robust and accurate simulation tool. Nonsmooth simulation models for single mixed refrigerant processes already exist in the literature. However, these processes are relatively simple, and are normally only considered for small-scale production or floating operations. Other processes, such as dual mixed refrigerant processes, are therefore normally considered for large-scale production of LNG. It is necessary to investigate whether the nonsmooth flowsheeting strategy is capable of also handling these more complex liquefaction processes. This article describes a simulation model for a dual mixed refrigerant process. The model is solved for two cases using the Peng-Robinson equation of state, each solving for a different set of unknown variables. Both cases converged within a few iterations, showing nearly identical results to simulations run in Aspen Plus.

Keywords: LNG, DMR process, nonsmooth modelling, process simulation.

1. Introduction

Liquefied natural gas (LNG) plays an important role as an energy carrier in the shift towards green energy sources. It is considered a cleaner alternative to other fossil fuels such as coal and oil, particularly due to a low sulfur content, no particle emissions during combustion and lower specific CO_2 emissions. Moreover, LNG can be distributed to a global market unlike traditional pipeline gas that relies on pre-existing infrastructure. However, liquefaction of natural gas is a very energy intensive process that requires the natural gas to be cooled down to cryogenic temperatures. At these low temperatures, thermodynamic losses due to irreversible heat transfer become significant, thus motivating designs with small temperature differences particularly at the cold end of the heat exchangers. Temperature differences are minimized by using refrigerant cascades and/or refrigerant mixtures, as well as splitting of streams to provide cooling at different temperature levels. Multistream heat exchangers (MHEXs) thus represent critical unit operations in the liquefaction process. Besides from being very energy intensive, LNG processes also entail significant capital costs through investments in expensive and

proprietary equipment such as cryogenic heat exchangers and turbomachinery. Hwang et al. (2013) report that natural gas liquefaction represents about 30-40 % of the total cost in the LNG chain.

The high operating and capital costs motivate the need for optimization to develop energy-efficient designs. However, state-of-the-art process simulators such as Aspen Plus have significant limitations in their ability to model MHEXs. For instance, Aspen Plus lacks rigorous checks for avoiding temperature crossovers, instead requiring the user to rely on an iterative trial-and-error approach to LNG simulation. A multistream heat exchanger model that performs these checks, as well as area calculations for economic analysis has previously been developed using a nonsmooth approach (Watson et al., 2015). The model is formulated using a partially equation-oriented framework where the flash calculations are removed from the overall flowsheet and instead solved sequentially as nested subroutines (Watson et al., 2017a-b and Watson and Barton, 2017b). The results are a reduced model size and increased robustness compared to using a fully equation-oriented framework. The MHEX model was used to simulate successfully the PRICO process (Watson et al., 2017b) and three other single mixed refrigerant (SMR) processes of varying complexity (Vikse et al., 2017). The same processes were later optimized in an article by Watson et al. (2017c). However, single mixed refrigerant processes are normally considered only for small-scale or floating production units for which capital investment and space limitations are the main concerns. In the case of base-load plants, the high production volume and high operating costs advocate designs that are more efficient. One such design is the AP-DMR^{TM} (dual mixed refrigerant) process, which offers an attractive alternative for large-scale LNG production, due to its high efficiency and flexibility in design. However, the additional flexibility comes at a cost, as the AP-DMR^{TM} process is significantly more complex to model than the SMR processes. This paper models a version of the AP-DMR^{TM} process with single stage compression using the nonsmooth approach to flowsheet modelling (Watson et al., 2017b) and the MHEX model by Watson et al. (2015). Two case studies are presented and the results are validated using Aspen Plus.

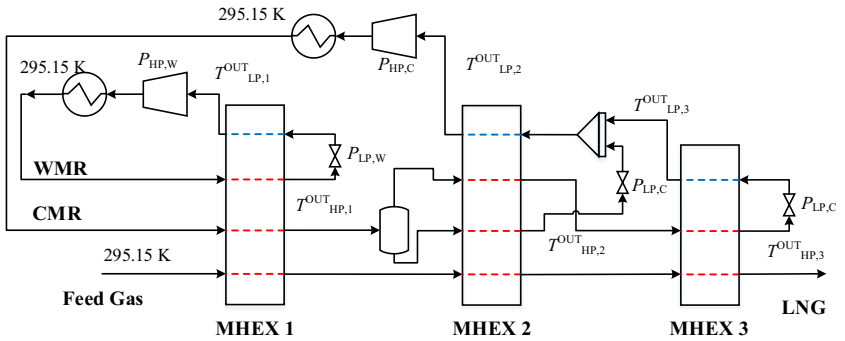


Figure 1: Flowsheet of the DMR process.

2. The DMR model

The flowsheet of the DMR process modelled in this paper is shown in Figure 1. The model, which is an extension of the hybrid process presented in Example 3 in Vikse et al. (2017), features three separate MHEXs. A warm mixed refrigerant (WMR) cycle is used

for precooling both the natural gas (NG) and the cold mixed refrigerant (CMR). An adiabatic phase separator is included after MHEX 1 to avoid freeze-out of heavier hydrocarbons as well as a high circulation rate in MHEX 3. The liquid product is subcooled and used for cooling in MHEX 2, whereas the vapour stream proceeds to MHEX 3 where it is condensed and subcooled to provide cooling at the cold end of the process.

The variables indicated in Figure 1 are as follows: $P_{\mathrm{HP,C/W}}$ and $P_{\mathrm{LP,C/W}}$ are the high (HP) and low pressure level (LP) of the CMR (C) or WMR (W), $T^{\mathrm{OUT}}_{\mathrm{HP,1/2/3}}$ and $T^{\mathrm{OUT}}_{\mathrm{LP,1/2/3}}$ are the temperatures of the HP and LP refrigerant streams out of MHEXs 1-3, whereas $F_{\mathrm{MR,C/W}}$ are the refrigerant flowrates. Moreover, the minimum approach temperature and heat exchanger conductance in the MHEXs are denoted as $\Delta T_{\mathrm{min,1/2/3}}$ and $UA_{\mathrm{1/2/3}}$, respectively. Two cases are considered for the DMR model, each solving for a different set of unknown variables:

- Case I – variable set: $T^{\mathrm{OUT}}_{\mathrm{HP,1}}$, $T^{\mathrm{OUT}}_{\mathrm{HP,2}}$, $T^{\mathrm{OUT}}_{\mathrm{LP,3}}$, $P_{\mathrm{LP,C}}$, $F_{\mathrm{MR,C}}$, $f_{\mathrm{W,ethane}}$, $\Delta T_{\mathrm{min,2}}$.
- Case II – variable set: $T^{\mathrm{OUT}}_{\mathrm{HP,1}}$, $T^{\mathrm{OUT}}_{\mathrm{HP,2}}$, $T^{\mathrm{OUT}}_{\mathrm{LP,3}}$, $P_{\mathrm{HP,C}}$, $P_{\mathrm{LP,W}}$, $P_{\mathrm{LP,C}}$, UA_3.

Here, $f_{\mathrm{W,ethane}}$ is the component molar flowrate of ethane in the WMR, which is used to solve for the mole fraction of ethane $z_{\mathrm{W,ethane}}$ from $f_{\mathrm{W,ethane}} = z_{\mathrm{W,ethane}} F_{\mathrm{MR,W}}$. The initial guesses for the unknown variables as well as the parameter values in the model are provided in Table 1. A heavy composition is used for the WMR consisting of 47.83 % ethane, 34.17 % propane and 18.00 % n-butane. The corresponding refrigerant composition for the CMR is 7.00 % nitrogen, 41.80 % methane, 33.20 % ethane and 18.00 % n-butane. An NG composition with 1 % nitrogen, 91.60% methane 4.93 % ethane, 1.71 % propane, 0.35 % n-butane, 0.40 % iso-butane and 0.01 % n-pentane is used in both cases. A natural gas molar flowrate of 1 kmol/s at 5.5 MPa pressure is used in the simulations. In addition, the outlet temperatures of the natural gas in the MHEXs are set equal to the temperatures of the high-pressure refrigerant streams.

The model subdivides the process streams into three sub-streams, corresponding to the two-phase, subcooled and superheated phase regions (Watson et al., 2017a). In addition, each sub-stream is further partitioned into stream segments with constant heat capacity flowrate to ensure accurate representation of the property models. In the simulations, five stream segments are used for both the subcooled and superheated regions in each MHEX, whereas 20 segments are used for representing the two-phase region. However, as the stream segments for the two-phase sub-stream are solved separately through nested flash routines, they do not affect the model size. Consequently, the model is represented by 95 unknown variables, 88 of which are intermediate temperatures of the superheated and subcooled stream segments in the MHEXs. These variables are initialized through an automatic initialization procedure that assumes a linear relationship between enthalpy and temperature. The model was solved with the Peng-Robinson equation of state with parameters taken from Aspen Plus, using the nonsmooth Newton solver described in more detail in Vikse et al. (2017).

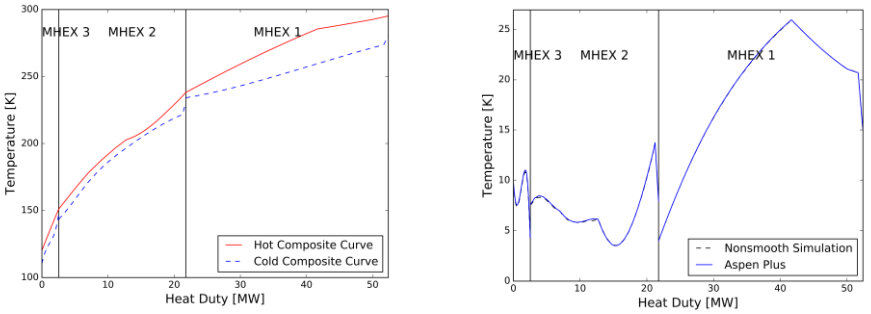
Table 1: MHEX and refrigerant stream data for the DMR model. The initial guesses for the unknown variables are placed in brackets.

Property	Case I	Case II	Property	Case I	Case II
$\Delta T_{\min,1}$ [K]	4.0	4.0	$P_{\mathrm{HP,W}}$ [MPa]	1.667	1.667
$\Delta T_{\min,2}$ [K]	[5.0]	5.0	$P_{\mathrm{LP,W}}$ [MPa]	0.420	[0.420]
$\Delta T_{\min,3}$ [K]	4.0	4.0	$T_{\mathrm{HP,1}}^{\mathrm{OUT}}$ [K]	[240.15]	[240.15]
UA_3 [MW/K]	0.3	[0.3]	$T_{\mathrm{LP,1}}^{\mathrm{OUT}}$ [K]	280.15	280.15
$F_{\mathrm{MR,C}}$ [kmol/s]	[1.45]	1.45	$T_{\mathrm{HP,2}}^{\mathrm{OUT}}$ [K]	[160.15]	[160.15]
$P_{\mathrm{HP,C}}$ [MPa]	4.850	[4.850]	$T_{\mathrm{LP,2}}^{\mathrm{OUT}}$ [K]	230.15	230.15
$P_{\mathrm{LP,C}}$ [MPa]	[0.250]	[0.250]	$T_{\mathrm{HP,3}}^{\mathrm{OUT}}$ [K]	120.15	120.15
$F_{\mathrm{MR,W}}$ [kmol/s]	[1.55]	1.55	$T_{\mathrm{LP,3}}^{\mathrm{OUT}}$ [K]	[145.15]	[145.15]

3. Results

3.1. Case I

A solution with $T_{\mathrm{HP,1}}^{\mathrm{OUT}} = 238.11$ K, $T_{\mathrm{HP,2}}^{\mathrm{OUT}} = 150.96$ K, $T_{\mathrm{LP,3}}^{\mathrm{OUT}} = 146.96$ K, $P_{\mathrm{LP,C}} = 0.310$ MPa, $F_{\mathrm{MR,C}} = 1.370$ kmol/s and $\Delta T_{\min,2} = 3.55$ K was obtained after three iterations. The new WMR composition consists of 50.16 % ethane, 32.64 % propane and 17.20 % n-butane. The molar flowrate of the WMR also changed to $F_{\mathrm{MR,W}} = 1.62$ kmol/s as a result of varying the component molar flowrate of ethane. Heat exchanger conductance values of MHEXs 1 and 2 were calculated to $UA_1 = 1.928$ MW/K and $UA_2 = 3.16$ MW/K during post-processing. The solution resulted in an isentropic compression power of 5.09 MW and 8.78 MW for the WMR and CMR, respectively.



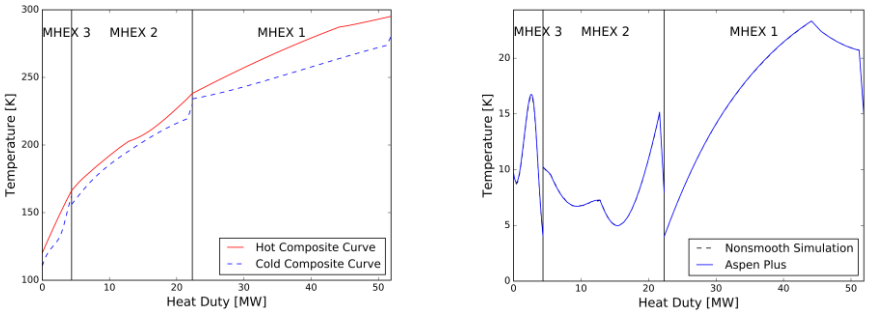
(a) (b)

Figure 2: (a.) Composite curves for Case I. (b.) Corresponding driving force plot.

Figure 2 presents the composite curves and the driving force plot for Case I. The driving force plot also includes results from simulations in Aspen Plus v9. As Aspen Plus can only solve for a single unknown variable in the MHEX modules, the above solution can only be obtained using a manual iterative procedure. Instead, the simulation was performed by using the solution from the nonsmooth model, only allowing the outlet temperatures of the LPR stream in each MHEX to vary. Aspen Plus and the nonsmooth model obtain the same driving force distribution at the solution. Moreover, the calculated isentropic compression power, UA_1 and UA_2 all lay within 1 % of the results of the nonsmooth model. The UA_3 -value deviates < 3.5 % of the nonsmooth model.

3.2. Case II

The second case converges after 5 iterations to a solution that requires 5.01 MW and 9.04 MW of isentropic compression power for the WMR and CMR, respectively. The composite curves and driving forces for the three MHEXs are presented in Figure 3. In addition, the solution in Aspen Plus is provided in Figure 3b. Again, the driving force curves are identical, indicating that Aspen Plus indeed obtains the same feasible solution. In addition, the isentropic compression power as well as heat conductance values for MHEX 1 and 2 are all within 1% of the nonsmooth model, whereas the deviation in UA_3 was calculated to < 2.5 % of the nonsmooth model.



(a) (b)

Figure 3: (a.) Composite curves for Case II. (b.) Corresponding driving force plot.

4. Conclusions

This article uses the nonsmooth flowsheeting strategy developed by Watson et al. (2017b) to model a dual mixed refrigerant process for LNG. Two simulation cases were analysed with the Peng-Robinson equation of state, employing a nonsmooth Newton solver for the system of equations. Both cases converged after a few iterations and the results correlate well with solutions obtained in Aspen Plus. In addition, the nonsmooth model contains additional degrees of freedom, making it easier to obtain feasible solutions without having to rely on a manual iterative procedure to avoid temperature crossovers in the MHEXs. The dual mixed refrigerant process is significantly more complex than the single mixed refrigerant processes previously studied by Vikse et al. (2017) and Watson et al. (2017b). Obtaining feasible solutions after just a few iterations demonstrate the model's capability of handling large and complex liquefaction processes. The next step will be to incorporate the model in an optimization framework.

5. Acknowledgements

This publication has been funded by HighEFF – Centre for an Energy Efficient and Competitive Industry for the Future. The authors gratefully acknowledge the financial support from the Research Council of Norway and user partners of HighEFF, an 8 year Research Centre under the FME-scheme (Centre for Environment-friendly Energy Research, 257632/E20). Harry A. J. Watson acknowledges Statoil for providing funding for this project.

References

J. H. Hwang, N. K. Ku, M. I. Roh, K. Y. Lee, 2013, Optimal Design of Liquefaction Cycles of Liquefied Natural Gas Floating, Production, Storage, and Offloading Unit Considering Optimal Synthesis, Industrial & Engineering Chemistry Research, 52, 15, 5341-5356.

M. Vikse, H. A. J. Watson, T. Gundersen, P. I. Barton, 2017, Versatile Simulation Methods for Complex Single Mixed Refrigerant Natural Gas Liquefaction Processes, Submitted.

H. A. J. Watson, K. A. Khan, P. I. Barton, 2015, Multistream heat exchanger modeling and design, AIChE Journal, 61, 10, 3390-3403.

H. A. J. Watson, P. I. Barton, 2017a, Modeling phase changes in multistream heat exchangers. International Journal of Heat and Mass Transfer, 105, 207–219.

H. A. J . Watson, M. Vikse, T. Gundersen, P. I. Barton, 2017a, Reliable Flash Calculations: Part 1. Nonsmooth Inside-Out Algorithms, Industrial & Engineering Chemistry Research, 56, 4, 960–973.

H. A. J. Watson, M. Vikse, T. Gundersen, P. I. Barton, 2017b, Reliable Flash Calculations: Part 2. Process flowsheeting with nonsmooth models and generalized derivatives, in press, Industrial & Engineering Chemistry Research.

H. A. J. Watson, P. I. Barton, 2017b, Reliable Flash Calculations: Part 3. A nonsmooth approach to density extrapolation and pseudoproperty evaluation, in press, Industrial & Engineering Chemistry Research.

H. A. J. Watson, M. Vikse, T. Gundersen, P. I. Barton, 2017c, Optimization of single mixed-refrigerant natural gas liquefaction processes described by nondifferentiable models, Submitted.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50061-6

Synthesis of Value Added Product Processes from Residual Biomass

A.I. Casoni[b], V.S. Gutierrez[b], M.A. Volpe[b], P.M. Hoch [a,b*]

[a]*Depto de Ing. Química, Universidad Nacional del Sur, 8000 Bahia Blanca, Argentina*
[b]*Planta Piloto de Ingeniería Química PLAPIQUI –UNS – CONICET, 8000 Bahia Blanca, Argentina*
p.hoch@plapiqui.edu.ar

Abstract

In this work, the preliminary synthesis of a fast pyrolysis plus catalytic hydrogenation process for the production of value added products from sunflower seed hulls will be presented at the conceptual stage. A General Disjunctive Programming (GDP) problem is solved, proposing a superstructure generated by the different ways of treating the raw material, posing different alternatives for the separation process and desired products. The Net Present Value is used as objective function. The solution of this problem allows for the identification of promising alternatives for the sensible and profitable use of residual biomass.

Keywords: Bioprocess, Mixed Integer Optimization, Sunflower Seed Hulls, Furfural, Furfuryl Alcohol.

1. Introduction

Residual biomass can be found in production processes or as an undesired component in the environment. As an example of the first, sunflower seed hulls are a residue from the edible oil industry, with thousands of tons a year, and as example of the latter, some types of algae present in water bodies. It is common practice to burn remnants of biomass in production processes as fuel for energy production, causing several environmental issues as the fumes and odor are very unpleasant. For this reason, there is a growing interest in using these biomass residues as feedstock for obtaining high value added (VAP) products. There are several technologies, which can be classified into three categories: thermochemical, physicochemical and biochemical. Thermochemical waste-to-energy (WTE) technologies are characterized for using high to very high temperatures to convert waste into energy or a VAP. Several thermochemical processes can be proposed for the conversion of biomass as feedstock as shown in Patel and Liu, (2017) for pyrolysis, gasification and Fischer-Tropsch synthesis, and Tozlu et al, (2016) also included incineration). Pyrolysis proved to be a high efficiency process, with an efficiency of nearly 80% for the production of energy, with a high potential to be used in large scale setups because of its promising characteristics for the production of VAPs (Higman and Tam, 2014). This is the reason pyrolysis was chosen as the first reaction process for the conversion of seed hulls into bio-oil in this work. Following this path, it is possible to obtain fuels from the bio-oil, bio char, syngas (Laird *et al*, 2009), and by separation from bio-oils, some other products as furfural and then, by hydrogenation, furfuryl alcohol. Studies have been performed lab and bench-scale to characterize the bio-oils and bio-char obtained (Sanchez *et al*, 2009). When residual biomass is used as feedstock, two objectives are met, solving an environmental issue and having added value products from

otherwise a residue, with the benefit of having a free raw material. Acid hydrolysis has been the traditional way of producing furfural (Quaker Oats patent) but it poses several environmental problems because of the very acidic waste produced, which is difficult to treat. In Argentina this has been the traditional route for obtaining furfural from wood.

For this work, several fast pyrolysis alternatives were developed at lab and bench scale by the research group (Casoni *et al*, 2015, 2017), which allowed the characterization of the raw material as well as the possible products following different process paths, with different catalysts and conditions. A preliminary flowsheet for the pyrolysis of sunflower seed hulls is presented at a conceptual stage in Casoni *et al* (2018) without evaluating all the alternatives for the different blocks of the process. It is desirable to scale up these processes so as to provide a solution for the treatment of the growing amount of residual biomass, and having a profit out of it. To the authors' knowledge, there is no industrial application reported using sunflower seed hulls with alternatives aligned with Green Chemistry principles so this work is a pioneer in the field.

2. Experimental

The production of furfural via fast pyrolysis and furfuryl alcohol via catalytic hydrogenation at laboratory scale were first studied. Details can be found elsewhere (Casoni *et al* 2015, 2018). The experimental results are used to predict the selectivity and conversion of the reactants considering different pretreatments of the sunflower seed hulls (SS) (untreated or pretreated using aqueous solutions with either 5% H_3PO_4. or 5% $ZnCl_2$). Results of SS characterization and pyrolysis yields are shown in tables 1 to 3. These results are used as a base to build the conceptual models for the pyrolysis and hydrogenation reactors, including mass and energy balances.

From the different biochars obtained post pyrolysis, catalysts are synthesized. The catalysts will be referred as Pd/BC, Pd/BC_{Ac} and Pd/BC_{Zn} for the catalyst prepared with Pd supported on biochar obtained from untreated hulls (BC); H_3PO_4 pretreated hulls (BC_{Ac}) and $ZnCl_2$ pretreated hulls (BC_{Zn}) respectively (Casoni *et al*, 2017). The catalysts are then used for the hydrogenation of furfural to furfuryl alcohol.

Table 1. Sunflower Seed Hulls (SS) characterization (Proximate Analysis and biomass contents).

	SS (untreated)	SS_{Ac} 5% H_3PO_4	SS_{Zn} 5% $ZnCl_2$
Moisture (%, dry basis)	6.1[a]	5.9	5.7
Volatile Matter (%, dry basis)	79.8[a]	83.8	83.1
Ashes (%, dry basis)	2.1[a]	2.3	1.9
Hemicellulose %	18.4	17.3	19.8
Cellulose %	39.1	33.8	36.7
Lignin %	20.4	16.9	18.7
Others [b] %	22.1	32.0	34.8

[a] Casoni *et al*., 2015. [b] Pectins, resins, extractives, etc.

Table 2. Yield to different products of pyrolysis

	SS (untreated)	SS_{Ac} 5% H_3PO_4	SS_{Zn} 5% $ZnCl_2$
Bio-oil % $_{ww}$	45	39	33
Bio-char % $_{ww}$	27	26	31
Gas % $_{ww}$	28	35	36

Regarding the yield to products, there are several subproducts present in the outlet stream of the reactor, depending on the pre-treatment of sunflower seed hulls used prior to the pyrolysis. A simplified account of products for the different conditions is shown in Table 3. A complete distribution of products in the bio-oil for several pretreatments can be found in Casoni et al (2015).

Table 3. Bio-oils chemical composition.

$\%_{ww}$ in bio oil	SS (untreated)	SS_{Ac} 5% H_3PO_4	SS_{Zn} 5% $ZnCl_2$
Acetic acid	43	17	15
Furfural	10	24	73
Levoglucosenone	-	-	10
Others[a]	47	59	2

[a] mainly phenolic compounds – See Casoni et al. (2015)

3. Model

A superstructure considering the different pretreatment stages (Untreated hulls, or pretreated with either H_3PO_4 or $ZnCl_2$), pyrolysis conditions, catalysts used and separation alternatives is presented in a very simplified way in figure 1. Models for the different processes are proposed using first principles and yields are used for the reactors.

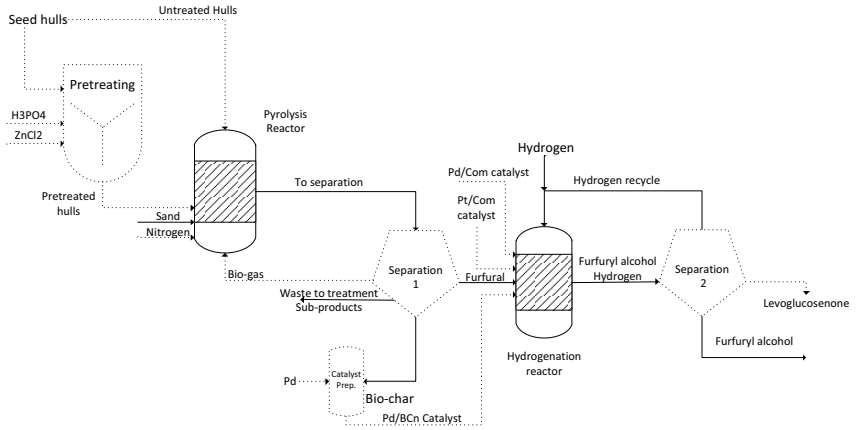


Figure 1: Simplified superstructure of the furfural and furfuryl alcohol production process. Equipment in dotted lines (together with associated streams) are represented by sets of logic propositions, each one represented by a Boolean variable. Separation sectors include several alternatives for the separation of products, being furfural, furfuryl alcohol and levoglucosenone (if applicable) the desired products.

3.1. Superstructure description

The sunflower seed hulls are transported to the facility and grounded using hammer mills. Only transportation costs are taken into account for the cost of the hulls, because they are a waste from sunflower edible oil industries. Hulls are stored in containers until transported to pyrolysis reactor or previously, to stirred tanks where they can be pretreated with either H_3PO_4 5% (880 u$s/ton H_3PO_4) or $ZnCl_2$ 5% (1950 u$s/ton $ZnCl_2$) if applicable. Diluted acid is previously prepared on site. Then, the liquid is discarded (Water and phosphoric acid can be disposed untreated, while diluted $ZnCl_2$ must be treated before disposal) and the wet solids need to be washed with water to eliminate the remaining acid and salts. The use of untreated sunflower seed hulls was also considered. Laboratory tests were performed to characterize the untreated hulls using these pretreatments, shown in Table 1.

The grounded (and pretreated if applicable) hulls must be dried before being sent to the pyrolysis step. For drying the hulls a flash drier is used.

The untreated or pretreated hulls are sent to the pyrolysis reactor, where they are mixed with hot sand (at temperatures from 650 to 700 °C) and fluidized with either the bio-gas from the reaction or nitrogen, where the proportions of sand/biomass should be between 10-20 and fluidization gas/biomass between 2-9. Type of reactor is circulating bed. The reactor is isothermic operating at 500 °C The product is sent to a series of cyclones where sand and bio-char are separated, and vapors are sent to a cooling system to condense the bio-oil with furfural. Biochar is either used for the preparation of the catalyst used in the subsequent hydrogenation step or burnt for energy generation.

A separation train (Separation 1) is then used for obtaining the liquid products from pyrolysis (bio-oil, furfural, levoglucosenone, acetic acid and water). Alternatives are: several single equilibrium stages or a single distillation unit with several side streams.

Furfural is sent to the second reaction step for hydrogenation, using the prepared catalyst (Catalyst prep). The possibility of using commercial Pd/Com and Pt/Com catalysts can also be included in the superstructure, Com standing for commercial catalyst support.

A final separation sector (Separation 2) is proposed to obtain all the products, evaluating the use of either distillation, single equilibrium steps, condensation or pervaporation membranes for obtaining high purity levoglucosenone. There is a huge interest in obtaining the latter because its selling price is 300 u$s/g. However, the separation steps required for attaining the high purity needed in the pharmaceutical industry are difficult and costly. With the model the feasibility of producing high purity LG will be evaluated at a preliminary level.

Treatment of residues is also proposed for all applicable cases and their costs taken into account.

3.2. Model formulation highlights

The problem is posed as a hybrid GDP using LOGMIP within GAMS. Due to the size of the problem it is not possible to include all the equations, so a few highlights are shown. The formulation of a hybrid GDP can be found in Vecchietti and Grossmann (1999)

Objective function is Net Present Value (*NPV*). The internal revenue rate (*IRR*) is also calculated as a way of evaluating the profitability of the project. *NPV* is calculated as:

$$NPV = -C_0 + \sum_{i=1}^{T} \frac{C_i}{(1+r)^i} \quad (1)$$

Where *Co* is the initial investment, C_i the cash flow in each year *i* (assumed to be annual *NB*), *T* is the lifespan (*T*=15 years) and r the discount rate (*r*=0.1).

Disjunctions are posed for the different equipment in each sector where alternatives can be used. For example, in the pretreatment sector, the unit will be present if pretreated hulls are selected by the model. Boolean variables Y_k are associated to untreated hulls, H_3PO_4 and $ZnCl_2$ and constraints on flow rates are written.

$$Y_{Untreated} \Rightarrow \neg Y_{Ac} \wedge \neg Y_{Zn} \tag{2}$$

$$Y_{Ac} \Rightarrow \neg Y_{Zn} \tag{3}$$

$$\begin{bmatrix} Y_{Untreated} \\ F_{Pyr} = F_{Hulls} \end{bmatrix} \vee \begin{bmatrix} \neg Y_{Untreated} \\ F_{Pyr} = F_{Hulls} + F_{Pret} \end{bmatrix} \tag{4}$$

$$\begin{bmatrix} Y_{Ac} \\ F_{Pret} = F_{H3PO4} \end{bmatrix} \vee \begin{bmatrix} \neg Y_{Ac} \\ F_{Pret} = F_{ZnCl2} \end{bmatrix} \tag{5}$$

Supported on-site prepared Pd catalyst used is linked to pretreatment, meaning for example that if untreated hulls are used, Pd/BC catalyst or any of the commercial catalysts will be used.

$$Y_i \Rightarrow Y_{Pd/i} \qquad i = Ac, Zn \tag{6}$$

$$Y_{Untreated} \Rightarrow Y_{Pd/Com} \wedge Y_{Pd/BC} \tag{7}$$

The preparation unit will be present only if Pd catalyst is used. This information will be used to define the yields in each reactor only, as conceptual models (based on laboratory test yields) for the reactors are posed. Volume is estimated using the residence time needed to complete the reaction for the different catalysts used (Casoni et al, 2018) so as to predict a preliminary reactor cost.

For separation sector 1 disjunction sets include a series of condensing stages or a distillation column with side streams. For separation sector 2, alternatives considered are single separation stages, distillation, pervaporation, or a hybrid distillation/pervaporation process. All the models are posed at a conceptual level using mass and energy balances, a shortcut model for the case of distillation and the membrane area for pervaporation is calculated using a correlation generated from a rigorous simulation.

For the scaling up of the reactors it was assumed that the same laboratory scale selectivity and conversion are reached in the industrial scale setup, which can provide a theoretical maximum. The desired product in the hydrogenation reactor is furfuryl alcohol > 80% $_{ww}$. Levoglucosenone 93% is obtained as subproduct if pretreated hulls are pyrolyzed. This can be further purified or sold at a low price to a producer who can purify it. Selling price of levoglucosenone <99.9% (pharma purity) is considered 1% of the selling price.

4. Results

Table 4 shows the most relevant results of the optimization problem. Riser temperature adopts the lower limit to decrease the pyrolysis cost. Combustion of bio-char provides enough energy to regenerate the sand. Levoglucosenone resulted to be the highest VAP obtained with this process, and so the pretreatment of seed hulls was selected, using H_3PO_4 because it does not need treatment for disposal, together with the preparation and use of catalyst Pd/BC_{Ac}. The purification stage of LG was not selected as the cost of

purification was anti-economic. This reduces the selling price of LG and then benefits are highly reduced. However, optimization results showed that the NPV of this process considering a lifespan of 15 years is 52 MMu$s with an IRR of 24%, making this a profitable project. A sensitivity analysis (Not shown) indicates that levoglucosenone selling price is key for the economics of the project.

Table 4. Optimization results for a production of 7000 t/y furfuryl alcohol with 3 pretreatment tanks and 3 condensing stages for separating furfural from bio-oil,

Sunflower seed hulls (ton/y)	20118	Hydrogen - fresh feed (Nm^3/h)	391.7
H_3PO_4 for pretreatment (kg/h)	1344	Residence time pyrolysis (s)	2
Riser temperature (°C)	500	Furfural produced (kg/h)	1130
Sand flow rate (kg/h)	40888	Total installation cost (MMu$s)	30.6
Bio-Char (kg/h)	393.9	Operating cost (MMu$s/y)	48.6
Bio Gas recycle (kg/h)	2775	NPV (MMu$s)	52
Bio-Oil (kg/h)	10177	IRR (%)	24

5. Conclusions

The preliminary synthesis of a novel waste to VAP industrial process based on pyrolysis of sunflower seed hulls and hydrogenation of furfural was presented, with the initial objective of producing furfuryl alcohol, but results suggested that the possibility of producing levoglucosenone as a subproduct should be considered. Provided that this is a conceptual stage study using simplified models including mass and energy balances, a rigorous design would be desirable to take account of more detailed instances of the process. For example, a nonlinear optimization programming technique would provide details on sizing and operating variables more accurately. Especially for LG purification stages this could dramatically increase the *NPV*, because a fine tuning of the separation process could lead to the decision of producing high purity LG, which is being studied.

References

Casoni A.I., M. Bidegain, M.A. Cubitto, N. Curvetto, M.A. Volpe (2015) Pyrolysis of sunflower seed hulls for obtaining bio-oils. Bioresource Technology, 177, pp. 406-409.

Casoni A.I., P.M. Hoch, M.A. Volpe, V.S. Gutiérrez (2018) Catalytic conversion of furfural from pyrolysis of sunflower seed hulls for producing bio-based furfuryl alcohol, Journal of Cleaner Production 178, pp. 237-246.

Higman C.A.A., S. Tam (2014). Advances in coal gasification, hydrogenation and gas treating for the production of chemicals and fuels. Chem. Rev. 114 1673-1708.

Laird D.A., R.C. Brown, J.E. Amonette, J. Lehmann (2009) Review of the pyrolysis platform for coproducing bio-oil and biochar, Biofuels, Bioprod. Bioref. 3, pp. 547–562.

Sanchez M.E., E. Lindao, D. Margaleff, O. Martınez, A. Moran (2009), Pyrolysis of agricultural residues from rape and sunflowers: Production and characterization of bio-fuels and biochar soil management, J. Anal. Appl. Pyrolysis 85 pp. 142–144

Patel B., X Liu (2017) Synthesis of a Biomass-to-Liquids (BTL) Process using a Hybrid Pyrolysis-Gasification System, Computer Aided Chemical Engineering, 40, 691-696.

Tozlu O.A., E. Özahi, A. Abuşoğlu (2016). Waste to energy technologies for municipal solid waste management. Renew. Sust. Energ. Rev. 54, 809-15.

Vecchietti A.R., I.E. Grossmann (1999) LOGMIP: A disjunctive 0-1 non-linear optimizer for process system models, Comp. and Chem. Eng. 23, 555-565

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50062-8

Exploiting the Synergy between Work and Heat for Holistic Energy Integration

Sajitha K. Nair, Iftekhar A. Karimi*

Department of Chemical & Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 11758.
cheiak@nus.edu.sg

Abstract

Work and heat are the two main forms of energy consumption in chemical processes. These work and heat requirements of the processes can be synergized to increase the overall energy efficiency. In this work, we simultaneously synthesize a network of work-exchange units (single-shaft-turbine-compressor) and heat exchange units (heat exchangers) to minimize the total annualized costs. The process streams are not pre-classified as high-pressure or low-pressure streams in work integration, or as hot or cold streams in heat integration. Furthermore, the model accommodates phase and pressure-temperature dependent properties, allowing liquid, vapor, and 2-phase streams. Finally, we illustrate the benefit of simultaneous work-heat integration through a case study of a propylene-propane separation process.

Keywords: work-heat integration, heat exchange network synthesis, work-heat exchange network synthesis, mixed-integer non-linear programming.

1. Introduction

The world's population is growing and fossil fuels, the main source of energy, are depleting. Moreover, the combustion of fossil fuels release greenhouse gases and contribute to global warming. These concerns are encouraging global efforts for increased energy efficiency and conservation. Heat integration is a well-established tool in process industries to reduce the overall utility consumption. Apart from heat, work is another form of energy consumption required for pressure changes. As pressure changes are often accompanied by temperature changes; compressors, turbines, and valves which bring about these pressure changes strongly interlink pressure (work) and temperature (heat). Thus, considerable energy savings can be achieved by simultaneously integrating both forms of energy requirements of the process streams.

The fundamental principles of heat exchange network synthesis can be extended to work-heat exchange network synthesis (WHENS). As heat exchanger transfers heat indirectly between two streams, a single-shaft-turbine-compressor (SSTC) can be used to transfer work between multiple streams on the same shaft. However, WHENS is a challenging problem as the required or generated work is a highly non-linear function of operating conditions. Wechsung et al. (2011) combined the pinch analysis, exergy analysis, and mathematical programming to minimize the external work/utilities or their costs. Razib et al. (2012) proposed an mixed-integer non-linear programming (MINLP) formulation for work exchange network (WEN) synthesis of SSTCs, ignoring heat integration among process streams. Onishi et al. (2014) presented an MINLP model for simultaneous heat and work exchange network synthesis; classifying streams as LP (low

pressure streams that undergo compression)/HP (high pressure streams that undergo expansion). They allowed multiple SSTCs and heat integration between WEN stages. Du et al. (2015) proposed a transshipment model to obtain the minimum utility consumption in WENs. Huang and Karimi (2016) presented a mathematical model for WHENS with a single SSTC and no pre-fixed end heater/coolers.

The existing literature uses predefined pressure manipulation routes or fixed identity of streams as LP/HP and hot/cold. However, in a real case such as refrigeration cycle, the same stream can undergo both compression and expansion. Furthermore, the existing models apply only to gaseous streams with constant thermo-physical properties such as heat capacity and Joule-Thomson (JT) coefficient. This limits the practicality of currently available models. We have developed a novel framework for work-heat integration in which the streams are not classified as HP or LP streams in the WHENS. This flexibility allows pressure change for streams with net zero pressure change. Additionally, the identities of the streams are not fixed a priori in heat integration. We optimize the design and operating variables to minimize the total annualized costs. To this end, we apply the developed MINLP model to study energy requirements of a C3 splitter column in various cases and exhibit the benefit of simultaneous work-heat integration.

2. Model

2.1. Problem Statement

Given S streams with known flow rates, inlet/outlet temperatures and pressures, we synthesize a work-heat exchange network (WHEN) with minimum total annualized cost (TAC) that may include 2-stream exchangers, utility heaters/coolers, Joule-Thomson (JT) expansion valves, SSTC with electric generator or motor.

2.2. Assumptions

- Pressure drops in pipes and exchanger are zero.
- Isothermal mixing of substreams in heat integration.
- Turbines, compressors, and valves are not operated in parallel.
- The cost of valves is zero compared to compressors and turbines.
- Only one zone of a stream (liquid, vapor, or 2-phase) can exist in an exchanger.

2.3. WHEN Framework

The WHEN employs two types of equipment. SSTC and Joule-Thomson (JT) expansion valves change stream pressures, hence are called pressure-changing (PC) equipment. 2-stream exchangers and heaters/coolers change stream temperatures without changing pressures, hence are called non-pressure changing (NPC) equipment. The NPC equipment together make up one single Heat Exchange Network (HEN). SSTC has one or more compressors or turbines. The turbines generate the work to run the compressor. If the turbine work is insufficient, then the SSTC has an electric motor that uses the deficit power from the grid. If the turbine work is in excess, then an electricity generator produces the power for the grid. A JT valve may be used to reduce a stream's pressure.

2.4. Superstructure

Each stream s $(s = 1,2,..,S)$ begins at initial conditions and passes through a series of N_s stages $(n = 1,2,\dots,N_s)$ to reach its final pressure. In every stage n, it first enters the

HEN for temperature adjustment and then, it passes through one PC equipment (either a valve or a compressor/turbine on the SSTC). During this pressure change, its temperature also changes. After exiting stage N_s, the stream enters the HEN for the last time, and leaves at outlet conditions. For a stream s that cannot change pressure in the WHEN, we set $N_s = 0$. Thus, each stream s enters the HEN $(N_s + 1)$ times, once in every stage n and then after stage N_s.

The HEN consists of K stages $(k = 1,2, \dots, K)$ of 2-stream exchangers. Each stage may have one or more parallel 2-stream counter-current exchangers that heat/cool process streams using other process streams. In the conventional HEN synthesis literature, streams are pre-classified as either hot or cold. In contrast, stream identities are unknown in this WHEN. The HEN involves heat exchange among all $S + \sum_s N_s$ streams. For simplicity, we assume isothermal mixing after each stage as proposed by Yee and Grossmann (1990) stagewise superstructure. Furthermore, a stream may use utility only before it exits the HEN.

2.5. Model formulation

We develop an MINLP model for minimizing the TAC of WHENs. In a stage, a stream can use any of the pressure-changing (PC) equipment. Furthermore, binary variables are used to define the identify of stream as hot or cold and to detect the liquid and vapor zones of a stream. These binary variables are also used for calculating zone-based heat transfer coefficients. The capital expense (CAPEX) for the WHEN includes the costs of SSTC, valves, heat exchangers, and utility heaters/coolers in the HEN. The operating expense (OPEX) includes the costs of running the helper motors, and the costs for heating/cooling utilities. The power generated by the SSTC provides the revenue. A stream s $(1 \leq s \leq S)$ with $N_s = 0$ uses no PC equipment and directly enters the final HEN. Thus, the same model can be used for synthesizing HEN, WEN and WHEN by changing the N_s and identity for each stream.

3. Propylene-Propane Separation

Propylene is one of the important starting product in petrochemical industry, due to its large use in manufacture of polypropylene. It is important to separate propane from propylene as the polypropylene process requires polymer-grade propylene (≥ 99.5 wt. % propylene). This separation is traditionally carried out in a distillation column, called C3 splitter, and is one of the most energy-intensive commercial distillation processes. Numerous studies have explored mechanical vapor recompression as an alternative for the traditional reboiler and condenser.

In this case study, we study various configurations for feed preheating, top vapor condensation and vaporization of bottom liquid for a fixed design of distillation column through WHENS and obtain the optimal configuration. The flowsheet for the process is as presented in Figure 1. There are 3 streams; the feed stream, top vapor stream and the bottom reboiling stream. We assume linear variation of enthalpies with pressure and temperature in different zones. We generate enthalpy, BPT, DPT data using Peng-Robinson in Aspen HYSYS, and regress them linearly. We used GAMS 24.8.5 for optimization with BARON as the MINLP solver with default parameters.

3.1. Case 1: Only heat integration

In the first case, we synthesize the heat exchange network with a minimum approach temperature of 5 K, ignoring the work exchange. Since the feed stream is changing

pressure from 2.1 MPa to 1.8 MPa, a valve is used to bring about this pressure change. We set $N_s = 0$ for all the streams and find the optimal heat integrated configuration. As can be seen in Figure 1(a), the heat integration is not economically attractive, and all the streams use utility for the desired enthalpy change. The temperature level of each stream prohibits heat exchange among them. Thus, the total annualized cost is 5.579 million \$/y with OPEX of 5.506 million \$/y as listed in Table 1.

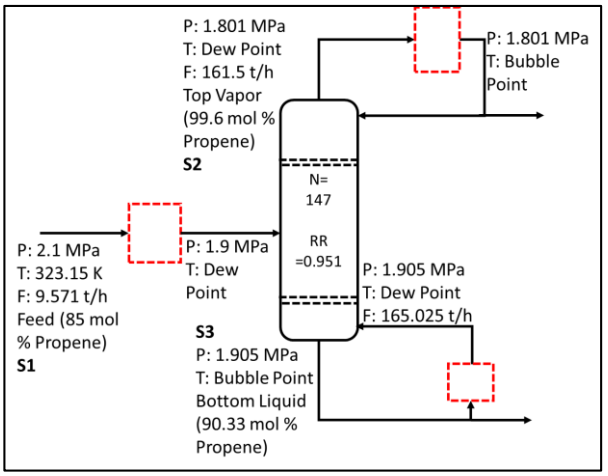


Figure 1. Flowsheet for the C3 splitter

3.2. Case 2: Work and heat integration with fixed identities

We use the same model to study work and heat integration done sequentially for mechanical vapor recompression process. In this case, there are 2 streams for pressure change (top vapor stream as HP and feed stream as LP) and 3 streams for temperature change (top vapor and feed stream as hot streams, and bottom liquid stream as cold stream). The outlet pressure and temperature of the top vapor stream are optimized to improve the heat integration, such that the total annualized cost is minimum. The identity of all the streams are fixed in heat and work integration. In this case, the optimal network is as shown in Figure 2 (b) with TAC of 1.228 million \$/y. The compressed top vapor is used to reboil the bottom liquid stream, thus drastically reducing both the hot and cold utility consumption and the OPEX associated with it.

3.3. Case 3: Simultaneous work and heat integration without pre-fixed identities.

We use the model for simultaneous WHENS with no pre-fixed identities for the top vapor and feed stream. Although this case considers the configurations from earlier two cases, the final optimal configuration as shown in Figure 2 (c) is different from the earlier two configurations. The top vapor stream changes identity several times. First, it is a cold stream which is heated before sending to a compressor where it acts as a LP stream. Later, it acts as a hot stream and subsequently as a HP stream. Although it is a stream with zero net pressure change, our model allows it to change pressure. In contrast to the intuition of cooling a stream before compression to reduce power consumption, the model preheats the compressor inlet stream. If this stream is not preheated, then a higher pressure-ratio is needed in the compressor to use the outlet stream as a hot stream and maintain the minimum approach temperature (MAT) in the reboiler. If the preheater is placed after the compressor, external hot utility is needed as

it is the highest temperature stream in the process. Thus, the current configuration is optimal, which the earlier two cases did not include. This affirms the benefit of simultaneous work-heat integration without prefixed identities.

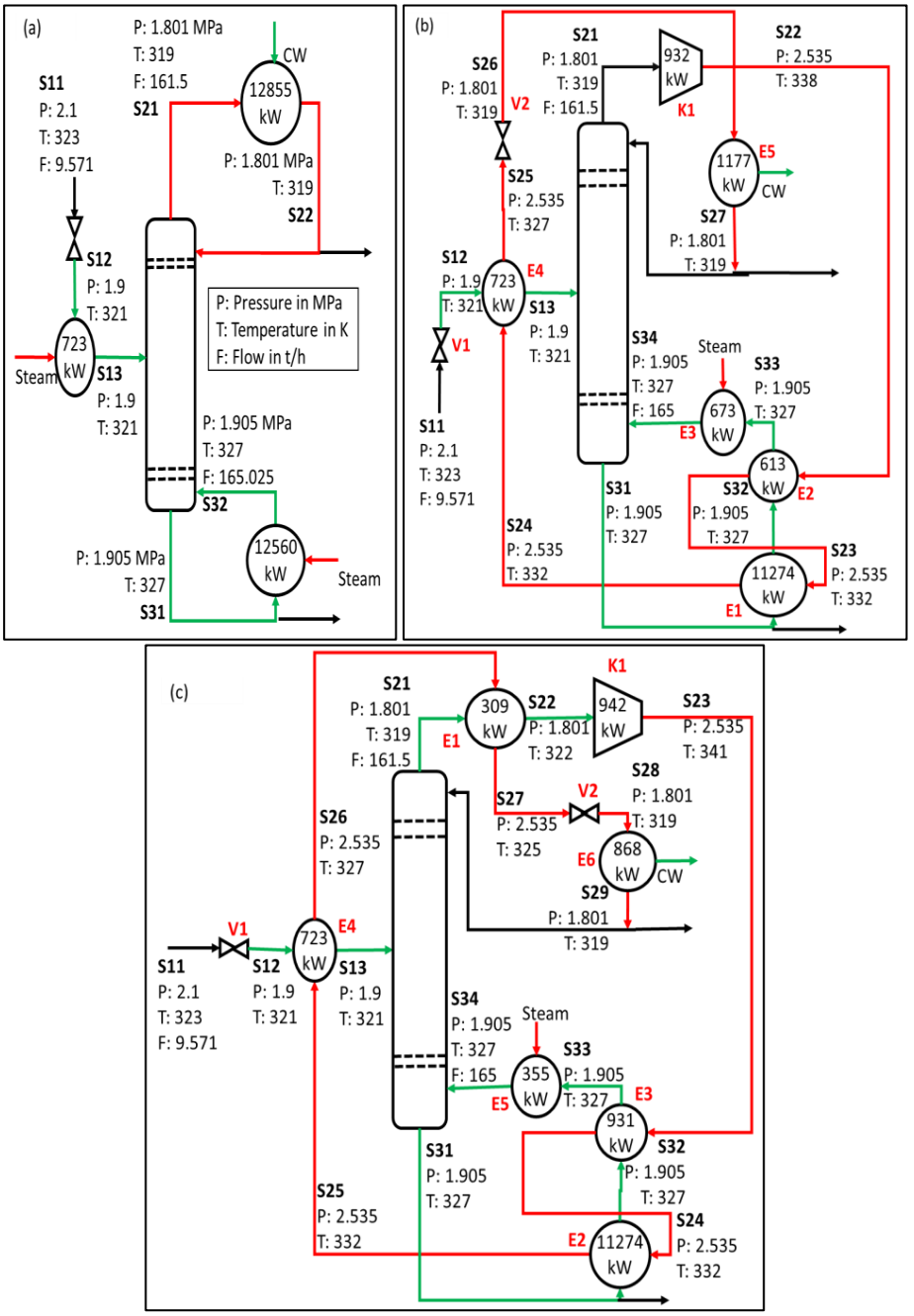


Figure 2. (a) Optimal configuration for Case 1, (b) Optimal configuration for Case 2, (c) Optimal configuration for Case 3

Table 1. Results for Case 1, Case 2, and Case 3.

	Case 1	Case 2	Case 3
Hot utility demand (kW)	12855	673	355
Cold utility demand (kW)	13284	1177	868
Power (kW)	0	932	942
CAPEX (million $)	0.275	1.881	2.015
OPEX (million $/y)	5.506	0.732	0.604
TAC (million $/y)	5.579	1.228	1.136

4. Conclusions

We have provided a general framework and developed an MINLP model for WHENS that does not pre-classify streams into hot/cold or HP/LP. We obtain the network of SSTCs, valves, and heat exchangers with the corresponding operating conditions and associated costs. The proposed model is generic and practical with a wide range of applications in process industries. We have illustrated its applicability through a case study on energy consumption in C3 separation process. Through this case study, we highlight the advantages of holistic energy integration through WHENS with unknown identity of streams over HENS and WHENS with fixed identity of streams.

Acknowledgement

Sajitha K. Nair acknowledges the financial support under the President's Graduate Fellowship from the National University of Singapore. The work was also funded in part by the National University of Singapore through a seed grant (R-261-508-001-646/733) for CENGas (Center of Excellence for Natural Gas)

References

J. Du, Y. Zhuang, L. Liu, J. Li, J. Fan, Q. Meng, 2015, Synthesis of Indirect Work Exchanger Network Based on Transshipment Model, Computer Aided Chemical Engineering, 37, 1139-1144.

K. Huang, I. A. Karimi, 2016, Work-Heat Exchanger Network Synthesis (WHENS), Energy, 113, 1006-1017.

V. C. Onishi, M. A. S. S. Ravagnani, J. A. Caballero, 2014, MINLP Optimization Algorithm for the Synthesis of Heat and Work Exchange Networks, Computer Aided Chemical Engineering, 33, 115-120.

M. S. Razib, M. M. F. Hasan, I. A. Karimi, 2012, Preliminary Synthesis of Work Exchange Networks, Computers and Chemical Engineering, 37, 262-277.

A. Wechsung, A. Aspelund, T. Gundersen, P. I. Barton, 2011, Synthesis of Heat Exchanger Networks at Subambient Conditions with Compression and Expansion of Process Streams, AIChE Journal, 57, 8, 2090-2108.

T. F. Yee, I. E. Grossmann, 1990, Simultaneous Optimization Models for Heat Integration-II Heat Exchanger Network Synthesis, Computers and Chemical Engineering, 14, 10, 1165-1184.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50063-X

Mechanical Energy Recovery through Work Exchanger Network Integration: Challenges and Opportunities

Aida Amini-Rankouhi, Yinlun Huang*

Department of Chemical Engineering and Materials Science, Wayne State University, Detroit, MI 48202, USA
yhuang@wayne.edu

Abstract

Heat and mass integration are two types of process integration technologies that have been practiced for improving energy and material efficiency in the process industries. Heat integration, frequently through integrating a heat exchanger network (HEN) into a process system, is for the recovery of thermal energy. However, recovery of mechanical energy, another form of energy, has not been fully explored. In this paper, we discuss the concept of work integration and its fundamentals. This will be followed by a general review of the frontier research on work exchanger network (WEN) synthesis, which is a system approach to implementing work integration. Challenges in WEN synthesis, such as energy targeting, equipment innovation and costing, and system configuration when heat integration is incorporated, are discussed. Preliminary studies on energy targeting, heat-incorporated WEN synthesis, and cost estimation are presented. Future research opportunities in WEN design and deployment are also discussed.

Keywords: Work integration, work exchange network, and mechanical energy recovery

1. Introduction

Process sustainability has become a main concern in industries, for which energy efficiency is a key indicator. Over the past decades, the chemical process industry has shown a great success in energy recovery in process systems through applying heat integration technologies. In chemical plants, thermal and mechanical energy are two common forms of energy. While the former can be effectively recovered by heat exchanger networks (HEN's), the recovery of the latter, however, has not drawn sufficient attention. Note that process work is more expensive than process heat, but recovery of mechanical energy is much more challenging.

From the thermodynamics point of view, heat flow where temperature is the state variable can be systematically managed to improve thermal energy efficiency, while work flow where pressure is the state variable must be carefully characterized so that opportunities for recovering mechanical energy can be identified. It is known that a large number of chemical plants have process streams to be pressurized, which require work for compression, or depressurized, which can produce work through expansion. Naturally, work exchange among process streams through synthesizing work exchanger network (WEN) should be a feasible approach for mechanical energy recovery.

In this paper, the state-of-the art research on work integration is reviewed, and some key challenges and opportunities of this area are discussed. We will then present a preliminary research on energy target setting and basic economic analysis that are necessary for synthesizing a cost-effective WEN. Furthermore, issues of heat integration

in WEN synthesis and evolution of current devices are also discussed.

2. WEN Synthesis-Progress Overview

It is recognized that the utilization of the mechanical energy available in a set of high-pressure streams for pressurizing a set of lower pressure streams in a process system may greatly reduce energy cost for compression operation. The pressure driven mechanical energy can be recovered using two types of work transfer units (WTU), the direct or indirect recovery devices. The former is called work exchanger (WE), which was first introduced for seawater reverse osmosis desalination systems (to replace energy-intensive pumps and turbines) by Cheng et al. (1967). The device was built using two displacement vessels configured in parallel that could simultaneously pressurize one fluid stream in one vessel and depressurize an equivalent volume of another stream in the other vessel in each operational cycle. Figure 1(a) is a sketch of one vessel, where the stream flows are controlled by four valves (Cheng et al., 1967). As a comparison, an indirect WTU, namely single-shaft-turbine-compressor (SSTC), is sketched in Fig. 1(b). This type of unit exchanges work in two steps: the pressure energy of a high-pressure stream is firstly converted to mechanical energy using an expander (turbine), and then to a compressor to pressurize a low-pressure stream (Chen and Wang, 2012). This type of device, however, has a low operational efficiency.

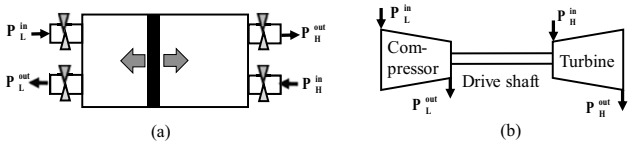


Figure 1. Sketch of work transfer units: (a) a (flow) WE and (b) a SSTC.

Inspired by the heat integration and methodological development for HEN synthesis, Huang and Fan (1996) introduced a notion of work integration through synthesizing WEN's. In their work, the similarities and differences between HEN and WEN synthesis were deliberated and the P-W (pressure-work) diagram was introduced to synthesize WEN.

A basic thermodynamic analysis on gas-gas work exchange was later conducted by Deng et al. (2010). Chen and Feng (2012) studied a design targeting issue for an ammonia synthesis example. Liu et al. (2014) proposed a graphical method using the P-W diagram. However, their method is somehow simplistic, and economic evaluation is restricted to the amount of energy recovered without counting capital cost. Amini-Rankouhi and Huang (2017) introduced a thermodynamic analysis method to evaluate rigorously the maximum amount of recoverable mechanical energy prior to WEN synthesis. Case studies showed that the energy efficiency was higher than the known studies. Du et al. (2017) presented a direct WEN synthesis using transhipment model under an isothermal condition, and the maximum energy recovery is not achieved. Razib et al. (2012) proposed a superstructure-based modelling approach using simple 2-stream SSTC units to design WEN with MINLP technique. Onishi et al. (2014) described a new MINLP model for WEN synthesis with hypothetic heat integration for optimal pressure recovery of process gaseous using SSTC units. Huang and Karimi (2016) included a HEN in a WEN system to minimize total annualized cost. In WEN synthesis, isentropic efficiency of both types of units has been considered. Heat and work integration has been also studied through placement of heat engines and heat

pumps (Townsend and Linnhoff, 1983), and compressors and expanders (Fu and Gundersen, 2016) during HEN synthesis.

3. Challenges and Opportunities

The known studies have shown that WEN synthesis is a new type of process integration technology, and WEN can be integrated into process systems to recover mechanical energy that is consumed by compressors, pumps, turbines, and other types of pressure vessels in the process industries.

There are various similarities between HEN and WEN syntheses, as fundamentally in each type of network, a set of high-potential streams (hot streams in HEN or high-pressure streams in WEN) transfer energy to a set of low-potential streams (cold or low-pressure streams) due to the existence of a driving force (ΔT in HEN and ΔP in WEN). In HEN, heat transfer units are easy to operate. By contrast, the compressors and expanders in WEN may operate in multiple stages, which could be under isothermal or non-isothermal condition. Thus, the shaft work either demanded for compression or provided by a work force may be significantly different. In addition, required compression and provided expansion energy is operating temperature dependent.

Methodologically, the Pinch Analysis technique successfully used in HEN synthesis cannot be directly used for WEN synthesis, because the basic notion of pinch point for heat exchange is not applicable for work exchange. As shown in the T-H diagram in Fig. 2(a), the temperature of the hot stream must be higher than that of the cold stream in the entire temperature range involved. However, this is not the case for work exchange between a high-pressure stream and a low-pressure stream as shown in the P-W diagram in Fig. 2(b). Therefore, the method for determining the maximum energy recovery by a HEN cannot be applied to a WEN problem. Besides, the basic formula for estimating the minimum number of heat transfer units in a HEN cannot be simply applied to the estimation of the minimum number of work transfer units in a WEN.

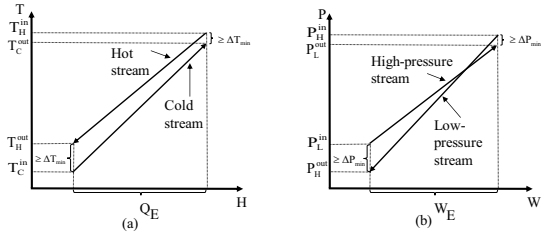


Figure 2. Comparison of (a) heat transfer in a HE and (b) work transfer in a WE.

It is noticed that WE, either direct or indirect, has only limited industrial applications. There is a serious lack of technological innovation in equipment design, especially for gas-gas or gas-liquid direct WE. This requires a more fundamental study on the designed operational behaviour of such a type of devices. Also note that if a WE involves a gas stream's pressurization or depressurization, the stream temperature can be changed considerably in operation. Therefore, such a WEN should be designed with heat integration technology incorporated; thereby leading to a hybrid exchange network.

WEN synthesis problems could be mathematically formulated and solved by MINLP techniques. Other type of synthesis methods could be also attractive, especially if a WEN design problem involves not too many high/low pressure streams, which is common. In such a case, heuristic based methods may demonstrate advantages, as a derived solution structure becomes explainable, which allows engineers to address some

practical design issues that could be difficult to formulate mathematically. Note that since heat-incorporated WEN system is structurally highly interacted, its operation could be sophisticated in terms of system dynamics, control, and process safety.

Energy Target Setting

For any chemical plant, a WEN is attractive if it is economically sound. Therefore, it is beneficial if the maximum amount of mechanical energy that can be feasibly recovered by a WEN can be determined prior to network design. This will facilitate justification of a need for WEN integration into an existing process.

Recently, we introduced a thermodynamic modelling and analysis method to accurately identify the maximum amount of recoverable mechanical energy of any process system of interest (Amini-Rankouhi and Huang, 2017). The energy targeting method includes four steps: (1) identification of all pressure intervals of known low-pressure streams for pressurization by given high-pressure streams, (2) evaluation of mechanical energy transfer from the high-pressure streams to the low-pressure streams, (3) determination of the minimum amount of external energy requirement, and (4) calculation of the maximum amount of recoverable mechanical energy. Each step is mathematically formulated, and a general computational procedure is developed for target setting of mechanical energy recovery of a WEN synthesis problem of any size. The detailed modelling and computational method is provided in Amini-Rankouhi and Huang (2017). By using that mathematical formulation, a WEN synthesis methodology is introduced, where heat integration is also incorporated to design a cost-effective hybrid network. A case study data in Razib et al. (2014) is used in this work to design a heat-incorporated WEN (HIWEN) by the introduced methodology. The synthesized network configuration is shown in Fig. 3. In this solution, a total of 61.35% mechanical energy of high-pressure streams and 49.91% energy for low-pressure streams are recovered. In addition, a total of 688 kW of thermal energy is recovered, too.

Table 1. Process Stream Data.

Stream No.	P^s (kPa)	P^t (kPa)	F (kg/s)	T^s (K)	T^t (K)	C_P (kJ/kg.k)
H_1	850	100	3	600	430	1.432
H_2	960	160	5	580	300	0.982
H_3	800	300	2	960	300	1.046
L_1	100	510	3	300	700	1.432
L_2	100	850	3	300	600	1.432

Basic Cost Analysis

The energy cost in a HIWEN can be readily estimated. Its capital cost of work transfer units, especially WE's, is determined by the number of units used, equipment structure, materials used, work transfer capacity, etc. Cheng and Fan (1968) studied the design of flow work exchanger for a desalination process, where the equipment structure and a basic equipment cost estimation were illustrated. We conducted a preliminary study on equipment design for gas-gas work exchange.

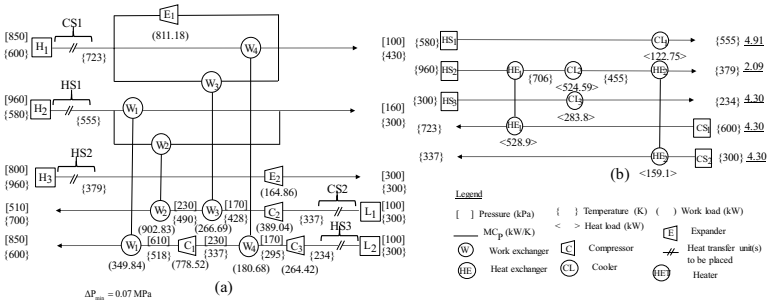


Figure 3. Flowsheet of heat-integrated work exchange network (a) WEN, and (b) HEN.

Similar to the cost estimation formula for heat transfer units, we propose the following formula: $C_B=\alpha S^{\beta}$, where S is the volume of a displacement vessel, and α and β are the parameters to be determined through experiment. As a case study, for one work exchanger unit, the cost data collected can be used to determine the parameters. Figure 4 gives the cost estimation of a WE based on the volume of the displacement vessel made by stainless steel, working in three different conditions with maximum pressure tolerance of the vessels and valves.

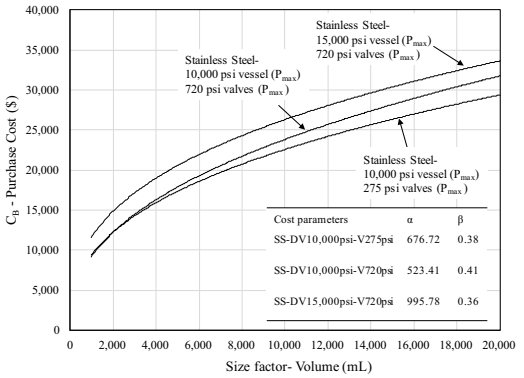


Figure 4. Cost estimation for one work exchanger unit.

Note that a comparison of the capital cost of a compressor or expander with a WE is mainly determined by equipment structure and capacity, materials used for the unit, process stream types, especially for gas-gas WE's, and operating condition. Thus, more reliable cost estimation methods are needed for WEN synthesis. Among different types of WE's, the Dual Work Exchange Energy Recovery Device (DWEER) has been widely used for seawater reverse osmosis (RO) desalination, which is one of the most efficient energy recovery systems developed to date (by Flowserve Corporation). This type of devices (dealing with liquid streams) has been reported to have low mixing and leakage losses, low maintenance cost, and self-adjustment capability to different flow rates and pressures. Despite that, WE's dealing with gas phase streams will demonstrate different operational characteristics. Note that operational safety related to leakage and mixing losses should be considered, especially when processing gas streams.

Another important concern is the operational performance of WE's, as the units may have a longer cycle time, depending on the operational mode, in comparison to compressors and expanders. Therefore, WEN dynamic control could be a challenge, as an effective operational coordination strategy is needed for operating different types of

units working in continuous or batch-like operational modes. This requires a more comprehensive study on system control design.

4. Conclusions

Improvement of energy efficiency in process systems has been always a key concern in the chemical process industry. However, the recovery of mechanical energy, as compared to thermal energy recovery, has not received sufficient attention. This renders a research need on work integration through designing WEN for chemical plants. In this work, a general review of the current research progress is provided, and the challenges and opportunities are presented. Preliminary studies have shown that WEN can be designed in a cost effective way. However, due to the special feature of work exchange, complexity in network design, and unique feature of work exchangers, WEN system design will be more sophisticated. In many cases, heat integration needs to be considered as well. It is expected that more systematic system design methodologies will be developed, and gas-to-gas work exchangers will be innovatively introduced in the near future.

Acknowledgment

This work is in part supported by the US NSF (Grant No. 1443912).

References

A. Amini-Rankouhi, Y. Huang, 2017, Work exchange network synthesis for mechanical energy recovery, AIChE J., 63, 4814-4826.
H. Chen, X. Feng, 2012, Graphical approach for targeting work exchange networks, Int'l J. of Chem., Mol., Nuclear, Mat. & Metall. Eng., 6, 716-720.
Z. Chen, J. Wang, 2012, Heat, mass and work exchange networks, Front. Chem. Sci. Eng, 6, 484-502.
C. Y. Cheng, S. W. Cheng, L. T. Fan, 1967, Flow work exchanger, AIChE J., 13, 438-442.
C. Y. Cheng, L.T. Fan, 1968, A flow work exchanger for desalination processes, US Dept. of the Interior, Office of Saline Water.
J. Q. Deng, J. Q. Shi, Z. X. Zhang, X. Feng, 2010, Thermodynamic analysis on work transfer process of two gas streams, Ind. & Eng. Chem. Res., 49, 12496-12502.
Flowserve Corp. DWEER. https://www.desalination.biz/59280/products/Dual-Work-Exchanger-Energy-Recovery-DWEER-by-Flowserve/4008. (Accessed 30.11.2017).
C. Fu, T. Gundersen, 2016, Heat and work integration: Fundamental insights and applications to carbon dioxide capture processes, Energy Conv. and Mgmt, 121, 36-48.
K. Huang, I. Karimi, 2016, Work-heat exchanger network synthesis (WHENS), Energy, 113, 1006-1017.
Y. L. Huang, L. T. Fan, 1996, Analysis of a work exchanger network, Ind. & Eng. Chem. Res., 35, 3528-3538.
G. L. Liu, H. Zhou, R. J. Shen, X. Feng, 2014, A graphical method for integrating work exchange network, Appl Energy, 114, 588-599.
V. C. Onishi, M. Ravagnani, J. A. Caballero, 2014, Simultaneous synthesis of work exchange networks with heat integration, Chem. Eng. Sci., 112, 87-107.
M. S. Razib, M. M. F. Hasan, I. A. Karimi, 2012, Preliminary synthesis of work exchange networks, Comp. & Chem. Eng., 37, 262-277.
D. W. Townsend, B. Linnhoff, 1983, Heat and power networks in process design. Part 1: Criteria for placement of heat engines and heat pumps in process networks. AIChE J., 29, 742-748.
Y. Zhuang, L. Liu, L. Zhang, J. Du, 2017, Direct work exchanger network synthesis of isothermal process based on improved transshipment model. J. Taiwan Inst. Chem. Eng., 81, 295-304.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50064-1

Process Synthesis under Seasonal and Daily Variability: Application on Concentrating Solar Power

Xinyue Peng, Thatcher W. Root, Christos T. Maravelias[*]

Department of Chemical & Biological Engineering, University of Wisconsin-Madison, WI 53706, USA
christos.maravelias@wisc.edu

Abstract

We propose an optimization-based framework for the synthesis of systems that are subject to variability in two frequencies. We use "scenarios" and "modes" to represent seasonal and daily variability respectively. We construct a multimode superstructure to consider different flowsheets alternatives for each mode. We formulate the optimization problem as a two-stage multimode stochastic programming approach with embedded models in each scenario. We illustrate the applicability of our framework through the design of concentrating solar power (CSP) plants.

Keywords: stochastic programming, superstructure optimization, solar power.

1. Introduction

Process synthesis under variability remains an important problem in process system engineering, especially nowadays with the increasing emphasis on the design of energy systems with intermittent renewable sources availability. The common approach to handle variability is to specify multiple operating conditions and then formulate the synthesis problem as a multiscenario/ multiperiod optimization problem (Grossmann et al., 1983). However, this approach is applicable when variability is present in one timescale. Many processes involve resource variability (e.g., wind, solar irradiance, electricity demand) on multiple timescales. To date, no previous approach has been proposed to systematically handle process synthesis under two-frequency variability. The majority of research in the area of chemical process synthesis has focused on processes which operate on a single "mode" (i.e., steady state) during a day and also have reasonably small variability, which, in these settings, is considered through "deviations" from the nominal operation and it is assumed that its effects can be compensated by control variables (Rooney and Biegler, 2003). Existing methods cannot be applied when the variability is so large and persistent that large subsystems of the facility are basically operating in completely different "modes" during, for example, the course of a day.

Accordingly, the goal of this paper is to outline an optimization-based framework for process synthesis under variability in two frequencies. Two key components of the framework is the use of scenarios to represent low frequency variability (e.g., seasonal) and modes, embedded within each scenario and linked with each other, to represent high frequency variability (e.g., daily). The framework allows us to formulate relatively accurate but yet tractable mixed-integer nonlinear (MINLP) models for process synthesis.

2. Modeling framework

In this paper, we focus on seasonal and daily variability (see Fig. 1), but the framework can be applied to systems under variability in any two frequencies. To efficiently capture the two frequencies, we adopt *scenarios* and *modes* to represent seasonal and daily variability, respectively. Scenarios ($s \in \mathbf{S}$) are representative days, each with a frequency of occurrence (π_s) within a year. For example, a year can be represented by three scenarios: a typical winter (25%), summer (25%) and mid-season (50%) day. Modes ($m \in \mathbf{M}$) are time periods within a day in which values of the varying parameters are notably different. For example, day and night modes may be used to describe daily solar irradiance variation.

All scenarios have the same types of modes occurring in the same sequence, which preserves the temporal connections between consecutive modes, but, to represent seasonal variations, the duration of the models can vary across scenarios (e.g., the day mode of a summer day scenario has longer duration than the day mode of a winter day scenario). Operation is linked across modes, which allows material and energy being stored and used in other modes. The main motivation behind using modes to represent daily variability is to allow the employment of different designs and operation strategies during the course of a day. Another motivation for the employment of modes, instead of refined time steps, is computationally tractability, since we aim to employ relatively detailed unit models (e.g., reactor models with kinetics or equilibrium relationships) and even consider, simultaneously, heat integration (which plays a key role in the development of renewable energy systems).

The framework consists of three key steps (see Fig. 1). First, we generate "scenarios" and "modes" from parameter data over a year. Second, we construct a multimode superstructure that contains different alternatives for each mode. Third, we formulate an design optimization problem as a multimode stochastic programming model. The design and associated operational decisions are obtained from the solution of the model. In the following subsections, we present the three steps in detail.

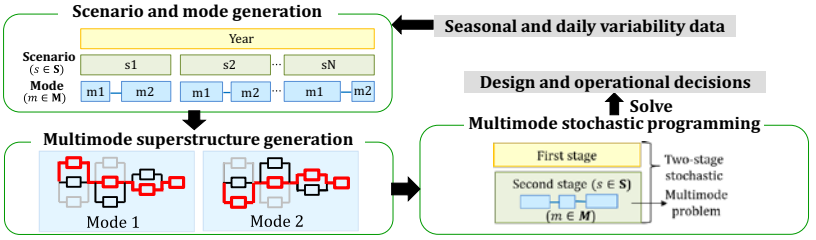


Figure 1. Overview of the proposed optimization framework.

2.1. Scenario and mode generation

The procedure of scenario and mode generation is shown in Fig. 2. We first identify M modes for each scenario based on patterns in daily profiles of varying parameters $\boldsymbol{\theta}$. For example, daily profiles of solar irradiance can be represented by day and night modes (see Fig. 3). Next, we develop a piecewise constant approximation to transform the actual daily profiles, into mode profiles, with $\boldsymbol{\theta}$ constant over M time intervals: for each model m and day l we have parameter $\boldsymbol{\theta}_{l,m}$ and duration $t_{l,m}$.

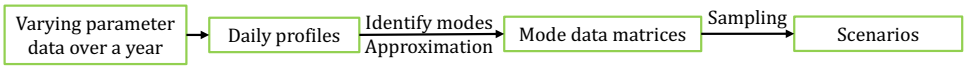


Figure 2. Scenario and mode generation procedure.

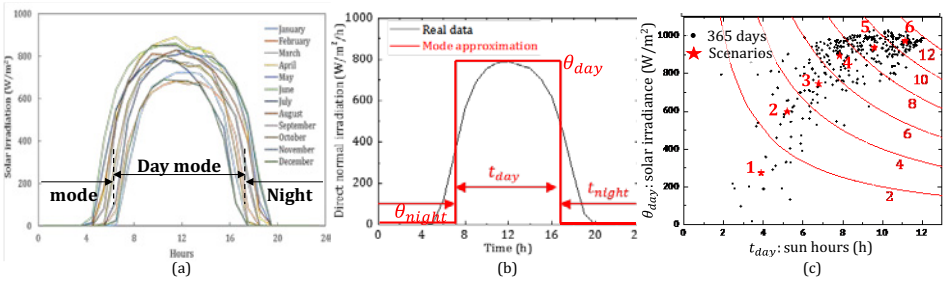


Figure 3. (a) Daily profiles of solar irradiance (b) mode-represented profiles from piecewise constant approximation and (c) 6 scenarios generated from 365 days.

For computational tractability, a small number (N) of representative days (i.e. scenarios) are selected to represent the low frequency variability (e.g., seasonal). This representation is reasonable because, in most cases, values of variability data over a year can be clustered into a small number of groups. For example, as shown in Fig. 3(c), solar conditions over 365 days (black dots) are divided into 6 groups based on the abundance of daily solar energy (energy abundance increases from bottom left to top right), and 6 scenarios are generated to represent these groups. The solar condition and occurrence frequency (π_s) of each scenario are then calculated as the average condition and the sum of occurrences of all days belonging to the corresponding group.

2.2. Multimode superstructure generation

Conventional superstructure optimization for process synthesis problems considers process alternatives at a *nominal* operating condition (Wu et al., 2016). This treatment is insufficient when the processes need to operate at different steady states as a result of variability. Accordingly, we propose a multimode superstructure that consists of different process flow sheets and unit alternatives for each mode.

To model the superstructure, we introduce three basic sets: (1) units $i \in \mathbf{I}$, (2) streams $j \in \mathbf{J}$, and (3) components $k \in \mathbf{K}$. Fig. 4 is the graphic representation of a general processing unit with all the relevant first and second stage variables.

Binary variables Z_i^d denote the selection of unit i for the overall design, while binary variables $Z_{i,m}^o$ denote the operational status (on/off) of unit i in mode m. If i is off in m ($Z_{i,m}^o = 0$), all component flows of its inlet ($j \in \mathbf{J}_{i,m}^{\mathbf{IN}}$) and outlet streams ($j' \in \mathbf{J}_{i,m}^{\mathbf{OUT}}$) are forced to zero. Streams are indexed over (i, m) because different streams can connect to the same unit in different modes. Unit i is installed if it is turned on in any mode.

$$Z_i^d \geq Z_{i,m}^o \quad i \in \mathbf{I}, m \in \mathbf{M} \tag{1}$$

The unit operation over scenarios and modes can then be formulated using a flow-based approach given as, in its general form, as,

$$(\boldsymbol{y}_{i,s,m}^{\mathbf{3}}, \boldsymbol{y}_{i,s,m}^{\mathbf{4}}) = f(\boldsymbol{y}_{i,s,m}^{\mathbf{1}}, \boldsymbol{y}_{i,s,m}^{\mathbf{2}}) \quad s \in \mathbf{S}, m \in \mathbf{M} \tag{2}$$

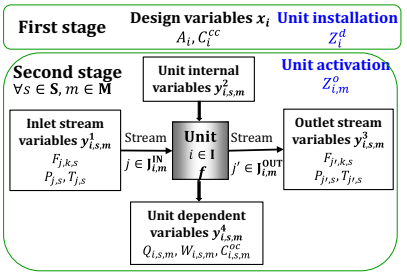


Figure 4. A general processing unit.

where $\boldsymbol{y_{i,s,m}^1}, \boldsymbol{y_{i,s,m}^2}$ are input operational variables and $\boldsymbol{y_{i,s,m}^3}, \boldsymbol{y_{i,s,m}^4}$ are output operational variables. Specifically, $\boldsymbol{y_{i,s,m}^1}$ are inlet stream variables (component flow rates $F_{j,k,s}$, stream temperatures $T_{j,s}$ and pressures $P_{j,s}$); $\boldsymbol{y_{i,s,m}^2}$ are unit internal variables (e.g. distillation reflux ratio); $\boldsymbol{y_{i,s,m}^3}$ are outlet stream variables ($F_{j',k,s}, T_{j',s}, P_{j',s}$); and $\boldsymbol{y_{i,s,m}^4}$ are unit dependent variables (e.g., heat duty $Q_{i,s,m}$, work $W_{i,s,m}$, operational cost $C_{i,s,m}^{oc}$).

While operational variables are different across modes and scenarios, design variables $\boldsymbol{x_i}$ (equipment size A_i, capital cost C_i^{cc}) are decided at the first stage, that is, remain fixed during operation. To guarantee operational feasibility, the equipment size is chosen as the maximum required size among all scenarios and modes:

$$A_i \geq f_i^{sizing}(\boldsymbol{y_{i,s,m}^1}, \boldsymbol{y_{i,s,m}^2}, t_{s,m}) \;\; i \in \mathbf{I}, s \in \mathbf{S}, m \in \mathbf{M} \tag{3}$$

where f_i^{sizing} is the unit sizing function, which carefully selected so that when unit inlet component flows are equal to zero, f_i^{sizing} is also forced to zero.

2.3. Multimode stochastic programming

The optimization problem is formulated as a two-stage stochastic programming approach with multimode subproblems inside each scenario, leading to a multimode stochastic programming model given as follows,

$$\begin{aligned}
&\min_{x} f^1(\boldsymbol{x}) + \textstyle\sum_s \pi_s \cdot Q_s\left(\boldsymbol{x}, \boldsymbol{\theta}_{s,m}, t_{s,m}\right) \\
&\quad \boldsymbol{x} = [\,\boldsymbol{x}^{\mathbb{B}}, \boldsymbol{x}^{\mathbb{C}}]^{\mathrm{T}}, \boldsymbol{x}^{\mathbb{B}} \in \{0,1\}^{n1}, \;\; \boldsymbol{x}^{\mathbb{C}} \in \mathbf{R}^{n2} \\
&\quad Q_s = \min_{\boldsymbol{y_{s,m}}} f^2(\boldsymbol{x}, \boldsymbol{y_{s,m}}, \boldsymbol{\theta}_{s,m}, t_{s,m}) \qquad \forall\, s \in \mathbf{S} \\
&\quad \text{s. t.} \left.\begin{array}{c} h_{s,m}(\boldsymbol{x}, \boldsymbol{y_{s,m}}, \boldsymbol{\theta}_{s,m}) = 0 \\ g_{s,m}(\boldsymbol{x}, \boldsymbol{y_{s,m}}, \boldsymbol{\theta}_{s,m}) \leq 0 \\ r_{s,m}(\boldsymbol{x}, \boldsymbol{y_{s,m}}, t_{s,m}, \boldsymbol{y_{s,m+1}}, t_{s,m+1}) = 0 \\ \boldsymbol{y_{s,m}} = \left[\,\boldsymbol{y_m^{\mathbb{B}}}, \boldsymbol{y_{s,m}^{\mathbb{C}}}\right]^{\mathrm{T}}, \boldsymbol{y_m^{\mathbb{B}}} \in \{0,1\}^{n1}, \boldsymbol{y_{s,m}^{\mathbb{C}}} \in \mathbf{R}^{n3} \end{array}\right\} \forall\, m \in \mathbf{M}
\end{aligned} \tag{4}$$

where $\boldsymbol{x}$ are design variables, including discrete $\boldsymbol{x}^{\mathbb{B}}$ (i.e. Z_i^d) and continuous $\boldsymbol{x}^{\mathbb{C}}$ (i.e. A_i, C_i^{cc}) decisions; $\boldsymbol{y_{s,m}}$ are operational variables, including discrete $\boldsymbol{y_m^{\mathbb{B}}}$ (i.e. $Z_{i,m}^o$) and continuous $\boldsymbol{y_{s,m}^{\mathbb{C}}}$ (i.e. $\boldsymbol{y_{i,s,m}^1}, \boldsymbol{y_{i,s,m}^2}, \boldsymbol{y_{i,s,m}^3}, \boldsymbol{y_{i,s,m}^4}$); $h_{s,m}$ are equality constraints (e.g. unit mass and energy balances); $g_{s,m}$ are inequality constraints including equipment sizing equations and design specifications; and $r_{s,m}$ are coupling constraints (e.g. storage or inventory level tracking) that connect the operation across modes.

3. Case study: Concentrating solar power plants

We apply the proposed framework to concentrating solar power (CSP) plants with thermochemical energy storage (TCES) (shown in Fig. 5). CSP converts concentrated sunlight to heat, which then drives a turbine to generate electricity. TCES stores the solar-heat for night power generation by reversibly converting it to chemical energy. The plant comprises four subsystems, the collector, receiver, TCES and turbine. Note that different process configurations are adopted in day and night modes.

In previous work by the authors, a deterministic optimization model for CSP+TCES processes operating at a nominal annual average solar condition has been developed (Peng et al., 2017). Building upon that work, we consider the seasonal and daily variability in solar irradiance (θ: direct normal irradiance W/m^2).

We chose Daggett in southern California as the CSP plant site and obtain its hourly solar irradiance data over 1-year from the National Solar Radiation Data Base. We generate six scenarios, each has two modes (shown in Fig. 3(c)).

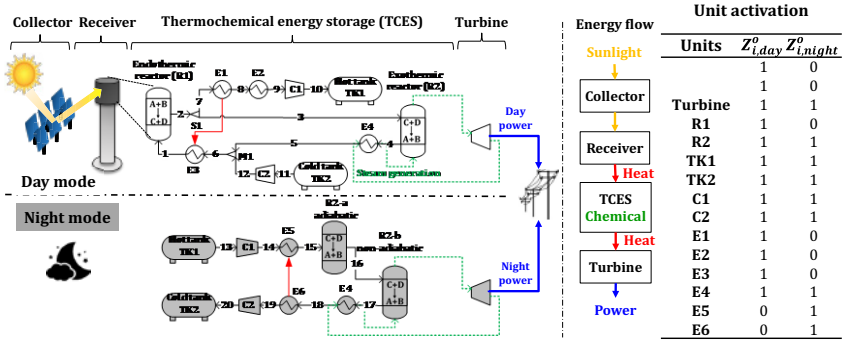


Figure 5. Process flow sheet of CSP plants with fluid phase TCES systems.

3.1. Model of CSP plants

Our goal is to simultaneously optimize design and operation to minimize the overall levelized cost of electricity (LCOE). The first-stage design variables ($\boldsymbol{x}$) include solar field area, receiver size and TCES equipment sizes. The second-stage operational variables ($\boldsymbol{y_{s,m}}$) include turbine output levels, stream flow rates and unit operating conditions. Equality constraints arise from the physical model of subsystems (i.e. collector, receiver and turbine), unit mass and energy balances, as well as cost calculations. Inequality constraints arise from equipment sizing and design specifications. Coupling constraints arise from material balances around storage tanks. We specify the plant nameplate capacity as 100 MW and generate bounds for all design and operational variables through a propagation algorithm.

3.2. Results

The resulting MINLP model has 1095 variables and 1461 constraints. It is formulated in GAMS (24.2.3) and solved using global optimization solver BARON (12.7.7) (Tawarmalani and Sahinidis, 2005) on an Intel® .

The design and LCOE obtained from the proposed stochastic approach based on six scenarios and the deterministic approach based on an annual average day are compared in Table 1. The stochastic design achieves a 0.5 $¢/kWh$ LCOE reduction over the

deterministic design (i.e., the value of stochastic solution). While the turbine capacity in both designs is fixed to 100 MW, the deterministic approach leads to a design where the other subsystems are larger (i.e., higher capital cost). This is because the average day assumption overestimates the plant performance. This *larger* (deterministic) design is inferior when the seasonal variability is considered.

Table 1: Comparison of stochastic and deterministic designs

Approach	LCOE (¢/kWh)	Capital cost (M$)				
		Total	Collector	Receiver	TCES	Turbine
Proposed	12.7	801	381	192	49	179
Deterministic	13.2	892	451	211	51	179

Power output levels and compression demand in all scenarios are shown in Fig. 6. In *sunny* scenarios 5 and 6, the plant maintains the full load for both day and night. In less sunny scenarios, the daytime production is high but night drops notably. The reason is that night power generation requires extra compression for reaction gas storage, the amount of which depends on both the night turbine output level and the night duration. Finally, in scenario 1, there is no night production because the total solar input is low.

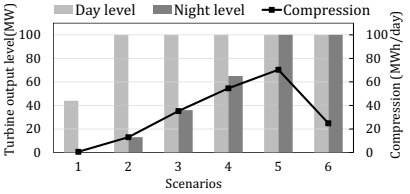


Figure 6. Turbine output levels and compression demand for storage in all scenarios

4. Conclusions

We developed a framework for process synthesis under seasonal and daily variability, represented by scenarios and modes respectively. A multimode superstructure is constructed to consider different process alternatives for each mode. The optimization design model was formulated as a multimode stochastic programming model. The proposed framework allows us to develop better designs of systems that are subject to two variability in two frequencies, notably, renewable energy systems converting solar energy to liquid fuels and power.

References

I. E. Grossmann, K. P. Halemane and R. E. Swaney, 1983, Optimization strategies for flexible chemical processes, Comput. Chem. Eng. 7, 439–462.

X. Peng, T. W. Root and C. T. Maravelias, 2017, Storing solar energy with chemistry: the role of thermochemical storage in concentrating solar power, Green Chem. 2427–2438.

W. C. Rooney and L. T. Biegler, 2003, Optimal process design with model parameter uncertainty and process variability, AIChE J. 49, 438–449.

M. Tawarmalani and N. V. Sahinidis, Math. Program., 2005. A polyhedral branch-and-cut approach to global optimization, Math. Program. 103, 225–249.

W. Wu, C.A. Henao, C. T. Maravelias, 2016, A superstructure representation, generation, and modeling framework for chemical process synthesis, AIChE J. 62, 3199–3214.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50065-3

A simple PLS-based approach for the construction of compact surrogate models

Jose E. A. Graciano[a,b], Reinaldo Giudici[b], Rita M. B. Alves[b], Benoît Chachuat[a*]

[a] *Centre for Process Systems Engineering, Department of chemical Engineering, Imperial College London, South Kensington Campus, London, SW7 2AZ, UK*
[b] *Department of Chemical Engineering, Escola Politécnica da Universidade de São Paulo, Av. Prof. Lineu Prestes, São Paulo, 05434-070, BR*
[*] *b.chachuat@imperial.ac.uk*

Abstract

This work describes a simple algorithm based on partial least square (PLS) to enable the construction of surrogate models using a single tuning parameter. The proposed algorithm is illustrated with the case study of a membrane module for natural gas sweetening, where a mechanistic model is used as data generator. The effect of the tuning parameter is analysed, and it is shown that this parameter captures the trade-off between the surrogates' accuracy and their complexity (number of terms). The algorithm performance is also compared with four different approaches from the literature, showing a similar performance.

Keywords: surrogate modelling, variable selection, optimization, natural gas sweetening.

1. Introduction

The development of efficient numerical methods alongside faster computers has triggered a step-change in the process synthesis area, where heuristic approaches are being replaced by superstructure optimization approaches, due to their better capacity to account for a large number of design alternatives (Henao and Maravelias, 2011). However, optimization-based process synthesis methods lead to complex mathematical programs, which even nowadays, prove difficult to solve when detailed, mechanistic models are used to represent each process in the superstructure. One approach to overcoming this computational burden is to replace these detailed models by simpler models (also called, surrogates, emulators or reduced-order models). The later can be fitted with data generated by the former, and thereby remain accurate in a certain operational region. Over the last decade, the use of these surrogate models in the area of process design and synthesis has been reported by a number of authors; see, e.g., Caballero and Grossmann (2008), Huang et al. (2012).

The surrogate-based optimization approach follows four basic steps: (i) a sample set of independent variables is taken within a certain operational region; (ii) the mechanistic model is simulated for this sample set; (iii) the surrogate is fitted using the input/output information provided by the mechanistic model; and (iv) optimization studies are conducted based on the surrogate, and the results are checked against the mechanistic model. Of course, the major issue faced by any surrogate modeling approach is making sure that the optima of both the surrogate and mechanistic models match. In other words, it is necessary to have a compromise between the possible underfit and overfit

that may occur during parameter estimation in step (iii). To overcome this problem, it is necessary to select a model that is flexible enough to fit the data, yet not too flexible to prevent over-fitting. The review paper by Bhosekar and Ierapetritou (2018) emphasize the importance of model selection in surrogate-based optimization, and lists a number of methods to perform this selection, including heuristics, integer programming, methods relying on model fitness, Bayesian variable selection, and methods based on correlation between input-output variables.

Special attention has been given to surrogates comprised of algebraic equations since they can easily be incorporated into optimization platforms such as GAMS and AMPL. There, a local or global optimization solver can be used to solve the MINLP problems resulting from the superstructure formulation. Recently, Cozad et al. (2014) proposed an algorithm for the proper construction of algebraic surrogates using a reformulated Akaike Information Criterion. The resulting MILP is solved a number of times, by increasing the total number of input variables until the information criterion gets worse. Boukouvala and Floudas (2017) presented an algorithm for global optimization based on surrogate models. In their approach the model selection step is performed by the Sure Independence Screening method (SIS) implemented in the R® language, which applies a ranking criterion to sort the input variables with respect to the vector of marginal correlations to the outputs, and then selects a suitable surrogate based on standard regularization methods for variable selection (Fan and Lv, 2008).

Of the alternatives to develop surrogates are algorithms for dimensionality reduction using partial least square (PLS). This family of methods analyzes the estimation problem in a projected space, performing the variable selection by adding informative variables to construct a new reduced model or by removing non-informative variables from a complete initial model. The goodness-of-fit may be assessed by cross-validation, which is a model validation technique used to estimate the model prediction capacity over an independent data set. The review paper by Mehood et al. (2012) on model selection methods using PLS cites a relatively large number of algorithms developed for this problem. In the present paper we describe a simple implementation of such model reduction algorithms in MatLab®, based on the standard *plsregress* function. This algorithm relies on the wrapper methods class presented by Mehmood et al. (2012) and presents the advantage of using a single tuning parameter as the stopping criterion. We compare our algorithm with four other alternatives implemented either in MatLab® and R®, using the case study of a membrane module for natural gas sweetening.

2. Problem Statement

Consider a vector of measured output variables $y \in \mathbb{R}^{n_{exp}}$, with n_{exp} the number of sampling points; and a set of input measurements $\boldsymbol{X} \in \mathbb{R}^{n_{exp} \times n_v}$, with n_v the number of possible combinations for the input variables (or basis functions such as polynomial, radial or exponential forms). The problem is to determine a compact surrogate model in the form of Eq.(1), where $\boldsymbol{X_s} \in \mathbb{R}^{n_{exp} \times n_s}$ is a subset of $\boldsymbol{X}$, $\beta \in \mathbb{R}^{n_s}$ is the vector of predicted parameters, $\varepsilon \in \mathbb{R}^{n_{exp}}$ is the vector of errors, and n_s is the number of selected terms from the initial model.

$$y = \boldsymbol{X_s}\beta + \varepsilon \tag{1}$$

The objective is to obtain a surrogate model comprising the least number of terms (or basis functions) n_s, as well as values for its parameters β, in order to: (i) minimize the

error vector ε (model underfit); and (ii) minimize the prediction error as estimated by cross-validation (model overfit). Another constraint in this problem is that the number of selected variables should remain lower than the number of experiments, $n_s < n_{exp}$, in order to avoid a direct interpolation of the available data by the selected surrogate.

3. An algorithm for model reduction based on standard PLS regression

The proposed algorithm is similar to the COVPROC algorithm proposed by Reinikainen and Höskuldsson (2003). There, the surrogate model is constructed variable by variable, until it reaches the stopping criterion that balances the quality of fit and prediction. Instead, we compute the stopping criterion herein based on the best-case scenario obtained with the complete model $\boldsymbol{X}$, which tends to overfit the set of measurements. Specifically, by using cross-validation, it is possible to obtain the minimum value of the objective function (mean square error of cross-validation, MSECV), which is used as the stopping criterion in turn. Basically, the algorithm proceeds in 6 steps:

Step 1: *Determine the stopping criterion value for the best-case scenario* – The best-case scenario for the PLS regression corresponds to the complete model $\boldsymbol{X}$, and entails determining the number of latent values which minimizes the MSECV. Figure 1 illustrates a typical MSECV profile with respect to the number of latent variables: the MSECV decreases with an increasing number of latent variables until a minimal value is reached (latent variable number 53), after which the MSECV starts to increase due to model overfitting. The stopping criterion is calculated as $sc = wcv - (wcv - bcv)t$, where wcv is the worst-case value (minimum number of latent variables), bcv is the best-case value (ideal number of latent variables), and $0 \leq t \leq 1$ is the tuning parameter accounting for the model accuracy. The best accuracy is obtained with $t = 1$, yet this comes at the price of a greater number of latent variables.

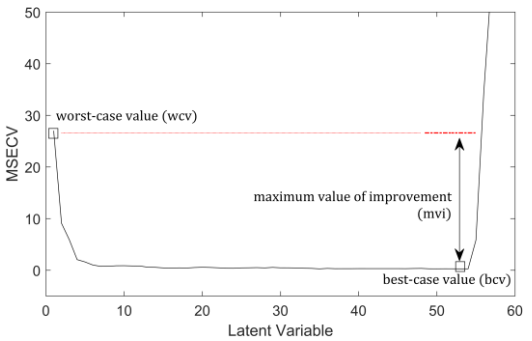


Figure 1. Illustrative example of a typical MSECV curve

Step 2: *Determine the initial number of latent variables* – This entails finding the minimum number of latent variables such that MSECV $\leq sc$. In Figure 1 for instance, if t is set as 0.95, the algorithm retains the first 6 latent variables (n_{LV}) as an initial guess.

Step 3: *Reduce the model* $\boldsymbol{X}$ *to a subset of selected variables* $\boldsymbol{X_s}$ *for the given* n_{LV} – We use the MatLab® function *plsregress* to compute the set of loading weights (w), as a measure of the importance for each model term on the latent variables (Mehmood et al., 2012). This information is used to select the model variable from $\boldsymbol{X}$ having the largest absolute value for each latent variable in order to build the reduced model $\boldsymbol{X_s}$.

Step 4: *Perform a PLS regression with the reduced model* $\boldsymbol{X_s}$ – A new curve of MSECV is computed with the reduced model.

Step 5: *Find a set of latent variables for the reduced model* $\boldsymbol{X_s}$ – The MSECV curve is analyzed with an increasing number of latent variables for finding the minimal number of components satisfying the stopping criterion MSECV$\leq sc$. If the latter stopping criterion is reached, the algorithm stops with a decision for $\boldsymbol{X_s}$ and the correspondent number of latent variables. Otherwise, it continues to Step 6.

Step 6: *Increase the surrogate dimensionality* n_{LV} – If the number of latent variables n_{LV} is less than the number of experiments n_{exp}, the algorithm sets $n_{LV} \leftarrow n_{LV} + 1$ and returns to Step 3. Otherwise, it stops and returns the best possible $\boldsymbol{X_s}$ for which the number of variables in the surrogate is less than the number of experiments $(n_s < n_{exp})$.

It is worth noting that the algorithm may be used to detect the need for more experimental data. When the algorithm terminates in step 6, and the reduced model does not present an acceptable level of adjustment, it would indeed be necessary to obtain more experimental data to properly fit the output variable y; one may also increase the complexity of the initial set of basis functions $\boldsymbol{X}$ to promote a better fit.

4. Case study

The proposed algorithm is compared with four different alternatives available in the literature. Table 1 lists these algorithms and summarizes their main features.

Table 1. List of algorithms for variable selection

Description	Reference
GA-PLS 1 (Matlab®) Genetic algorithm combined with PLS. In this class of algorithms, different subsets of input variables are used to fit the model. The selection occurs in the way that variables that show a better fit have a higher probability to continue in the sets of selected variables in the subsequent iterations.	(Leardi, 2000)
GA-PLS 2 (R®) Genetic algorithm combined with PLS. A different implementation of the genetic algorithm combined with PLS for variable selection.	(Hasegawa et al., 1997)
SIS (R®) Sure independent screening. This method proceeds in two main steps. First, the input set $\boldsymbol{X}$ is reduced to a number of input variables less than the number of experiments. Next, the algorithm uses a penalization method to refine the surrogate model.	(Fan and Lv, 2008)
MC-UVE-PLS – R® – Monte Carlo based uninformative variable elimination. In this approach, the input set $\boldsymbol{X}$ is augmented by noise variables. After that, the variables that have less "importance" than the artificial noise are eliminated from the model. This procedure is repeated until the prediction error get worse.	(Centner et al., 1996)

The input-output data used to fit the surrogates are generated from a detailed mechanistic model of a membrane module used in natural gas sweetening (Scholz et al., 2013). This model describes a number of non-ideal effects (concentration polarization, pressure loss, real gas behaviour and Joule-Thomson effect) and predicts the concentration of 5 species (CO_2, H_2S, CH_4, C_2H_6, and C_3H_8), resulting into a large and complex set of equations. In this particular case, the use of surrogate models to optimize

and/or integrate this system into a superstructure for membrane network design may be an attractive choice. The detailed model comprises 9 input variables (feed flow rate, temperature, pressure, composition and pressure of the permeate side) that are sampled within a given operational region using Latin Hypercubes Sampling to generate 60 experimental points. Then, the mechanistic model is used to simulate the corresponding set of output variables. The objective of the surrogate model presented hereafter is to predict the behaviour of the permeated stream temperature. Due to the presence of CO_2 and the large pressure drop across the membrane, the temperature of permeate stream decreases as a result of the Joule-Thomson effect, which modifies the membrane permeability and its overall separation performance in turn. The complete initial model $\boldsymbol{X}$ is given by a second-order polynomial model of the 9 input variables plus an artificial variable (membrane pressure drop), resulting in a complete model with a total of 65 basis functions (n_V). 80 percent of the set of input-output variables are used to select and calibrate the surrogate models, while the remaining 20% are used to evaluate the performance of the resulting surrogate models. Table 2 summarizes the results obtained with each approach.

Table 2 – Results obtained for different model reduction algorithms

Algorithm	Tuning factor	n_s	n_{LV}	r^2	RMSEV [°C]
This work	0.950	4	3	0.9454	1.1602
	0.980	5	4	0.9775	0.7329
	0.990	6	5	0.9817	0.6713
	0.995	8	7	0.9955	0.3322
	1.000	25	21	0.9981	0.2147
GA-PLS 1	n/a	27	10	0.9971	0.2665
GA-PLS 2	n/a	10	4	0.9965	0.2936
SIS	n/a	7	n/a	0.9986	0.1838
MCUVE-PLS	n/a	22	6	0.9967	0.2851

RMSEV – root mean square error of validation; r^2 – correlation factor

Increasing the tuning factor t up to 0.995 results in a moderate increase in the number of selected basis functions n_s, yet in a significant increase in the correlation factor r^2. Then, increasing t further from 0.995 to 1 results in a sharp increase in the number of selected variables, yet a small increase in r^2, which is consistent with the MSECV profile shown in Figure 1. It is difficult to determine the optimal value for the tuning parameter t a priori since the MSCEV profile is problem specific, but understanding this tuning parameter could be used as a means to evaluate the surrogate's accuracy with respect to the number of selected terms. The algorithms GA-PLS 1, GA-PLS 2 and MCUVE-PLS present a large variability in the resulting surrogates, changing the number and the selected basis functions to compose the final surrogate model. This behaviour may be partly attributed to the stochastic nature of these algorithms. However, this stochasticity is not necessarily seen as a problem here, as long as the algorithm is able to identify a compact surrogate model with good predictive capabilities, as observed for instance in the results of GA-PLS 2. It is worth noting that, for results obtaining errors less than 0.5°C, only three out six cases selected a number of basis functions lower than or equal to 10. In all the other cases, the number of selected basis functions is, at least, twice as large. The algorithm SIS presented the best result among the tested algorithms, since it provides the smallest surrogate model with the best prediction capacity within the validation set.

5. Conclusion

This paper has presented an algorithm for construction of surrogate models based on PLS techniques. This algorithm features a simplified term selection, by employing a single tuning parameter to capture the trade-off between accuracy and complexity of the candidate surrogate models. An application of this algorithm has been presented for the surrogate modelling of a membrane module used in natural gas sweetening. The results have been compared with four other approaches from the literature, showing that the proposed algorithm can exhibit a comparable performance to these existing approaches for suitable values of the tuning parameter. Naturally, it is not possible to decide which algorithm is better by analysing a single case study, but the approach is nonetheless useful to identify key features and limitations of the proposed algorithm. In practice, the variability in the algorithm outcomes, associated with their fast processing, suggests that several algorithms could be run concurrently, and only the best surrogate models would then be retained amongst the resulting candidates.

Acknowledgements

The authors would like to thank the sponsorship of Shell and FAPESP through the "Research Centre for Gas Innovation - RCGI" (FAPESP Proc. 2014/50279-4), hosted by the University of Sao Paulo, and the strategic importance of the support given by ANP (Brazil's National Oil, Natural Gas and Biofuels Agency) through the R&D levy regulation, as well the fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq Proc.: 200470/2017-5).

References

A. Bhosekar, M. Ierapetritou, 2018, Advances in surrogate based modeling, feasibility analysis, and optimization: A review. *Comp. & Chem. Eng.*, 108: 250–267

F. Boukouvala, C.A. Floudas, 2017, Argonaut: AlgoRithms for Global Optimization of coNstrAined grey-box compUTational problems. *Optim. Lett.*, v. 11(5): 895–913

J.A. Caballero, I.E. Grossmann, 2008, An algorithm for the use of surrogate models in modular flowsheet optimization. *AIChE J.*, 54(10): 2633–2650

V. Centner, et al., 1996, Elimination of Uninformative Variables for Multivariate Calibration. *Anal. Chem.*, 68(21): 3851–3858

A. Cozad, N.V. Sahinidis, D.C. Miller, 2014, Learning surrogate models for simulation-based optimization. *AIChE J.*, 60(6): 2211–2227

J. Fan, J. Lv, 2008, Sure independence screening for ultrahigh dimensional feature space. *J. R. Stat. Soc. Series B*, 70(5): 849–911

K. Hasegawa, Y. Miyashita, K. Funatsu, 1997, GA Strategy for Variable Selection in QSAR Studies: GA-Based PLS Analysis of Calcium Channel Antagonists. *J. Chem. Inf. Comp. Sci.*, 37(2): 306–310

C.A. Henao, C.T. Maravelias, 2011, Surrogate-based superstructure optimization framework. *AIChE J.*, 57(5): 1216–1232

D. Huang, T.T. Allen, W.I. Notz, N. Zheng, 2012, Global optimization of stochastic black-box systems via sequential kriging meta-models. *J.f Global Optim.*, 54(2): 431–431, 2012

R. Leardi, 2000, Application of genetic algorithm-PLS for feature selection in spectral data sets. *J. Chem.*, 14(5–6): 643–655

T. Mehmood, K.H. Liland, L. Snipen, S. Sæbø, 2012, A review of variable selection methods in Partial Least Squares Regression. *Chemom. Intell. Lab. Syst*, 118: 62–69

S.P. Reinikainen, A. Höskuldsson, 2003, COVPROC method: strategy in modeling dynamic systems. *J. Chemo.*, 17(2): 30–139

M. Scholz, T. Harlacher, T. Melin, M. Wessling, 2013, Modeling Gas Permeation by Linking Nonideal Effects. *Industrial & Engineering Chemistry Research*, 52(3): 1079–1088

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50066-5

A Comprehensive Approach for the Design of Solvent-based Adhesive Products using Generalized Disjunctive Programming

Jingyue Cui[1], Suela Jonuzaj[1] and Claire S. Adjiman[1,*]

[1]*Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, London SW7 2AZ, U.K.*
c.adjiman@imperial.ac.uk

Abstract

In this work, we present a comprehensive and systematic methodology for the design of optimal adhesive products within the computer-aided product design (CAPD) framework. In the proposed approach, the optimal number, identities and compositions of active ingredients and solvents in the final product are determined simultaneously. Generalised Disjunctive Programming (GDP) is employed to formulate the main design decisions of the problem (i.e., how many ingredients should be included, which active ingredients and solvents compounds should be used and in what proportions). The design methodology has been applied to identifying cheap and environmentally friendly acrylic adhesives, which are commonly used in construction.

Keywords: Adhesives, Product design, CAPD, GDP.

1. Introduction

There has been growing interest in the chemical industry towards the manufacturing of improved adhesive products that meet specified target properties over the last decades. Adhesives are formulations that can join materials together when applied on their surfaces, and are widely used in consumer goods, paper and packaging industry, construction, and transportation [Kinloch, 1987; Ceresana Market Research, 2017]. Nearly 14 million tonnes of adhesives are used worldwide and market demand is expected to increase in the next decade, resulting in a substantial revenue increase of 3.6 % per year [Ceresana Market Research, 2017]. Therefore, the design of environmentally benign adhesive products at minimum cost and time is important to maintain sustainability in a highly competitive field.

Although extensive work has been conducted within the Computer-Aided Product Design (CAPD) framework [Gani, 2004] for designing commodity products, pharmaceuticals and agrochemicals [Conte et al., 2011; Mattei et al., 2014; Gani and Ng, 2015], only a few model-based studies have been reported in the literature for the design of adhesives, including dental adhesives [Abedin et al., 2017; Spencer et al., 2010] and die attach adhesives [Fung et al., 2016]. Product formulations contain blends of ingredients from different chemical categories (e.g., polymers, solvents or solvent mixtures and aromas) which are often challenging to determine simultaneously due to the combinatorial explosion of the problems (many possible combinations of chemicals), leading to hierarchical design methodologies where each type of ingredients is determined in sequential steps [Conte et al., 2011; Mattei et al., 2014]. Within the decomposition-based methods smaller sequential subproblems are usually posed and

solved, where the active ingredient is first determined, and then solvents or binary solvent mixtures and additives are identified.

In this work, we propose the first comprehensive CAPD formulation for the design of acrylic adhesives. We develop this within the framework of a generic methodology in which the number of components, their identities and compositions are optimized simultaneously to maximize performance [Jonuzaj et al., 2016; Jonuzaj and Adjiman, 2017]. The optimal active ingredient and solvent mixtures are selected from given lists of candidate compounds. The design methodology is applied successfully to the design of environmentally friendly solvent-based acrylic putty.

2. Adhesive product design

2.1. Problem definition

The CAPD problem involves the generic formulation of the product design problem where the optimal number of ingredients, the optimal identities of all components and their compositions are identified, such that all given specifications are satisfied and the performance objective is optimised. The adhesives to be designed consist of three main classes of chemicals, the active ingredient (AI), solvent mixtures and additives. AI (usually a polymer) is the most important substance that defines the main function of the product; the solvent mixtures are usually in high concentration and are used to dissolve the AI and additives; small quantities of additives are present in the formulation to enhance the product quality and attributes. Here, additives in small concentrations (traces) that may be included in the final product are not taken into account and are not included in the design formulation. In order to develop the formulations, we define several index sets. The first is the set of components in the formulation, $I=\{1,\ldots,N_c\}$, where N_c is the total number of product ingredients. $N=\{1,\ldots,N_{max}\}$ is the set of components to be determined, whereas the sets $A=\{1,\ldots,N_a\}$ and $S=\{1,\ldots,N_s\}$ define the lists of compounds from which the unknown active ingredients and solvent components must be chosen, respectively.

2.2. GDP formulations

GDP techniques and different solution strategies have recently been employed to formulate mixture design problems [Jonuzaj et al., 2016]. The GDP formulation of the generalized product design problem is given as follows:

$$\begin{aligned}
\min_{x,Y} \quad & f(x) \\
\text{s.t.} \quad & g(x) \le 0 \\
& \bigvee_{a \in A} \begin{bmatrix} \hat{Y}_{i,a} \\ r_{i,a}(x) \le 0 \end{bmatrix}, i \in I \\
& \bigvee_{s \in S} \begin{bmatrix} Y_{i,s} \\ h_{i,s}(x) \le 0 \end{bmatrix}, i \in I \\
& \bigvee_{n \in N} \begin{bmatrix} \tilde{Y}_n \\ \tilde{F}_n(x) \le 0 \\ x_i \ge x_i^L, i = 3, \ldots, n \\ x_i \ge 0, i = n+1, \ldots, N_c \end{bmatrix}
\end{aligned}$$

$$\Omega\left(\hat{Y}, Y, \tilde{Y}\right) = \text{True}$$
$$x \in [x^L, x^U] \subset \mathbb{R}^m$$
$$\hat{Y}, Y, \tilde{Y} \in \{\text{True}, \text{False}\} \qquad (GDP)$$

where *f* and *g* represent the objective function and general constraints that must hold regardless of the discrete choices, respectively. Three different disjunctive sets are included in the general model. The first two sets involve disjunctions for assigning an active ingredient *a* and a solvent *s* from the lists (*A* and *S*) to components *i* in the product formulation. Conditional constraints $r_{i,a}$ and $h_{i,s}$ are active when the corresponding Boolean variables $\hat{Y}_{i,a}$ and $Y_{i,s}$ are true, respectively. The third disjunction includes constraints, $\tilde{F}_n$, that depend on the number of product ingredients and are active when the Boolean variable $\tilde{Y}_n$ is true. The mole fraction, x_i, of a component i is greater than a user-specified threshold value x_i^L if the component is present in the product formulation and is zero otherwise. It is noted that the mole fraction of the AI and the first solvent component are always greater than zero (the AI and at least one solvent should be present in the formulation). Logic propositions $\Omega\left(\hat{Y}, Y, \tilde{Y}\right)$ are derived to avoid degenerate solutions and relate the Boolean variable for the identities and number of components. The GDP formulation is converted into an MINLP problem by replacing the Boolean variables with binary ones and converting the logic propositions into linear inequalities. Conditional constraints inside the disjunctions are formulated using the big-M approach. A detailed description of the logic relations and the resulting MINLP model is given in Jonuzaj et al. [2016].

3. Case study: the design of acrylic putty

3.1. Problem description

The objective of the design problem is to identify an optimal AI and solvents or solvent mixtures that participate in acrylic putty in order to minimise the toxicity of the product. The problem includes phase stability and phase equilibrium relations which increase the complexity of the models. In particular, the mixture problem consists of nonlinear and nonconvex solid-liquid equilibrium relations for calculating the solubility of the AI in solvent mixtures, and the miscibility function to ensure that the generated solvent mixtures are in one liquid phase, as shown in Table 1. The liquid phase activity coefficients are calculated with the UNIFAC model. Pure component and mixture properties, such as toxicity (τ), viscosity (η), surface tension (σ) and evaporation time (T^{90}), are evaluated using experimental values [Martin and Young, 2001] or group contribution methods and linear mixing rules (when no experimental property values exist). The optimal AI and solvent components are selected from lists of 4 and 8 candidates, respectively. The candidate polymers and solvents used in this case study have been selected based on the effectiveness of the chemicals to achieve product attributes, their cost, health-and-safety issues and environment impact. In the application of the proposed methodology, we consider two formulations: (i) a restricted problem (models P1-P5), where we fix the number of ingredients, and products with one AI and one, two, three, four or five solvents are determined (i.e., *N*=1-6); (ii) the generalized problem (P6), where the number of components is not fixed but bounded by a maximum N_{max}, which takes a value of 6 here (i.e., 1 AI and up to 5 solvents).

Table 1: Problem specifications.

Description	$N \leq 6$
Components in the mixture	1 active ingredient (a) and up to 5 solvents (s_1 - s_5)
Active ingredients (AI) [Warson, 1993]	3 units of methyl methacrylate (3MMA), 3 units of propyl methacrylate (3PMA), 4 units of butyl acrylate (4BA), 4 units of ethyl acrylate (4EA)
Candidate solvents [Ebnesajjad and Landrock, 2015]	ethyl acetate, butyl acetate, hexane, benzene, ethyl benzene, propyl benzene, toluene, xylene
Model equations [Conte et al., 2011; Martin and Young, 2001]	$\ln x_a + \ln \gamma_a = \frac{\Delta H_{fus,a}}{R}\left[\frac{1}{T_{m,a}} - \frac{1}{T}\right]$, $\frac{\partial \ln \gamma_i^{i,j}}{\partial x_i^{i,j}} + \frac{1}{x_i^{i,j}} \geq 0$, UNIFAC model, $0.6 \leq \eta \leq 0.9$, $26.5 \leq \sigma \leq 29.5$, $55 \leq T^{90} \leq 570$, linear mixing rule: $\xi = \sum_{i=1}^{N_c} \xi_i x_i$
Problem size	1,402 eqns., 819 continuous and 109 discrete vars.

3.2. Results and discussion

All models (P1-P6) were implemented and solved in GAMS version 24.8.3, running on a single core of a dual 6 core Intel Xeon E5-1660 machine at 3.30 GHz. SBB [Bussieck and Drud, 2001], a local branch-and-bound MINLP solver, was used to solve all problem cases presented. GAMS files for all MINLP models can be found at *https://zenodo.org/record/1163603* and the results are summarized in Table 2.

Table 2: Optimal product toxicity (τ), identities and composition (x_i) of AI and solvent mixtures, obtained when solving the restricted and generalised problems of the case study. The lower bound of solvent mole fractions in the mixture is set to $x_i^L = 0.01$.

Problem	τ	Ingredients	Components	x_i	CPU (s)	$N_{\equiv}^{o}$ of nodes
P1 (N=2)	4.114	a	3PMA	0.207	0.24	22
		s_1	Ethyl acetate	0.793		
P2 (N=3)	4.097	a	3PMA	0.100	0.50	48
		s_1	Toluene	0.853		
		s_2	Butyl acetate	0.047		
P3 (N=4)	4.018	a	3PMA	0.104	2.29	142
		s_1	Ethyl benzene	0.563		
		s_2	Ethyl acetate	0.219		
		s_3	Toluene	0.114		
P4 (N=5)	**3.976**	a	3PMA	0.100	22.60	716
		s_1	Ethyl benzene	0.447		
		s_2	Ethyl acetate	0.228		
		s_3	Toluene	0.187		
		s_4	Propyl benzene	0.038		
P5 (N=6)	3.980	a	3PMA	0.100	43.07	663
		s_1	Ethyl benzene	0.475		
		s_2	Ethyl acetate	0.229		
		s_3	Toluene	0.158		
		s_4	Propyl benzene	0.028		
		s_5	Butyl acetate	0.010		

P6 ($N \leq 6$)	**3.976**	a	3PMA	0.100	153.66	1034
		s_1	Ethyl benzene	0.447		
		s_2	Ethyl acetate	0.228		
		s_3	Toluene	0.187		
		s_4	Propyl benzene	0.038		

Overall, the optimal product with the lowest toxicity is found in models P4 and P6, and consists of a poly-propyl methacrylate (3PMA) and 4 solvents (ethyl benzene, propyl benzene, ethyl acetate and toluene). The use of a solvent mixture, rather than a single solvent, leads to a product with low toxicity yet acceptable AI content. In addition to the above formulations, the traditional decomposition-based problem, where the AI is first determined and then the optimal solvents are identified in smaller subproblems, was solved; different more toxic formulations were obtained with 3MMA being identified *a priori* as the optimal AI. CPU time increases with the number of product ingredients due to the increased size of the problems. In the studied example, the restricted models with fixed number of components require less computational time than the general problem. In practice, however, formulating a single generic problem (P6) requires less user input than formulating and solving a series of restricted problems. The smallest problem (P1), where one active ingredient and one solvent are determined, is also solved globally using BARON version 17.10.16 [Tawarmalani and Sahinidis, 2005] in 3.94 CPU seconds. The results verify the optimal solution obtained with the SBB MINLP solver. Convergence to global optimality was not reached in 100,000 CPU seconds when larger problems (P2-P6) were solved.

4. Conclusions

This study provides a general and comprehensive mathematical formulation for the design of optimal adhesive products based on a computer aided product design (CAPD) framework. Within this systematic generalized approach the number of product ingredients, the identities of the components (i.e., active ingredients and solvent compounds) and their compositions were determined simultaneously. The general formulation makes it possible to identify the best AI and the optimal solvent mixture at the same time, without specifying in sequential steps each type of chemicals which may lead to suboptimal solutions. The general methodology was applied successfully to the design of acrylic adhesives, where optimal blends with one active ingredient and up to 5 solvents were determined. The simultaneous design of the AI and solvent mixtures can lead to better results than the ones obtained when fixing the active ingredient in advance. The proposed approach could be used to generate a ranked list of promising designs by adding integer cuts and would thus serve as a guide to experiments. Further work is required to evaluate the impact of uncertainty in the property prediction models and the linear mixing rule used to calculate mixture properties. Furthermore, the development of suitable algorithms to achieve global solutions should be considered.

Acknowledgments: The authors gratefully acknowledge financial support from the EPSRC of UK via a Leadership Fellowship (EP/J003840/1), a platform grant (EP/J014958/1) and a Doctoral Prize to SJ.

Data statement: Data underlying this article can be accessed on Zenodo at *https://zenodo.org/record/1163603*, and used under the Creative Commons Attribution license.

References

F. Abedin, B. Roughton, Q. Ye, P. Spencer, K. Camarda, 2017. Computer-aided molecular design of water compatible visible light photosensitizers for dental adhesive. *Ch.Eng.Sc.*159,131–139.

M.R. Bussieck, A.S. Drud, 2001. SBB: Simple Branch and Bound algorithm for mixed integer nonlinear programming. URL: *https://www.gams.com/latest/docs/solvers/sbb/*.

Ceresana Market Research. Market Study: Adhesives - World (3rd edition). Available online: *http://www.ceresana.com/en/market-studies/industry/adhesives-world*. Accessed: Sep 2017.

E. Conte, R. Gani, K.M. Ng, 2011. Design of Formulated Products: A Systematic Methodology. *AIChE J.* 57, 2431–2449.

S. Ebnesajjad A.H. Landrock, 2015. Adhesives Technology Handbook (3rd Edition). Elsevier Inc. San Diego, USA.

K.Y. Fung, K.M. Ng, L. Zhang, R. Gani, 2016. A grand model for chemical product design. *Com. & Chem. Eng.* 91, 15–27.

R. Gani, 2004. Chemical product design: challenges and opportunities. *Com. & Chem. Eng.* 28, 2441–2457.

R. Gani, K.M. Ng, 2015. Product design - Molecules, devices, functional products, and formulated products. *Com. & Chem. Eng.* 81, 70–79.

S. Jonuzaj, C.S. Adjiman, 2017. Designing optimal mixtures using Generalized Disjunctive Programming: Hull relaxations. *Chem. Eng. Sci.* 159, 106–130.

S. Jonuzaj, P.T. Akula, P.M. Kleniati, C.S. Adjiman, 2016. The formulation of optimal mixtures with Generalized Disjunctive Programming: A solvent design case study. *AIChE J.* 62, 1616–1633.

A.J. Kinloch, 1987. Adhesion and Adhesives: Sci. and Techn. *Sprin. Sci. + Bus. Media, B.V*.

Martin, T. M.; Young, D. M. Prediction of the acute toxicity (96-h LC50) of organic compounds to the fathead minnow (Pimephalespromelas) using group contribution method. Chem. Res. Toxicol. 2001, 14, 1378-1385

M. Mattei, G.M. Kontogeorgis, R. Gani, 2014. A comprehensive framework for surfactant selection and designfor emulsion based chemical product design. *Fluid Ph. Eq.* 362, 288-299.

R. Raman, I.E. Grossmann, 1994. Modelling and computational techniques for logic based integer programming. *Com. & Chem. Eng.* 18, 563–578.

P. Spencer, Q. Ye, J. Park, E. Topp, A. Misra, O. Marangos, Y. Wang, B. Bohaty, V. Singh, F. Sene, J. Eslick, K. Camarda, J.L. Katz, 2010. Adhesive/dentin interface: the weak link in the composite restoration. *Ann. of Biomed. Eng.* 38, 1989–2003.

M. Tawarmalani, N.V. Sahinidis, 2005. A polyhedral branch-and-cut approach to global optimization. *Mathem. Program.* 103, 225–249.

H. Warson, 1993, Surface coatings, raw materials and their usage, vol. 1. Edited by the Surface Coatings Association of Australia. Chapman & Hall, London, p. ix + 622.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50067-7

Sustainable carbon constrained natural gas monetization networks in industrial parks

Dhabia M. Al-Mohannadi[a,b], Jamileh Fouladi[a], Patrick Linke[a]

[a] *Department of Chemical Engineering, Texas A&M University at Qatar, Education City, PO Box 23874, Doha, Qata*

[b] *Texas A&M University, Artie McFerrin Department of Chemical Engineering, College Station, United States*

Abstract

Natural gas is a key resource for global energy supply and serves a feedstock for many processes. Each of the natural gas utilization or monetization routes, whether to value added products or use as fuel, is associated with certain emissions and cost. Therefore, the development of effective strategies for natural resource monetization is important for profitable resource utilization. However, the global community commitment to ambitious CO_2 emission reduction targets thus there is also a need to address the environmental concerns associated with natural gas monetization. Al-Mohannadi et al (2017) introduced a method that identifies efficient CO_2 reduction options in industrial parks with a natural gas allocation model. The approach produces natural gas and CO_2 networks for an industrial city that adheres for a given CO_2 target. While, the method aims at maximizing profitability of natural gas monetization, and reduces CO_2 through the use of carbon capture and utilization (CCUS) and the application of renewable energy, it does not consider the sustainability of the networks. This work will incorporate different indicators to understand the impact of the carbon constrained natural gas allocation networks on a natural gas central economy. This insight into sustainability performance will help better design for overall environmental protection and better use of natural resources. The aim of the method is to aid policymakers to quantify and understand the economic, environment and social impact of emission regulation.

Keywords: Natural gas monetization, CO_2 reduction, Sustainable process design, Optimization, Renewable Energy

1. Introduction

The demand for natural gas is increasing for its diversities and low footprint. Recent advances in hydraulic fracking have significantly boosted proven natural gas reserves and resulted in increased natural gas processing capacities in many parts of the world (American Petrolum Institute, 2015). Natural gas can be processed into a variety of products, prominent examples of which include liquid fuels, fertilizer and methanol (Al-Douri et al, 2017). This process of producing value added products out of natural gas, known as natural gas monetization, has become an important pillar of many economies. One natural gas centric economy is State of Qatar which is the leading exporter of Liquefied Natural Gas (LNG), and a major producer of fertilizer and other basic materials over the past two decades (U.S. Energy, 2015). In the meantime, 195 countries

adopted universal, legally binding global climate to avoid dangerous climate change by limiting global warming to well below 2°C in Paris Untied Nations Conference of Parties (COP 21). Thus, there is a need to averting climate change and still conitute to sustaible growth and development.

Research contributions focused reducing the carbon footprints of existing industrial complexes with process integration approaches such as the concept of CO_2 reuse between industrial processes for example Von der Assen et al (2016) look into CO_2 utilization potential through heuristics and metric system. Abdul Aziz et al (2017) created a pinch analysis graphical method for CO_2 reuse in an industrial city that takes into account heat and power demand of the industrial site. Other research was focused on identifying highly profitable natural gas utilization schemes such as Al-Sobhi and Elkamel (2015). The method does not consider carbon dioxide reduction effect on or possible carbon dioxide utilization options. Recenetly, Al-Mohannadi et al (2017) developed an apporach to identify the most promising configurations from a number of possible monetization alternatives and carbon management options that could be applied in the industrial cluster, that exploits synergies between natural gas conversion and CO_2 management.

While, Al-Mohannadi et al (2017) consider economic and environmental aspects, social aspect is not covered leading to incomplete assessment of the designs. Thus, including sustainable evaluation (economic, environmental and social elements) into the Al-Mohannadi et al, (2017) method could assist with the design and development of low carbon emission gas utilization strategies and would reduce the probability of overlooking solutions that could come from non-systematic trial-and-error approaches. This work aims to incorporate the three dimensions of sustainability into an existing optimization framework accounting for profitability, environmental impacts and evaluate social issues such as health as explained in the next section.

2. Method

This work aims to identify sustainable strategies for natural gas utilization in an industrial clusters. It builds on the approach and model proposed by Al-Mohannadi et al (2017; 2016). In Al-Mohannadi et al (2017) Mixed Integer Linear Program (MILP) formulation, the attention is limited to clusters that utilize natural gas as the primary feedstock for its plants, where the maximum supply of natural gas is limited.

The materials exchanged within the industrial cluster are natural gas (methane) and CO_2. Each plant receives natural gas supply from a common distribution infrastructure. Each plant is connected to the existing electricity grid for power export or import. Each plant is associated with emissions and contributes to the overall footprint of the cluster. Adjusting Al-Mohannadi et al, (2017) formulation to include each plant's associated with a waste emission substance that have environmental and social impacts measured using an appropriate corresponding factor.

Plants may convert natural gas to alternative products; convert or sequestrate CO_2 through carbon capture utilization and storage, produce power at reduced or eliminated specific natural gas requirements, which feeds directly into the electricity grid, e.g. wind, solar power, or other renewable power generation. The plants however, are assumed to not process other waste substances and thus the waste emission substances will be emitted to the atmosphere. The produced network is then optimized first for

given CO_2 target and second for sustainability defined target. An example is analysed in the next section and solving using WhatsBest solver (Lindo, 2006). Illustrated Example

In this work the integration and optimization is applied on a set industrial city with known processes and layout. The analysis is focused on economic performance and emissions to the air. The main optimization objective is to maximize the annual industrial city profit adhering to carbon dioxide target, which mean the economic and environmental metrics has been covered in the objective. Using the Tool for the Reduction and Assessment of Chemical and Other Environmental Impacts (TRACI) method (EPA, 2012), the social impact is quantified. TRACI is an environmental impact assessment software available with different characterization factors to assess life cycle assessment and sustainability metrics. Characterization factors measure the potential impacts following the equation below.

$$I^i = \sum_{xm} CF^i_{xm} * M_{xm} \quad (1)$$

Where:

I^i= The potential impact of all chemicals (x) for a specific impact category of concern (i)

CF^i_{xm}= The characterization factor of chemical (x) emitted for impact category (i)

M_{xm}= The mass of chemical (x) emitted

The TRACI software can be divided into four main groups: inventory of stressors, impact categories, characterization and overall effect. The impact categories studied in this work are Global Warming impact quantified using CO_2 emissions and the social metric Human Health effect quantified using particulate emissions.

3. 1. Industrial Cluster Information

The industrial cluster studied includes a set of plants, namely a Gas-to-Liquid (GTL) facility, a Cement plant, a Natural Gas Fired Power Plant, a Renewable Solar Photovoltaic Plant (PV), Enhanced Oil Recovery (EOR), Saline Storage, Methanol plant (both a standard (A), and a CO_2 -receiving plants(B)) and a Greenhouse. The plants main products and approach required information are adopted from Al-Mohannadi et al (2017) including the required plants economic information. Table 1 shows a summary of the most important parameters in addition to the sustainability metrics used for the natural gas consuming plants.

Table 1: Industrial city data

Plant	CO_2 Factor	NO_x Factor
Methanol (A)	0.50 t/t Methanol	6.3E-5kg/t Methanol
Cement	0.54 t/t Cement	2.1 kg/t Cement
Gas-to-Liquid	4.02 t / t GTL (total)	0.3 kg /t GTL
Natural gas fired power plant	0.40 t/MWh	0.5 kg/MWh

Nitrogen oxides (NO_x) data were obtained from Oryx (2015), Cochez et al (2010) and Seebregts et al (2010). Nitrogen oxide substance effect on human health is quantified in the TRACI method using the factor (0.7 PM2.5 eq/kg NO_x). While most plants in this city emit coarse particulate matter (PM10) from gas-fired turbine fine particulate matter (PM2.5) are also produced, which can cause severe human health risk (EPA, 2012).

3. 2. Results and Discussion

It was found that when the flow of methane in the city was restricted to 30E03 t/d and no CO_2 emission was imposed, the optimization selected Methanol, GTL and Cement plants to be operated to the full capacity of products. The city maintained a profit of 3E09 $/y. It was noticed that the natural gas fired power plant met the power requirement and exports by consuming natural gas. The optimization did not activate the renewable solar power option to satisfy the demand as burning methane as fuel was more profitable. Moreover, the network chose to re-use CO_2 from GTL plant to generate extra profit from the EOR sink. When a 20 % target reduction of CO_2 was imposed on the city, the network did not lose much of its profit, this was due to revenue generating CO_2 sink namely enhanced oil recovery and methanol. Both sinks were supplied by high purity CO_2 from GTL, which did not require further treatment. When the capture target was increased to 30% CO_2 reduction, the cement plant was shut off to reduce emissions and maximum PV capacity was activated. The profit reduced 2.9E09 $/y. When the network was set to maximum CO_2 reduction, the network cost was at 5.9E05 $/y. All CO_2 was allocated to EOR and Storage. All natural gas receiving plants were switched off, except the power plant as there is a constraint on the minimum power requirement needed to export outside the city.

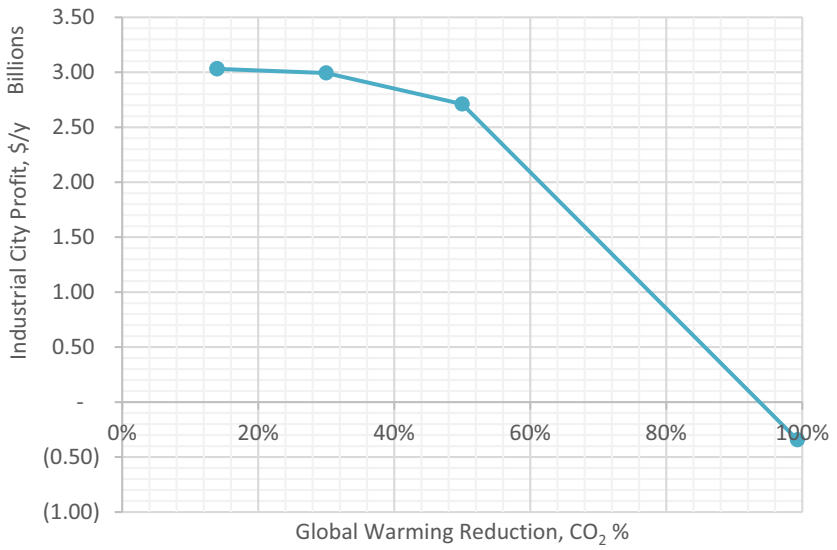


Figure 1: Industrial city's Profit vs. Global Warming Reduction

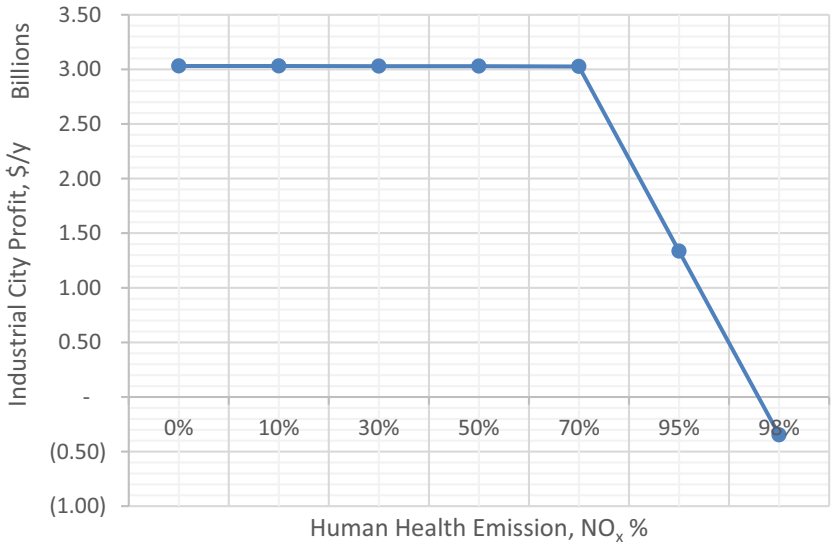


Figure 2: Industrial city's Profit vs. Negative Human Health Emission Reduction

The results of the analysis have been presented in form of a Pareto curve in Figure 1 elucidating the variation of CO_2 reduction target with the industrial city's profit. The network was also analysed to quantify the impact of human health of the city by varying the allowed NO_x emissions. The results of the analysis have been presented in form of a Pareto curve in Figure 2. When a target of 10 % reduction of NO_x emissions was set, the network allocated natural gas to maximize the capacity of the GTL and methanol plants while reducing the amount allocated to the cement plant, which was operated at 87%. When the target was increased to 30% reduction of NO_x emissions, the optimization reduced the cement output further and activated the renewable power option. When the network was solved to minimize NO_x emissions output, the network cost was -3.5E05 $/y. All natural gas plants where shut off, with the exception of the power plants due to the power export requirement.

The global warming impact and human health impact follow the same pattern when optimized against the industrial city profit. It was noticed when human health factor was reduced it also resulted in reducing CO_2 emissions. This shows that multi-objective optimization will give better insights into the trade-offs between the metrics. Furthermore, this study looked into emissions into air, the inclusion of water effluents would also enrich the assessment.

3. Conclusions

This paper introduced the concept of systematic screening of sustainable natural gas monetization in an industrial cluster. The resulting optimization-based approach can synthesize integrated natural gas, CO_2 and evaluate different defined sustainability metrics. An example is solved illustrating the applicability of the approach and showed that need to systematically screen options for different impacts in order before implementing a network design. Future work would look into multi-objective

optimization in order to get a more insightful network simultaneously instead of a stepwise approach.

Acknowledgements

Dhabia Al-Mohannadi gratefully acknowledge the support of Qatar Research Leardership Program (QRLP) under grant of Qatar National Research Fund (QRLP8-G-3330082).

References

E. Abdul Aziz, S. Wan Alwi, J. Lim, Z. Abdul Manan, J. Klemes, 2017, An integrated pinch analysis framework for low CO2 emissions industrial site planning. Journal of Cleaner Production, 146, 125-138

A. Al-Douri, D. Sengupta, M. El-Halwagi, 2017, Shale gas monetization – A review of downstream processing to chemicals and fuels, Journal Of Natural Gas Science and Engineering, 42, 436-455.

D. M. Al-Mohannadi, K. Abdulaziz, S.Y. Alnouri, P. Linke, 2017, On the synthesis of carbon constrained natural gas monetization networks, Journal of Cleaner Production, https://doi.org/10.1016/j.jclepro.2017.09.012.

D. M. Al-Mohannadi, P. Linke, 2016, On the Systematic Carbon Integration of Industrial Parks for Climate Footprint Reduction Journal of Cleaner Production, Volume, Issue, 4053-4064

S.A. Al-Sobhi, A. Elkamel, 2015, Simulation and optimization of natural gas processing and production network consisting of LNG, GTL, and methanol facilities. Journal of Natural Gas Science and Engineering. 23, 500-508

American Petroleum Institute. Facts about Shale Gas. Available at: http://www.api.org/policy and-issues/policy-items/exploration/facts_about_shale_gas. Accessed December 15, 2015.

E. Cochez, W. Nijs, G. Simbolotti, G. Tosato, 2010, IEA ETSAP: Cement Production, International Energy Agency, Energy Technology System Analysis Program, Energy Technology Network, Technology Brief I03, June 2010

J. Fouladi, 2017, A systematic approach for desigining industrial park integration networks across the water energy nexus, MS Thesis, Texas A&M University at Qatar, Doha, Qatar.

Lindo, What'sBest! 9.0,2006, Excel Add-In for Linear, Nonlinear, and Integer Modeling and Optimization; Lindo Systems: Chicago, IL

A.J. Seebregts, G. Simbolotti, G. Tosato, 2010, IEA Technology Brief: Gas Fired Power Plant, Technology Brief E02 – April 2010, International Energy Agency, Energy Technology Systems Analysis Programme

D. Tanziland, B. Beloff, 2006, Assessing Impacts: Overview on Sustainability Indicators and Metrics, Environmental Quality Management 15,4,41-56.

N. von der Assen, L. Müller, A. Steingrube, P. Voll, A. Bardow, 2016, Selecting CO2 Sources for CO2 Utilization by Environmental-Merit Order Curves, Environtal Science and Technology, 50, 1093−1101

Oryx, 2015, Sustainbulity Report 2015, Available at: http://www.oryxgtl.com.qa/sustainability-reports/2015/files/mobile/index.html#39, Accessed on November 28 2017

United States Environmental Protection Agency (EPA), 2012, Tool for the Reduction and Assessment of Chemical and other Environmental Impacts (TRACI).

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50068-9

Challenges in Replacing Heuristics-Based Trial-and-Error Procedures by Mathematical Optimization for Basic Equipment Design

André L. H. Costa[a,*], Miguel J. Bagajewicz[b]

[a]*Rio de Janeiro State University (UERJ), Rua São Francisco Xavier, 524, Maracanã, CEP 20550-900, Rio de Janeiro, RJ, Brazil.*
andrehc@uerj.br
[b]*School of Chemical, Biological and Materials Engineering, University of Oklahoma, Norman, Oklahoma USA 73019.*
bagajewicz@ou.edu

Abstract

Even though the last decades have seen an explosion of simulation and optimization tools in Process Systems Engineering, this has not had a significant impact in the basic design of equipment such as heat exchangers, condensers, reboilers, distillation columns, absorption, adsorption columns, etc. To this date most mainstream textbooks advocate the use of trial and error procedures, while scattered work has been done using MINLP methods. In addition, these methods are most of the time local and/or do not converge if no good initial points are provided. We provide here alternatives to formulate global and robust rigorous models, sometimes linear. In this paper, we discuss the challenges that this quest is presenting.

Keywords: Process Equipment Basic Design, Optimization Models

1. Introduction

Practitioners employ computational tools for equipment design, but they are limited to analyze the behavior of proposed solution alternatives in a task that is highly dependent on the engineer experience. Despite the wide utilization of modern simulation tools, the rationale of the basic design procedure is usually based on trial-and-error procedures guided by heuristic rules and rarely search optimality. This philosophy is the established mainstream paradigm in design books for education and in many specialized texts. For example, in the area of heat exchanger design, the procedures recommended in current textbooks are based on the aforementioned trial-and-error schemes. These procedures contain rules of thumb or heuristic options that usually end when a solution unit performs the desired heat transfer task with an acceptable pressure drop, but without a focus on cost minimization. These procedures require the intervention of knowledgeable personnel to make good choices during the search. Heat exchanger rating software employ distributed models and up-to-date correlations, but the design solutions are obtained using simple search algorithms with several limitations. The same can be said about the design of separation vessels and distillation columns.

Advances of optimization models and mathematical programming algorithms have been made by the PSE community, but they are seldom used in the solution of equipment design problems in practice and remain difficult to solve in general. In the area of heat

exchangers is where MINLP models, in many cases difficult to solve, even locally, have proliferated.

This paper presents a discussion about the drawbacks that restrain the utilization of mathematical programming for the solution of design problems, such as: a) The utilization of simplified models obtained through analytical solutions; b) Convergence problems originated from nonlinear (MINLP) formulations. Aiming to circumvent these problems we propose some possible solutions involving: 1) The introduction of discrete variables, and reformulations to attain simpler models; 2) The need for models that represent transport properties as a function of local conditions expressed through state variables (temperature, pressure, concentration), as opposed to overall averages through the equipment. To illustrate the ideas, we show examples of simple MINLP models, MILP and IP models for the design of heat exchangers as compared to heuristics, and how global optimization can be implemented in the case where the resulting models cannot be completely and rigorously reformulated to be linear.

Finally, we clarify that the reformulations we are proposing are such that no approximations, linearizations by truncating Taylor series or other simplifications are made. In other words, the reformulated model is such that every solution of one model is feasible in the other, that is, no loss of rigor takes place.

2. Heuristic-based Trial and Error Procedures

Consider the case of the design of shell-and-tube heat exchangers. Heuristic procedures, presented in most design books (Towler and Sinnot, 2012, for example) and heat transfer books (Serth, 2007; Cao, 2010), are based on trial-and-error schemes, such as the set of steps presented below:

(1) Make initial specifications, such as fluid allocation (tube-side x shell-side), shell and head types, and heat exchanger configuration (e.g. number of tube and shell passes). Pick a combination of tube diameters, tube lengths, baffle spacing, etc; (2) Start estimating the area using some good estimate of the overall heat transfer coefficient (U^{est}), that is

$$A^{est} = \frac{\hat{Q}}{U^{est}\widehat{\Delta T}_{lm}F} \quad (1)$$

where $\hat{Q}$ and $\widehat{\Delta T}_{lm}$ are given according to the thermal task; (3) Pick a number of tubes, compatible shell diameter, such that the area matches closely to A^{est}; (4) Calculate heat transfer coefficients for the tube side and the shell side and with a recommended fouling coefficient, calculate a new overall heat transfer coefficient, and the pressure drops resulting from using the above geometric choices; (5) If the overdesign is acceptable, and the pressure drops are below recommended maximums, then the design is usually finished. Otherwise, some clever adjustments are made based on the analysis of the current trial, aiming at fixing the problems observed. After that, a new evaluation of the new solution candidate is executed (return to Step 4).

Little is said by these procedures, as to how the iteration is handled, when a new trial is needed, neither recommendations are given for the cases where the new calculated value of overall heat transfer coefficient is excessively larger that the value estimated. This is handled by somebody with expertise. One can see a similar approach for the case of designing vertical and horizontal flash-units, distillation column tray design, etc.

In conclusion: most recommended procedures do not incentivize iterative procedures, only trial and verification, that is they provide heuristics and require expertise to obtain a feasible answer, not an optimal one.

The limitations of the traditional design approaches present an opportunity for mathematical programming to contribute with solutions. Despite the number of papers published that tries to fill this gap, there are important limitations that hinder the utilization of such approaches for practical problems, as discussed in the next section.

3. Optimization Models

In the last few decades, several authors started to present different optimization approaches to the design of heat exchangers. Nowadays, the two main approaches employed are stochastic methods and mathematical programming. The utilization of mathematical programming efforts has been dominated by nonconvex formulations, encompassing older nonlinear programming approaches (Jegede and Polley, 1992) or more recent mixed-integer nonlinear programming solutions (Ravagnani and Caballero, 2007). Of all the above approaches, mathematical programming is only one that has any chance of achieving and guarantee globally optimal solutions.

A more detailed analysis of the literature about the utilization of mathematical programming techniques for solving the design problem of shell-and-tube heat exchangers indicates several aspects that limit the utilization of these approaches for practical applications:

a) The utilization of nonlinear formulations that are associated to convergence problems and multiple local optima with different values of objective function. The lack of robustness is an important barrier for the acceptance of this kind of tool by a broader audience, including practitioners;
b) Utilization of heat exchanger models based on analytical solutions where the heat transfer coefficients have a uniform value along the heat transfer area. This problem is particularly important for services with phase change (vaporization or condensation), where the conditions along the flow can change considerably.In the next sections, we discuss how these challenges have been addressed in our group.

4. Use of Discrete Variables

Our proposal is to use a set of discrete values for geometric variables, thus allowing a representation of the design solution closer to the engineering practice. Additionally, having the design variables in a discrete space allows reformulation to a linear form.

For example, heat exchanger tubes come in discrete options (3/4", 1", 1 ¼", 1 ½", 2") with different wall thickness dictated by pressure (BWG 12, 14, 16, etc.) and different discrete options for length. The same can be said for shell diameters and other geometric variables that have been standardized (TEMA). A similar case can be made for other equipment. Thus, the diameter is expressed in terms of discrete options as follows:

$$dti = \sum_{sd=1}^{sdmax} \widehat{pdti}_{sd}\, yd_{sd} \tag{2}$$

where yd_{sd} is a binary variable and $\widehat{pdti}_{sd}$ is the corresponding discrete value. The expression also requires to be accompanied by:

$$\sum_{sd=1}^{sdmax} yd_{sd} = 1 \tag{3}$$

Consider now for example the calculation of the heat transfer coefficient for the tube-side flow using the Dittus-Boelter correlation:

$$ht = \frac{0.023\, Ret^{0.8} \widehat{Prt}^{n}\, \widehat{kt}}{dti} \tag{4}$$

Assuming one tube-side pass, the expression of the Reynolds number in relation to the mass flow rate is:

$$Ret = \frac{dti\, v\, \widehat{\rho t}}{\widehat{\mu t}} = \frac{4\ \widehat{mt}}{Ntt\ \pi\ \widehat{\mu t}\ dti} \tag{5}$$

Thus, substituting, one obtains:

$$ht = \frac{0.023 \left(\frac{4\ \widehat{mt}}{\left[\sum_{sNtt=1}^{sNttmax} \widehat{pNtt}_{sNtt}\, yNtt_{sNtt}\right] \pi\ \hat{\mu}\ \sum_{sd=1}^{sdmax} \widehat{pdti}_{sd}\, yd_{sd}} \right)^{0.8} \widehat{Prt}^{n}\, \widehat{kt}}{\sum_{sd=1}^{sdmax} \widehat{pdti}_{sd}\, yd_{sd}} \tag{6}$$

Note that we replaced the total number of tubes by a sum of discrete options. The above expression is nonlinear in the binary variables and requires reformulation if it is run by most commercial MINLP.

5. Reformulation

We first realize that the product of summations elevated at a certain exponent, even negative, can be rewritten in form that contains products of integers, as follows:

$$[\sum_i \widehat{pd}_i\, yp_i]^{n1}\, [\sum_j \widehat{qd}_j\, yq_j]^{n2} \ldots [\sum_k \widehat{zd}_k\, yz_k]^{nm} = p^{n1} q^{n2} \cdots z^{nm} = \sum_{i,j,\ldots k} \widehat{pd}_i^{\,n1}\, \widehat{qd}_j^{\,n2} \ldots \widehat{qd}_k^{\,nm}\, yp_i\ yq_j \ldots yz_k \tag{7}$$

Thus Eq.(6) can be rewritten as follows:

$$ht = 0.023\, \widehat{Prt}^{n}\, \widehat{kt}\, \left[\frac{4\ \widehat{mt}}{\pi\ \hat{\mu}}\right]^{0.8} \sum_{sd=1}^{sdmax} \sum_{sNtt=1}^{sNttmax} \frac{yNtt_{sNtt} yd_{sd}}{\widehat{pdti}_{sd}^{\,1.8}\, \widehat{pNtt}_{sNtt}^{\,0.8}} \tag{8}$$

A generalized expression for this equation, valid for any number of tube passes, including the relations associated to the tube count in relation to the shell diameter can be find in Gonçalves et al. (2016). We emphasize again that we arrive at Eq.(8) by rigorous reformulation, that is without making any approximation of any sort.

In general, we expect, products of binaries and continuous variables as well as fractional terms, which we know can be easily and rigorously linearized (Williams, 2013). The resulting formulation can be MILP or even IP (Gonçalves et al., 2017). Now, we are exploring this approach for other alternatives of heat exchangers, such as double-pipe, air coolers, plate-and-frame, and plate-and-fin; there are also investigations using a similar approach to the design of the internals of distillation columns, flash units and separators.

Another alternative for representation of the discrete variables is the utilization of a table that contains all combinations of the discrete values, i.e. each row of the table is a solution candidate. Therefore, instead of using several sets of binary variables, this

formulation employs a single set, associated to the table rows. Numerical tests indicated that this approach can bring large reductions of the computational effort (Gonçalves et al., 2017).

6. Global Optimization

Sometimes, the resulting model contains continuous variables participating in non-linear terms that are not amenable to be discretized. We illustrate this with the case of an air cooler. Hitherto, we solved this problem using a fixed air flow rate, which yielded an IP problem (Souza et al., 2017). Now, we are exploring the case of variable air flow, which leads to a model that is nonlinear in only one continuous variable (the air flow rate). The air flow rate participates in several expressions used to calculate the Reynolds number and consequently the overall heat transfer coefficient. This can be solved using any global optimizer (we tried Baron, Antigone and Rysia (Faria and Bagajewicz, 2012)). One can think of just using different flow rates and solve the IP fixed flow problem repeatedly, by using several discrete values of flow, or incorporate such discretization to the model and reformulate. Both options are mathematically not rigorous, but practically sound.

We also find the same situation in the case of heat exchangers with fouling being modelled. Traditionally, the design of heat exchangers addresses the fouling problem using fixed values of fouling resistances. However, the literature presents fouling models that can be embedded in the design equations, thus allowing the problem of fouling mitigation to be inserted into the optimal design problem. Sometimes, it is possible to include the fouling model and still generate a linear problem (Lemos et al., 2017). However, in more complex systems, these models may bring to the problem continuous variables involved in nonlinear terms that cannot be eliminated (e.g. fouling layer thickness), which demands MINLP optimization algorithms.

7. Local models

One of the challenges ahead is to move away from simplified models based on analytical solutions (e.g. LMTD method) and address the optimization problem using discretized portions of the equipment where local properties are calculated using conservation equations. This issue becomes especially important for phase change services where the nature of the streams present a large variation along the flow path (e.g., in a total condenser, the inlet stream is saturated vapour and the outlet stream is saturated liquid). Instead of only use the end temperatures, the model will contain equations related to the set of temperatures distributed along the flow path.

8. Implications for Process Design/Retrofit

The current paradigm in PSE regarding process design and/or retrofit is one where one first use approximate models to determine structure and then eventually perform basic equipment design. The alternative is to incorporate detailed equipment models into the whole optimization. We believe that the PSE community needs to move in the latter direction and increasing computing power will render better designs/retrofits can be accomplished. For example, consider Heat Exchanger Network design/retrofit: With a few exceptions, all existing methods optimize using a-priori selected values of heat transfer coefficients. We believe that even structural changes can come from using

different values. The only chance address it is to include the detailed exchanger design into the modelling. This thought is extended to the whole process design.

9. Conclusions

PSE has been attained remarkable advances in the last decades, providing a myriad of computational tools to solve chemical engineering problems. For example, process simulators are widely employed to provide solutions in chemical process industries. However, despite the advances obtained in problem formulation and solution algorithms, the design problems are still solved in practice using the same trial-and-error approach.

Possible reasons that hinder the direct utilization of PSE tools to solve real design problems were discussed along this paper. The lack of robustness and the simplified physical modelling are two aspects that hinder the utilization of the optimization tools for solving design problems. We are trying to address the former issue using reformulation techniques to generate linear optimization problems and, when this approach is not sufficient, we suggest the utilization of global optimization solvers. We are also investigating solutions for the latter issue based on the adoption of discretized models for the formulation of the design problem, therefore overcoming the lack of accuracy of analytical solutions in certain situations (e.g. condensers and vaporizers).

References

E. Cao, 2010, Heat Transfer in Process Engineering, McGraw-Hill.

D. Faria, M.J. Bagajewicz, 2012. A New Approach for Global Optimization of a Class of MINLP Problems with Applications to Water Management and Pooling Problems. AIChE J. 58, 8, 2320-2335.

C.O. Gonçalves, A.L.H. Costa, M. J. Bagajewicz, 2016, Shell and Tube Heat Exchanger Design Using Mixed-Integer Linear Programming , AIChE J., 63, 1907-1922.

C.O. Gonçalves, A.L.H. Costa, M.J. Bagajewicz, 2017, Alternative MILP Formulations for Shell and Tube Heat Exchanger Optimal Design, Ind. Eng. Chem. Res., 56, 5970-5979.

F.O. Jegede, G.T. Polley, 1992, Optimum Heat Exchanger Design, Trans Inst Chem Eng, 70(A2), 133–141

J. C. Lemos, A. L. H. Costa, M. J. Bagajewicz, 2017, Linear method for the design of shell and tube heat exchangers including fouling modeling, App. Therm. Eng., 125, 1345-1353.

M.A.S.S. Ravagnani, J.A. Caballero, 2007, A MINLP Model for the Rigorous Design of Shell and Tube Heat Exchanegrs Using TEMA Standards, Chem. Eng. Res. Des., 85, 1423-1435.

R.W. Serth, 2007, Process Heat Transfer - Principles and Applications, Academic Press.

P.A. Souza, A. L. H. Costa, M. J. Bagajewicz, 2017, Globally Optimal Linear Approach for the Design of Process Equipment: The Case of Air Coolers, AIChE J., accepted for publication.

G. Towler, R.K. Sinnot, 2012, Chemical Engineering Design: Principles, Practice and Economics of Plant and Process Design, Butterworth-Heinemann, 2nd ed.

H. Williams, 2013, Model Building in Mathematical Programming, 5th ed., Wiley.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50069-0

A General Framework for Process Synthesis, Integration and Intensification

Salih Emre Demirel, Jianping Li, M. M. Faruque Hasan*

Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77845, USA.
**Corresponding Author Email: hasan@tamu.edu.*

Abstract

Block-based superstructure (Demirel et al., 2017; Li et al., 2017) offers advantages over classical unit operation-based process synthesis approaches as it leverages upon building blocks to represent different physicochemical phenomena, processing tasks and equipment configurations. In this work, we extend this approach to demonstrate that the same block superstructure representation can be used to perform simultaneous synthesis, integration and intensification in a unified framework. The block-based representation also enables automatic generation of intensified flowsheets without *a priori* postulation of connectivity. Moreover, it is generic enough to be reduced to various process integration network problems, such as heat exchanger networks, fuel gas networks, water networks, etc. A literature example is used to demonstrate the benefits of block superstructure for the particular case of flowsheet generation and heat exchanger network synthesis (HENS).

Keywords: Block Superstructure, Process Synthesis, Intensification, Integration.

1. Introduction

Optimization-based process synthesis provides a systematic pathway for identifying optimal process flowsheets from numerous alternatives. However, most synthesis methods are currently limited in scope for "out-of-the-box" design solutions as they require pre-postulated superstructures. This also prohibits innovative equipment design which plays a critical role in process intensification (Stankiewicz and Moulijn, 2000). To this end, we have proposed building block superstructure (Demirel et al., 2017; Li et al., 2017) which eliminates the need for pre-specified equipment types, and automates the search for intensified process alternatives.

In this work, further benefits of the building block superstructure are presented via extending the formulation to include simultaneous heat integration and to form a comprehensive framework for intensified conceptual process design. As the proposed superstructure automatically generates the flowsheet alternatives without any *a priori* information on the final flowsheet, position and the identity (i.e. hot/cold) of the integrated streams are not known beforehand. Accordingly, proposed formulation is for simultaneous process synthesis and heat integration network synthesis with unclassified process streams. This opens up the potential of utilizing the block superstructure as a general framework for process synthesis, heat integration and intensification. The benefits of the proposed approach is demonstrated via a literature example and it is shown that significant advantages can be obtained by utilizing building block superstructure in conceptual process design.

2. Block Superstructure in a Grid Formation

Block superstructure is shown in Figure 1. Here, each block can be used to represent different chemical phenomena, tasks, materials and equipment. A detailed discussion on how to use building blocks to represent various phenomena are given in Demirel et al. (2017). Each block is connected with neighbouring blocks via bidirectional inter-block streams. Bidirectional flow representation enables incorporation of the recycle connections into the superstructure. Furthermore, heater/cooler and expander/compressor equipment are positioned on each stream so as to quantify the utility requirements in the superstructure and to consider operating and capital costs for these equipment. Each block can also have feed and product streams to facilitate interaction with external environment.

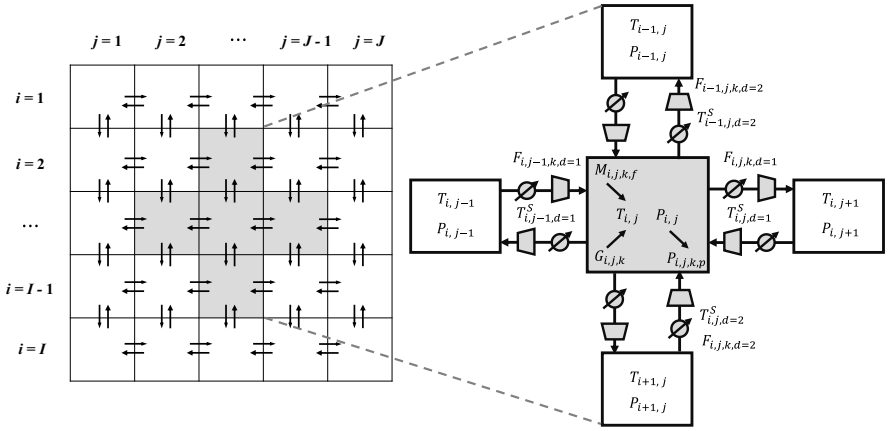


Figure 1. A schematic of the block superstructure.

With this representation, many process synthesis superstructures can be incorporated into the block superstructure. An example superstructure taken from literature (Kong et al., 2017) and its block superstructure representation are illustrated in Figure 2. Here, reactors are represented as single blocks while separators are represented via block boundaries between two blocks to indicate the position of the separated stream.

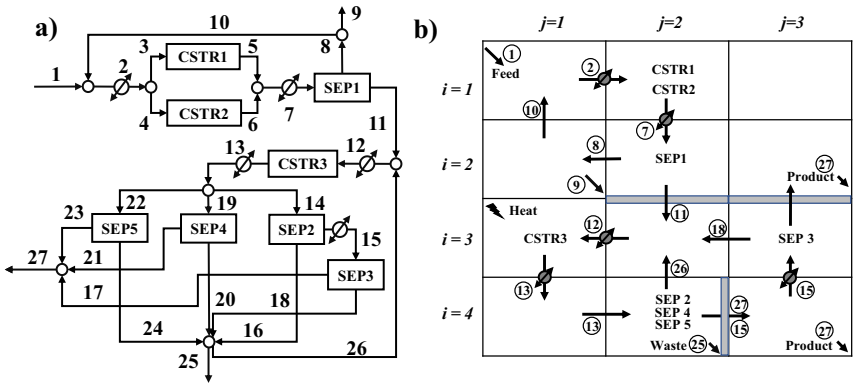


Figure 2. Superstructure representations (a) unit-operation based, (b) block based.

Different process integration networks can be also included in the same superstructure representation. Each stream contains a heater/cooler and the identity of the stream (i.e. hot or cold) and corresponding utility consumption can be determined via an energy balance around each stream. If the utility requirement of a hot and cold stream at two different positions are the same, then they can be integrated to each other. If no such stream available, then the heating/cooling requirement is satisfied via external utilities. This is illustrated in Figure 3a. Although each stream is allowed to match with any other stream in the superstructure, these matches can be restricted to certain positions to yield the same superstructure with the one given by Yee and Grossmann (1990), in which case, each column represents one stage. In Figure 3b, this is illustrated for 2 hot and 2 cold streams. Stream splitting can also be achieved via increasing the number of rows that a stream can go through. For instance, if all the streams in Figure 3b can go through splitting, then 8×5 superstructure with each stream occupying two rows would be used.

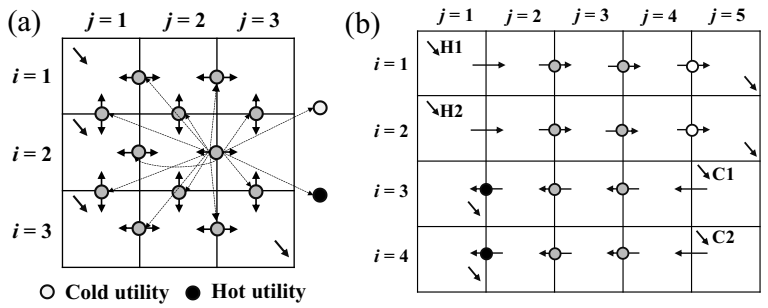


Figure 3. Heat exchanger network representation using block superstructure.

3. Superstructure Model for Synthesis, Heat Integration and Intensification

The overall model for process synthesis, integration and intensification is based on a single mixed integer nonlinear programming (MINLP) and summarized below. Model contains block superstructure constraints for generating the process flowsheet and HENS constraints for determining the corresponding HEN.

$$\min_{x,x\prime,z,z\prime,y,w} f(x,x',z,z') + h(y,w) \tag{1}$$

$$g_{i,j}(x,x') = 0 \quad i \in I, j \in J, \quad g'_{i,j,d}(x,x') = 0 \quad i \in I, j \in J, d \in D \tag{2}$$

$$q_{i,j}(z) \le 0 \quad i \in I, j \in J, \quad q'_{i,j,d}(z') \le 0 \quad i \in I, j \in J, d \in D \tag{3}$$

$$r_{i,j}(x,x',z) \le 0 \quad i \in I, j \in J, \quad r'_{i,j,d}(x,x',z') \le 0 \quad i \in I, j \in J, d \in D \tag{4}$$

$$G_{l,l\prime}(x',w) = 0 \quad l,l' \in L \; l < l' \tag{5}$$

$$Q_l(y) \le 0 \quad l \in L \tag{6}$$

$$R_{l,l\prime}(w,y) \le 0 \quad l,l' \in L \; l < l' \tag{7}$$

$$x'^L \le x' \le x'^U \quad x^L \le x \le x^U \quad w^L \le w \le w^U \quad x \in R^{I\times J}, x' \in R^{I\times J\times D}, w \in R^{L\times L}$$
$$z = \{0,1\}^{I\times J} \; z' = \{0,1\}^{I\times J\times D} \; y = \{0,1\}^{L\times L}$$

The overall superstructure model is comprised of objective function, Eq.(1), which minimizes the annualized cost while considering operating and capital cost for the process, i.e. $f(x, x', z, z')$, and corresponding heat exchanger network, i.e. $h(y, w)$, process related constraints Eqs.(2)-(4), and HENS model constraints Eqs.(5)-(7). Each block is designated with its position in the superstructure as $B_{i,j}$, in which $i = 1, \dots, I$ and $j = 1, \dots, J$ are row and column indices, respectively. There are mainly two type of variables in the superstructure model: block, i.e. x, z, and boundary, i.e. x', z', variables. Block variables, for instance, include $P_{i,j}$ and $T_{i,j}$ which stands for block pressure and temperature, respectively. Boundary variables, on the other hand, are defined to represent flow rate of the interblock streams, $F_{i,j,d}$, and stream temperatures, $T^s_{i,j,d}$ in which $d = \{1,2\}$ designates the orientation of the boundary variables as horizontal ($d = 1$) or vertical ($d = 2$). While block binary variables, z, are used to assign reaction related phenomena to the blocks, boundary binary variables, z', are used to assign separation related phenomena or operation to the boundaries in the superstructure. In process related constraints, Eqs.(2) indicate block material and energy balances and stream energy balances, Eqs.(3) designate logical constraints, and Eqs.(4) relate continuous and binary variables.

In the HENS model, block indices are mapped into a single index, $i, j, d \rightarrow l$ where $l = 1, \dots, L$ is an ordered set and contains all the streams in the superstructure, i.e. $L = 2 \times I \times J$. Note that, as the model is posed as a simultaneous synthesis and heat integration model, and the identity of the streams, i.e. hot/cold, in the superstructure are not known beforehand, binary variables are used to determine stream identities. Accordingly, HENS model is similar to the one proposed by Yee and Grossmann (1990) except that it is extended for unclassified process streams. While continuous variables,w, are used to represent the heat duty for each matching stream in the superstructure and determine approach temperatures for these matches, binary variables, y, are used to determine the stream matches within the superstructure, and classifying streams as hot or cold. In HENS model related constraints, Eq. (5) stands for heat balances for each stream, Eq. (6) designates logical constraints for determining the existence of a match and classification of a stream, and Eq. (7) relates HENS model binary variables with continuous variables.

4. Case Study

An example problem adapted from Kong et al. (2017) is used to demonstrate the use of block superstructure for automatic flowsheet generation first and, then for determining its corresponding HEN. The original superstructure given by Kong et al. (2017) and its block representation is given in Figure 2. There are four components in the process: A and B are used as raw materials to produce C (R1: A + B → C), which is further utilized in a second reaction to obtain the final product D (R2: C ↔ D). There are two isothermal reactor alternatives for R1: CSTR1 (500 K) and CSTR2 (400 K). R2 is an equilibrium reaction carried out in an isothermal equilibrium reactor: CSTR3 (330-400 K). There are five separator alternatives. While Sep1 (430 K) is used to separate C from R1 reactants, Sep2-5 are used to separate product D from C. Note that Sep1 and CSTR3 requires hot utility. The aim of the problem is to synthesize a process for maximizing the total annual profit considering product revenue, unit capital costs and utility costs. All the equipment parameters along with their cost parameters are taken from Kong et al. (2017). First, the problem is solved via fixing the binary variables related with the unit operations and flow directions, as shown in Figure 2 and the original superstructure

is obtained. When this fixed superstructure solved with ANTIGONE, it yields the global solution with an objective value of $14,004,425 in 152 CPU s.

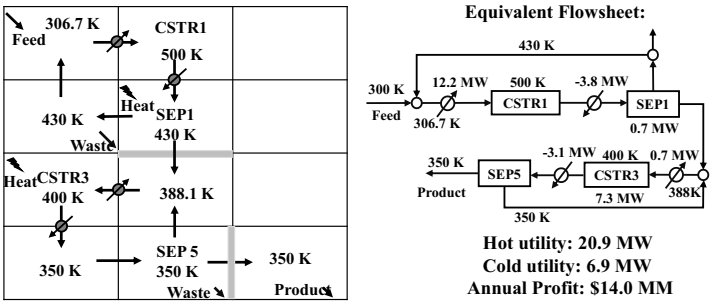


Figure 4. Solution obtained from the original superstructure.

Resulting block superstructure solution and its equivalent flowsheet are shown in Figure 4. Then, the same problem is solved for 22 h with ANTIGONE for automatic flowsheet generation without any fixing and with an initial solution provided by the strategy outlined by Li et al. (2018). The flowsheet identified by block superstructure has an annual profit of $14,171,800 which is $167,300 higher than the process flowsheet obtained via the original superstructure. This result is shown in Figure 5. The improvement is due to the change in the position of the unreacted recycle stream from Sep1. As Sep1 operates a lower temperature than the CSTR1, the recycle stream from top of the Sep1 needs to be heated before entering into CSTR1 (see Figure 4). However, recycling reactor outlet directly facilitates a higher mixing temperature at the feed mixer and decreases the cost associated with the recycle stream. Accordingly, flow rate of the recycle stream can be increased which contributes to higher overall conversion. Note that this structural alternative is not included in the original superstructure and block superstructure could identify this connection without pre-specifying its existence.
.

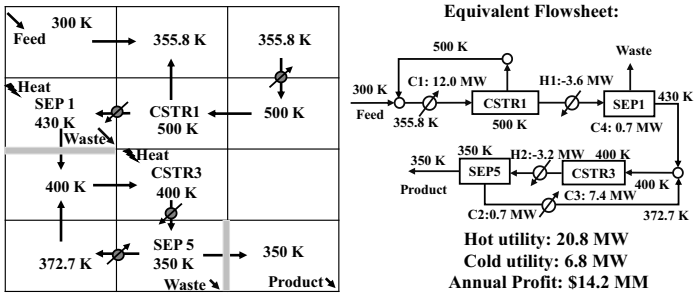


Figure 5. Final block superstructure result.

HEN for this result can also be addressed via block superstructure. The final flowsheet contains 4 cold streams (C1-C4 while C3 and C4 are isothermal) and 2 hot streams (H1 and H2). Accordingly, 6×7 block superstructure can be used to represent HEN superstructure as shown in Figure 6. Note that in Figure 6, the resultant block structure and only the matched streams are shown with a heater/cooler symbol. While first two rows are used for two hot streams, other rows are used for cold streams. First and last columns are used to represent hot and cold utility consuming heat exchangers,

respectively. Isothermal cold streams are handled with 1 K fictitious temperature difference. As the identity of the streams are known, binary variables associated with the stream classification are fixed accordingly. The objective is to synthesize a HEN with minimum annual cost while considering fixed and variable capital costs and operating costs. The optimal block superstructure result is obtained in 7 CPU s with ANTIGONE and corresponding HEN is also shown in Figure 6. Resulting network has a total annual cost of $1,566,440 and contains two heat exchangers (between C1-H1 and C1-H2), four heaters and one cooler. Since the heat content of the hot streams are low, three cold streams (C2, C3, and C4) could not be integrated and, they require hot utility.

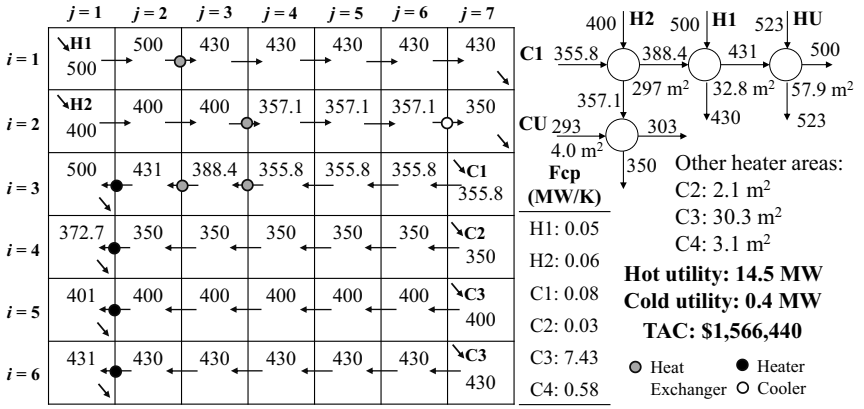


Figure 6. Optimized HEN for the generated flowsheet.

5. Conclusions

In this work, block superstructure model is extended to include simultaneous HENS. The use of the model is illustrated via a literature example first for automated flowsheet generation and then constructing the corresponding HEN. Note that superstructure model developed in this work is also applicable for simultaneous HENS. Future work will focus on utilizing the full model and demonstrating the capabilities of block superstructure for simultaneous process synthesis, heat integration and intensification.

References

S. E. Demirel, J. Li, M. M. F. Hasan, 2017, Systematic process intensification using building blocks, Computers & Chemical Engineering, 105, 2–38.

J. Li, S. E. Demirel, M. M. F. Hasan, 2017, Simultaneous process synthesis and process intensification using building blocks, Computer Aided Chemical Engineering, 40, 1171–1176.

J. Li, S. E. Demirel, M. M. F. Hasan, 2018, Process Synthesis using block superstructure with automated flowsheet generation and optimization , Submitted.

L. Kong, V. Avadiappan, K. Huang, and C. T. Maravelias, 2017, Simultaneous chemical process synthesis and heat integration with unclassified hot/cold process streams, Computers & Chemical Engineering, 101, 210-225.

A. I. Stankiewicz, and J. A. Moulijn, 2000, Process Intensification: Transforming Chemical Engineering. Chemical Engineering Progress, 1, 22–34.

T.F. Yee, and I.E. Grossmann, 1990, Simultaneous optimization models for heat integration—II. Heat exchanger network synthesis, Computers & Chemical Engineering, 14(10), 1165-1184.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50070-7

Design and optimization of plate heat exchanger networks

Kexin Xu*, Robin Smith

Centre for Process Integration, School of Chemical Engineering and Analytical Science, The University of Manchester, M13 9PL, UK.
kexin.xu@manchester.ac.uk

Abstract

Heat transfer enhancement technologies have been widely used in heat exchanger network (HEN) retrofit due to its relatively low retrofit cost. However, for enhancement to be effective, one of the side film coefficients must be controlling on one of the side that presents a higher magnitude. Alternatively, plate heat exchangers (PHE) have relatively small minimum approach temperature and high heat transfer coefficient, and they do not rely on the differences between coefficients. Applying PHE into retrofit design of HEN is considered as an attractive option as it can enhance heat recovery and reduce energy consumption. The first part of this paper proposes an optimal design of two-stream multi-pass plate heat exchangers, including gasket plate heat exchangers and welded plate heat exchangers. An MINLP model is formulated to minimize the total area of a PHE by selecting the best plate geometries, chevron angle and flow arrangement. The second part of this paper presents a new cost-effective methodology of applying the optimized PHE model in HEN retrofit with a fixed network structure. Sensitivity analysis is used to identify the best heat exchangers to replace. A non-linear optimization based model is formulate to deal with downstream effect after replacement. A case study highlights the benefits of the new approach of applying PHE in HEN retrofit.
Keywords: plate heat exchanger, heat exchanger network, retrofit, optimization,

1. Introduction

With the growth of energy consumption and increase of environmental degradation, improving heat transfer efficiency became an important task. HENs retrofit is a cost-effective way to increase heat recovery based on the existing heat exchanger network in process industries. Adding additional heat transfer area and topology modifications are the two main methods to improve energy efficiency in retrofit over the past years. In practice, these conventional methods can lead to high cost retrofit due to amount of pipe work and shutdown time. Heat transfer enhancement is a relatively cost-effective technique for HENs retrofit as it can be applied during the maintenance period. However, for the enhancement to be effective, at least 50% of overall heat transfer resistance must be with either the tube-side film coefficient or shell-side coefficient. Besides, energy saving is restricted by the limitation of the heat transfer coefficient.

PHEs are one of the most efficient types of heat transfer equipment over other types of heat exchangers (Kakac et al. 2012). Compared with conventional shell-and-tube heat exchangers, PHEs could significantly increase energy efficiency with minimum approach temperatures as low as 2° C, decrease the possibility of fouling, reduce the fuel consumption and CO_2 emissions. PHEs are widely used in energy-intensive

process industries, such as food, pharmaceutical, petrochemical plants, and refineries. Based on these distinct advantages, applying PHEs in traditional shell-and-tube heat exchanger networks retrofit can be an attractive option as it allows high energy saving without the need of topology modifications and additional heat transfer area.

Plate geometry, chevron angle and flow arrangement are the three main factors that affect heat transfer in PHEs. However, most previous studies proposed different methods to investigate the effect of no more than two factors on heat transfer. Arsenyeva et al. (2011) proposed a relatively complete design method for plate heat exchangers. However, the method is time-consuming and more suitable for a rating problem by employing the ϵ-NTU method. Thus, this paper presents a computer aided approach to automatically optimize a single multi-pass plate heat exchanger by using logarithmic mean temperature difference (LMTD) method in thermal-hydraulic design to reduce computation time, with the consideration of various flow arrangement selections, plate pattern selection and pressure drop constraints.

Pinch analysis, mathematical programming and hybrid methods are the three main retrofit techniques in existing HENs. Sreepathi and Rangaiah (2014) presented a detailed review of different methods for HENs retrofit. Heat transfer enhancement is a relatively cost effective retrofit option compared to adding area or topology modifications as existing HENs structure can be maintained. A novel MILP based iterative method was proposed by Pan et al. (2014) to retrofit HENs with heat transfer enhancement. The work done by Akpomiemie and Smith (2015) extended this methodology to account for the downstream effects on the network after the application of heat transfer enhancement. However, for heat transfer enhancement to be effective, one of the film coefficients must be controlling. Besides, although heat transfer enhancement techniques can be performed during the normal shutdown time, the potential risk of damaging the heat exchangers when adding inserts or fins is unneglectable. Thus, to overcome these drawbacks, this work presents a new retrofit method that provides insights into the application of PHEs into HENs to reduce the network energy consumption and increase heat recovery with fixed network structure.

2. Methodology

2.1. Design a single plate heat exchanger

The proposed method aims to find the optimal design for a multi-pass plate heat exchanger with required process conditions. The variables include plate geometries, chevron angle, and the pass number of each stream. The objective is to minimize the total area of a single PHE since there is a direct proportion between heat transfer area and capital cost of a single plate heat exchanger. The heat load and pressure drop for both sides are the two main constrains.

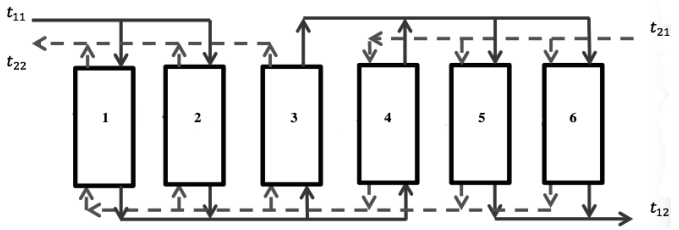


Figure 1. An example of multi-pass (X_1=3, X_2=2) flow arrangement with 6 blocks

Table 1. Geometrical parameters of Alfa Laval plates (Arsenyeva et al. 2011)

Plate type	δ,mm	d_e, mm	b, mm	A, m^2	D_p, mm	$f_{ch}\times10^3$, mm	L_p, mm
M3	2.4	4.8	100	0.03	36	0.240	320
M6	2.0	4.0	216	0.15	50	0.432	694
M6M	3.0	6.0	210	0.14	50	0.630	666
M10B	2.5	5.0	334	0.24	100	0.835	719
M15B	2.5	5.0	449	0.62	150	1.123	1381

*δ is the inter-plate gap, d_e is the equivalent diameter, b is the plate width, A is the plate area, D_p is the connection diameter, f_{ch} is the cross-section area, L_p is effective length.

Large number of plates give freedom for flow arrangement to achieve the required heat load within the specified pressure drop. For a two-stream multi-pass plate heat exchanger, plates can be separated into several parallel blocks and each block can be regarded as a one-pass PHE. For a X_1-X_2 PHE, the total number of blocks is $N = X1 \cdot X2$, where X_1 is the number of passes of hot stream and X_2 is the number of passes for cold stream. Figure 1 shows an example of 3-2 multi-pass flow arrangement. The LMTD method is employed for the thermal design of a single PHE. The key point for thermal design for a multi-pass PHE is to set up linear correlations of temperature among blocks. Each flow arrangement can be considered as a possible individual design. Enumeration technology is used to compare different flow arrangement.

Plate geometries and plate type are the other two main factors that affect heat transfer in plate heat exchangers. For practical consideration, standardized plate sizes from Alfa Laval are considered in this design. The detailed geometries are listed in Table 1. The plate selection problem is formulated as a mixed integer nonlinear programming (MINLP). To account for the effect of chevron angle on heat transfer efficiency, H, L and M type channel with different chevron angles are considered. The empirical correlation of three different channels and Nusselt number is stated in Eq.(1).

$$Nu = A \cdot Re^n \cdot Pr^{0.4} \cdot \left(\frac{\mu}{\mu_w}\right)^{0.14} \quad (1)$$

Where Pr is Prandtl number, μ and μ_w are the dynamic viscosities at stream and wall temperature respectively.

The values of the coefficient A and the exponent n under different plate geometries derived from many industrial manufacture plates are listed in Table 2. In total, there are five different plate types and three different chevron angles. To select plate type and chevron angle, binary variables are introduced to derive optimal solution automatically. Thus, a MINLP model is formulated in GAMS by using ANTIGONE solver to optimize a single multi-pass PHE unit.

Table 2. Parameter model of different plate geometries from Alfa Laval (Arsenyeva et al. 2011)

Plate type	M3			M6			M6M			M10B			M15B		
Channel type	H	L	M	H	L	M	H	L	M	H	L	M	H	L	M
A	0.265	0.12	0.18	0.25	0.12	0.165	0.27	0.11	0.14	0.224	0.126	0.117	0.26	0.085	0.13
n	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.71	0.73	0.713	0.693	0.748	0.7	0.74	0.74

2.2. Apply plate heat exchanger in HEN retrofit with fixed structure

This section presents the methodology of application of PHEs into retrofit design of an existing fixed structure HEN. The steps of methods, including identify the best candidate heat exchangers to replace, apply optimized single plate heat exchanger model and set up non-linear model to rebalance network, are detailed in this section.

2.2.1. Identification of heat exchanger candidates to replace

Utility path and sensitivity analysis are used to identify candidate heat exchangers and determine the rank of identified exchangers. To develop a cost effective retrofit method of an existing HEN without structure modifications and additional heat transfer area, the exchangers to be replaced must be on a utility path so that heat loads can be shifted along the path. For a simple network, utility paths can be specified by inspection. For a complex HEN, Incidence Matrix Approach is used to identify utility paths in the HEN. Sensitivity analysis is used to identify the best heat exchangers, which can bring the most energy saving among others. It also helps users to select the sequence of replacement.

2.2.2. Apply plate heat exchanger and optimization of HEN

After the best heat exchangers for replacement are identified, plate heat exchangers need to be applied in the replacement of selected shell-and-tube heat exchangers. The input process stream data are derived from candidate heat exchangers. The maximum heat transfer coefficient of the single plate heat exchanger can be derived from the optimization model in GAMS.

To rebalance the HEN, the heat loads of heat exchangers need to be relocated. The heat loads can be shifted through the utility paths. Application of PHEs can decrease energy consumption. However, the high installation cost of new heat exchanger is unneglectable. Thus, to propose a cost effective retrofit design, the objective of this optimization problem is to maximize retrofit profit. The total cost of retrofit including the implementation cost of bypass, the installation cost of plate heat exchangers, the cost of adding heat transfer area only if the heat exchangers are not on the utility path. Reaching target temperatures and maintaining the existing heat exchangers area are the two main constraints of the design.

Objective Function: *Maximize Retrofit Profit* = Profit from energy saving – Total cost of retrofit

Constraints: maintain the areas of existing heat exchanger (excluding the candidate heat exchanges) and reach target temperatures

Variables: heat load for all exchangers on a utility path
heat transfer coefficient for candidate heat exchanger

Table 3. The overall process data

	T_{in}	T_{out}	V	ρ	cp	λ	μ
Units	°C	°C	m^3/h	kg/m^3	kJ/(kg·K)	W/(m·K)	kg/(m·h)
Cold stream	28	89	5	978.4	3.18	0.66	51.3
Hot stream	95	79.36	15	960	4.21	0.67	1.07

* V is flow rate of streams, ρ is density of streams, cp is heat capacity of streams, λ is heat conductivity of streams, μ is dynamic viscosities of streams.

Flow arrangement	1-1	2-1	3-1	4-1	1-2	2-2	3-2	4-2
Plate type	M6	M10B	M6	M15B	M10B	M6	M10B	M6M
Number of plates	60	136	135	36	34	44	30	64
Total area (m^2)	8.76	32.64	21.00	22.32	8.09	6.60	7.20	8.96
Flow arrangement	1-3	2-3	3-3	4-3	1-4	2-4	3-4	4-4
Plate type	M6M	M6	M6M	M10B	M10B	M6	M6	M10B
Number of plates	39	30	39	24	84	32	36	32
Total area (m^2)	5.46	4.50	5.46	5.76	19.6	4.80	5.4	7.68

Table 4. The results of optimization a single plate heat exchanger

3. An example of single plate heat exchanger

An example from published literature (Arsenyeva et al. 2011) is studied to verify the effectiveness and accuracy of the new proposed design methodology. The overall process data are listed in Table 3. It can be clearly seen from Table 4 that the optimal solution is obtained for 2-3 flow arrangement with 30 plates (M6 plate type). The minimum total area of the plate heat exchanger is 4.50m^2. As a comparison, the minimal area from literature is 5.04 m^2 with 38 plates. Thus, it is possible to say that the proposed methodology can provide a solution with better heat transfer behavior and it significantly reduces the computation time for optimization a single PHE from several hours to 20 minutes.

4. Case study

A simplified crude preheat train is used to illustrate the benefit of application of plate heat exchanger into retrofit design of HEN in a fixed network structure. The existing structure of the network is shown in Figure 2. To quantify the retrofit profit, the operating time is fixed as one year.

5. Results and discussions

Incidence Matrix Approach is used to identify utility paths in the HEN. Only process heat exchanger 7 is not on the utility path. According to results from sensitivity analysis,

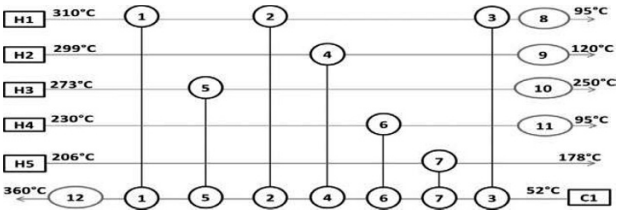


Figure 2. A simplified preheat train

	Parameter	Replacement	Enhancement
Retrofit Cost	Replacement	$76,630.58	$3,981.16
	Increasing Area	$0	$0
	Implementing By-pass	$15,000	$15,000
	Total Cost	$91,630.58	$18,981.16
Retrofit Profit	Utility Savings	$296,684.13	$237,252.68
	Net Saving	$205,053.55	$216,271.52

Table 5. A comparison between replacement and enhancement

heat exchanger 5 is the most sensitive exchanger. After replacing heat exchanger 5 by the optimized PHE, the optimization procedure is conducted to rebalance the network by using global solver on LINDO Systems What's Best!. The comparisons between applying PHEs and enhancement (Akpomiemie and Smith 2015) are highlighted in Table 5. The retrofit profit of replacement is a bit smaller than enhancement due to the high installation cost of plate heat exchanger. However, the replacement can save more energy than enhancement.

6. Conclusions

A design and optimization method for plate heat exchangers is proposed, which addresses plate pattern selection, determination of flow arrangement and pressure drop constraints. The computation time and total area are considerably decreased. Plate heat exchanger can be used as an alternative to heat transfer enhancement. The benefit of applying plate heat exchanger into retrofit design of HEN with fixed structure is highlighted by the case study.

References

Akpomiemie, M. O. & Smith, R. (2015). Retrofit of heat exchanger networks without topology modifications and additional heat transfer area. Applied energy, 159, 381-390.

Arsenyeva, O. P., Tovazhnyansky, L. L., Kapustenko, P. O. & Khavin, G. L. (2011). Optimal design of plate-and-frame heat exchangers for efficient heat recovery in process industries. Energy, 36(8), 4588-4598.

Kakac, S., Liu, H. & Pramuanjaroenkij, A. (2012). Heat exchangers: selection, rating, and thermal design: CRC press.

Pan, M., Bulatov, I. & Smith, R. (2014). Efficient retrofitting approach for improving heat recovery in heat exchanger networks with heat transfer intensification. Industrial & engineering chemistry research, 53(27), 11107-11120.

Sreepathi, B. K. & Rangaiah, G. (2014). Review of heat exchanger network retrofitting methodologies and their applications. Industrial & engineering chemistry research, 53(28), 11205-11220.

Smith, R. (2005). Chemical process: design and integration: John Wiley & Sons.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50071-9

Synthesis of Heat-integrated Water Network with Interception Unit

Dominic C. Foo[a*], Gopal C. Sahu[b], Shweta Kamat[c], Santanu Bandyopadhyay[c]

[a]*Department of Chemical and Environmental Engineering/Centre of Excellence for Green Technologies, University of Nottingham Malaysia, Broga Road, 43500 Semenyih, Selangor, Malaysia*
[b]*Now with Aditya Birla Science & Technology Centre, PFIC (unit of Grasim Industries Limited), Plot No. 1 & 1- A/1, MIDC Taloja , Tal. Panvel, Dist. Raigad 410208, India*
[c]*Department of Energy Science and Engineering, Indian Institute of Technology Bombay, Mumbai, 400076, India*
Dominic.Foo@nottingham.edu.my

Abstract

Water and energy are two of the important elements in ensuring sustainability in the process industry. In the past decade, the synthesis of heat-integrated water network has gained good attention among the research community of process system engineering. Various techniques ranging from pinch analysis and mathematical programming have been proposed to solve this problem. Note however that most works have only considered direct reuse/recycle scheme, where water sources are heated/cooled prior to their recovery to the water sinks. In this work, the placement of water interception units will be considered. Apart from direct water reuse/recycle, water sources may be subjected to partial purification with the interception unit, where impurity loads are to be removed prior to their recovery. This is commonly known as water regeneration. A hybrid approach consisting of superstructural and transshipment models is proposed for the solution strategy. Even though the approach is based on mathematical programming framework, it emphasizes the targeting philosophy of pinch analysis. A literature example is solved, with the objective being to minimize water flowrates and cost targets. It is shown that water regeneration cost is an important factor that dictates the overall water flowrates and costs.

Keywords: Process integration; simultaneous energy and water reduction; waste minimization; resource conservation; optimization.

1. Introduction

Energy and water are the key resources used in process industries. The need to conserve these resources has led to the development of heat integrated water regeneration networks (HIWRNs). Water consumption is reduced through the strategies of re-use, recycle and regeneration; while energy is conserved by maximum heat transfer. HIWRN synthesis has been carried out through various mathematical optimization models. One of such was a mixed integer non-linear programming (MINLP) formulation, where water allocation networks (WANs) were synthesized prior to the minimization of the total annual costs (TAC) by using sequential (Chen et al., 2010) or simultaneous techniques (Dong et al., 2008). Other formulations involved the minimization of total operating cost (TOC) as the first step through MINLP (Ibric et al., 2014) or used non-linear programming (NLP) formulations to set bounds (Ibric et al., 2016) followed by

the minimization of TAC using MINLP. Several sequential and simultaneous optimization strategies have been proposed and compared. However, their interchangeability has not been formally reported. The proposed model overcomes the drawbacks of simultaneous and sequential optimization techniques. The optimization problem was simplified to an NLP formulation (Ataei et al., 2009). The simplest form of the optimization problem was found to be an NLP formulation from a review on the heat integrated WANs (Ahmetović et al., 2015). This paper proposes a linear programming framework to synthesise HIWRN in a novel way for fixed flowrate problems.

2. Problem Definition

The essential parameters in the synthesis of HIWRN are flow rate, *F*, contaminant concentration, *C*, and temperature, *T*. For a fixed flow rate problem, the processes are divided into water-consuming (sink) and water-producing (source) processes. Water can be conserved through the recovery among water sources and sinks. Due to contaminant concentration constraint on the sinks, and/or water availability of the sources, fresh water may be used to supplement water requirement of the HIWRN. Energy consumption, on the other hand, can be minimized through heat recovery between hot and cold process streams. Hot and/or cold utilities are used to satisfy the unmet energy needs. A typical HIWRN is shown in Figure 1.

Subscripts *s* and *d* denote the source and sink. Subscripts *i*, *j*, *r*, *t*, *k*, *hu* and *cu* represent indices for source, sink, interception unit, contaminant, temperature level, hot utility, and cold utility. In this formulation, *i* and *j* equal to zero for a fresh water source and waste water. T_{d0} cannot exceed the discharge limit, but no limit is imposed on C_{d0}. *f* denotes the variable flow rate. There are *M* contaminants. *N* and *A* denote the number of units and unit cost. The transshipment model is used with *T*I temperature intervals (George et al., 2011). *H* and *G* are the amounts of heat supplied and gained by hot and cold streams, while *HS* and *HD* are the heat surplus and heat deficit. *Q* corresponds to the utility and *R* denotes the residual heat.

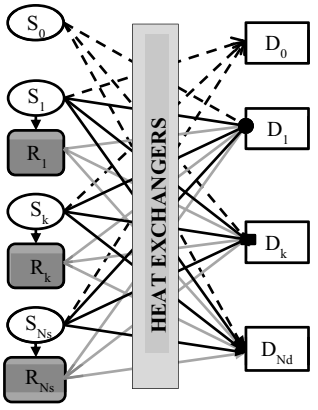


Figure 1 A typical HIWRN

Each water source is regenerated at its temperature, thereby requiring 1 to N_s interception units. Isothermal mixing of water sources and regenerated sources occurs at the sinks. Single pass interception units are considered for this model. The interception unit acts as a source (outlet), a sink (inlet) and is also classified into a fixed outlet and fixed removal ratio (RR) type (Foo, 2012). F_{Reg} is the amount of water regenerated. The contaminant concentrations leaving the fixed outlet and RR type interception units are denoted by $C_{out,t,}$ and $C_{rr,t}$. The removal ratio (*rr*) for an RR type of interception unit is the ratio of contaminant load removed to the amount of contaminant load entering it. As a result of no flow loss, the *rr* for each interception unit may be simplified as shown in Eq. (1).

$$rr_r = \frac{C_{rr,t} - C_{si,t}}{C_{si,t}} \quad \forall\, r = i,\ where\ r = 1\ to\ N_s \tag{1}$$

As the *rr* is specified and $C_{si,t}$ is a known quantity, $C_{rr,t}$ is fixed. The problem turns out to be the same as the fixed outlet type. Any methodology applicable to the fixed outlet type of interception unit can be used for the fixed RR type as well provided that the temperature of the interception unit equals the temperature of the source that is regenerated. Thus the problem is formulated for the fixed outlet interception unit.

3. Mathematical formulation

3.1. Fresh Water Minimization (M1)

The fresh water is minimized in Eq. (2) subjected to the flow balance constraints for sources, sinks and interception units expressed as Eqs. (3)-(5) and contaminant load constraint is given by Eq. (6).

$$Minimize\ F_{s0} = \sum_{j=1}^{N_d} f_{0j} \tag{2}$$

$$\text{Subject to: } \sum_{j=1}^{N_d} f_{ij} + f_{i0} + \sum_{r=1}^{N_r} f_{ir} = F_{si}\ \forall\ i\ \in \{1,2,\dots N_s\} \tag{3}$$

$$\sum_{i=1}^{N_s} f_{ij} + f_{0j} + \sum_{r=1}^{N_r} f_{rj} = F_{dj}\ \forall\ j \in \{1,2,\dots N_d\} \tag{4}$$

$$\sum_{r=1}^{N_r} f_{ir} = \begin{cases} \sum_{j=1}^{N_d} f_{rj} & if\ i = r \\ 0 & otherwise \end{cases} \qquad \forall\ r\ (1\ to\ N_s) \tag{5}$$

$$\sum_{i=1}^{N_s} f_{ij}\ C_{si,t} + f_{0j} C_{s0,t} + \sum_{r=1}^{N_r} f_{rj}\ C_{out,t} \le F_{dj} C_{dj,t}\ \ \forall\ j \in \{1,2,\dots N_d\}\ and\ t \in \{1,2,\dots M\} \tag{6}$$

3.2. Regeneration flowrate minimization (M2)

The regeneration, given in Eq. (7), is minimized subjected to either a fixed amount of fresh water or waste water expressed by Eqs. (8) and (9) respectively and Eqs. (2)-(6).

$$Minimize\ F_{Reg} = \sum_{r=1}^{N_r} \sum_{j=1}^{N_d} f_{rj} \tag{7}$$

$$\text{Subject to: } \sum_{j=1}^{N_d} f_{0j} = F_{s0} \tag{8}$$

$$\sum_{i=1}^{N_s} f_{i0} = F_{d0} \tag{9}$$

3.3. Utility minimization (M3)

The hot utility is minimized in Eq. (10) subject to the flow rates given in Eqs. (8), (9) and (11) and the regeneration contaminant concentrations. The other constraints are heat surplus, heat deficit, energy balance in each *TI*, the upper and lower limit of heat transfer and the non-negativity constraints expressed by Eqs. (12)-(16) respectively.

$$\text{Minimize } Q_{hu} = \sum_{k=1}^{TI} Q_{hu,k} \tag{10}$$

$$\text{Subject to: } \sum_{i=1}^{N_s} f_{ir} = \sum_{j=1}^{N_d} f_{rj} = F_{Reg} \qquad \forall\, re(1\ to\ N_s) \tag{11}$$

$$HS_k = \sum_{j=0}^{N_d}\sum_{i=0}^{N_s} H_{i,k,j} + \sum_{j=1}^{N_d}\sum_{r=1}^{N_r} H_{r,k,j} \quad \forall\, k \in \{1, 2, \ldots TI\} \tag{12}$$

$$HD_k = \sum_{j=0}^{N_d}\sum_{i=0}^{N_s} G_{i,k,j} + \sum_{j=1}^{N_d}\sum_{re=1}^{N_{re}} G_{re,k,j} \quad \forall\, k \in \{1, 2, \ldots TI\} \tag{13}$$

$$R_k - R_{k-1} - Q_{hu,k} + Q_{cu,k} = HS_k - HD_k \qquad \forall\, k \in \{1, 2, \ldots TI\} \tag{14}$$

$$R_0 = R_{TI+1} \tag{15}$$

$$R_k, HS_k, HD_k, Q_{hu,k}, Q_{cu,k} \geq 0 \qquad \forall\, k \in \{1, 2, \ldots TI\} \tag{16}$$

4. Solution Strategies

The mathematical models for the minimization of fresh water, regenerated water and utility are linear programming (LP) formulations. Simultaneous optimization and three-stage sequential strategy are proposed based on the above models.

4.1. Simultaneous strategy

The simultaneous optimization strategy involves the minimization of the TOC given in Eq. (17) subject to Eqs. (3)-(6) and Eqs. (12)-(16). A_{fw} and A_{ww} are the unit costs of fresh water and waste water respectively.

$$\begin{aligned} Minimize\ TOC = {} & A_{fw}\sum_{j=1}^{N_d} f_{0j} + A_{ww}\sum_{j=1}^{N_d} f_{i0} + A_r\sum_{r=1}^{N_r}\sum_{j=1}^{N_d} f_{rj} + \\ & A_{hu}\sum_{k=1}^{TI} Q_{hu,k} + A_{cu}\sum_{k=1}^{TI} Q_{cu,k} \end{aligned} \tag{17}$$

4.2. Three stage sequential strategy

Models M1, M2, and M3 are solved in each step of the three stage sequential strategy. In Eqs.(8) and (9), F_{s0} and F_{d0} values will be set to the minimum fresh water and waste water requirement obtained from M1 and in Eq.(11), the F_{Reg} value will be set to the amount of regeneration carried out by each interception achieved by solving M2.

5. Example for HIWRN synthesis

Table 1 represents the limiting water data (George et al., 2011) for the HIWRN synthesis. A single pass interception unit with fixed contaminant concentration of 80 ppm and without any constraint in its operating temperature is used. It is further assumed that the fresh water enters at 20 °C and the waste is to be discharged at 30 °C, while the specific heat capacity of water is taken as 4.2 kJ/kg °C. GAMS 24.2.2 software is used for solving the optimization models.

The three-stage sequential strategy was implemented. The fresh water (FW) and waste water (WW) flowrates resulting from M1 were found to be 52.5 kg/s. With constant FW requirement, the minimum regeneration flowrate was identified as 87.5 kg/s through M2. The flow rates, temperatures and contaminant concentrations of the regenerated streams obtained from M2 serve as an input to M3 along with the FW requirement from M1. The minimum hot ($Q_{H,\,min}$) and cold ($Q_{C,\,min}$) utility requirements are 6615 kW and

4410 kW. Without regeneration, the FW consumption of 77.28 kg/s and hot and cold utility of 10963.62 kW and 7717.78 kW is observed (George et al., 2011). The WAN and energy targeting results are shown in Tables 2 and 3. No flow occurs through the interception unit R2. As the temperature and outlet contaminant concentration is same for R1 and R3, they can be combined into a single interception unit, with the regeneration of 87.5 kg/s.

Table 1 Limiting process data for Example (George et al., 2011)

Sinks (Dj)	**Flow rate (kg/s)**	**Concentration (ppm)**	**Temperature (℃)**
D1	100	50	100
D2	40	50	75
D3	166.67	800	100
Sources (Si)	**Flow rate (kg/s)**	**Concentration (ppm)**	**Temperature (℃)**
S1	100	100	100
S2	40	800	75
S3	166.67	1100	100

Table 2 WAN from three stage sequential methodology

			C (ppm)	50	50	800	100	-	1100	433.33
			F (kg/s)	100	40	166.7	25	0	62.5	52.5
			T(°C)	100	75	100	100	75	100	30
C (ppm)	*F* (kg/s)	T(°C)		**D1**	**D2**	**D3**	**R1**	**R2**	**R3**	**WW**
0	52.50	20	**FW**	37.5	15					
100	100	100	**S1**			47.5	25			27.5
800	40	75	**S2**			15				25
1100	166.7	100	**S3**			104.2			62.5	
80	25	100	**R1**		25					
80	0	-	**R2**							
80	62.5	100	**R3**	62.5						

Table 3 Energy targeting for WAN in Table 2

T_H (°C)	F_{R12}	F_{WW1}	F_{WW2}	ΔH_H (kW)	T_C (°C)	F_{FW1}	F_{FW2}	F_{23}	ΔH_C (kW)	Cum ΔH_H (kW)
	(25)	(27.5)	(25)			(37.5)	(15)	(15)		
130					100	↑		↑		6615 ($Q_{H,\,min}$)
				0		↑		↑	5512.5	
105					75	↑	↑	↑		1102.5
				0		↑	↑		1102.5	
100	↓	↓			70	↑	↑			0(pinch)
	↓	↓		5512.5		↑	↑		5512.5	
75	↓	↓	↓		45	↑	↑			0 (pinch)
		↓	↓	5512.5		↑	↑		5512.5	
50		↓	↓		20	↑	↑			0 (pinch)
		↓	↓	4410					0	
30		↓	↓		0					4410 ($Q_{C,\,min}$)

The costs of FW, regenerated water, WW, hot utility and cold utility are \$1/t, \$0.5/t, \$1/t, \$120/kW and \$ 10/kW. The FW requirement, WW discharge, amount of water regenerated, hot and cold utilities were found to be same from the simultaneous and three stage sequential strategy. The TOCs with and without regeneration were \$5.5 million/y and \$6.3 million/y. The WAN and energy targeting can be done for simultaneous optimization (results in two regeneration units). The regeneration cost is varied from \$0.5/t to \$2/t and its effect on water flow rates and utilities is depicted in

Figure 2. With the increase in the unit cost, the amount of regeneration decreases and the utilities increase; while the FW consumption first decreases and then increases. The energy cost increases with the increase in the unit cost (Figure 3). There is a decrease followed by an increase in the water and total costs. The proposed methodology is applicable to multiple contaminants problems (not demonstrated due to brevity).

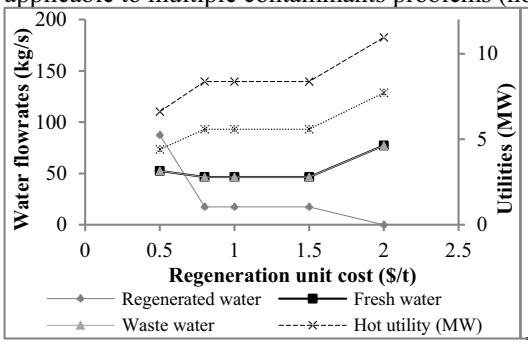


Figure 2 Effect of regeneration cost on flow rate and utilities

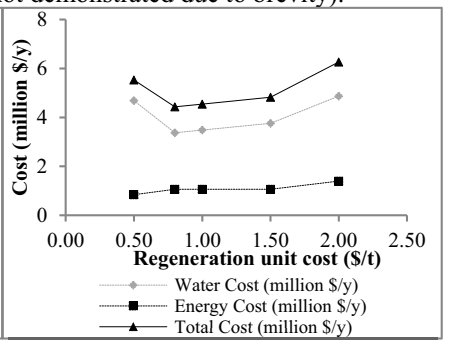


Figure 3 Effect of regeneration cost on water, utility and total cost

6. Conclusion

A simultaneous optimization approach with an objective to minimize the TOC is compared with a three stage sequential methodology, where the fresh water, regenerated water, and utility are minimized separately. Both the approaches are LPs and provide the same results, which implies that both methods can be used interchangeably. The effect of change in unit cost of regeneration on water flow rates, utilities, their respective costs and the total cost was studied. Due to the dependence of convergence efficiency on the number of variables, the three-stage sequential optimization formulation ensures better convergence efficiency as compared to the simultaneous technique in case of large data.

References

A. Ataei, M.H. Panjeshahi, S. Karbassian, 2009, Simultaneous energy and water minimization approach for systems with optimum regeneration of wastewater, Res J Environ Sci, 3, 604-618.

C.L. Chen, H.L. Liao, X.P. Jia, Y.J. Ciou, J.Y. Lee, 2010, Synthesis of heat-integrated water-using networks in process plants, J Taiwan Inst Chem Eng, 41, 512-521.

D.C.Y. Foo, 2012, Process Integration for Resource Conservation, CRC Press.

E. Ahmetović, N. Ibrić, Z. Kravanja, I. Grossmann, 2015. Water and energy integration: A comprehensive literature review of non-isothermal water network synthesis. Comps and Chem Eng, 82, 144–171.

H.G. Dong, C.Y. Lin, C.T. Chang, 2008, Simultaneous optimization approach for integrated water-allocation and heat-exchange networks, Chem Eng Sci, 63, 3664-3678.

J. George, G. C. Sahu, S. Bandyopadhyay, 2011, Heat integration in process water networks, Ind Eng Chem Res, 50, 3695-3704.

N. Ibrié, E. Ahmetović, Z. Kravanja, 2014, Simultaneous optimization of water and energy within integrated water networks, Appl Therm Eng, 70, 1097.

N. Ibrié, E. Ahmetović, Z. Kravanja, 2016, Mathematical programming synthesis of non-isothermal water networks by using a compact/reduced superstructure and an MINLP model, Clean Techn Environ Policy, 18, 1779–1813.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50072-0

Computer-aided process simulation, design and analysis: lactic acid production from lignocellulosic residues

José A. Méndez-Alva[a], Eduardo S. Perez-Cisneros[b], Divanery Rodriguez-Gomez[c], Oscar A. Prado-Rubio[d], Beatriz Ruiz-Camacho[a], Ricardo Morales-Rodriguez[a]*

[a]*Departamento de Ingeniería Química,Universidad de Guanajuato, Noria Alta s/n Col. Noria Alta, Guanajuato, Guanajuato, 36050, México.*
[b]*Departamento de Ingeniería de Procesos e Hidráulica, Universidad Autónoma Metropolitana-Iztapalapa, Av. San Rafael Atlixco 186, 09340, México, D.F., México.*
[c]*Coordinación de Ingeniería Bioquímica, Instituto Tecnológico Superior de Irapuato, Carretera Irapuato-Silao km 12.5. Col. El Colpal, Irapuato, Guanajuato, 36821, México.*
[d]*Departamento de Ingeniería Química, Universidad Nacional de Colombia, Campus La Nubia, Bloque L, Manizales, Caldas, Colombia.*
ricardo.morales@ugto.mx

Abstract

Lactic acid (LA) is a compound with a considerable range of application in industry, where its use as precursor of diverse products is currently relevant. Thus, it is important to complement experimental efforts to investigate the design of the lactic acid production relying in computer-aided tools using a process system engineering approach. This work presents the implementation of a computer-aided processing platform and the analysis of a possible industrial production route of LA from the waste of sugar industry. The process configuration consisted on a pretreatment stage to extract the raw material from the residues. Subsequently, the LA fermentation is included using a more realistic design and analysis considering the combination of batch and continuous operations based on a scheduling plan. Finally, the separation and purification stages using reactive distillation are included, aiming to obtain an LA solution above 95 % mass fraction and a production of 14,448 kg/hr from 28,886 kg/h of sugarcane bagasse. The process was evaluated to determine its economic feasibility based on a price of 1.5 US/kg of LA. It was estimated that the process has a recovery time of 3.6 years.

Keywords: Lactic acid production; renewable materials; aspen plus; reactive distillation.

1. Introduction.

Currently important issues such as, environmental protection, sustainability and oil reserves decline have motivated the search of alternatives to produce some high value substances from renewable raw materials rather than fossil sources. Among the different basic chemicals, the lactic acid (2-hydroxypropionic acid, $CH_3CHOHCOOH$) is interesting due to its wide range of applications to produce and manufacture commodities, mainly in food industry, chemical products, cosmetics and pharmaceuticals sector. More recently, the interest to produce lactic acid has grown since it is the basis to produce

polylactic acid (PLA). PLA is a biodegradable polymer classified as bio-plastic and it can be used for different products due to their properties, for example: high tensile strength, high modulus, thermoplastic properties, biodegradability and bioenvironmental compatibility/absorbability, in addition with the application in the medicine area (Pérez-Cisneros et al., 2015). Therefore, LA is a strategic product to be manufacture in the near future since it can reduce our dependency on fossil raw materials. It has been estimated that the demand for lactic acid increases annually by 5-8 % (Yadav et al, 2011). The annual world market for lactic acid production was forecasted in a demand of 259,000 metric tons for the year 2012 (Martinez et al, 2013), and was expected to reach 367,300 metric tons in 2017.

The production of lactic acid can be achieved either by chemical synthesis or through fermentative production routes. Between the options, fermentative production offers advantages such as, the use of renewable and cheaper substrates, low production temperatures and low energy consumption. The fermentation-based production of lactic acid involves the typical sections of a lignocellulosic bio-based configuration: pretreatment to breakdown the lignocellulosic matrix (PT), enzymatic hydrolysis to release the glucose from cellulose (EH) and fermentation where glucose and xylose are converted to lactic acid (F) (Morales-Rodriguez et al., 2016). The produced fermentation broth contains diluted lactate, which must be treated in a downstream processing to obtain higher lactic acid concentration for the polymer production. For the separation and purification different procedures have been explored, for instance, electro-enhanced dialysis (Prado-Rubio et al., 2009), liquid-liquid extraction, ultrafiltration, osmosis, adsorption, reactive extraction and reactive distillation (Pérez-Cisneros et al., 2015). However, some technologies have shown some disadvantages, for example, ultrafiltration and osmosis are unviable for a large amount of raw material. Adsorption and reactive extraction have some problems regarding to species particle selectivity, capacity and regeneration. More importantly, all these methods suffer from low recovery of lactic acid (Prado-Rubio, et al., 2016). Due to these limitations, up to now, the most attractive method for the purification of lactic acid from fermentation broth is a two-stage reactive distillation process. There, the lactic acid is first reacted with an alcohol to form a more volatile ester that it can be purified by distillation. Subsequently, the pure lactic acid is recovered by hydrolysis of the ester. This method of purification cannot only handle a large capacity of raw material, but can also achieve a high recovery of lactic acid (Su et al, 2013).

Thus, the objective of this work is to present process design and simulation to investigate the lactic acid production from lignocellulosic materials, as well as analysing the techno-economic feasibility of the process using a computer-aided tool. The process configuration includes the PT, EH, F and downstream process relying on reactive distillation.

2. Computer-aided process design: lactic acid production

2.1. Simulation approach

The LA production process was implemented in Aspen plus process simulator allowing the design, and analysis of the proposed process feasibility. In order to perform the evaluation of the LA process production, the sugarcane bagasse was used as lignocellulosic raw material. This study considers a feeding flowrate of 28,886.6 kg/hr consisting of cellulose (68.22%), lignin (29.19%) and xylan (2.59%). According to the

involved compounds and process conditions, the thermodynamic models employed in the simulations were the NRTL and the Hayden-O'Connell models. The NRTL model parameters were obtained from the National Renewable Energy Laboratory (NREL) database (Wooley and Putsche, 1996).

The process topology was built relying on some previous work. For instance, the bioproduction that include the pretreatment, enzymatic hydrolysis and fermentation was taken from Morales-Rodriguez et al. (2016) and the downstream processes was based on the study proposed by Pérez-Cisneros et al. (2015).

2.2. Process description

The LA process configuration is illustrated in Figure 1, which consists of 6 sections: 1) pretreatment, 2) neutralization and nutrient production, 3) enzymatic hydrolysis, 4) inoculation, 5) LA fermentation and 6) separation and purification. The process starts mixing the bagasse (30 wt/wt%) with a sulfuric acid solution. This mixture is fed to reactor R-101 where the lignocellulosic matrix is broken-down breakdown while some glucose and xylose are also produced. The output stream of reactor R-101 is a mixture of solid, liquid and vapours, which is sent to separator F-201 to separate steam and liquid in two streams (201 and 202, respectively).

For the nutrient production, a fraction of the liquid phase (2% of total flow) is divided (S-201) and fed to reactor R-201 together with a stream of ammonia (207) to produce ammonium sulphate, which is used as nitrogen source for the microorganisms at stage 5. The reaction in R-201 is carried out at 1 atm, 50°C with a sulfuric acid conversion of 89%. For the neutralization process, the output stream of R-201 is mixed (M-202) with the rest of the vapour and liquid coming from the M-201 mixer. This stream passes through the heat exchanger H-201 to reach the temperature of 25°C and it is fed to the neutralization reactor (R-202) together with a stoichiometric flowrate stream of sodium hydroxide (211). As preparation for the hydrolysis, the neutralized stream (212) is mixed (M-301) with the stream of solids (203) from the separator F-201. The solid-liquid flow (302) enters the divider (S-301) to generate two flows, 10% of the flow (303) is sent to the inoculation section, while the remaining 90% (304) is sent to the enzymatic hydrolysis section. Before entering the enzymatic reactors, the stream temperature is raised to 50°C in the H-302 heat exchanger. Besides, a stream (307) of diluted enzyme is mixed with the stream leaving the exchanger in the M-302 and subsequently feeds the reactors (R-301-R-305). The feeding ratio of the enzyme should be 20 mg per 1 g of cellulose and the aqueous dilution should be 30% enzyme. The enzymatic reactors breakdown cellulose to produce glucose with a conversion of 66%. The stream 303 reaches a temperature of 37°C in an exchanger, then mixed with a diluted enzyme solution and directed to the inoculation reactors (R-401-R-406) to release glucose molecules from the solids. The enzyme/glucose ratio is the same as the hydrolysis section. The streams 417 and 320 are sent to solid-liquid separators (S-404 and S-402, respectively), where the liquid stream of both separators are mixed (M-404) and the mixture (420) is sent to the fermentation stage of lactic acid, while the solid phase (323) is available to be combustion and generate energy. The lactic acid fermentation reactors (R-501-R-506) are fed by stream 420 to carry out the conversion of glucose (80 %) and xylose (60%) into lactic acid as shown in the following reactions:

$$C_6H_{12}O_6 \rightarrow 2\ C_3H_6O_3 \quad (1)$$

$$3C_5H_{10}O_5 \rightarrow 5C_3H_6O_3 \quad (2)$$

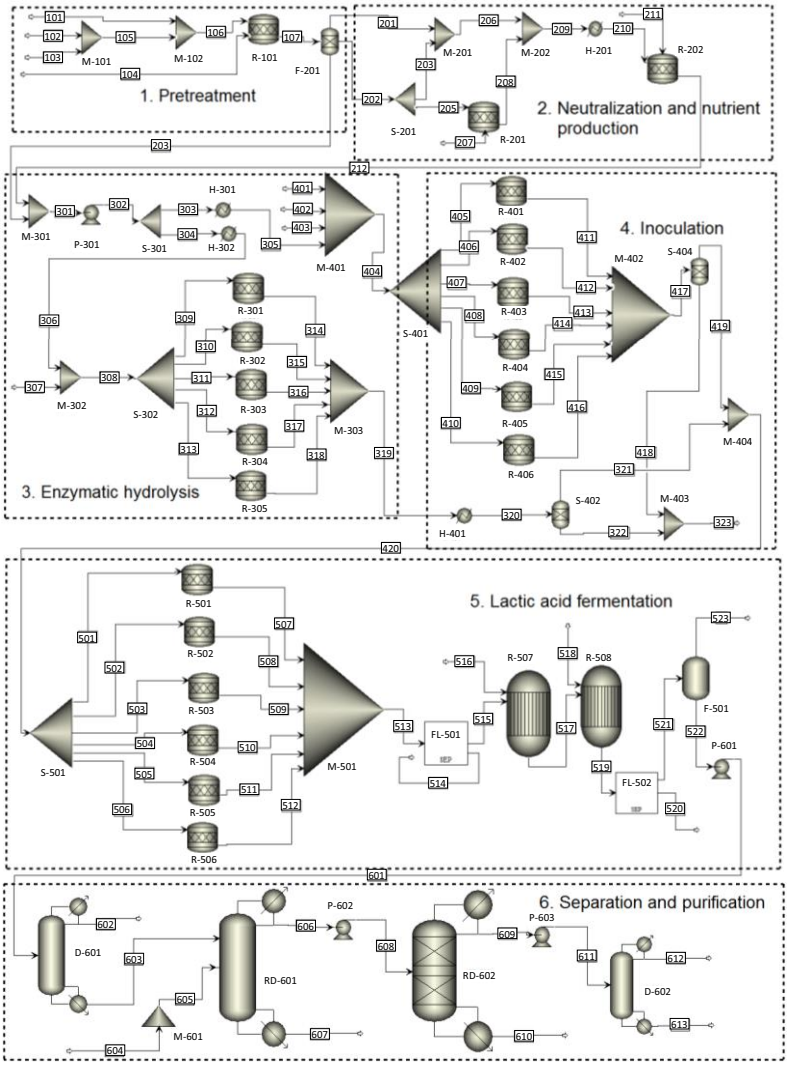


Figure 1. Process flowsheet for the acid lactic production

The lactic acid (515) is precipitated from the fermentation broth as calcium lactate in the reactor R-507 where a 15% mass fraction solution of calcium hydroxide (516) is also fed. Stream 517 feeds reactor R-508 together with a 15% mass fraction sulfuric acid solution to remove the calcium salt and to obtain lactic acid. The calcium sulphate is separated by filtration (FI-502), while the liquid mixture (521) is sent to a flash tank (F-501) to separate a large part of the water from the mixture. The reactions of R-507 and R-508 are presented below, respectively:

$$2C_3H_6O_3 + Ca(OH)_2 \rightarrow Ca(C_3H_6O_3)_2 + 2H_2O \quad (3)$$

$$Ca(C_3H_6O_3)_2 + H_2SO_4 \rightarrow 2C_3H_6O_3 + CaSO_4 \quad (4)$$

The purification section stars with stream 601, where this flow reaches a distillation column (D-601) to remove the largest amount of water (602) and send the flow of lactic

acid (603) to the first column of reactive distillation (RD-601) together with a stoichiometric flow of methanol (605) to carry out the esterification reaction (Eq. 5):

$$C_3H_6O_3 + CH_3OH \rightarrow C_4H_8O_3 + H_2O \quad (5)$$

The stream leaving the dome of the column (RD-601) is pumped (P-602) to the next reactive column (RD-602) where the hydrolysis (Eq. 6) is carried out to obtain lactic acid:

$$C_4H_8O_3 + H_2O \rightarrow C_3H_6O_3 + CH_3OH \quad (6)$$

The lactic acid is obtained from the RD-602 in the stream 610 with flowrate of 14,448.3 kg/hr and an acid mass fraction of 98%. On the other hand, the stream 609 is sent to another conventional distillation column (D-602) to recover the methanol from the process.

2.3. Scheduling for combining batch and continuous operation

The process configuration combined the continuous and batch operation modes. In order to perform a reliable analysis of the process, it is necessary to propose an operation scheduling, which must include the filling, reaction, drawing and idle that are repeated periodically. The schedule for enzymatic hydrolysis and seed (see left column) could be understood as following: for reactor number one the periodic operation last 60 hours. It starts with the loading of 12 h, followed by a reaction time (36 h) and finally it ends with 12 h of drawing/emptying the content of the reactor. Once the first cycle is completed the next cycle starts again by repeating the same schedule. The first lactic acid fermentation reactor starts after 48 h then following for a similar schedule 12, 48 and 12 h. There is a difference in the reaction time for the fermentation section, because this operation requires of 12 extra hours to reach the desired titter. In addition, it is necessary to add an extra fermentation unit to keep the same schedule.

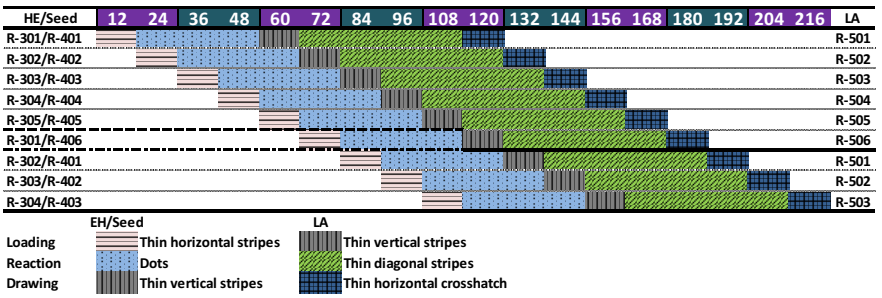


Figure 2. Process flowsheet for the acid lactic production

3. Findings using the computer-aided simulation

The computer-aided simulation allowed performing the economic evaluation of the process using Aspen Economic Analyzer, as well as the potential energy integration employing The Aspen Energy Analysis. The economic analysis consisted on performing a sensitivity analysis on the LA selling price. Table 1 illustrates the different evaluated prices and the economic evaluation. The results illustrates that it is possible to have competitive selling prices with still positive economic metrics. Considering a commercial price of 1.5 USD/kg, the result showed a short payback period (PBP) that allowed having a satisfactory net present value (NPV) in 10 years. The results also showed a great

opportunity for the use of agroindustrial residues in the production of LA, promoting the valorisation of the biomass.

Table 1. Economic analysis for lactic acid production

Selling price, [US$ kg^{-1}]	NPV [MUS$] at 10 years	PBP [years]
1.5 (Alibaba.com)	255.08	3.13
1.2	109.99	4.81
1.1	78.15	5.62

The energy integration analysis was also performed in this work, which allowed identifying that 60.46% of the energy could be integrated reducing the even more the utilities cost.

4. Conclusions

The development of the computer-aided simulation allowed the tecno-economical evaluation of LA bioproduction. The results illustrated that it is feasible to have LA production following a biotechnological route from lignocellulosic residues. This work gives the basis for the complete design of the lactic acid production plant and open the opportunity to have further comparison with different process technologies, such as membrane-based processes, liquid-liquid extraction and in situ product removal configurations.

5. References

F.A.C. Martinez, E.M. Balciunas, J.M. Salgado, J.M.D. González, A. Converti, R.P.S. Oliveira, 2013, Lactic acid properties, applications and production: a review. Trends Food Sci Technol, 30, 70–83.

R. Morales-Rodriguez, E.S. Perez-Cisneros, J.A. de Los Reyes-Hereida, D. Rodriguez, 2016, Evaluation of biorefinery configuration through a dynamic model-based platform: Integral operation for bioethanol and Xylitol Co-Production from lignocellulose, Renewable Energy, 89, 135-143

E.S. Pérez-Cisneros, L. Avilés-Cabrera, V. Medina-Bañuelos, M. Sales Cruz, A. Ochoa-Tapia, T. Viveros-García, R. Lobo-Ohemichen, 2015, A Computational Platform for Simulation, Design and Analysis of a Poly(Lactic) Acid Production Process From Different Lignocellulosic Raw Materials, Comput.-Aided Chem. Eng., 37, 1187-1192.

O.A. Prado-Rubio, S.B. Jørgensen, G. Jonsson, 2009, Lactic Acid Recovery in Electro-Enhanced Dialysis: Modelling and Validation. Comput.-Aided Chem. Eng., 26, 1406-1410.

O.A. Prado-Rubio, R. Morales-Rodriguez, P. Andrade-Santacoloma, H. Hernández-Escoto, 2016, Process Intensification in Biotechnology Applications. In book “Process Intensification in Chemical Engineering”, J.G. Segovia-Hernández, A. Bonilla-Petriciolet (eds.). Springer editorial ISBN: 978-3-319-28392-0. pp. 183-219. DOI 10.1007/978-3-319-28392-0_7

C.-Y. Su, C.-C. Yu, I-L. Chien, J.D. Ward, 2013, Plant-Wide Economic Comparison of Lactic Acid Recovery Processes by Reactive Distillation with Different Alcohols. Ind. Eng. Chem. Res., 52, 11070-11076.

R.J. Wooley, V. Putsche, 1996, Development of an ASPEN PLUS Physical Property Database for Biofuels Components. NREL/MP-425-20685, National Renewable Energy Laboratory, Golden, Colorado, USA.

A.K. Yadav, A.B. Chaudhari, R.M. Kothari, 2011, Bioconversion of renewable resources into lactic acid: an industrial view. Crit Rev Biotechnol, 31, 1-19.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50073-2

ProCAPD – A Computer-Aided Model-Based Tool for Chemical Product Design and Analysis

Sawitree Kalakul[a], Lei Zhang[b], Hanif A. Choudhury[c], Nimir O. Elbashir[c], Mario R. Eden[a]*, Rafiqul Gani[d]

[a]*Department of Chemical Engineering, Auburn University, USA*
[b]*Institute of Process Systems Engineering, Dalian University of Technology, China*
[c]*Department of Chemical Engineering, Texas A&M University at Qatar, Qatar*
[d]*PSE for SPEED, Skyttemosen 6, Allerød, DK-3450, Denmark*
edenmar@auburn.edu

Abstract

In this work, a systematic framework for a computer-aided mixture/blend design (CAM^{b}D) developed by Kalakul et al. (2017) is integrated into a computer-aided model-based tool for chemical product design and analysis called ProCAPD. Within the framework as implemented, a software architecture, database management, model development, design methodologies and design algorithms are integrated. The software employs a template approach, where each template follows the same common steps in the workflow for design of blended products, but has the option to employ different blend design solution approaches, product specific property models, data and calculation routines, when necessary. With the new additions, the software is able to support the design and analysis of a wide range of blended products, such as, jet-fuels, diesel fuels, gasoline fuels, solvent mixtures and lubricants. The decision making process is supported by dedicated property models and structured databases, specifically developed for each blend design problem scenario. Output from the software is a small set of most promising product candidates and a short list of recommended experiments that can validate and further fine-tune the product compositions. The application of the new feature is highlighted through case studies involving a diesel surrogate design and a tailor-made jet-fuel blend.

Keywords: Chemical product design, Computer-aided mixture/blend design (CAM^{b}D).

1. Introduction

Mixture/blend design problems are considered when single molecules are unable to satisfy all the desired product specifications. It has been widely accepted that use of computer-aided mixture/blend design (CAM^{b}D) methods helps speed up development times and reduce consumption of experimental resources. CAM^{b}D is defined as follows: given a set of chemicals and a set of property constraints, determine the optimal mixture and/or blend consisting of a sub-set of chemicals (Gani, 2004). CAM^{b}D problems are usually formulated and solved as Mixed Integer Non Linear Programming (MINLP) or Mixed Integer Linear Programming (MILP). The main efforts have been directed to the design of solvent mixtures (Jonuzaj et al., 2016), polymer blends and formulated liquid products (Conte et al., 2011). Different solution strategies have been employed to solve CAM^{b}D problems such as Artificial Neural Networks (ANN) and Analytical Hierarchy Process (AHP). Zhang et al. (2015) formulated and solved MINLP problems to design working fluid mixtures for Organic Rankine Cycles, while Jonuzaj et al. (2016) did the

same for solvent mixture design. Yunus et al. (2014) decomposed the overall MINLP problem into sub-problems to tailor-make gasoline and lubricant blends.

Even though many CAM^{b}D problems have been solved and different solution approaches have been proposed, due to the lack of the needed property models and data for the target properties, the currently available methods and tools can only solve a small percentage of blend design problems. Also, the process needs, inclusion of sustainability issues as well as product performance verification add to the complexity of the problem due to constraints that need to be satisfied. This is a challenging task requiring the integration of many tasks, such as, data acquisition, data testing, model development, and multi-scale modeling in a computer-aided framework. Therefore, the objective of this paper is to present a hybrid computer-aided framework for CAM^{b}D that is integrated with experimental verification options and is flexible, robust and versatile. The objective of the framework with its associated data, models, methods and tools is to provide the means to design, analyze and verify blended products in a fast, efficient and systematic manner. The application of the developed framework, implemented as a new option (template) in an extended version of the computer-aided tool for chemical product design called ProCAPD (Kalakul et al, 2017), is illustrated through two case studies. With respect to the CAM^{b}D template, more property models for fuel and lubricant blends (e.g. Cetane index, pour point, carbon dioxide emission from the combustion engines and friction coefficient) and more algorithms to solve CAM^{b}D problems are added. Thus, the software is able to solve wider ranges of CAM^{b}D problems in a fast and efficient way.

2. CAM^{b}D Template within ProCAPD

The framework (architecture) of the CAM^{b}D template in the ProCAPD software is highlighted in Figure 1 in terms of the main steps (workflow), the methods and tools employed in each step and the dataflow from one step to the next.

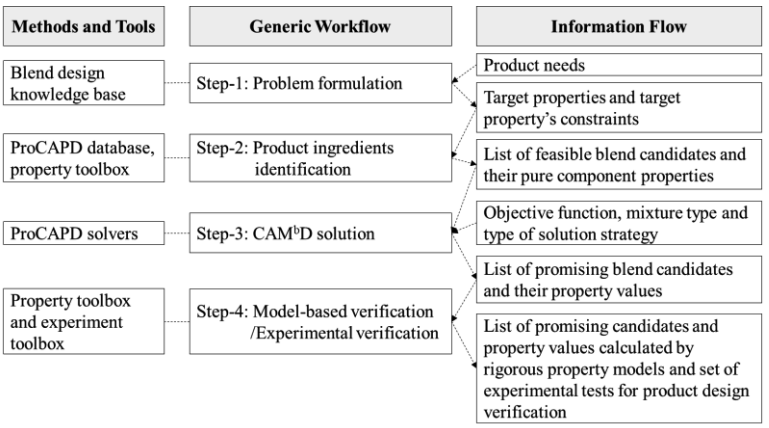


Figure 1: Architecture (framework) of the CAM^{b}D option in the ProCAPD software.

In Step-1, a product design problem is defined through the definition of needs and their translation to target properties for the desired products. Chemical products such as mixtures and blends are obtained by mixing several selected ingredients (chemicals) to obtain the desired product attributes (target properties). Therefore, in Step-2, the list of

feasible blend candidates (ingredients) is generated for the defined product. The fastest way to create this list is to simply search the available chemicals from a database. However, as databases are usually incomplete, another alternative is to employ CAMD to generate this list. CAMD methods such as "generate and test" (Harper et al., 2000) and mathematical optimization methods are commonly employed for the generation of ingredient lists. In Step-3, the mixture and blend design problems are formulated and solved. The blend design problem represents an MILP/MINLP model, depending on the specific property model used. In this paper, a two-step solution approach (Zhang et al., 2015) and a decomposition-based algorithm (Yunus et al., 2014) are employed to solve the MINLP problems. Finally, in Step-4, each feasible blend from Step-3 is verified by either rigorous model-based tests or experimental tests.

3. ProCAPD Application Examples

Two application examples involving the design of a diesel surrogate and a tailor made jet-fuel blend are illustrated through the use of the CAM^{b}D template in ProCAPD.

3.1 Design of a Diesel Surrogate

Conventional diesel fuels comprise of a large number of compounds that are extremely difficult to be simulated for better understanding of the fuel behavior in combustion engines. Therefore, the objective of this case study is to design a diesel surrogate that has fewer compounds and emulates certain important physicochemical properties of the conventional diesel.

Step-1: Problem Definition – The new diesel surrogate should have good fuel performance and meet requirements for the ASTM D975 standard as listed in Table 1.

Table 1. List of product needs and their corresponding target property constraints

Need	Property	Unit	Target Property	
			Diesel Surrogate	Tailor-made Jet-Fuel Blend
Ability to be burned	Reid vapor pressure	kPa	$RVP \leq 1.38$	$RVP \leq 1$
	ASTM distillation temp	°C	$T_{90} \leq 288$	$T_{10} \leq 205$
Flammability	Flash point	°C	$T_f \geq 52$	$T_f \geq 38$
Engine efficiency	Higher heating value	MJ/kg	$HHV > 35$	$HHV \geq 43$
	Density at 15 °C	g/mL	$0.79 \leq \rho \leq 0.87$	$0.77 \leq \rho \leq 0.84$
	Cetane Index	-	$50 \leq CI \leq 60$	
Consistency of fuel flow	Kinematic viscosity at 20°C	mm^2/s	$1.3 \leq V \leq 4.1$	
	Kinematic viscosity at -20°C	mm^2/s		$0 \leq V \leq 8$
	Melting point	°C		$T_m \leq -40$
Environmental impacts	$-log(LC_{50})$	mol/L	$-log(LC_{50}) < 5$	$-log(LC_{50}) \leq 4.58$
	CO_2 emission	kgCO_2/mile		$CO_2E \leq 25.36$

Step-2: Product Ingredients Identification – A set of chemicals (ingredients) that can be blended together in order to match the target properties is generated using the CAMD option provided in ProCAPD. 20 chemicals are generated according to the benchmark of the existing diesel fuel such as boiling point, type of chemicals, melting point and flash point. Pure compound and mixture property models of the chemicals are retrieved from the ProCAPD database. Missing properties are estimated through the property toolbox.

Step-3: CAMʰD Solution – The surrogate design problem is formulated as a MINLP problem using the decomposition approach, where ρ is maximized, subject to target properties. Initially, the software lists 20 chemicals as feasible candidates, which can formulate 15,504 quinary mixtures (5 different chemicals mixed together). Out of which, 2,713 mixtures are excluded due to their linear property constraints and a further 6,254 mixtures are excluded because they form immiscible blends. Subsequently, 6,537 mixtures are left which are evaluated to find the optimal mixture compositions that satisfy all constraints, linear as well as non-linear. The composition (in vol%) of the most promising surrogate consists of n-dodecane (15%), n-tetradecane (31%), cyclo-octane (44%), iso-cetane (9%) and toluene (1%). The properties of the diesel surrogate are listed in Table 2.

Step-4: Model-Based Verification/Experimental Verification – Rigorous property models for *V*, *CI* and ASTM distillation temperature (T_{10}, T_{50} and T_{90}) are applied in this step (see Table 2). The results predicted by ProCAPD are in good agreement with the experimentally measured data (Choudhury et al., 2017). For the engine performance analysis, net torques of the surrogate and the conventional diesel are similar. *CI* value of the surrogate (52) is lower than the conventional diesel (57) which leads to better complete combustion along with ignition delay characteristics. Therefore, the selected diesel surrogate satisfies all target properties and has similar combustion and emission characteristic as the conventional diesel.

Table 2. List of target property values of the diesel surrogate (*: Experiment)

RVP	*HHV*	*V*	*ρ*	*-logLC50*	T_f	T_{10}	T_{50}	T_{90}	*CI*
1.1	47	2.0	0.8	4.21	60	156	184	232	52
1.1*	47*	1.8*	0.8*	-	-	-	-	-	53*

3.2 Design of a Tailor-Made Jet-Fuel Blend

PK fuels (Paraffinic Kerosene) can be used as alternative fuels for commercial aviation. However, PK fuels lack certain chemical constituents, which although they are benign to the environment, offer a trade-off in performance when used in a jet-fuel engine. Therefore, the objective of this example is to improve properties of a GTL PK fuel by blending with feasible additives.

Step-1: Problem Definition – The tailor-made jet-fuel blend should have good fuel performance and meet or exceed stringent requirements for the ASTM D1655 standard as listed in Table 1.

Step-2: Product Ingredients Identification – The tailor-made jet-fuel blend needs to include the GTL PK fuel plus appropriate additives. The composition of the GTL PK fuel is obtained from GC analysis provided by the TEES Gas & Fuels Research Center, Texas A&M University at Qatar. 50 feasible additives are generated using the CAMD technique in ProCAPD according to the benchmark of the existing jet-fuel.

Step-3: CAMʰD Solution – The surrogate design problem is formulated as a MINLP problem using the decomposition approach and the two-step approach, where the composition of the GTL PK fuel is maximized, subject to target properties. In the two-step solution appraoch, 143 equations are generated with 30,099 continuous variables and

51 discrete variables. In the decomposition-based algorithm, initially, a total of 50 chemicals are selected as additives from which 1,225 binary mixtures can be formed, out of which 903 mixtures are excluded due to their linear property constraints. The remaining 332 ternary mixtures are all miscible and subsequently all 332 mixtures are evaluated to determine the optimal mixture compositions that satisfy all target property constraints, linear as well as non-linear. The most promising blends with the maximum GTL PK composition obtained from the decomposition-based algorithm and the two step solution approach are listed in Table 3. The calculated target properties of the promising blends are listed in Table 4.

Table 3. List of blends matching the jet-fuel target properties ((DC: the decomposition-base algorithm; TS: The two step solution approach))

ID	Composition (vol%)
1 (DC)	GTL PK (80.2) Decalin (8.3) Butylbenzene (11.5)
2 (DC)	GTL PK (78.6) Decalin (12.4) Pentylbenzene (9)
3 (DC)	GTL PK (77) Decalin (6) Hexylbenzene (17)
4 (TS)	GTL PK (77) Decalin (18) Hexylbenzene (5)

Table 4. List of blends matching the jet-fuel target properties and their properties

ID	RVP	HHV	V	ρ	CO_2E	$-logLC_{50}$	T_f	T_{10}	T_{50}	T_{90}	T_m	WSD
1	0.57	46.6	3.5	0.78	21.4	4.41	56	183	197	222		
		46.9*	3.8*	0.78*			56*	181*	195*	220*	-55*	0.70*
2	0.55	46.5	4.2	0.79	20	4.32	56	186	202	221		
		46.8*	4.2*	0.78*			58*	184*	199*	219*	-55*	0.73*
3	0.52	46.5	4.2	0.78	22	4.51	57	190	209	228		
		46.9*	4.2*	0.78*			60*	187*	205*	222*	-55*	0.64*
4	0.50	43.1	4.1	0.78	24	4.42	51	183	197	222		
GTL PK	0.64	47.6	4.0	0.756	23	4.58	58	184	204	228		
		47.0*	4.1*	0.752*			56*	181*	201*	223*	-49	0.712

Step-4: Model-Vased Verification/Experimental Verification – Rigorous property models to predict η and ASTM distillation temperatures are applied in this step (see Table 4). The results predicted by ProCAPD are in good agreement with the experimentally measured data. All blends satisfied the target properties constraints. Blending aromatics with the GTL PK fuel increases the freezing points along with the better pumpability of the blends. The large wear scar diameter (WSD) implies poor lubricity. The upper bound of WSD is 0.85 mm. Blend ID1 and Blend ID3 are promising candidates because they improve all target properties. In addition, the volumetric heating values of Blend ID1 (36.5 MJ/mL) and Blend ID3 (36.4 MJ/mL) are higher than the GTL PK (33.5 MJ/mL). Although the lubricity of Blend ID3 is better than Blend ID1, its gives higher aromatic content, this may generate more particulate matter (PM) during engine operation that reduce engine lifetime and can cause serious respiratory and cardiovascular health problems. Therefore, Blend ID1 is the best blend candidate.

4. Conclusions

A systematic framework for CAM^{b}D has been developed and implemented as the part of the CAM^{b}D template of the ProCAPD. The application of the CAM^{b}D template is

highlighted through examples involving the design of a diesel surrogate and a tailor-made jet-fuel blend. In order to save time and avoid combinatorial explosions within a very large search space of solving CAM^{b}D problems, the large mixed-integer non-linear problems of the three fuels are solved using the decomposition-based and the two step solution approach. The results from the case studies reveal the uncertainty of choosing the methods to solve CAM^{b}D problems. Both methods identify promising blends with familiar additives, however, the decomposition based algorithm tends to give better solutions. It helps to obtain the maximum GTL PK compositions with better target properties for the design of the jet-fuel blend. Both methods help to reduce the search space and provide promising chemical candidates that are competitive, and environmentally feasible, making it flexible and capable of solving a wide range of CAM^{b}D problems. Therefore, ProCAPD enhances the future development of chemical product design as huge amounts of data, models, knowledge, methodologies and algorithms are integrated and managed in a systematic and efficient way, increasing the possibility to capture past experiences and provide better guidelines for future chemical products (Gani, 2004).

References

E. Conte, R. Gani, T.I. Malik, 2011, The virtual Product-Process Design laboratory to manage the complexity in the verification of formulated products, Fluid Phase Equilibria, 302, 294-304.

H.A. Choudhury, S. Intikhab, S. Kalakul, R. Gani, N.O. Elbashir, 2017. Integration of computational modeling and experimental techniques to design fuel surrogates, Journal of Natural Gas Science and Engineering, ISSN 1875-5100.

L. Zhang, D.K. Babi, R. Gani, 2016, New Vistas in Chemical Product and Process Design, Annual Review of Chemical and Biomolecular Engineering, 7, 557-582.

N.A. Yunus, K.V. Gernaey, J.M. Woodley, R. Gani, 2014. A systematic methodology for design of tailor-made blended products. Computers & Chemical Engineering, 66, 201-213.

P.M. Harper, R. Gani, 2000, A multi-step and multi-level approach for computer aided molecular design, Computers & Chemical Engineering, 24, 677-683.

R. Gani, 2004, Chemical Product Design: Challenges & Opportunities, Computers and Chemical Engineering, 28, 2441-2457.

S. Jonuzaj, P.T. Akula, P.M. Kleniati, C.S. Adjiman, 2016. The formulation of optimal mixtures with generalized disjunctive programming: A solvent design case study, AIChE Journal, 62(5), 1616-1633.

S. Kalakul, M.R. Eden, R. Gani, 2017, The Chemical Product Simulator - ProCAPD, Computer Aided Chemical Engineering, 40, 979-984.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50074-4

Hybrid Method/Tool for Sustainable Process Synthesis, Design, Analysis, and Improvement

Anjan K. Tula[a], Mario R. Eden[a*], Rafiqul Gani[b]

[a]*Department of Chemical Engineering, Auburn University, Auburn, AL 36849, USA*
[b]*PSE for SPEED, Skyttemosen 6, Allerød, DK-3450, Denmark*
edenmar@auburn.edu

Abstract

Due to increased concerns regarding environment, public health and safety through regulatory enforcements and growing competitiveness through globalization, there is an incentive and a quest for innovative designs that are efficient, profitable as well as sustainable. To generate a truly optimal process flowsheet a systematic method to identify the types of tasks-operations that need to be performed, the corresponding design of the operation-equipment, their configuration, mass-energy flows, environmental impacts, etc., need to be employed. In this case, development and use of a hybrid method that combines the best features of knowledge-based and optimization-based approaches into one method is an attractive option. This work focuses on the development of an integrated method and computer aided software tool capable of enumerating all feasible alternatives within the entire search space defined by a process synthesis problem. Solving this type of large and complex synthesis problems using mathematical formulations is usually cumbersome and difficult. In this work we integrated ProCAFD (computer-aided synthesis tool used for enumeration) with Super-O, which automates the mathematical formulation of the synthesis problem thereby leading to very quick and reliable solutions. An overview of the key concepts and the method is presented along with details of the developed computer-aided tool.

Keywords: Process synthesis, Mathematical programming, ProCAFD, Super-O.

1. Introduction

In process synthesis the objective is to find the best processing route that can convert a given set of raw materials to desired products. In recent years, diminishing natural resources and increasing environmental awareness has driven the process industry towards development of more sustainable and innovative processes. To generate truly innovative and optimal solutions it is imperative that one considers the entire search space of alternatives and solve the resulting superstructure based optimization problem. In this work a superstructure of all feasible alternatives is created using a technique similar to computer aided molecular design (CAMD) (Tula et al., 2015). According to this technique, chemical process flowsheets are synthesized in the same way as atoms or groups of atoms are combined to form molecules in CAMD techniques. The building blocks in the process flowsheet synthesis problem are called process-groups (PG), which represent a single or set of unit operations that are selected by employing a thermodynamic insights based method (Jaksland et al. 1995). These building blocks are then combined using connectivity rules to enumerate all feasible process flowsheet alternatives.

In general, two solution strategies are commonly employed to solve superstructure based synthesis problems: simultaneous approach and decomposition-based approach. In the simultaneous approach, the synthesis problem is solved by considering all the constraints and components of the process synthesis simultaneously. On the other hand, in the decomposition approach the synthesis problem is divided into a set of sub problems that are solved sequentially in a predefined order. Mathematical optimization methods are employed by the simultaneous approach to determine the best processing route from a superstructure of alternatives. Industrially relevant problems in general pose mathematical challenges due to their size and complexity in formulation. Consequently, the main focus of most of the research in this area has been related to the development of solution methods, aiming at solving progressively larger and more complex problems, while at the same time minimizing the computational power required. In this work we have integrated ProCAFD and Super-O, where ProCAFD was used to enumerate all the possible flowsheets while Super-O automates the mathematical formulation of the generated alternatives to solve the synthesis problem using a mathematical optimization approach.

2. Methodology and Dataflow

The developed synthesis, design, analysis and improvement method is based on the three-stage approach (Babi et al., 2015) for innovation as shown in Figure 1.

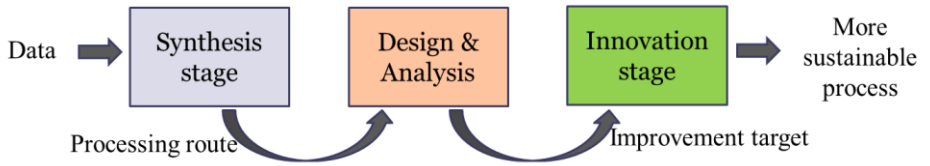


Figure 1: Three stage approach for sustainable synthesis, design, analysis of flowsheets

In stage 1, process synthesis is performed, and all the feasible alternatives are generated for a given synthesis problem. This superstructure of alternatives is solved to identify the best processing route based on process constraints and performance criteria. In stage 2 additional design details are added to the process tasks-operations so that simulations with rigorous process models can be performed. The simulation results serve as input for equipment sizing and utilities requirement calculations, which in turn serve as input for process analysis (economic, sustainability factors, LCA factors, etc.). Based on this analysis process hot-spots are identified, which are translated into design targets, that is, if satisfied, eliminates the process hot-spots and generates sustainable flowsheet alternatives. In stage 3, the optimal processing route is further improved by applying various synthesis-intensification methods to determine process alternatives that match the targets for improvement. Any match of the design targets corresponds to a more sustainable process alternative. In this paper, only stage 1 is discussed in detail as the methodology has been extended for this stage.

2.1. Synthesis Stage

The synthesis method consists of 6 steps: (1) problem definition; (2) problem analysis; (3) process-groups selection; (4) generation of flowsheets; (5) superstructure generation; and (6) selection of optimal flowsheet.

2.1.1. Problem Definition: The objective here is to define the synthesis problem, design constraints and the performance criteria based on which the generated alternatives are to be benchmarked. The structural description of the synthesis problem is defined through data on raw materials (input streams), desired products (outputs stream) and process specifications (for example, product purity, product recovery, etc.)

2.1.2. Problem Analysis: The objective of this very important step is to generate the information required for solving the synthesis problem. Analysis is performed to further define the synthesis problem by knowledge bases and physical insight based methods. First, reaction analysis is performed to identify reaction tasks needed to produce the desired product as defined in the previous step. A database search is performed to find the reaction mechanisms yielding the desired product. Next, pure component and mixture property analysis is performed for all the chemical species listed in the problem to generate data that can be used for identification of separation tasks and corresponding feasible separation techniques. Analysis of 22 pure component properties for the system chemicals is performed. Based on this analysis a table of property ratios is generated for all the binary pairs of chemicals. Mixture property analysis identifies the existence of possible azeotropes, eutectic points and the available driving forces for all identified binary pairs. Based on the mixture analysis and the property ratios feasible separation techniques are identified using the extended Jaksland et al. (1995) physical insights-based method.

2.1.3. Process-Groups Selection: Here, all the process-groups that are applicable for the synthesis problem are selected and initialized based on the problem definition and analysis steps. Information from the problem definition step is used to select and initialize the inlet and outlet process-groups with corresponding components. From the problem analysis step, corresponding reaction process-groups and separation process-groups are selected and initialized.

2.1.4. Generation of Flowsheets: In this step, the initialized process-groups are combined according to a set of rules and specifications to generate feasible flowsheet structures. This is achieved through the flowsheet generation method based on a PG-combinatorial algorithm, which is governed by a set of connectivity and logical rules that ensure the generation of only feasible solutions, thereby avoiding any combinatorial explosion.

2.1.5. Superstructure Generation: Based on all the generated process alternatives a superstructure is generated and represented in the form of a Processing Step-Interval Network (PSIN). Figure 2 represents an illustration of a superstructure linking raw materials and products using process intervals. Here process steps, which are represented as columns correspond to different sequences of processing tasks to convert raw materials to products. Process intervals represent different techniques to achieve the particular processing task. The main advantage of representing the superstructure in the form of PSIN is that a generic model (Quaglia et al., 2015) for the process interval can be used to perform optimization on the superstructure. The optimal processing route determined this way, also provides mass and energy balance results for each processing step. Through the generic model, each process interval of the superstructure is represented by a sequence of tasks (Figure 3): (1) mixing, (2) reaction, (3) waste separation, and (4) product separation. The model allows all the interval blocks to be defined by the same set of equations related to the tasks allowing simplified data flow and time effective formulation of large problems. The model is sufficiently flexible to allow multiple inlets to and outlets from the interval, including recycle streams from downstream intervals and bypasses.

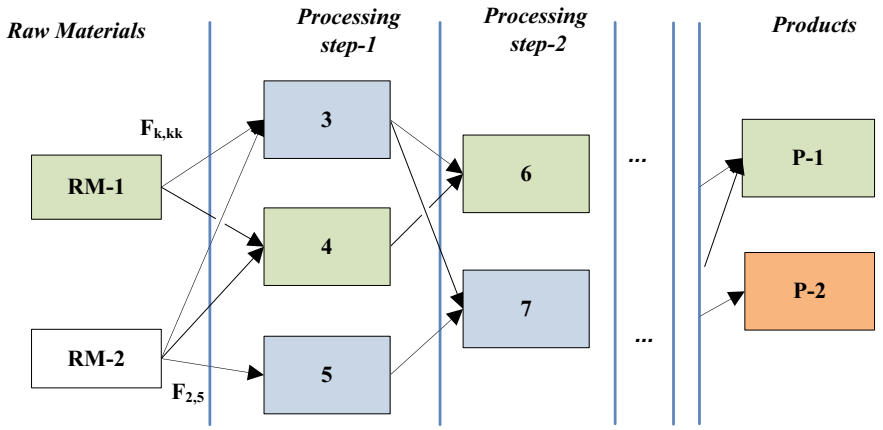


Figure 2: Generic superstructure representation (PSIN)

The model parameters required are given by the definition of the process-groups. The raw material compositions are obtained from the inlet process-groups, reaction process groups provide the conversion factors while the separation split factors are obtained from the separation process-groups. For instance, in the case of the distillation process-group, the recovery of the components lighter than the light key is set to 100% in the overhead product and the recovery of the components heavier than the heavy key is set to 100% in the bottom product. The recovery of the key components is greater than or equal to 99.5%. Similarly, separation factors based on different driving forces available for each process-group are defined.

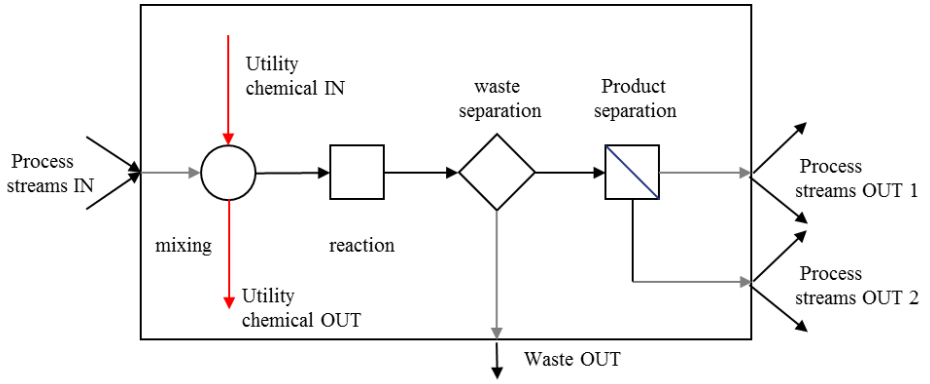


Figure 3: Generic model for process interval

2.1.6. Selection of Optimal Flowsheet: The optimization problem formulated in step 5 is solved by direct approach (that is, all the equations are solved simultaneously). The problem formulation is in the form of a MI(N)LP which is solved here with an appropriate solver (using GAMS). The solution gives the optimal process alternative to convert the given raw material to products.

3. Software implementation

ProCAFD (Tula et al., 2017) is a computer-aided software developed in C# based on the CAMD approach which uses decomposition strategy to solve the synthesis problem. However, in this work the scope of the software is further expanded to solve the synthesis problem in a simultaneous approach using mathematical programming. Figure 4 shows the schematic representation of the integrated software based on the developed method. ProCAFD is integrated with structured databases, methods, in-house tools and external simulation tools like Aspen Plus, PRO/II, which can be accessed through the main user interface, and interact to provide sufficient information to solve the synthesis-design problem.

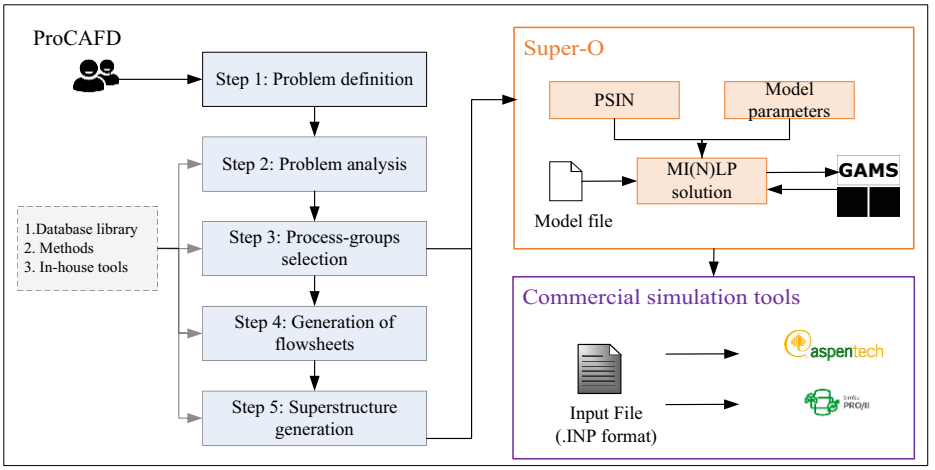


Figure 4: Schematic representation of integrated computer-aided tool for process synthesis

3.1. Software Workflow: In ProCAFD, the user defines the synthesis problem and all the necessary information is retrieved from the integrated database libraries. Based on the data provided, ProCAFD generates information to create the superstructure of all feasible alternatives using different methods and integrated in-house tools. The generated superstructure is represented in the form of Processing Step-Interval Network and sent to Super-O along with the model parameters. Based on this information Super-O generates a GAMS-readable file containing all problem data and generic equations using the generic model file. Super-O which is integrated with GAMS, directly calls the appropriate solver, to solve the optimization problem and generate an output file containing the results. ProCAFD is also integrated with commercial simulators to verify the design alternatives using rigorous models (stage 2). Based on the selected optimal flowsheet from Super-O, ProCAFD calculates the operational design parameters and an input file (.INP) is generated which can be accessed directly in commercial simulation packages like Aspen Plus or PRO/II.

3.2. Case Studies for Testing: This integrated tool has been tested by solving various synthesis problems. The problem size and model size of three cases solved using ProCAFD are listed in Table 1. More details on the solution can be obtained from the authors.

Table 1: List of case studies solved.

Case	NPI	NP	NC	NR	NEQ	NV(NDV)	TA
Hydrodealkylation of toluene	48	2	5	2	51978	50560(106)	340
DMC from propylene carbonate	31	2	4	2	20534	19763(72)	53
Cumene from benzene and propylene	56	2	5	2	68125	66488(122)	240

Abbreviations are as follows: NPI is the number of processing intervals; NP is the number of products; NC is the number of components; NR is the number of reactions; NEQ is the number of equations; NV is the number of variables; NDV is the number of discrete variables; TA is total number of alternatives.

4. Conclusions

In this work an integrated hybrid method is proposed for sustainable process synthesis which combines physical insights based methods with optimization based approach. While the physical insights methods make sure that all feasible flowsheet alternatives are generated, the optimization based approach ensures optimal solution from the generated superstructure. There is a limit on the number of flowsheets that can be handled (generated and analyzed) using a manual approach for synthesis methods. However, by application of a computer-aided software-tool, multiple flowsheet alternatives can be generated and screened efficiently and reliably. A computer-aided software-tool has been presented together with a software architecture that incorporates multiple methods, algorithms and computer-aided tools. Through the introduction of a generic model to represent process intervals, faster and systematic formulation of large synthesis-design problems, which are solved through different options in ProCAFD, are made possible. In this work the integration of ProCAFD with external simulation tools like Aspen Plus and PRO/II has been made, thereby enabling the user to perform rigorous simulations of any selected process alternative.

References

D.K. Babi, J. Holtbruegge, P. Lutze, A. Gorak, J.M. Woodley, R. Gani, 2015, Sustainable process synthesis–intensification, Computers Chem Eng, 81, 218-244

M.-O. Bertran, R. Frauzem, A.-S. Sanchez-Arcilla, L. Zhang, J.M. Woodley, R. Gani, 2017, A generic methodology for processing route synthesis and design based on superstructure optimization, Computers Chem Eng, 106, 892-910

C. Jaksland, R. Gani, K. M. Lien, 1995, Separation process design and synthesis based on thermodynamic insights. Chem Eng Sci, 50, 511-530

A. Quaglia, C.L. Gargalo, S. Chairakwongsa, G. Sin, R. Gani, 2015, Systematic network synthesis and design: problem formulation, superstructure generation, data management and solution, Computers Chem Eng, 72, 68-86

A.K. Tula, M.R. Eden, R. Gani, 2015, Process synthesis, design and analysis using a process-group contribution method, Computers Chem Eng, 81, 245-259

A.K. Tula, J. Bottlaender, M.R. Eden, R. Gani, 2017, A computer-aided software-tool for sustainable process synthesis-intensification, Computers Chem Eng, 105, 74-95

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50075-6

Work and Heat Exchange Networks – Opportunities and Challenges

Haoshui Yu[a], Chao Fu[b], Truls Gundersen[a*]

[a]*Department of Energy and Process Engineering, Norwegian University of Science and Technology, Kolbjoern Hejes v. 1A, NO-7491 Trondheim, Norway*
[b]*SINTEF Energy Research, Kolbjoern Hejes v. 1A, NO-7491 Trondheim, Norway*
truls.gundersen@ntnu.no

Abstract

Design and optimization of Heat Exchanger Networks (HENs) is well established and has been heavily applied in the process industries since its pioneering period in the 1970s and 1980s. While temperature and thermal energy (heat) obviously are key elements in process plants and power stations, pressure and mechanical energy (work) are equally important. By adding expanders and compressors as equipment and power as a utility to the classical HEN problem, the considerably more challenging Work and Heat Exchange Network (WHEN) problem has been formulated as a new and growing discipline within Process Systems Engineering (PSE). Considerable opportunities exist for improving energy efficiency in process plants and power stations by using the WHEN methodology. Recent applications include (I) design of LNG processes, (II) design of CO_2 capture processes, and (III) an industrial sensible heat pump. In fact, WHENs have strong similarities to Heat Pumps and Refrigeration Cycles, with the additional advantage of a PSE approach. Challenges in the graphical methodology for WHENs compared to HENs include (a) the shape of Composite and Grand Composite Curves will change with pressure manipulations, (b) the Pinch points may change, (c) the hot and cold utility demands will change, (d) the stream identity (hot or cold) may temporarily change, (e) process streams may be used as utilities, and (f) work and heat have different energy quality. To overcome the above challenges, the WHENs problem could be addressed by developing Optimization models using Total Annual Cost as the objective function. However, new challenges will appear if Mathematical Programming is used, such as (i) developing a sufficiently rich yet efficient superstructure, (ii) non-convexities in the model that may result in local optima, (iii) potentially a large number of binary variables that may give a combinatorial explosion, and (iv) discontinuities in the process models that will cause numerical problems and slow down convergence, unless recent developments in nonsmooth analysis can be adopted for these problems.

Keywords: Work and Heat Exchange Networks, Process Integration, Optimization

1. Introduction

The objectives of this paper are (i) to introduce the new research field WHENs to the PSE community, (ii) to provide a brief review of this research area, and (iii) to discuss challenges and opportunities. The main contributions of the paper are thus (ii) and (iii).

In the process industry, a significant amount of energy can be saved by Process Integration. Heat exchange between heat sources and heat sinks to reduce hot and cold utility consumption is the key idea of heat integration, which has been a successful

application of process integration. The Heat Exchanger Networks (HENs) problem has been well defined and studied since the 1970s (Gundersen and Naess, 1988). In the process industry, pressure is also an important parameter. It is obvious that work exchange between high and low-pressure streams will reduce work consumption (shaft work or electricity), which is another example of Process Integration (Huang and Fan, 1996). Razib et al. (2012) developed a superstructure for Work Exchange Network (WEN) configurations minimizing total annualized cost. Significant energy savings verified the efficacy of the model. Thus, WENs aiming at reducing work consumption is emerging in Process System Engineering. If HENs and WENs are considered simultaneously, energy efficiency can be further improved by considering heating from compression and cooling from expansion in the heat recovery problem. This gives rise to a new problem referred to as Work and Heat Exchange Networks (WHENs).

Aspelund et al. (2007) proposed a graphical design methodology based on Pinch Analysis referred to as ExPAnD. Their study considers the pressure change of process streams and extends the heat integration methodology to the WHENs problem. The key contribution from their work is that compression (expansion) adds heating (cooling) and should be done above (below) the process Pinch. Fu and Gundersen (2015a) presented a systematic graphical design procedure based on four theorems for integration of compressors in HENs above ambient temperature. Similarly, four theorems were proposed to integrate compressors into heat exchanger networks below ambient temperature (Fu and Gundersen, 2015b). The integration of expanders into heat exchanger networks above and below ambient temperature are also investigated (Fu and Gundersen, 2015c,d). It is concluded that compression/expansion above/below ambient temperature should start at distinct temperatures, such as Pinch, ambient or hot/cold utility depending on the actual case. No other temperatures will give lower exergy consumption. Later, a systematic methodology was developed by Fu and Gundersen (2016a) to integrate both compressors and expanders into heat exchanger networks above ambient temperature. Fu and Gundersen (2016b) summarized the fundamental thermodynamic insights and applied them to three carbon capture processes. Considerable energy savings have been achieved by heat and work integration based on these thermodynamic insights. However, there are still many challenges in WHENs. This paper mainly focuses on the realized and potential applications of WHENs and the challenges in this field.

2. Problem definition

The WHENs problem in its most general form can be stated as follows: “Given a set of process streams with fixed supply and target states (pressure and temperature), as well as utilities for heating, cooling and power; design a Work and Heat Exchange Network consisting of heat transfer equipment such as heat exchangers, heaters and coolers, as well as pressure changing equipment such as compressors, expanders, pumps and valves, in such a way that Exergy consumption or Total Annual Cost is minimized.”

What makes WHENs different from HENs is that the thermodynamic path from supply to target state depends on the sequence and the extent of changes in pressure and temperature. As a result, the shape of the Composite and Grand Composite Curves will change, and the design problem becomes considerably more difficult. In addition, the identity of streams (hot/cold) may temporarily change. Even a stream with constant pressure could temporarily have its pressure changed by compression and expansion. In that way streams would act as a utility, which is often done below ambient temperature.

The WHENs problem has a lot in common with heat pumps and refrigeration cycles, however, one important difference is that WHENs offer a systematic and systems-oriented approach to solving the design problem. Heat pumps and refrigeration cycles have been addressed as standalone units with the exception that economizers are used in refrigeration cycles, i.e. an integration of heat between the cycle and the background process. Another difference is that the WHENs methodology also considers compressors and expanders that are not part of cycles.

3. Opportunities for Work and Heat Exchange Networks

There are many possible applications of WHENs in the process industry. These opportunities arise in both above and below ambient temperature processes. WHENs become more important below ambient temperature due to the fact that refrigeration is obtained by compression/expansion. In this section, some realized and potential applications of WHENs will be presented by examples.

3.1. Offshore LNG process

Aspelund and Gundersen (2009) proposed a novel energy and cost-effective transport chain for stranded natural gas. Using liquefied N_2 and CO_2 as carriers of cold exergy, natural gas is liquefied and transported by ship to an onshore power plant using oxy-combustion for CO_2 capture. Figure 1 shows the process streams that take part in the offshore process WHEN design. Wechsung et al. (2011) combined models for Pinch operator, exergy operator and pressure operator to synthesize the WHEN in the offshore LNG process. The results show that pressure manipulation can significantly reduce the irreversibilities of the system. Huang and Karimi (2016) also applied their WHENs model to synthesize the off-shore LNG process. However, they treat natural gas and nitrogen as simplified streams with constant heat capacity flowrate, which is a rather crude assumption. Onishi et al. (2015) also developed a mathematical model for the retrofit of WHENs and verified the model by this offshore LNG process. Considerable energy savings from work and heat integration are reported in these studies.

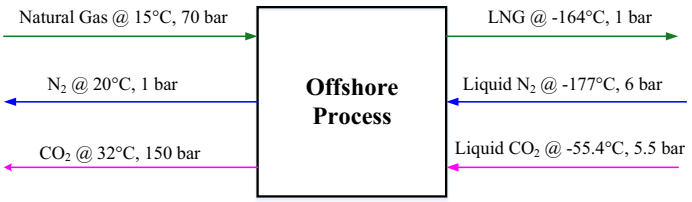


Figure 1 WHENs problem in an offshore LNG process (Aspelund and Gundersen, 2009)

3.2. Post-combustion carbon capture processes

Carbon capture and storage is a competitive alternative to reduce CO_2 emissions. Zhao et al. (2010) proposed a post-combustion carbon capture process using multistage CO_2/N_2 gas separation membranes. There are two membranes in this process. The retentate from both membranes at 8 bar can be expanded to 1 bar to recover shaft work. If the retentate streams are integrated with flue gas and air in the air preheater, more expansion work can be recovered. Fu and Gundersen (2016b) applied the

thermodynamic insights for WHENs to this carbon capture process. Work and heat integration can reduce the specific work consumption by 12.9%.

3.3. Sensible heat pump for industrial heat recovery

Recently, Fu and Gundersen (2016c) proposed a novel sensible heat pump based on a reversed Brayton cycle. Their study is an extension of the appropriate placement concept for compressors and expanders by Fu and Gundersen (2016a). In order to recover the compression heat and expansion work efficiently, a sensible heat pump is developed based on work and heat integration. The sensible heat pump is more competitive for cases where there is no sharp Pinch point on the Grand Composite Curve (GCC), the temperature range is large, and energy saving is a primary concern.

Holiastos and Manousiouthakis (2002) also applied a heat and work integration methodology to crude oil pre-heat train successfully. Their methodology is, however, limited to the integration of heat engines and heat pumps, thus it does not address heat integration opportunities from pressure changes to process streams. The above-mentioned studies are successful applications of WHEN methodologies. There is still a large number of potential applications of WHENs in the process industry, such as cascade refrigeration cycles in ethylene plants (Dinh et al., 2015) and low-temperature waste heat recovery by organic Rankine cycles in oil refineries (Yu et al., 2017).

4. Challenges in Work and Heat Exchange Networks

Although the WHENs problem has received increasing attention in the last decade and some methodologies have been developed, there are still considerable challenges. These challenges exist in practical issues, Pinch Analysis based methodologies, optimization based approaches, and selecting the appropriate objective function.

In realistic applications, more issues need to be considered, which are not addressed in the majority of current research: Multiple hot/cold utilities, gliding temperature utilities, multi-stage compression/expansion, phase change of the process streams, pressure ratio determination, and design issues related to operating compressors and expanders at high and low temperatures, respectively.

The methodologies based on Pinch Analysis rely on Composite Curves (CCs) and the GCC. Even though CCs and the GCC are powerful and reliable tools for HENs, they cannot be easily applied to WHENs for the following reasons: Pressure manipulations may change the identity (hot/cold) of the process streams and the CCs and GCC will definitely change after pressure manipulation. Consequently, the Pinch point may change correspondingly. Thermodynamic insights proposed by Fu and Gundersen (2016a) actually solve the issue of Pinch changes, however, their work mainly focuses on one or two pressure changing streams and single stage pressure manipulation. If multiple pressure manipulating streams are considered or the Pinch point changes after pressure change, the manual procedure will be very tedious for design of WHENs.

To overcome the above challenges, the WHENs problem could be addressed by optimization. Mathematical programming can deal with multiple streams, multiple thermal utilities and multiple stages of pressure manipulation, in most cases based on the idea of superstructures. However, new challenges will appear if mathematical programming is used. The models are commonly highly nonconvex nonlinear programming (NLP) or mixed integer nonlinear programming (MINLP) problems. The formulation and solution of MINLP optimization problems are not likely to reach the

same stage of maturity and reliability as linear and mixed integer programming. To the extent that integer (binary) variables are required to formulate the design problem, this represents another major challenge in optimization based models. The problem size and search tree in a Branch and Bound approach grow exponentially with the number of such discrete variables. A superstructure should be rich enough to consider all possible configurations, but this makes the problem difficult to solve efficiently. Therefore, a rich yet efficient superstructure is urgently needed for WHENs. It is not trivial to propose a superstructure that can take into account all the factors concerning operating cost, equipment cost, operability and flexibility. A few superstructures tailor made for WHENs have been proposed, however, a detailed discussion of these is beyond the scope of this paper. A new more general representation using fundamental building blocks aiming at process integration and intensification presented by Demirel et al. (2017) could have the potential to be used as a superstructure for WHEN synthesis. As far as discontinuities are concerned, it is hoped that recent developments in nonsmooth analysis can be adopted to solve these problems (Watson et al., 2015).

Since both heat and work are involved in WHENs, it is challenging to set up an appropriate objective function. There are generally two categories of objective functions; energy-related objective functions and economic indicators. Exergy is increasingly adopted as the energy-related objective function. Exergy can provide a unified treatment of different energy forms, such as work and heat. Equipment cost varies constantly and has large uncertainties. Thermodynamics will give sound and reliable criteria for process design. However, the exergy of heat is calculated based on reversible processes (i.e. the Carnot cycle), which results in an overestimation of the quality of heat. Valves are favorable because of their negligible capital cost compared with turbines, however, if exergy is the objective function, valves are not likely to be selected due to the exergy destruction in such units. To overcome the limitations of energy-related objective functions, a techno-economic objective function can be used in order to balance energy cost and equipment cost. However, accurate equipment cost correlations are not easy to get. Correlations for equipment cost in the literature are simplified, which may lead to unreliable results. A detailed economic analysis is still a challenge when techno-economic optimization is performed.

5. Conclusions

Work and Heat Exchange Networks (WHENs) is an emerging new topic in Process System Engineering. There have been some promising applications that illustrate the significant energy savings that can be obtained by synergizing work and heat integration. However, the methodology for WHENs is still under development and many challenges need to be addressed and overcome. Due to the respective disadvantages of Pinch based methods and Mathematical Programming, it is better to combine these methods taking advantage of the merit of each approach. The Pinch based methods can provide fundamental thermodynamic insights, which can help in developing more efficient superstructures and as a result, the size of the mathematical model can be reduced and make it easier to find optimal solutions. In addition, practical issues should also be taken into account in the future work.

Acknowledgements

This publication has been funded by HighEFF—Centre for an Energy Efficient and Competitive Industry for the Future. The authors gratefully acknowledge the financial

support from the Research Council of Norway and user partners of HighEFF, an 8 year Research Centre under the FME-scheme (Centre for Environment-friendly Energy Research, 257632/E20).

References

A. Aspelund, D.O. Berstad, T. Gundersen, 2007, An Extended Pinch Analysis and Design procedure utilizing pressure based exergy for subambient cooling, Applied Thermal Engineering, 27(16), 2633-2649.

A. Aspelund, T. Gundersen, 2009, A liquefied energy chain for transport and utilization of natural gas for power production with CO_2 capture and storage–Part 1, Applied Energy, 86(6), 781-792.

H. Dinh, J. Zhang, Q. Xu, 2015, Process synthesis for cascade refrigeration system based on exergy analysis, AIChE Journal, 61(8), 2471-2488.

S.E. Demirel, J. Li, M.M.F. Hasan, 2017, Systematic process intensification using building blocks, Computers & Chemical Engineering, 105, 2-38.

C. Fu, T. Gundersen, 2015a, Integrating compressors into heat exchanger networks above ambient temperature, AIChE Journal, 61(11), 3770-3785.

C. Fu, T. Gundersen, 2015b, Sub-ambient heat exchanger network design including compressors, Chemical Engineering Science, 137, 631-645.

C. Fu, T. Gundersen, 2015c, Integrating expanders into heat exchanger networks above ambient temperature, AIChE Journal, 61(10), 3404-3422.

C. Fu, T. Gundersen, 2015d, Sub-ambient heat exchanger network design including expanders, Chemical Engineering Science, 138, 712-729.

C. Fu, T. Gundersen, 2016a, Correct integration of compressors and expanders in above ambient heat exchanger networks, Energy, 116, Part 2, 1282-1293.

C. Fu, T. Gundersen, 2016b, Heat and work integration: Fundamental insights and applications to carbon dioxide capture processes, Energy Conversion and Management, 121, 36-48.

C. Fu, T. Gundersen, 2016c, A Novel Sensible Heat Pump Scheme for Industrial Heat Recovery, Industrial & Engineering Chemistry Research, 55(4), 967-977.

T. Gundersen, L. Naess, 1988, The synthesis of cost optimal heat exchanger networks: an industrial review of the state of the art, Computers & Chemical Engineering, 12(6), 503-530.

K. Holiastos, V. Manousiouthakis, 2002, Minimum hot/cold/electric utility cost for heat exchange networks, Computers & Chemical Engineering, 26(1), 3-16.

K. Huang, I.A. Karimi, 2016, Work-heat exchanger network synthesis (WHENS), Energy, 113, 1006-1017.

Y. Huang, L. Fan, 1996, Analysis of a work exchanger network, Industrial & Engineering Chemistry Research, 35(10), 3528-3538.

V.C. Onishi, M.A. Ravagnani, J.A. Caballero, 2015, Retrofit of heat exchanger networks with pressure recovery of process streams at sub-ambient conditions, Energy Conversion and Management, 94, 377-393.

M.S. Razib, M.M.F. Hasan, I.A. Karimi, 2012, Preliminary synthesis of work exchange networks, Computers & Chemical Engineering, 37, 262-277.

H.A.J. Watson, K.A. Khan, P.I. Barton, 2015, Multistream heat exchanger modeling and design, AIChE Journal, 61(10), 3390-3403.

A. Wechsung, A. Aspelund, T. Gundersen, P.I. Barton, 2011, Synthesis of heat exchanger networks at subambient conditions with compression and expansion of process streams, AIChE Journal, 57(8), 2090-2108.

H. Yu, J. Eason, L.T. Biegler, X. Feng, 2017, Process integration and superstructure optimization of Organic Rankine Cycles (ORCs) with heat exchanger network synthesis, Computers & Chemical Engineering, 107, 257-271.

L. Zhao, E. Riensche, L. Blum, D. Stolten, 2010, Multi-stage gas separation membrane processes used in post-combustion capture: Energetic and economic analyses, Journal of Membrane Science, 359(1), 160-172.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50076-8

Analysis of Evaporation Effects in Hydrotreating TBR models

Niels Frithjof Lomholdt*, Carlos Eduardo Ramírez-Castelán,
Jakob Kjøbsted Huusom

Department of Chemical and Biochemical Engineering, Technical University of Denmark (DTU), Building 229, Kgs. Lyngby DK-2800, Denmark
s134465@student.dtu.dk

Abstract

The effect of evaporation on the output of literature dynamic models for hydrotreating trickle-bed reactors was examined. The model system, presented by Mederos & Ancheyta (2007), was applied to a given oil treating case. In order to account for evaporation, Pro/II flash calculations were implemented over the length of the reactor in cocurrent operation. It was found that, given the model and the chosen reaction network, the heat of reaction would result in a large fraction of the liquid feed evaporating in the reactor. It was also found that the reaction rate of the feed influenced the effect of the evaporation. A higher reaction rate was found to widen the range of outputs attainable between high and low mass transfer rates.

Keywords: Hydrotreating, trickle-bed reactor, evaporation, adiabatic model, Pro/II

1. Introduction

Trickle-bed reactors (TBR) are catalytic reactors used to convert sulfur, nitrogen and aromatic compounds in hydrocarbon fuels. In the reactor, hydrogen is contacted with the fuel over the catalytic particles at high pressures resulting in highly exothermic reactions. In order to properly design and control TBRs, representative, rate-based, adiabatic and dynamic models are appreciated. Such a model was presented by Mederos & Ancheyta (2007). The model assumes that only normal gases (H_2, H_2S and NH_3) are subject to transfer between the gas and liquid phase, where they may be dissolved. This means that evaporation of the fuel is completely neglected from the mass and energy balances. A high level of evaporation, could cause incomplete wetting of the catalyst as well as reactor hot spots resulting in catalyst degradation. Rigorous modeling of rate-based evaporation and phase equilibria is deemed to add a very high level of complexity to the model, but neglecting it may be a serious deficiency. Therefore, in order to examine the potential effects of evaporation on the aforementioned model, it was combined with process simulation software to calculate the equilibrium compositions of the phases. This is based on the novel concept presented by Ramírez-Castelán et al. (2017). The inclusion of flash calculations was used to compare the model outputs at different rates of mass transfer with the output neglecting evaporation.

2. Methodology

The model presented by Mederos & Ancheyta is a first-principle partial differential equation system. The model was, in this work, modified by taking into account the change in gas velocity as demonstrated by Alvarez & Ancheyta (2012). The model equations are

given in Eq. (1) to (6) with the gas velocity derivative given in Eq. (7). The differential equation system was discretized in the axial direction using a backwards finite-difference method with 100 space steps.

$$\frac{\varepsilon_G}{R\,T_G}\frac{\partial p_j}{\partial t} = -\frac{1}{R\,T_G}\frac{\partial(u_G\,p_j)}{\partial z} - k_j^{GL}\,a_{GL}\left(\frac{p_j}{H_j} - c_j^L\right) \tag{1}$$

$$\varepsilon_L\frac{\partial c_j^L}{\partial t} = -u_L\frac{\partial c_j^L}{\partial z} + k_j^{GL}a_{GL}\left(\frac{p_j}{H_j} - c_j^L\right) - k_j^{LS}a_{LS}\left(c_j^L - c_j^S\right) \tag{2}$$

$$(1-\epsilon)\,\varepsilon_p\frac{\partial c_j^S}{\partial t} = k_j^{LS}a_{LS}\left(c_j^L - c_j^S\right) + \rho_B(\eta\,r)_j \tag{3}$$

$$\varepsilon_G\rho_G c_{pG}\frac{\partial T_G}{\partial t} = -u_G\rho_G c_{pG}\frac{\partial\,T_G}{\partial z} - U_{GL}a_{GL}(T_G - T_L) \tag{4}$$

$$\varepsilon_L\rho_L c_{pL}\frac{\partial T_L}{\partial t} = -u_L\rho_L c_{pL}\frac{\partial T_L}{\partial z} + U_{GL}a_{GL}(T_G - T_L) - U_{LS}a_{LS}\,(T_L - T_S) \tag{5}$$

$$(1-\epsilon)\rho_S c_{pS}\frac{\partial T_S}{\partial t} = U_{LS}a_{LS}(T_L - T_S) + \textstyle\sum_i \rho_B(\eta\,r)_i(-\Delta H_i) \tag{6}$$

$$\frac{\partial u_G}{\partial z} = \frac{RT_G}{P}\left(\frac{\varepsilon_G P}{RT_G^2}\frac{\partial T_G}{\partial t} - \textstyle\sum_j k_j^{GL}a_{GL}\left(\frac{p_j}{H_j} - c_j^L\right)\right) + \frac{u_G}{T_G}\frac{\partial T_G}{\partial z} \tag{7}$$

Table 1 Nomenclature

ε_k	Volumetric holdup in phase k	k	Mass transfer coefficient
ϵ	Void fraction of particle bed	a	Interfacial area
ε_p	Particle porosity	ρ_k	Density of phase k
c_j^k	Concentration of compound j in phase k	$c_{p,k}$	Specific heat of phase k
p_j	Partial pressure of compound j	U	Heat transfer coefficient
u_k	Superficial velocity of fluid phase k	T_k	Temperature of phase k
H_j	Henry's coefficient of compound j	P	Reactor pressure
ΔH_i	Reaction enthalpy of reaction i	ρ_b	Bulk density of particle bed
$(\eta\,r)_j$	Apparent reaction rate of compound j	$(\eta\,r)_i$	Apparent reaction rate of reaction i

The specific oil treating case considered is given by De La Paz-Zavala et al. (2013), which gives a boiling point curve found in Table 2. Their study uses a simplified reaction scheme to describe the combined reactions of sulfur and aromatic compounds. Here, the conversion of a sulfur-compound molecule results in the creation of a polyaromatic molecule, which can react to form a monoaromatic molecule, etc. The representative compounds are divided into regular compounds and methylated analogues as defined in Table 3. The methylated compounds react slower, and thus by changing the ratio of methylation, the reaction rate can be altered. The reaction network and its true rate

expressions have been documented by Vanrysselberghe et al. (1996,1998). Diffusion limitations in the catalyst particles were accounted for using a method presented by Chacon et al. (2012) utilizing the catalyst properties given by Mederos & Ancheyta. The vapor-liquid equilibrium of saturated species was included using a pseudo-component stream defined in Pro/II with the boiling point curve as given by De La Paz-Zavala et al. The gas-phase was assumed ideal and Pro/II library models were used for the liquid.

Table 2 ASMT D86-boiling point curve for the fuel considered in the simulations with specific gravity and molar mass (De La Paz-Zavala et al. (2013)).

Distilled fraction [vol.-%]	Temperature [˚C]
IBP	278.1
10	293.9
20	300.1
30	305.3
50	315.6
70	327.3
80	334.3
90	343.4
FBP	358.3
Specific gravity at 20 ˚C / 4 ˚C	0.8676
Molar mass [g / mol]	260

Table 3 Real components in the feed as used in the simulations (De La Paz-Zavala et al. (2013)).

Compound group	Representative compound	Methylated analogue	Total content [wt.%]
Sulfur	Dibenzothiophene	4,6'-dimethyl dibenzothiophene	1.96
Polyaromatic	Biphenyl	3,3'-dimethyl biphenyl	13.4
Monoaromatic	Cyclohexyl benzene	3'-methyl cyclohexyl toluene	20.1

The model was applied in two different scenarios. Only cocurrent operation was considered. Firstly, the model was run to a steady state at different degrees of methylation. Isothermal flashes were calculated over the reactor to record the vapor recovery of fuel species. Secondly, flash calculations were implemented at regular increments in order to take into account the effect of the fuel species evaporation on the process. One isothermal flash was set in the inlet, meaning the feed streams enter the reactor at equilibrium. The in-reactor flashes were adiabatic. The methylation of the feed was altered. Also, the number of flashes was altered between no flashes in the reactor and a high number of flashes throughout the reactor. This allowed examining the possible range of model outputs compared to the model without evaporation over the range of possible mass transfer rates.

3. Results

The simulations were performed using a feed inlet temperature of 643.15 K and a pressure of 7.8 MPa. The results for the potential vapor recoveries are illustrated in Fig. 1. Generally, the aromatic components are estimated to be more volatile than the fuel mixture. With decreasing methylation, the vapor recovery increases greatly as the reaction rate is higher, thus developing heat faster. At 25 wt.-% methylated feed, the vapor recovery increases with more than 30 mol-% over the reactor profile. Thus, implementing evaporation in the model would have a significant effect on the

composition of the liquid phase in the reactor. The flash calculations were implemented in the model at steady state using 19 in-reactor flashes. The temperature profiles are found in Fig. 2. For low mass transfer, the temperature is lower, caused by displacement of the substrates to the gas phase. For the high mass transfer cases, the temperature increase is slower. However, depending on the feed, the end temperature is either the same or higher than when neglecting evaporation. Thus, the potential conversion is higher for the case with evaporation than for the one without.

Calculating the molar flow profiles for the simulations, the effect of the mass transfer on the individual component conversions can be estimated. The sulfur molar flow is seen in Fig. 3. The low mass transfer cases have remaining sulfur in the gas phase while the other cases have complete conversion. For the high mass transfer cases, the flow profiles are very similar to the case without evaporation suggesting the high reaction rate diminishes the effect of evaporation. The molar flow profiles for the polyaromatic components are found in Fig. 4. For both degrees of methylation at high mass transfer, the reaction rates are lower than the no-evaporation case, but, over the length of the reactor, the molar flow becomes lower. This suggest the attainable conversion is higher when including mass transfer. The reason seems to be that evaporation of monoaromatics decreases the product inhibition of polyaromatic reactions. The highly methylated feed has a higher conversion than the low methylated feed in this case. This is an effect of the temperature difference

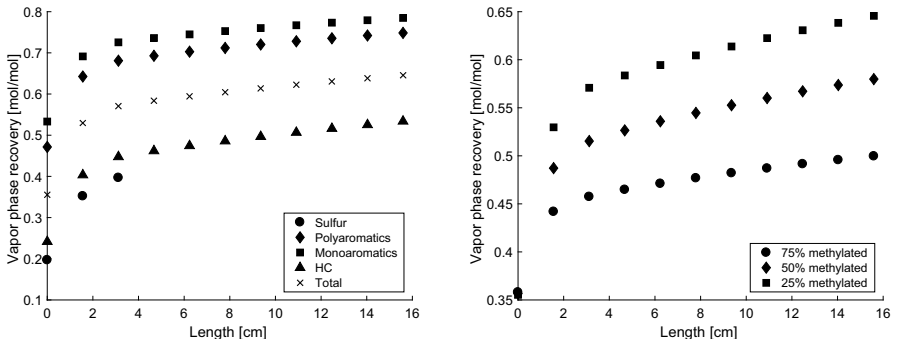


Figure 1 Calculated vapor recovery profiles of oil species in simulations neglecting evaporation. The plots present an overview of the individual species at 25 wt.-% methylated feed (left) and the recovery of oil species at different levels of methylation (right).

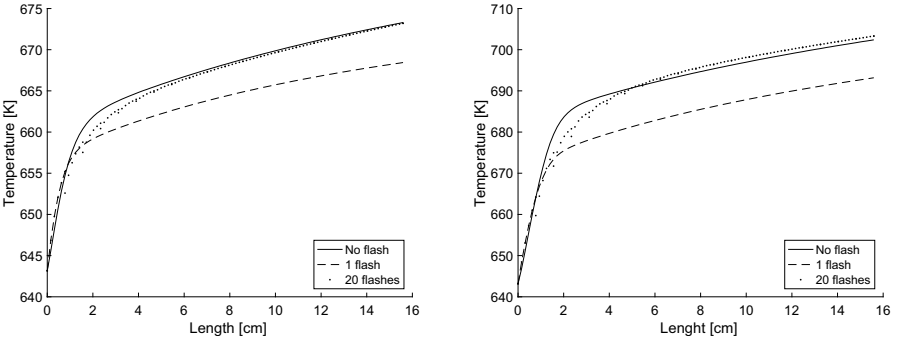


Figure 2 Liquid temperature profiles with or without flashes. The plots present the feed with 75 wt.-% (left) and 25 wt.-% methylated substrate (right).

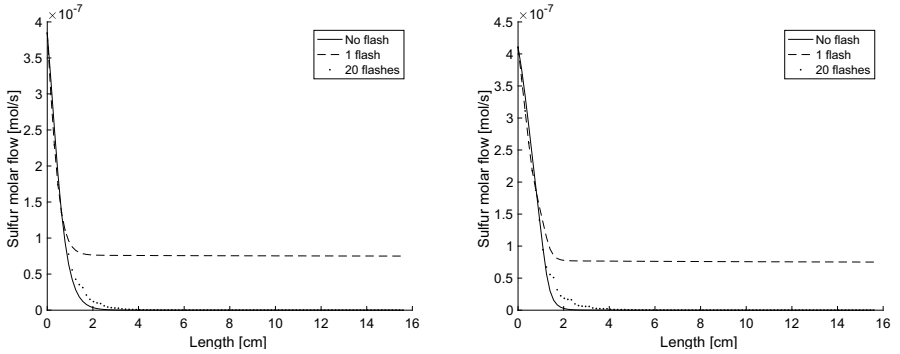


Figure 3 Molar flow profile of sulfur compounds with or without flashes. The plots present the feed with 75 wt.-% (left) and 25 wt.-% methylated substrate (right).

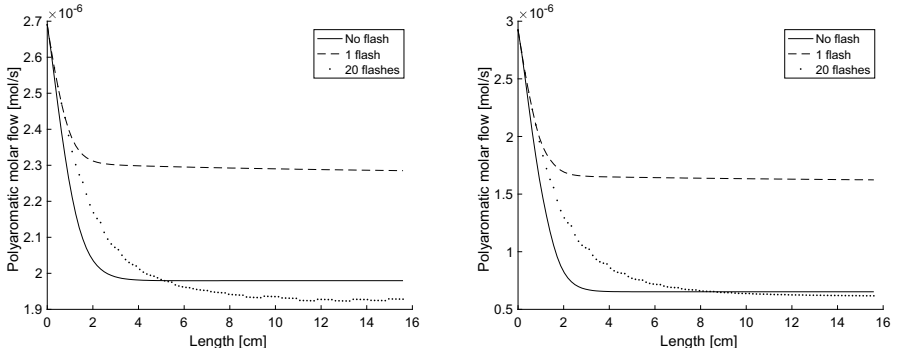


Figure 4 Molar flow profile of polyaromatic compounds with or without flashes. The plots present the feed with 75 wt.-% (left) and 25 wt.-% methylated substrate (right).

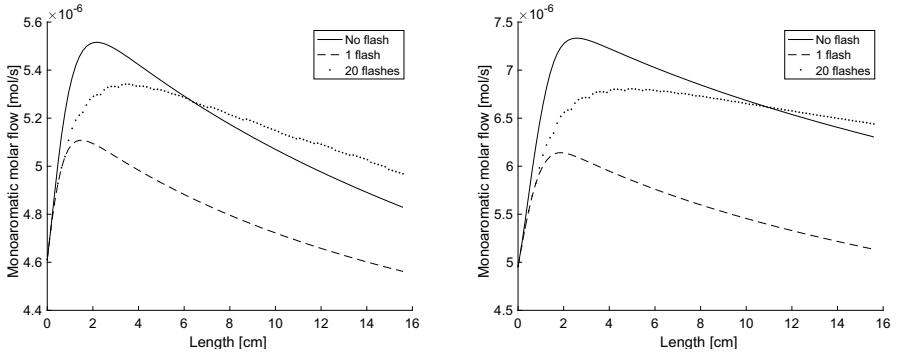


Figure 5 Molar flow profile of monoaromatics with or without flashes. The plots present the inlet with 75 wt.-% (left) and 25 wt.-% methylated inlet (right).

causing a lower substrate displacement in the case of lower reaction rates. The molar flow profiles for the monoaromatic compounds are given in Fig. 5. Generally, when including evaporation, the reaction rate and conversion is lower caused by the displacement of monoaromatics to the vapor. Comparing the different feeds, the high mass transfer case has a higher outlet flow corresponding to the higher conversion rate of the polyaromatics. Overall, it is seen that the inclusion of evaporation has a significant effect on the reaction

rate and conversion of the fuel, especially for the species which have slower reaction rates or are equilibrium limited. The presented approach therefore has value in showing the presence of these effect at the limits of mass transfer, even if it cannot predict the actual mass transfer rates of the oil species.

4. Conclusion

Using Pro/II to make flash calculations for the literature model, it was found that using the conditions given by De La Paz-Zavala et al. (2013) the equilibrium vapor recovery of fuel species is overall high. Implementing flashes over the length of the reactor, the mass transfer of fuel species was simulated, at either high or low mass transfer rates. It was found that the reaction rate of the fuel has a significant influence on the effect of the oil species mass transfer. At lower reaction rates, the high mass transfer simulation generally approximates the reactor profiles for temperature, sulfur and polyaromatic components assuming no evaporation. Monoaromatic component profiles differ significantly from this because of the high volatility of the defined monoaromatic species. For polyaromatic species at higher reaction and mass transfer rates, the conversion generally exceeds that of the simulations neglecting evaporation. The reason is likely that the evaporation of monoaromatics decreases product inhibition. In these same cases, the conversion of monoaromatics is lower because of substrate displacement.

Generally, it was found that the evaporation has little effect on fast reactions, like the sulfur reactions, and a significant effect on reactions that are limited by equilibrium, like the polyaromatic reactions. The range of reactor profiles assuming evaporation depends on the reaction rate with higher reaction rates resulting in a wider range of possible profiles. It was found that the approach could be useful in elucidating effects of evaporation on TBR models while it has little value in predicting experimental results.

References

A. Alvarez, J. Ancheyta, 2012, “Transient behavior of residual oil front-end hydrodemetallization in a trickle-bed reactor”, Chem. Eng. Journal, vol. 197, pp. 204-214.

R. Chacón, A. Canale, A. Bouza, Y. Sánchez, 2012, “Modeling of a three-phase reactor for bitumen-derived gas oil hydrotreating”, Brazillian J. Chem. Eng., vol. 29, no. 1, pp. 135-146.

C. De la Paz-Zavala, E. Burgos-Vázquez, J. E. Rodríguez-Rodríguez, LF. Ramírez-Verduzco, 2013, “Ultra low sulfur diesel simulation. Application to commercial units”, Fuel, vol. 110, pp. 227-234.

F. S. Mederos, J. Ancheyta, 2007, “Mathematical modeling and simulation of hydrotreating reactors: Cocurrent versus countercurrent operations”, Appl. Cat. A Gen., vol. 332, no. 1, pp. 8-21.

C. E. Ramírez-Castelán, A. Hidalgo-Vivas, J. Brix, A. D. Jensen, J. K. Huusom, 2017, “Modelling of an adiabatic trickle-bed reactor with phase change”, 27th ESCAPE, vol. 40. pp. 115-120.

V. Vanrysselberghe, G. F. Froment, 1996, “Hydrodesulfurization of dibenzothiophene on a $CoMo/Al_2O_3$ catalyst: Reaction network and kinetics”, Ind. Eng. Chem. Res., vol. 35, no. 10, pp. 3311-3318.

V. Vanrysselberghe, R. Le Gall, G. F. Froment, 1998, “Hydrodesulfurization of 4-methyldibenzothiophene and 4,6-dimethyldibenzothiophene on a $CoMo/Al_2O_3$ catalyst: Reaction network and kinetics”, Ind. Chem. Res., vol. 37, no. 4, pp. 1235-1242.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50077-X

Research on faulty antibody library of dynamic artificial immune system for fault diagnosis of chemical process

Yiyang Dai*, Yi Qiu, Ziyun Feng

[a]*College of Chemistry and Chemical Engineering, Southwest Petroleum University, Chengdu, China*
daiyiyang1984@163.com

Abstract

Fault diagnosis is one of the most important methods to ensure the safety of chemical process. With the development of artificial intelligences, many new methods have been introduced into the research of process fault diagnosis in chemical devices. Dynamic artificial immune system (DAIS) is one of the artificial intelligence methodologies with strong ability of self-learning and self-adaptability. However, in traditional AIS-based fault diagnosis strategies, key variables used in antibody cloning phase like total number of antibodies and mutation parameters are determined by experience. In this paper, we propose an approach for faulty antibody library construction in DAIS. Fault antibodies were classified into different types by different fault mechanism, and the size of antibody and mutation parameters of each faulty type were optimized individually. The performance of the modified DAIS-based fault diagnosis strategies is illustrated through the benchmarked Tennessee Eastman process.

Keywords: fault diagnosis, chemical process, artificial immune system, faulty antibody library,

1. Introduction

In recent years, the concept of “smart manufacturing” is gradually accepted by the public as the fourth industrial revolution is about to take place right now. For the chemical process industry, the goal of Smart Manufacturing should not only maximize economic competitiveness, but also significantly reduce safety incidents (Dai et al., 2016). Fault diagnosis is one of the effective methods to control and mitigate process risk. Venkatasubramanian et al. (2003) classified fault diagnosis methods into three general categories including quantitative model-based methods, qualitative model-based methods, and process history based (data-based) methods. With the development of automation technology, large amounts of data could be easily recorded and collected. Meanwhile, a significant progress has been made in the data-mining and processing area, which can provide new technologies for the utilization of process fault diagnosis in past decades (Ge et al., 2013).

Among the data-based techniques, the multivariate statistical process monitoring (PSPM) techniques are widely used in the last few years, such as principal component analysis (PCA), independent component analysis (ICA), support vector machines (SVM), partial least square (PLS), among others. Recently, artificial intelligent technologies are developed rapidly, a variety of new approaches have been recently proposed in addition to traditional statistical methods, such as deep belief network

(DBN), convolutional neural network (CNN), hidden Markov model (HMM), and artificial immune system (AIS), et al. To adapt to fault diagnosis of complex chemical process, dynamic artificial immune system (DAIS), was proposed (Dai & Zhao, 2011; Zhao et al., 2014). In DAIS-based fault diagnosis strategy, fault antibodies generated from the cloning of historical fault samples are used to diagnose the fault type. However, traditional antibody cloning algorithm mostly depends on experience, while the total number of antibodies and mutation parameters would affect the performance of fault diagnosis directly.

In this paper, we propose a modified approach for faulty antibody library construction, and apply it to fault diagnosis based on DAIS with modified antibodies (MA). The rest of the paper is organized as follows. A brief introduction of DIAS is reviewed in section 2. Section 3 introduces our proposed approach of faulty antibody library construction. The results and analysis of case study on Tennessee Eastman process are reported in Section 4. Finally, we conclude the paper and presents some future work in section 5.

2. Dynamic Artificial Immune System

AIS is developed through the application of techniques such as mathematical and computational modeling of immunology, abstraction from those models into algorithm (and system) design and implementation in the context of engineering (Timmis et al., 2008). To apply AIS in fault diagnosis of chemical processes, Dai and Zhao (2011) integrated the clonal selection algorithm of AIS with dynamic time warping (DTW) and afterwards used the DAIS in an online fault diagnosis strategy for full operating cycles of chemical processes (Zhao et al., 2014).

Different from traditional AIS technique, antigens and antibodies in DAIS are represented by matrices of time-sampled data to reflect the dynamic trends of process. While antigens are created by real-time samples and antibodies are created by historical samples with mutation during cloning phase. The system diagnoses faults by calculating and analyzing the affinity between antigens and antibodies. Eq.(1) and Eq.(2) show the composition of antigen and antibody.

$$Ag^{l\times n} = [Ag(1), Ag(2), ..., Ag(k), ..., Ag(l)] \tag{1}$$

$Ag(k)$ represents the variable vector of selected s at the (l-k)th sampling time before now. n is the number of selected variables and l is the length of the antigen.

$$Ab_{fault}^{m\times n} = [Ab_{fault}(1), Ab_{fault}(2), ..., Ab_{fault}(k), ..., Ab_{fault}(m)] \tag{2}$$

$Ab_{fault}(k)$ represents the vector of selected variables at the kth sampling time after the fault encountered. n is the number of selected variables and m is the length of the antibody.

The main steps of DAIS-based fault diagnosis stagey can be described as below.

1) Initialization. Generate original antibodies from historical samples. Then clone these antibodies with mutation to construct antibody library. Also, threshold of each fault type is calculated in this step.
2) Fault detection. Generate antigen from the test data and calculate the difference between the antigen and each antibody in the normal antibody library. If all of the differences are greater than the normal threshold, then a fault is detected.

3) Fault diagnosis. After detecting a fault, the differences between the antigen and each antibody in faulty antibody library of each type are calculated. If there exist fault antibodies in only one of the faulty antibody libraries that have smaller difference degrees than the threshold of the faulty antibody library, the fault is recognized of the fault type of the corresponding faulty antibody library.
4) Self-learning. After the fault diagnosis of DAIS, a manual diagnosis should be conducted to double check the diagnosis results by DAIS. According to the manual diagnosis result, a new original fault antibody is then generated and stored into the faulty antibody library of the fault type of the manual diagnosis result.

To improve the performance of DAIS in fault diagnosis, lots of work have been down. Causal inference method (Shu & Zhao, 2015) and genetic algorithm based fault antibody feature selection optimization (FAFSO) algorithm (Liang & Zhao, 2017) were proposed for the selection of variables to generate antibodies and antigens. To overcome limitation of the scarcity of fault samples in DAIS-based fault diagnosis, vaccine transplant approach has been proposed (Shu & Zhao, 2016). In the next section, the issue of faulty antibody libraries construction will be discussed.

3. Faulty Antibody Libraries

In DAIS-based fault diagnosis strategy, original antibodies are generated by historical samples with normalization at first. After the original antibodies generated, mutation is employed to create new antibodies during the clone phase.

Let X_1 and X_2 represent the two randomly selected original antibodies and C represents a normal historical sample. Calculate the difference matrix φ_1 between X_1 and C and φ_2 between X_2 and C separately. A new difference matrix φ^* can be obtained with mutation of $\varphi 1$ and $\varphi 2$. Than randomly select a matrix X_N of the same length with X_1 from C, a new cloning antibody X^* is generated based on shown in Eq.(3):

$$X^* = X_N + \varphi^* = X_N + a\varphi_1 + b\left(\varphi_1 - \varphi_2\right) \tag{3}$$

where a is a random number between 0.5 and 2, b is a random number between -1 and 1.

However, in traditional mutation algorithm, the total number of antibodies and the mutation parameters like a and b are all determined by reference. A modified approach for faulty antibody library construction is proposed in this section. For each type of faults, the fault diagnosis rate (FDR) and false positive rate (FPR) defined as below are used for antibody construction. p_i is the count of type i samples that are classified to type i, $q_{i,j}$ is the count of type i samples that are classified to type j.

$$\mathrm{FDR}(i) = \frac{p_i}{\text{total count of type } i \text{ samples}} \tag{4}$$

$$\mathrm{FPR}(i,j) = \frac{q_{i,j}}{\text{total count of type j samples}} \tag{5}$$

The main steps of DAIS-based fault diagnosis stagey can be described as below.

The main steps of the modified approach can be described as below.

Step1: For each fault type, randomly divide original antibodies into two groups. The bigger group contains about three quarters of the original antibodies, and the smaller one contains other one quarter of them. And we name the antibodies in the bigger group as initial antibodies, and which in the smaller group as revision antibodies.

Step 2: Then set the range of parameters a as [0.5,2], b as [-1, 1]. And the total number of initial cloning antibodies is twice that of the original antibodies. Create cloning antibodies using initial antibodies with mutation using Eq.(3).

Step 3: For each fault type, use revision antibodies of every fault type to test the cloning antibodies. Let ε represents criteria. For fault type i, if there was a type j let FPR(i,j)>ε, remove the antibodies, whose distance with antigen type j is smaller than threshold, from fault i antibody library. If FDR(i)<1-ε, change the range of parameters a, b to twice of which in last step.

Step 4: Check all fault type, if there was any fault type i let FPR(i,j)>ε or FDR(i)<1-ε, repeat step 2 and 3 to revise fault i antibody library.

Compared with traditional procedure, original antibodies are divided into two groups. Revision antibodies can be used to optimize the cloning parameters a and b, while FPR and FDR are used as optimization basis. Also the total number will not be confirmed, until cloning procedure is finished. All the optimizations are set to improve the performance of the diagnosis result.

4. Case Study

4.1. Process description

The TE process is the benchmark of fault detection and diagnosis. A revised model was proposed (Bathelt et al., 2015) which exploded more variables and more types of faults (available at http://depts.washington.edu/control/LARRY/TE/download.html.). In this paper, the process data are simulated by this revised model (Fig. 1).

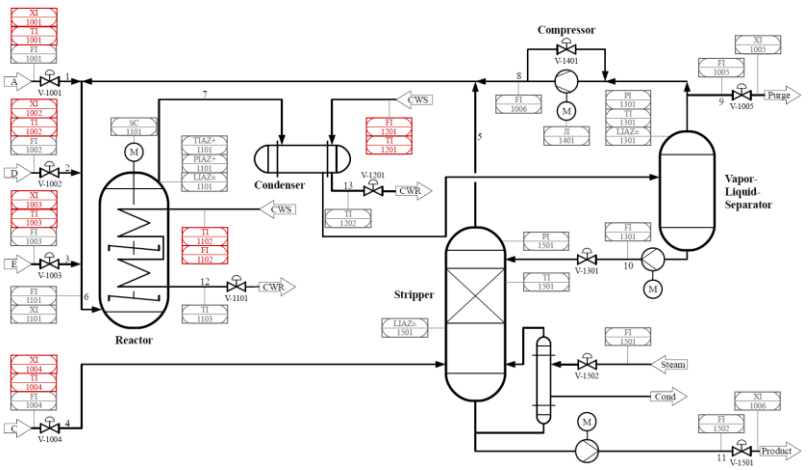


Fig. 1 P&ID of the revised TE process model;

The revised TE process model consists of 53 measured variables, including 22 process measurements, 19 composition measurements and 12 manipulated variables. For our case study, all variables are sampled at every 3 min and included in analysis. Only faults of step type, as faults 1-7 (see Table 1), are considered in this case study. Each type of fault is simulated for 12 times with different initial states, and the simulation time is set to 24 hours. The training dataset of each faulty antibody library consists of 4800 samples of that fault. The validation dataset has 480 samples of every fault. The training process is tended to construct the faulty antibody library based on AIS algorithm using the training dataset that has the best diagnostic performance on the validation dataset.

Table 1 Faults 1-7 defined in the TE process (Downs & Vogel, 1993).

Fault ID	Description	Type
IDV1	A/C Feed ratio, B Composition constant (stream 4)	Step
IDV2	B Composition, A/C Ratio constant (stream 4)	Step
IDV3	D Feed temperature (stream 2)	Step
IDV4	Reactor cooling water inlet temperature	Step
IDV5	Condenser cooling water inlet temperature	Step
IDV6	A Feed loss (stream 1)	Step
IDV7	C Header pressure loss – reduced availability (Stream 4)	Step

4.2. Diagnosis results

The FAFSO algorithm proposed by Liang and Zhao (2017) is used to select the fault antibodies' variables and optimize the faulty antibody libraries' thresholds. Then, antibody libraries are constructed with training dataset using proposed approach in section 3. While, 4000 samples are selected as initial data and 800 as revision data. After that, antigens are generated by validation dataset, and the differences between antigens and antibodies are calculated for fault diagnosis.

The fault diagnosis results of faults 1-7 are shown in Table 2 and the average testing FDR is 98.6%. Compared with fault diagnosis model in literatures (see Table 2), the performance of our proposed methods is improved significantly.

Table 2 Diagnosis performance comparison of different methods. (a) optimized variable selection based PCA (Ghosh et al., 2014); (b) supervised local neural network (Rad & Yazdanpanah, 2015); (c) Bayesian method (Jiang & Huang, 2016); (d) DBN based model (Zhang & Zhao, 2017); (e) AIS-FAFSO (Liang & Zhao, 2017); (f) DAIS-MA (proposed in this paper).

FDR/%	(a)	(b)	(c)	(d)	(e)	(f)
IDV1	99.87	96.37	100	100	100	100
IDV2	97.87	97.62	99	99	100	100
IDV3	2.37	20.62	6	95	93.7	94.3
IDV4	100	82.75	100	98	100	100
IDV5	99.87	96	100	86	90.23	95.6
IDV6	99.5	100	100	100	100	100
IDV7	100	100	100	100	100	100
Average	85.64	84.76	86.42	96.9	97.7	98.6

5. Conclusions

With the development of artificial intelligences and automation of chemical plant, many new data-based methods draw much attention from the chemical process fault diagnosis research. have been introduced into the research of process fault diagnosis in chemical

devices. Dynamic artificial immune system (DAIS) is one of the new artificial intelligence methodologies with strong ability of self-learning and self-adaptability. However, the construction of faulty antibody library is rarely studied. In this paper, an approach of faulty antibody libraries construction in DAIS is proposed. According to different performance of process data and fault mechanism, fault antibodies were classified into different types, and the size of antibody and mutation parameters of each faulty type were optimized individually. A case study on TE process is performed to show the improvement of DAIS when our faulty antibody construction approach is used.

Acknowledgement

The authors gratefully acknowledge financial support from the National Natural Science Foundation of China (NSFC) (No. 21706220).

References

Bathelt, A., Ricker, N. L., & Jelali, M. 2015. Revision of the tennessee eastman process model. IFAC-PapersOnLine, 48, 309-314.

Dai, Y. Y., Wang, H. Z., Khan, F., & Zhao, J. S. 2016. Abnormal situation management for smart chemical process operation. Curr. Opin. Chem. Eng., 14, 49-55.

Dai, Y. Y., & Zhao, J. S. 2011. Fault diagnosis of batch chemical processes using a dynamic time warping (dtw)-based artificial immune system. Ind. Eng. Chem. Res., 50, 4534-4544.

Downs, J. J., & Vogel, E. F. 1993. A plant-wide industrial process control problem. Comput. Chem. Eng., 17, 245-255.

Ge, Z. Q., Song, Z. H., & Gao, F. R. 2013. Review of recent research on data-based process monitoring. Ind. Eng. Chem. Res., 52, 3543-3562.

Ghosh, K., Ramteke, M., & Srinivasan, R. 2014. Optimal variable selection for effective statistical process monitoring. Comput. Chem. Eng., 60, 260-276.

Jiang, Q. C., & Huang, B. 2016. Distributed monitoring for large-scale processes based on multivariate statistical analysis and bayesian method. J Process Contr, 46, 75-83.

Liang, M., & Zhao, J. 2017. Feature selection for chemical process fault diagnosis by artificial immune systems. Chin. J. Chem. Eng. https://doi.org/10.1016/j.cjche.2017.09.023

Rad, M. A. A., & Yazdanpanah, M. J. 2015. Designing supervised local neural network classifiers based on em clustering for fault diagnosis of tennessee eastman process. Chemometrics Intellig. Lab. Syst., 146, 149-157.

Shu, Y. D., & Zhao, J. S. 2015. Dynamic artificial immune system with variable selection based on causal inference. Computer Aided Chemical Engineering, 37, 1793-1798.

Shu, Y. D., & Zhao, J. S. 2016. Fault diagnosis of chemical processes using artificial immune system with vaccine transplant. Ind. Eng. Chem. Res., 55, 3360-3371.

Timmis, J., Andrews, P., Owens, N., & Clark, E. 2008. An interdisciplinary perspective on artificial immune systems. Evolutionary Intelligence, 1, 5-26.

Venkatasubramanian, V., Rengaswamy, R., Yin, K., & Kavuri, S. N. 2003. A review of process fault detection and diagnosis: Part i: Quantitative model-based methods. Comput. Chem. Eng., 27, 293-311.

Zhang, Z. P., & Zhao, J. S. 2017. A deep belief network based fault diagnosis model for complex chemical processes. Comput. Chem. Eng., 107, 395-407.

Zhao, J. S., Shu, Y. D., Zhu, J. F., & Dai, Y. Y. 2014. An online fault diagnosis strategy for full operating cycles of chemical processes. Ind. Eng. Chem. Res., 53, 5015-5027.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50078-1

Optimal Enantiomer Crystallization Operation using Ternary Diagram Information

Caio Felippe Curitiba Marcellos[a], Helen Durand[b], Joseph Sang-Il Kwon[c], Amaro Gomes Barreto Jr.[d], Paulo Laranjeira da Cunha Lage[a], Maurício Bezerra de Souza Jr.[d], Argimiro Resende Secchi[a], Panagiotis D. Christofides[b,*]

[a] *Programa de Engenharia Química - COPPE, Universidade Federal do Rio de Janeiro, 21941-914 Rio de Janeiro, RJ, Brazil.*

[b]*Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA, 90095-1592, USA*

[c]*Department of Chemical Engineering, Texas A&M University, College Station, TX 77840, USA.*

[d]*Escola de Química, Universidade Federal do Rio de Janeiro, 21941-909 Rio de Janeiro, RJ, Brazil.*
pdc@seas.ucla.edu

Abstract

We recently developed (Curitiba Marcellos et al., 2017) a model predictive control (MPC) design for batch enantiomer crystallization that determines an optimal temperature profile in a crystallizer during crystallization of a single desired enantiomer from a solution of two enantiomers of a racemic compound forming system in solvent. This article analyzes that MPC design to highlight important features and assumptions of the constraints to generalize the underlying principles of their design.

Keywords: Racemic compound forming systems, Crystallization, Predictive Control.

1. Introduction

Crystallization can be used to separate enantiomers classified as racemic compound forming systems in the range of conditions (e.g., temperature and composition (Jacques et al., 1981; Kaspereit, 2006)) within which a single pure enantiomer can exist in equilibrium with a liquid solution containing both enantiomers and solvent according to solution thermodynamics. Our prior work (Curitiba Marcellos et al., 2017) utilized a mandelic acid (MA) in water example to motivate an operation and control (with MPC) strategy for a batch crystallizer for a racemic compound forming system that utilized a ternary diagram for the control formulation development. The motivation for this individual analysis of racemic compound forming systems in the framework of MPC for crystallization is that their specific thermodynamic properties impact MPC design for such systems. Because this work therefore had a strong focus on the underlying thermodynamics of a system for the MPC model and constraint design, there are underlying themes in the MPC development that should be interpreted as having broad implications for batch crystallization of systems with various thermodynamic properties, which this work seeks to analyze.

2. Mandelic acid in water example: ternary diagrams and MPC

We begin by revisiting the example from Curitiba Marcellos et al. (2017), which considers an MA in water system. Fig. 1 denotes the range of compositions of the desired and undesired enantiomers (denoted by *R* and *S*, respectively) which correspond to a saturated liquid in equilibrium with crystals of pure *R*. The triangles **BER**, **B₁E₁R**, and **B₂E₂R** represent the range of initial liquid compositions at temperatures T_0, T_1, and T_2, respectively, from which pure *R* crystals and a saturated liquid solution would be obtained if thermodynamic equilibrium were attained. This data indicates that a mixture with initial composition **P** can be cooled until it reaches the boundary of the two-phase region at a lower temperature (which would be **E₂** in the figure). After that temperature is reached, further cooling will produce a racemic compound (which is a solid containing both the *R* and *S* enantiomers), which we consider is undesirable.

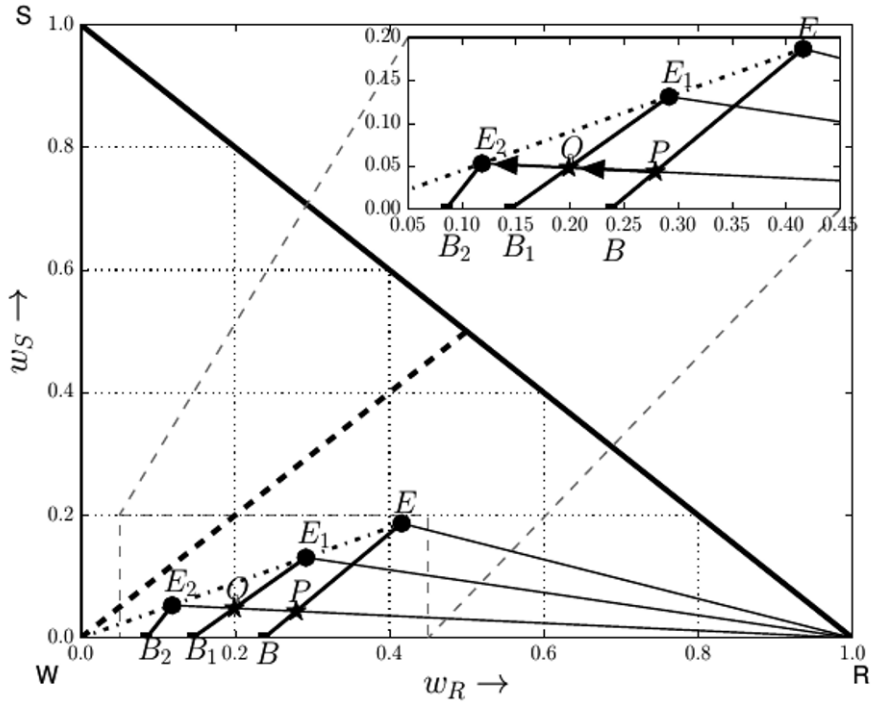


Figure 1: Right triangle ternary diagram for MA in water.

The batch crystallization process is modeled as a closed system containing two tanks (a well-mixed crystallizer and a well-mixed fines dissolution tank) between which the suspension in the crystallizer flows. The fines dissolution tank dissolves all crystals that enter it. A filter selective to smaller crystals is placed between the crystallizer output and the fines dissolution tank to prevent large crystals from entering the fines dissolution tank and dissolving. The motivation for including this fines dissolution loop is that it is desired to produce a greater mass of crystals from seeds than nuclei.

The population balance for the crystals, where the particle size distribution (PSD) $n = n(t, l)$ represents the number of crystals in the crystallizer of characteristic size l at time t per mass of solvent in the crystallizer (m_W), is as follows:

$$\frac{\partial n}{\partial t} = -\frac{\partial (Gn)}{\partial l} + B_0 \delta(l - l_0) - n\frac{d}{dt}\ln m_W + \left(\frac{1}{m_W}\right)\left(n_{in}\dot{m}_{W,in} - n_{out}\dot{m}_{W,out}\right) \quad (1)$$

where G is the crystal growth rate, B_0 is the nucleation rate (where nucleation occurs only at the crystal size $l_0 = 0$ m as enforced in Eq.(1) through the Dirac delta function), and n_{in}, n_{out}, $\dot{m}_{W,in}$, and $\dot{m}_{W,out}$ represent the population balances in the streams flowing into the crystallizer (from the fines dissolution tank) and out of the crystallizer (to the fines dissolution tank), as well as the mass flow rates of solvent into and out of the crystallizer, respectively. This equation is derived from a volume-integrated population balance equation for a well-mixed system (Randolph and Larson, 1988), in which the PSD was developed in terms of volume rather than solvent mass. An appropriate substitution for the volume-based PSD term (and related terms for the nucleation rate and also of the PSD's and flow rates into and out of the crystallizer) by terms based on solvent mass and required variables such as m_W and volume, and applying the chain rule and canceling terms, gives Eq.(1). In addition to the population balance equation, the mass balance on *R* in the crystallizer and fines dissolution tank and the temperature in the crystallizer were also modeled.

The resulting dynamic model was then incorporated within an MPC with the form:

$$\underset{\Delta T_{jkt,k}}{\text{minimize}} \quad J\left(T_{jkt,k}\right) := \frac{\mu_3^n(t_f)}{\mu_3^s(t_f)} \quad (2a)$$

$$\text{Process Model} \quad (2b)$$

$$\Delta T_{jkt,k} \leq 0, k = k_{it}, \ldots, k_{N_t - 1} \quad (2c)$$

$$T_{min} \leq T\left(t_f\right) \leq T_{max} \quad (2d)$$

$$\frac{P^i - P^e}{P^i(1 - P^e)} = 0.4435 \quad (2e)$$

where $T_{jkt,k}$ is the value of the cooling jacket temperature (manipulated input for the crystallizer) at the k-th sampling period, $\mu_3^i(t_f)$, $i = n, v$, represents the third moment (tied to crystal mass) of the PSD at the final batch time t_f for crystals derived from nuclei ($i = n$) or seeds ($i = s$), and $k_{it}, \ldots, k_{N_t - 1}$ represent the indices of the sampling periods in the prediction horizon. The values T_{min} and T_{max} represent the bounds on the temperature T at t_f, and the terms P^i and P^e represent the purities of the liquid in the crystallizer at the initial and eutectic compositions (where the purity is a ratio of the mass fraction of the *R* enantiomer to the sum of the mass fractions of the *R* and *S* enantiomers).

3. Analyzing MPC formulation for MA in water

We now seek to elucidate some of the implications of Eq.(2) for MPC for crystallization for other multicomponent systems, crystallizer designs, and operating objectives. First,

the formulation of the objective function of Eq.(2) reflects a desire to minimize, for the MA in water process, the ratio of the mass of crystals from nuclei to the mass from seeds. The constraint of Eq.(2b) is the model for the system (PSD and mass/energy balances), and the constraint of Eq.(2e) was derived as the maximum yield thermodynamically achievable for a batch crystallization process for MA in water with no fines dissolution loop. This constraint was utilized for simulations without and with the fines dissolution loop in Curitiba Marcellos et al. (2017) for consistency; however, the derivation of this equation in reference to yield is for the case with no fines trap. For the case with no fines trap, a thermodynamic analysis allowed the constraint of Eq.(2e) to be selected in combination with the objective function and Eq.(2b). Without a thermodynamic analysis, Eq.(2e) might be chosen as the objective function of the MPC to be maximized. However, for the MA in water process, there are multiple desired properties for the crystal product besides yield alone, such as the greater mass and number of crystals from seeds compared to nuclei. One method for attempting to account for all these considerations is by developing an objective function containing the various terms and weighting each of them. Another method, and that which is utilized in Eq.(2) for the case with no fines trap, is to examine the desired properties and to recognize that some of the desired properties are thermodynamically limited, meaning that they can be written as a constraint. In addition, changing the process design may help to improve some of the properties (which was the motivation in Curitiba Marcellos et al. (2017) for simulating the MA in water process with the fines dissolution loop after analyzing the number of crystals obtained from nuclei without it). Therefore, care should be taken in formulating an objective function when there are multiple desired operating goals to ensure that a consideration that can be accounted for in a constraint or in the process design is not being incorporated into the objective function. Furthermore, changes to the design may affect the meaning of constraints (e.g., the process design for which the constraint of Eq.(2e) was derived to represent the yield was changed to explore the benefits of a fines dissolution tank), and therefore care should be taken in ensuring that all constraints enforced in a batch crystallization process meaningfully apply to the design at hand and/or will not cause the crystallization process to proceed beyond the point at which crystals of a single desired species are formed.

Another principle exemplified by the development of Eq.(2) is that a visual representation of thermodynamics data such as ternary diagrams may help in developing thermodynamically-based equations for the process model. For example, in Eq.(2), the population balance model includes the saturation composition as a function of temperature within the growth and nucleation rate models that appear in Eq.(1). The saturation composition was modeled through a procedure that took advantage of work in Kaspereit (2006) involving the analysis of thermodynamic data available in an equilateral triangle ternary diagram for MA in water indicating that for various temperatures, the solubility curve can be approximated by a line between the eutectic and binary saturation compositions.

Visual displays of thermodynamic data like the ternary diagram can also be helpful for selecting thermodynamic constraints. For example, the ternary diagram makes the eutectic composition restriction visually apparent. However, because this is an equilibrium composition, if a constraint required that the solution in the crystallizer always be above that composition, it is possible that the solution would become cooled too much, depending on the rate of cooling computed by the MPC and the speed with which thermodynamic equilibrium is set up. The ternary diagram reveals that temperature

is another variable that is associated with the equilibrium condition and may be more effectively constrained. In general, care should be taken when working with thermodynamic data in defining MPC constraints, because thermodynamic equilibrium is a limiting condition and setting constraints that depend on any mass transfer-limited properties such as composition may not be appropriate.

Though the constraints in Eq.(2) allow control actions to be computed that drive the state to the limits of pure *R* crystallization in the yield and temperature constraints, MPC design for crystallization processes should consider that disturbances and plant-model mismatch are likely to occur, and even if control actions computed by the MPC cause thermodynamic constraints to be satisfied within the controller, they may not be satisfied by the actual process. Furthermore, the design in Eq.(2) considers that the final time of batch operation is included within the prediction horizon. Depending on the final batch time, this may be computationally prohibitive. However, the yield constraint as defined relies on this because the objective function and constraints both include the final time. For example, for the case with no fines dissolution loop, Eq.(2e) requires the process to produce the optimal yield by the end of the batch. If it is desired not to include the entire prediction horizon due to computation time reasons, one solution may be to adjust the objective function to allow it to also include the yield. However, a short prediction horizon may not allow the MPC to achieve as optimal of a cooling profile as including the entire batch time within the prediction horizon.

For any industrial crystallization process under MPC, all effects on the population balance that are significant should be accounted for (for example, Zhang et al. (2010) indicates that cooling rate may affect the crystallization growth rate) because unlike alternative operating policies such as a linear cooling profile in the crystallizer, it is difficult to know, a priori, what cooling profile an MPC will compute for a batch operation, even in the absence of disturbances, without performing closed-loop simulations. When disturbances occur, it may compute a different input trajectory than would be simulated without the disturbances. Therefore, it is not possible with MPC to assume that factors which may affect the crystallization process will be negligible, and they must be accounted for in the model.

The design in Eq.(2) considers that a feasible solution to the optimization problem exists. Due to the hard constraint on yield, a feasible solution is not guaranteed to exist within the batch time t_f chosen. Because this is a thermodynamic constraint, given a sufficiently large final batch time and no disturbances, it should be able to be met. However, in the presence of disturbances/plant-model mismatch, it is possible that the constraint may become infeasible due to t_f not being large enough or disturbances perturbing the equilibrium state.

The design in Eq.(2) does not operate a process at a steady-state in a traditional sense because a batch, rather than continuous, process is considered. Thus, stabilizing control designs like classical proportional-integral-derivative (PID) control or Lyapunov-based control would not apply, and optimization-based control will be an effective means for controlling the process. The MPC considered also does not enforce stabilization in a region around a steady-state (though stabilizing designs comprise a large portion of the MPC literature). However, if a thermodynamic equilibrium-based constraint, like that of Eq.(2e) for the case with no fines dissolution loop, is included, then at the end of the batch, the MPC may seek to stabilize the temperature in the crystallizer at the value corresponding to the equilibrium condition. Still, classical control designs that treat the

thermodynamic condition at the end of the batch as a steady-state and seek to drive the state to this condition are unable to ensure that critical thermodynamic constraints are respected or to optimize product characteristics during the crystallization process. Developing an MPC design for a continuous crystallization process (Kaspereit, 2006) would be fundamentally different than developing an MPC for a batch process.

Finally, the sampling period may be important. Increasing the sampling period length (reducing the number of sampling periods) can improve computation time, which may be beneficial, especially in the case in which the final time of the batch is important in developing a constraint as in Eq.(2e). However, to see the impact of reducing the number of sampling periods in the prediction horizon, consider a single sampling period in Eq.(2) in the case with no fines dissolution tank. In that case, to cause the constraint of Eq.(2e) to be satisfied at the end of the batch, the jacket temperature would need to drop to that at which the temperature in the crystallizer becomes fixed at the temperature corresponding to the eutectic purity for the given initial purity. This mode of operation then effectively drives the input to a steady-state value corresponding to the thermodynamic equilibrium condition at the end of the batch. It does not allow the trajectory along the way to this equilibrium condition to be optimized. If more sampling periods are utilized, the MPC has more flexibility to optimize the approach to thermodynamic equilibrium.

4. Conclusions

This article analyzed methods for MPC design for crystallization processes that take operating objectives as well as thermodynamic limitations into account.

References

1. C.F. Curitiba Marcellos, H. Durand, J.S.- I. Kwon, A. Gomes Barreto Jr., P. Laranjeira da Cunha Lage, M. Bezerra de Souza Jr., A.R. Secchi, and P.D. Christofides. "Optimal operation of batch enantiomer crystallization: From ternary diagrams to predictive control." *AIChE Journal*, in press.
2. J. Jacques, A. Collet, and S.H. Wilen. *Enantiomers, Racemates, and Resolutions*. John Wiley & Sons, Inc., New York, 1981.
3. M. Kaspereit. "Separation of Enantiomers by a Process Combination of Chromatography and Crystallisation." PhD thesis, Otto-von-Guericke-Universität Magdeburg, 2006.
4. A.D. Randolph and M.A. Larson. *Theory of Particulate Processes: Analysis and Techniques of Continuous Crystallization*. Academic Press, San Diego, California, 1988.
5. Y. Zhang, S. Mao, A.K. Ray, and S. Rohani. "Nucleation and growth kinetics of (R)-mandelic acid from aqueous solution in the presence of the opposite enantiomer." *Crystal Growth & Design*, 10: 2879-2887, 2010.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50079-3

Model Predictive Control for Process Operational Safety: Utilizing Safeness Index-Based Constraints and Control Lyapunov-Barrier Functions

Zhe Wu[a], Fahad Albalawi[c], Zhihao Zhang[a], Junfeng Zhang[a], Helen Durand[d], Panagiotis D. Christofides[a,b,*]

[a]*Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA, 90095-1592, USA*
[b]*Department of Electrical and Computer Engineering, University of California, Los Angeles, CA 90095-1592, USA*
[c]*Department of Electrical Engineering, Taif University, Taif, 21974, Saudi Arabia*
[d]*Department of Chemical Engineering and Materials Science, Wayne State University, Detroit, MI 48202, USA*
pdc@seas.ucla.edu

Abstract

In this paper, two different model predictive control (MPC) methodologies that can account for process operational safety are compared. Specifically, we first develop the Safeness Index, which represents the relative safeness of the process state in state-space, and is used to form a state constraint in MPC to guarantee process safety. Then, a Control Lyapunov-Barrier Function-based economic model predictive controller (CLBF-EMPC) is developed to account for the unsafe region in the design of the Control Lyapunov-Barrier Function (CLBF), and thus ensures closed-loop stability and safety simultaneously.

Keywords: Process operational safety; Model predictive control; Safeness Index; Control Lyapunov-Barrier Functions

1. Introduction

Process operational safety plays an important role in operating chemical processes in an economically-optimal manner. Recently, the U.S. Chemical Safety Board released a report of over 20 chemical incidents occurring in the United States since 2013 that caused severe consequences (Completed Investigations of Chemical Incidents, 2016). The catastrophic consequences remind us of the significance of process operational safety, especially in the design of process control systems. In particular, recent work (Leveson and Stephanopoulos, 2014) has described process safety as a problem with respect to multi-variable interactions of process variables instead of individual variables violating bounds. Motivated by the above considerations, the Safeness Index (Albalawi et al., 2017) was first constructed using first-principles modeling information and past operating data, which can coordinate the control system design with process safety. A control system associated with a constraint on the Safeness Index can enhance the process safety and guarantee closed-loop stability. On the other hand, an intriguing control method accounting for both stability and safety termed Control Lyapunov-Barrier Function-based control has been developed (Romdlony and Jayawardhana,

2016). Since unsafe state-space regions could be open and bounded sets in the operating region, it is promising to incorporate CLBFs, which inherit nice properties of Control Lyapunov Functions and Control Barrier Functions, respectively, into MPC design to guarantee closed-loop stability and safety simultaneously. In the present work, we describe similarities and differences between these two approaches in the context of economic MPC and demonstrate their application to a chemical process example.

2. Preliminaries

2.1. Notations

The notation $|\cdot|$ is used to signify the Euclidean norm of a vector, the notation $|\cdot|_Q$ denotes a weighted Euclidean norm of a vector (i.e., $|x|_Q = x^T Q x$ where Q is a positive definite matrix). x^T denotes the transpose of x. The notation $L_f V(x)$ denotes the standard Lie derivative. Set subtraction is denoted by "\", i.e., $A \setminus B := \{x \in \mathbf{R}^n \mid x \in A, x \notin B\}$. The function $f(\cdot)$ is of class $\mathcal{C}^1$ if it is continuously differentiable.

2.2. Class of Nonlinear Process Systems

The class of continuous-time nonlinear systems considered is described by the following state-space form:

$$\dot{x} = f(x) + g(x)u + h(x)w, x(t_0) = x_0 \tag{1}$$

where $x \in \mathbf{R}^n$ is the state vector, $u \in \mathbf{R}^m$ is the manipulated input vector, and $w \in \mathbf{W}$ is the disturbance vector, where $\mathbf{W} := \{w \in \mathbf{R}^q \mid |w| \le \theta, \theta \ge 0\}$. The control action constraint is defined by $u \in U := \{u^{\min} \le u \le u^{\max}\} \subset \mathbf{R}^m$. $f(\cdot), g(\cdot), h(\cdot)$ are sufficiently smooth vector and matrix functions of dimension $n \times 1$, $n \times m$, and $n \times q$, respectively. Without loss of generality, the initial time t_0 is taken to be zero ($t_0 = 0$), and it is assumed that $f(0) = 0$ and $g(0) = 0$, and thus, the origin is a steady-state of the nominal system of Eq.(1) with $w(t) \equiv 0$, (i.e., $(x_s^*, u_s^*) = (0,0)$).

3. Safeness Index-Based Control

In this section, we first present the Lyapunov-based controller assumption. Then, the process Safeness Index is introduced into the formulation of Lyapunov-based economic model predictive control (LEMPC), under which the closed-loop system can always be driven into the safe operating region.

3.1. Lyapunov-based Controller Assumption

Consider the nominal system of Eq.(1) with $w(t) \equiv 0$. We assume that there exists a feedback control law $\Phi(x)$ for the nominal closed-loop system (i.e., $w(t) \equiv 0$) of Eq.(1) that renders the origin of the nominal system asymptotically stable for all x in an open neighborhood of the origin, when applied continuously, in the sense that there exists a positive definite $\mathcal{C}^1$ control Lyapunov function V and class $\mathcal{K}$ functions $\alpha_i(\cdot), i = 1,2,3,4$ that satisfy the following inequalities:

$$\begin{aligned}
&\alpha_1(|x|) \le V(x) \le \alpha_2(|x|) \\
&\frac{\partial V(x)}{\partial x}(f(x) + g(x)\Phi(x)) \le -\alpha_3(|x|), \Phi(x) \in U \\
&\left|\frac{\partial V(x)}{\partial x}\right| \le \alpha_4(|x|)
\end{aligned} \tag{2}$$

The set of initial conditions under the feedback control law, from which the origin of the nominal system of Eq.(1) can be rendered asymptotically stable, is characterized as the set $\phi_s := \{x \in \mathbf{R}^n \mid \dot{V} + \kappa V(x) \le 0, u = \Phi(x) \in U, \kappa > 0\}$. The stability region is defined as a level set of the Lyapunov function (i.e., $\Omega_\rho := \{x \in \phi_s \mid V(x) \le \rho\}$).

3.2. Safeness Index-Based LEMPC Formulation

Since the process operational safety of a chemical plant results from multivariable interactions and interactions between units, it is desirable to integrate the process control system and process safety by designing a function of the process states (i.e., Safeness Index) that indicates the level of safety of a given state. The methods of determining the functional form of $S(x)$ include first-principles process models or systematic safety analysis tools such as HAZOP.

Based on the functional form of $S(x)$, the closed-loop state predictions can be required to be maintained within a safe region (where $S(x)$ is below the threshold on the Safeness Index S_{TH}) by using the index within the process control design. The safety systems can be triggered whenever the threshold S_{TH} is sufficiently exceeded, which represents the overall system state becoming unsafe.

A straightforward method for incorporating the Safeness Index into the process control system is to set it as a constraint in model predictive control. Therefore, the Safeness Index-Based LEMPC is formulated as follows:

$$\max_{u \in S(\Delta)} \int_{t_k}^{t_{k+N}} L_e(\tilde{x}(\tau), u(\tau)) d\tau \tag{3a}$$

$$\text{s.t. } \dot{\tilde{x}}(t) = f(\tilde{x}(t)) + g(\tilde{x}(t))u(t) \tag{3b}$$

$$\tilde{x}(t_k) = x(t_k) \tag{3c}$$

$$u(t) \in U, \forall t \in [t_k, t_{k+N}) \tag{3d}$$

$$V(\tilde{x}(t)) \le \rho_e, \forall t \in [t_k, t_{k+N}), \text{ if } x(t_k) \in \Omega_{\rho_e} \tag{3e}$$

$$S(\tilde{x}(t)) \le S_{TH}, \forall t \in [t_k, t_{k+N}) \text{ if } S(x(t_k)) \le S_{TH} \tag{3f}$$

$$\dot{V}(x(t_k), u(t_k)) \le \dot{V}(x(t_k), \Phi(x(t_k))), \text{ if } x(t_k) \in \Omega_\rho / \Omega_{\rho_e}, \text{ or } S(x(t_k)) > S_{TH} \tag{3g}$$

where $\tilde{x}(t)$ is the predicted state trajectory, $S(\Delta)$ is the set of piecewise constant functions with period Δ, and N is the number of sampling periods in the prediction horizon. The optimization problem of Eq.(3) optimizes the cost function $L_e(x(t), u(t))$ of Eq.(3a) subject to the nominal process model of Eq.(3b). Eq.(3c) defines the initial condition for the optimization problem of Eq.(3) using the measurement of the process state at the current time t_k. Eq.(3d) defines the input constraints applied during the prediction horizon. Under the Mode 1 constraint of Eq.(3e), the LEMPC optimizes the time integral of $L_e(x(t), u(t))$ in a time-varying fashion while maintaining the predicted closed-loop state within the set $\Omega_{\rho_e} := \{x \in \phi_s \mid V(x) \le \rho_e, \rho_e < \rho\}$ defined to make Ω_ρ a forward invariant set. The constraint of Eq.(3f) maintains the closed-loop state within the safe operating region if the current state is inside it. Under the Mode 2 constraint of Eq.(3g), the contractive constraint is activated only in the first sampling period to decrease the value of the Lyapunov function. The LEMPC is implemented in a sample-and-hold fashion, and only the first step of the optimized input trajectory will be applied over the next sampling period. It should be noted that the optimization problem of Eq.(3) may not be feasible sometimes due to the constraint of Eq.(3f). In this case, the explicit stabilizing controller $\Phi(x)$ will be applied in sample-and-hold over the next sampling period to guarantee closed-loop stability.

The theorem below states that under sample-and-hold implementation of the Safeness Index-Based LEMPC of Eq.(3), closed-loop stability is maintained in the sense that the state of the closed-loop system of Eq.(1) is always bounded in Ω_ρ, and process operational safety is enhanced in the sense that the state will enter the safe region in finite time if it exits this region. The proof is presented in Albalawi et al., 2017.

Theorem 1 *Consider the closed-loop nominal system of Eq.(1) with* $w(t) \equiv 0$ *under the implementation strategy of the Safeness Index-based LEMPC of Eq.(3). Given any initial state* $x_0 \in \Omega_\rho$, *it is guaranteed that* $x(t) \in \Omega_\rho, \forall t \geq 0$, *and* $x(t)$ *will enter the safe region in finite time.*

4. CLBF-Based EMPC

In this section, we first introduce the definition of the constrained Control Lyapunov-Barrier Function (CLBF), and then develop the CLBF-EMPC to attain guaranteed closed-loop stability and process operational safety.

4.1. Definition of constrained CLBF

Based on the original Control Lyapunov-Barrier Function (Romdlony and Jayawardhana, 2016) that was developed for the nominal system of Eq.(1) with $w(t) \equiv 0$, in this manuscript, we propose a constrained CLBF that accounts for the presence of input constraints in the system of Eq.(1). Specifically, the definition of a constrained CLBF is as follows:

Definition 1 *Given a set of unsafe points in state-space* $\mathcal{D}$, *a proper, lower-bounded and* $\mathcal{C}^1$ *function* $W_c(x): \mathbf{R}^n \to \mathbf{R}$ *is a constrained CLBF if it satisfies the following properties:*

$$W_c(x) > \rho_c, \quad \forall x \in \mathcal{D} \subset \phi_{uc} \tag{4a}$$

$$\left|\frac{\partial W_c(x)}{\partial x}\right| \leq r(|x|) \tag{4b}$$

$$L_f W_c(x) < 0, \forall x \in \{x \in \phi_{uc} \setminus (\mathcal{D} \cup \{0\}) \mid L_g W_c(x) = 0\} \tag{4c}$$

$$\mathcal{U}_{\rho_c} := \{x \in \phi_{uc} \mid W_c(x) \leq \rho_c\} \neq \varnothing \tag{4d}$$

where $\rho_c \in \mathbf{R}$ and $r(\cdot)$ is a class $\mathcal{K}$ function. Similar to Safeness Index-based control, we also assume that there exists a stabilizing control law $u = \Phi(x) \in U$ (e.g., the control law in Lin and Sontag, 1991) such that the origin of the nominal system of Eq.(1) can be rendered asymptotically stable. ϕ_{uc} is defined to be the union of the origin and the set where the time-derivative of $W_c(x)$ is negative with constrained input:

$$\phi_{uc} = \{x \in \mathbf{R}^n \mid \dot{W}_c(x(t), u(t)) = L_f W_c + L_g W_c u < -\alpha |W_c(x)|, u = \Phi(x) \in U\} \cup \{0\}$$

where α is a positive real number. The set of initial conditions is characterized as a level set $\mathcal{U}_{\rho_c}$ of $W_c(x)$ in Eq.(4d), from which it can be proved that the origin of the nominal system of Eq.(1) with $w(t) \equiv 0$ can be rendered asymptotically stable while maintaining the state in a safe region at all times.

4.2. CLBF-based EMPC formulation

The CLBF-EMPC design is represented by the following optimization problem:

$$\max_{u \in S(\Delta)} \int_{t_k}^{t_{k+N}} L_e(\tilde{x}(\tau), u(\tau)) d\tau \tag{5a}$$

$$\text{s.t. } \dot{\tilde{x}}(t) = f(\tilde{x}(t)) + g(\tilde{x}(t))u(t) \tag{5b}$$

$$\tilde{x}(t_k) = x(t_k) \tag{5c}$$

$$u(t) \in U, \forall t \in [t_k, t_{k+N}) \tag{5d}$$

$$W_c(\tilde{x}) \le \rho_e, \text{ if } x(t_k) \in \mathcal{U}_{\rho_e},\ \forall t \in [t_k, t_{k+N}) \tag{5e}$$
$$\dot{W}_c(x(t_k), u(t_k)) \le \dot{W}_c(x(t_k), \Phi(x(t_k))), \text{ if } x(t_k) \in \mathcal{U}_\rho \setminus \mathcal{U}_{\rho_e} \tag{5f}$$

where the notation and constraints of Eqs.(5a)-(5d) follow the notation and constraints of Eqs.(3a)-(3d), and $\mathcal{U}_\rho$ is a safe region which has no intersection with the unsafe region $\mathcal{D}$. Similarly, if the state at $t = t_k$ is inside $\mathcal{U}_{\rho_e}$, the Mode 1 constraint of Eq.(5e) is applied to maintain the predicted closed-loop state within the set $\mathcal{U}_{\rho_e} \subset \mathcal{U}_\rho$, which is designed to make the safe operating region $\mathcal{U}_\rho$ a forward invariant set. Under the Mode 2 constraint of Eq.(5f), the contractive constraint is activated only for the first sampling step to decrease the value of $W_c(x)$, such that the closed-loop state will move back into $\mathcal{U}_{\rho_e}$ within finite sampling steps. Based on the CLBF-EMPC of Eq.(5), the following theorem establishes that under the sample-and-hold implementation of the CLBF-EMPC of Eq.(5), the optimization problem is recursively feasible. Additionally, closed-loop stability and safety are guaranteed in the sense that the closed-loop state is always bounded in $\mathcal{U}_\rho$. Again, the proof is omitted for brevity.

Theorem 2 *Consider the closed-loop nominal system of Eq.(1) (i.e., $w(t) \equiv 0$) with a constrained CLBF $W_c(x): \mathbf{R}^n \to \mathbf{R}$ that has its minimum at the origin. Given any initial state $x_0 \in \mathcal{U}_\rho$, it is guaranteed that under the CLBF-EMPC of Eq.(5), $x(t) \in \mathcal{U}_\rho$, $\forall t \ge 0$ where $\mathcal{U}_\rho \subset \mathcal{U}_{\rho_e}$ and $\mathcal{U}_\rho \cap \mathcal{D} = \varnothing$.*

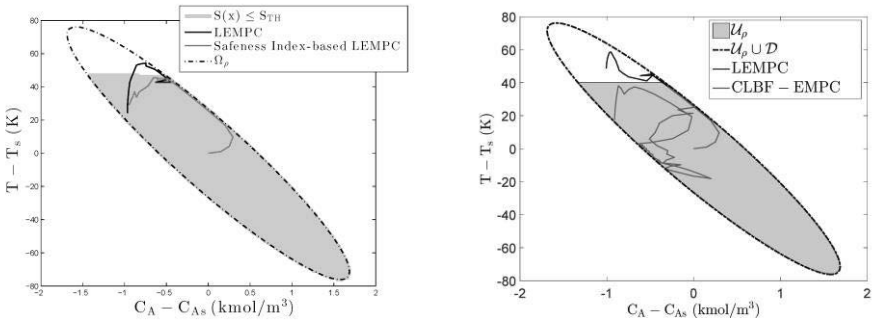
5. Application to a Chemical Process Example

A chemical process example is provided to illustrate the applications of the Safeness Index-based LEMPC and CLBF-EMPC to maintain the closed-loop state within the safety region defined by the Safeness Index function. Specifically, a well-mixed, non-isothermal continuous stirred tank reactor (CSTR) where an irreversible second-order exothermic reaction takes place is considered. The reaction transforms a reactant A to a product B ($A \to B$). The inlet concentration of A, the inlet temperature and feed volumetric flow rate of the reactor are C_{A0}, T_0 and F, respectively. The CSTR is equipped with a heating jacket that supplies/removes the heat at the rate Q. Based on material and energy balances, the CSTR dynamic model is of the following form:

$$\frac{dC_A}{dt} = \frac{F}{V}(C_{A0} - C_A) - k_0 e^{\frac{-E}{RT}} C_A^2 \tag{6a}$$
$$\frac{dT}{dt} = \frac{F}{V}(T_0 - T) + \frac{-\Delta H}{\rho_L C_p} k_0 e^{\frac{-E}{RT}} C_A^2 + \frac{Q}{\rho_L C_p V} \tag{6b}$$

where ΔH, k_0, E, and R represent the enthalpy of reaction, pre-exponential constant, activation energy, and ideal gas constant, respectively. The volume of the reacting liquid in the reactor V, heat capacity C_p, and fluid density ρ_L are assumed constant (process parameter values are in Albalawi et al., 2017). Additionally, a material constraint is introduced to make the averaged reactant material available over a given operating period $t_p = 1.0hr$ to be 0 (in deviation from the steady-state value, C_{A0s}). The explicit Euler method with an integration time step of $h_c = 10^{-5} hr$ is utilized to integrate numerically the dynamic model of Eq.(6). The states in the form of deviation variables are the concentration and temperature in the reactor and the manipulated inputs are the inlet concentration and heat rate, respectively (i.e., $x^T = [C_A - C_{As}\ T - T_s]$, $u^T = [C_{A0} - C_{A0s}\ Q - Q_s]$). The state-space profiles of the closed-loop system of Eq.(6) under the Safeness Index-based LEMPC of Eq.(3) and the CLBF-EMPC of Eq.(5) are shown in Fig. 1, where it is demonstrated that both the Safeness Index-based LEMPC

and CLBF-EMPC can maintain the state of the closed-loop system of Eq.(6) within the safe region (i.e., $S(x) \leq S_{TH}$, or $\mathcal{U}_\rho$); while under the standard LEMPC, the closed-loop states are only guaranteed to be bounded in the stability region (i.e., Ω_ρ, or $\mathcal{U}_\rho \cup \mathcal{D}$), but not within the safe (gray) regions. Additionally, the economic benefits under steady-state operation, Safeness Index-based LEMPC, and CLBF-EMPC are 13.9, 17.1, and 16.2, respectively, from which it is demonstrated that the proposed control schemes both economically outperform the tracking MPC (steady-state operation) and ensure process operational safety as well.



(a) Safeness Index-based LEMPC (b) CLBF-EMPC

Figure 1: The state-space profiles for the closed-loop CSTR under LEMPC (black trajectory) in both (a) and (b), and under the Safeness Index-based LEMPC of Eq.(3) (gray trajectory in (a)), and under the CLBF-EMPC of Eq.(5) (gray trajectory in (b)) for an initial condition (0,0).

6. Conclusion

In this work, we developed two MPC methodologies to deal with the process operational safety associated with economic model predictive control. Specifically, the problem of avoiding the unsafe region in state-space was first addressed with the notion of the Safeness Index and setting constraints within LEMPC using Safeness Index-defined regions. Then, we proposed a CLBF-based EMPC, which utilizes the constrained CLBF to design constraints in CLBF-EMPC, such that closed-loop stability, process safety, and the recursive feasibility are achieved simultaneously.

References

F. Albalawi, H. Durand, and P. D. Christofides. Process operational safety using model predictive control based on a process Safeness Index. Computers & Chemical Engineering, 104:76–88,2017.

Completed Investigations of Chemical Incidents. Final Report of the Investigations of Chemical Incidents. Technical report, U.S. Chemical Safety and Hazard Investigation Board, 2016.

N. G. Leveson and G. Stephanopoulos. A system-theoretic, control-inspired view and approach to process safety. AIChE Journal, 60:2–14, 2014.

Y. Lin and E. D. Sontag. A universal formula for stabilization with bounded controls. Systems & Control Letters, 16:393–397, 1991.

M. Z. Romdlony and B. Jayawardhana. Stabilization with guaranteed safety using control Lyapunov–barrier function. Automatica, 66:39–47, 2016.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50080-X

Run-to-Run Control of Film Thickness in PECVD: Application to a Multiscale CFD Model of Amorphous Silicon Deposition

Marquis Crose[a], Weiqi Zhang[a], Anh Tran[a], Gerassimos Orkoulas[b] and Panagiotis D. Christofides[a,c,*]

[a]*Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA, 90095-1592, USA*
[b]*Department of Chemical Engineering, Widener University, Chester, PA 19013, USA*
[c]*Department of Electrical Engineering, University of California, Los Angeles, CA 90095-1592, USA*
pdc@seas.ucla.edu

Abstract

In this work, we focus on the development of a multiscale computational fluid dynamics (CFD) model for plasma-enhanced chemical vapor deposition (PECVD) systems, and a corresponding run-to-run control framework with the purpose of improving product quality in the deposition of amorphous silicon thin film layers. A reactor scale model is linked with a microscopic, thin film growth model based on a hybrid kinetic Monte Carlo algorithm. Initially, open-loop operation suggests a product offset of greater than 8$\%$ from the set-point of 300 nm thickness. Application of the run-to-run control system, built around an exponentially-weighted moving average (EWMA) algorithm, is shown to reduce the thin film thickness non-uniformity to within 1$\%$ of the product set-point in less than 10 batches of operation.

Keywords: Multiscale modeling, computational fluid dynamics, run-to-run control.

1. Introduction

Uniform deposition of thin film layers remains a challenge in silicon processing industries due to the lack of in situ measurements and drift in the plasma composition caused by fouling (Gabriel et al., 2014). Specifically, in plasma-enhanced chemical vapor deposition (PECVD) of amorphous silicon (*a*-Si:H), growth rate non-uniformities greater than 20% across the surface of the wafer are common (Crose et al., 2016), which may lead to inefficient solar cells and microelectronics of poor quality. Process operators typically avoid this issue by preconditioning the reaction chamber and manually adjusting deposition conditions such that thickness non-uniformities may be reduced. Recently, Crose et al. (2016) developed a run-to-run (R2R) based, offline control scheme which has been demonstrated to reduce the offset in product thickness from 5% to less than 1% within ten batches of operation. While these results are promising, the multiscale model used relies on a first principles approach to solving the continuous mass, momentum and energy balances; however, recent advances in parallel computation strategies have made possible the application of multiscale computational fluid dynamics (CFD) to the modeling of PECVD (Crose et al., 2017).

Consequently, we propose the application of the newly developed run-to-run control strategy to the simulated deposition of amorphous silicon (*a*-Si:H) thin films using a

transient, multiscale CFD model. The two-dimensional axisymmetric model recently developed by Crose et al. (Crose et al., 2017) has been shown to accurately capture the complex behavior of the PECVD reactor and to reproduce experimentally observed non-uniformities with respect to the film thickness and porosity, and will be used throughout this work (Note: Fig. 3 demonstrates typical non-uniformity present in the process gas composition which leads to the aforementioned quality issues). A computationally efficient parallel processing scheme allows for the application of the R2R algorithm to 15 serial batch simulations which are shown to reduce the product offset from the 300 nm thickness set-point. It is important to clarify here that within each batch simulation 8 independent microscopic simulations are conducted which correspond to incremental radial wafer locations. Specifically, at each of $r = 0$, $r = 1$, ..., $r = 7$ cm, a kinetic Monte Carlo (kMC) algorithm is applied in order to capture both the exchange of mass and energy, as well as the character of the *a*-Si:H thin-film. The resulting multiscale model suggests that product offset from the desired thin-film thickness may be reduced to <1% for all zones.

The structure of this paper is as follows: first, the multiscale CFD model will briefly be presented including both the macroscopic reactor scale, as well as the microscopic thin film domain. Next, steady-state results will be discussed which motivate the use of an offline R2R control strategy. Finally, fifteen transient simulations demonstrate that the product quality, in terms of thin film thickness uniformity, may be greatly improved via the implementation of post-batch temperature control.

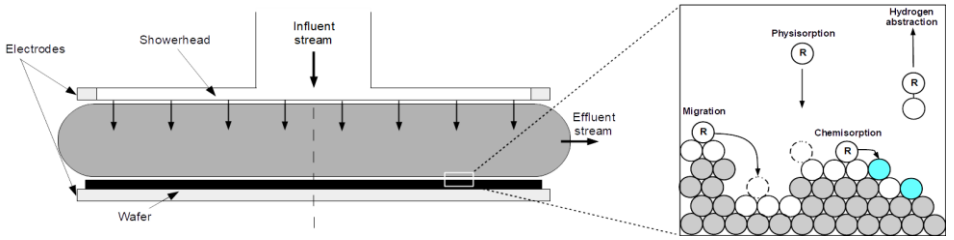


Figure 1. Macroscopic (left) and microscopic (right) PECVD simulation regimes.

2. Multiscale model

The PECVD process is inherently multiscale in that the macroscopic domain, which is defined as the gas phase contained within the reactor geometry, exchanges mass and energy with growing thin film layers which constitute the microscopic domain (see Fig. 1). As such, accurate modeling of PECVD processes requires the development of multiscale models which capture not only both domains, but the communication between these domains via a dynamic boundary. Recently, Crose et al. (Crose et al., 2017) demonstrated that multiscale models of this type are capable of reproducing accurate reactor operation and thin film growth; specifically, the spatial distribution of key deposition species such as SiH_3 and H are well characterized and highlight the need for batch-to-batch control strategies. The following subsections briefly present the development of the macro- and microscopic simulation domains; for a more in-depth discussion please refer to Crose et al. (Crose et al., 2016, 2017).

2.1. Macroscopic domain

In this work we utilize a two-dimensional, axisymmetric approximation of a typical chambered PECVD reactor (see Fig. 2). While the exact interior dimensions of PECVD

systems vary between manufactures, the dimensions shown here represent typical reactor geometries and may be considered as a representative base case. Additionally, the choice to use a two-dimensional (2D) axisymmetric approximation is motivated by both the axial symmetry of the cylindrical reactor as well as the desire to reduce computational demands. The results presented in the latter half of this work suggest a good agreement between the observed plasma characteristics and those reported experimentally; consequently, no additional geometries are explored.

In order to discretize a given geometry for use in CFD simulations, two general meshing strategies have emerged: (1) a collection of quadrilateral cells in a specific, repeating pattern is known as a structured mesh, and (2) unstructured meshes are characterized by a collection of polygons in an irregular pattern. Geometries lacking curvature or otherwise complex shapes benefit from the use of a structured mesh as they provide higher quality (e.g., orthogonality and aspect ratio) when compared to an unstructured mesh of equal cell count. Given the rectangular nature of the 2D axisymmetric geometry, a structured mesh composed of 120,000 cells is employed. It is important to note that throughout this work, ANSYS software is utilized for the creation of the geometric mesh (specifically, ICEM meshing) and as a solver for the partial differential equations which capture the gas phase mass, momentum and energy balances. An in-depth discussion of the finite volume methods used can be found in the ANSYS user manual (Ansys Inc., 2013).

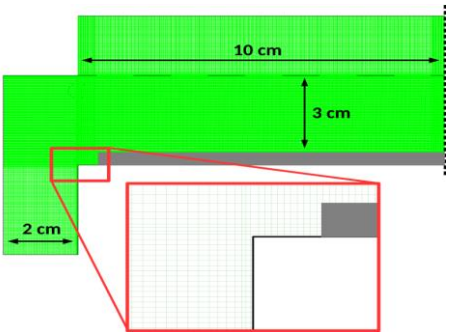


Figure 2. 2D axisymmetric geometry after discretization using a structured mesh containing 120,000 cells.

Three key user defined functions (UDFs) operate within the Fluent framework in order to tailor the solver to batch PECVD processing: the first involves a set of volumetric reactions which define the gas phase chemistry, the second details an electron density profile which enables plasma phase reactions, and the final UDF of considerable complexity allows for microscopic simulations (e.g., kinetic Monte Carlo algorithms) to operate in parallel to the macroscopic domain. Details on the various UDFs are available in the recent work of Crose et al. (Crose et al., 2017), while the UDF which defines the microscopic domain will be discussed in the following subsection.

2.2. Microscopic domain

Growth of amorphous silicon (*a*-Si:H) thin film layers along the surface of the wafer substrate (grey region in Fig. 2) defines the microscopic domain. Although tracking each individual particle which contributes to deposition remains a computationally infeasible task, growth of the lattice microstructure may be achieved using discrete kinetic Monte Carlo (kMC) algorithms. Specifically, at discrete locations along the

surface of the wafer from r = 0 cm to r = 8 cm, independent kMC simulations are conducted each of which utilize a triangular lattice of width 1200 particles (see Fig. 1, right). Regions along the wafer surface which are not explicitly contained within a kMC domain are assigned growth rates via interpolation between the known locations. In this manner, thin film layers may be grown along the complete span of the wafer substrate, and, of equal importance, the cells which border the wafer are provided with a dynamic boundary. Due to space limitations, the development of the kMC algorithm will not be provided here; however, a complete listing, including the relevant rate equations, may be found in Crose et al. (Crose et al., 2016). The transient boundary conditions of the multiscale model are of particular importance and are discussed in detail in the following subsection.

2.3. Multiscale workflow

Before the relevant steady-state and transient results can be discussed, it is important to clarify the methodology of the multiscale simulation presented here. At the start of a simulation, t = 0, the first time step for the macroscopic CFD domain is calculated within each cell of the reactor mesh using the ANSYS Fluent as discussed previously. Once the end of the first time step has been reached, $t = t_1$, species concentration, temperature and pressure information along the substrate boundary is fed to the microscopic domain. Discrete kMC simulations along the wafer surface are initialized using this information, and growth of the *a*-Si:H thin film begins. After the microscopic simulations have caught up to t_1, the boundary conditions for cells bordering the microscopic domain are updated based on the transfer of mass and energy to thin film growth. The macroscopic PDEs are once again solved such that the time progresses to $t = t_2$, and the cycle continues. The multiscale simulation progresses in this manner until the thin film product has reached its desired batch deposition time, t_{batch}.

3. Results

3.1. Steady-state species distribution

Once the reactor startup period has completed (i.e., after roughly the first three seconds of transient operation), distinct spatial non-uniformity becomes apparent in the concentration of SiH_3, and this non-uniform distribution persists throughout the remainder of the batch simulation. This phenomena is of particular importance in the red highlighted region of Fig. 3. Radial non-uniformity in the deposition species above the wafer substrate is the primary factor affecting the thickness profile of the thin film product. When operated in an open-loop fashion, without the use of batch-to-batch monitoring and control, the resulting thickness of the thin film layer near the edge of the wafer will undershoot the product specification of 300 nm by more than 8% (e.g., Fig. 4). At this time, there exists no direct means by which the species distribution within the PECVD reactor may be controlled; fortunately, run-to-run based algorithms have been shown to be effective in predicting optimal substrate temperatures which may reduce the thickness non-uniformity of amorphous silicon products (Crose et al., 2016).

3.2. Run-to-run control via EWMA

Given that the concentration and distribution of the deposition species (e.g., SiH_3) cannot be directly controlled, an alternative manipulated variable must be chosen in order to drive the thin film product to the desired set-point of 300 nm thickness. Recently, Crose et al. (Crose et al., 2016) demonstrated that the growth rates of *a*-Si:H thin film layers are a linear function of the substrate temperature in the neighborhood of

the nominal deposition condition of T = 475 K, for fixed species concentrations (namely, x_{SiH3} and x_H). Hence, a run-to-run control strategy is utilized in this work which requires only post-batch measurements of the product thickness in order to update the wafer temperature.

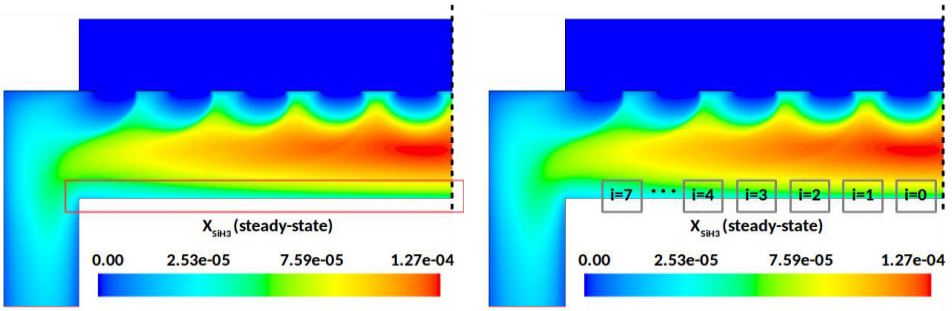


Figure 3. (left) Steady-state SiH$_3$ concentration within PECVD reactor. (right) Defined reactor zones which contain independent EWMA control algorithms.

Specifically, within each of the eight defined wafer zones (see Fig. 3, right) an exponentially-weighted moving average (EWMA) algorithm is applied in order to iteratively update the substrate temperature such that the *a*-Si:H thin film product may be driven to the desired thickness of 300 nm. The proposed EWMA algorithm is of the form:

$$\varepsilon_{k+1}^{i} = (1-\lambda)\varepsilon_{k+1}^{i} + \lambda(\tau_{s.p.} - \tau_{meas.}^{i}) \tag{1}$$

where ε^{i}_{k+1} is the parameter correction for the $k+1$ batch in zone i, λ is the learning factor, $\tau_{s.p.}$ is the thickness set-point and $\tau^{i}_{meas.}$ is the thickness measurement for zone i. The substrate temperature setting for each consecutive batch is then calculated via:

$$T_{k+1}^{i} = T_{k}^{i} + \varepsilon_{k+1}^{i}/(\alpha^{i} \cdot t_{batch}) \tag{2}$$

where T^{i}_{k+1} and T^{i}_{k} are the temperatures in zone i for the next and pervious batches, respectively, and α_i is a product of the growth rate linearization. Due to space limitations, further details on the development of the EWMA algorithm will not be provided here; please refer to Crose et al. (Crose et al., 2016).

3.3. Batch-to-batch operation

Fifteen data sets are shown in Fig. 4; each of which contains eight data points which correspond to the radial wafer zones. In the first data set (i.e., batch 1), all zones were maintained at the nominal deposition temperature, T = 475 K. The resulting thickness of the *a*-Si:H thin film demonstrates significant non-uniformity, most apparent in zones 6 and 7 which lie near the edge of the wafer substrate. This result is expected due to the observed loss in SiH_3 density near the outlet ports of the PECVD reactor (e.g., Fig. 4).

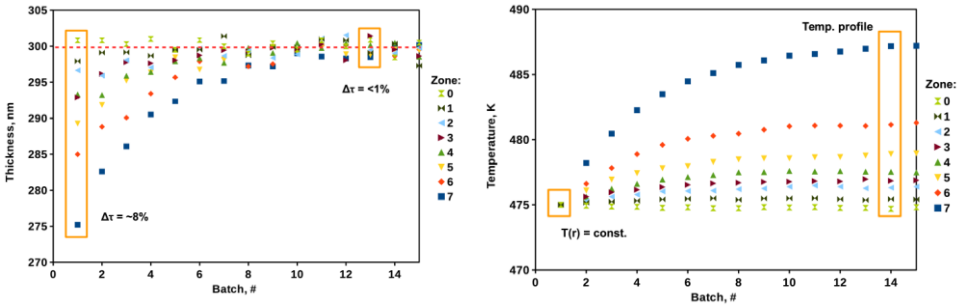


Figure 4. (left) Batch-to-batch thin film thickness. (right) Temperature profile developed via EWMA algorithm.

As the transient simulation progresses and additional batch deposition cycles complete, the EWMA algorithm begins to adjust the substrate temperature profile; see batches 2-15 in Fig. 4, right. Zone 0 maintains a near constant 475 K as its position is in the center of the reactor and receives the highest concentration of SiH_3. Conversely, the temperature of zone 7 is set to ~487 K by the R2R controller in an effort to drive the thin film thickness near the edge of the substrate to the set-point of 300 nm. By batch 8, all zones are within 1% of the product set-point and are maintained inside this margin thereafter.

4. Conclusions

A run-to-run control framework has been applied to batch deposition of amorphous silicon layers via PECVD. A multiscale CFD simulation which coordinates communication amongst the macroscopic, reactor scale model and the microscopic thin film growth domains has allowed for accurate growth of *a*-Si:H layers on a 16 cm wafer substrate. Radial concentration gradients in the deposition species, in particular SiH_3, have been shown to affect the uniformity of the thin film product. Through the application of the offline R2R control strategy which predicts optimized substrate temperatures, thickness offset in the product has been shown to be reduced from ~8% to less than 1% within 10 batches of operation.

References

ANSYS Inc. ANSYS Fluent Theory Guide 15.0 (November), November 2013.

M. Crose, Kwon, JSI., A. Tran, and P. D. Christofides. Multiscale modeling and run-to-run control of PECVD of thin film solar cells. Renewable Energy, 100:129–140,2016.

M. Crose, A. Tran, and P. D. Christofides. Multiscale computational fluid dynamics: Methodology and application to PECVD of thin film solar cells. Coatings, 7:23, 2017.

O. Gabriel, S. Kirner, M. Klick, B. Stannowski, and R. Schlatmann. Plasma monitoring and PECVD process control in thin film silicon-based solar cell manufacturing. EPJ Photovoltaics, 55202:1–9,2014.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50081-1

Distributed model predictive control of a system with multi-rate and delayed measurements

Arvind Ravi and Niket S. Kaisare*

Department of Chemical Engineering, Indian Institute of Technology Madras, Chennai 600036, India
nkaisare@iitm.ac.in

Abstract

The objective of this work is to implement a sequential distributed model predictive control (MPC) for a process with infrequent and delayed primary measurements in presence of disturbances in the system. The overall system comprises of multiple units. A centralized state estimator based on Kalman filter is used to estimate states for the entire system. The estimator uses sampled state augmentation approach for fusing the delayed information from primary measurements with the frequent secondary measurements to obtain better estimates for output feedback control. Distributed MPC is developed that uses these estimates, as well as selective information transfer, based on process knowledge, between different units of the system. An example of a reactor separator system is considered to demonstrate the applicability of the control formulation. This system also has a plant-model mismatch of some parameter values and a disturbance model framework is incorporated.

Keywords: Distributed MPC; Sampled-state augmentation; Centralized control; Disturbance modelling

1. Introduction

Large process plants generally comprise of various interconnected sub-units. Overall control of the entire plant, when accomplished by control of individual units, is known as decentralized control. However, the performance of decentralized controller deteriorates as the interaction between various sub-units becomes stronger. Best control performance can, in principle, be achieved in a centralized control framework which has the knowledge about all interactions and disturbances affecting the plant. However, difficulty in obtaining a combined model for the system as well as computational challenges make the task of establishing a robust centralized control intractable. Hence, distributed control, which establishes frequent exchange of process- and control-related information across various controllers in tune with the actual plant interaction, has been proposed (Venkat et al., 2005). Parallel computation of localized controller targets reduces the computational complexity and unit level distributed structure reduces the associated modelling rigor compared to its overall plant level control. Christofides et al. (2013) provide the conceptual base and exhaustive review of distributed MPC. The local controller structure is based on physical sub-units in the system, though automatic partitioning into local sub-controllers has been proposed (Rocha et al., 2016).

Some earlier works discuss the distributed control application for systems with delayed multirate measurements. Liu et al. (2010 and 2012) address this issue with a sequential and an iterative DMPC framework, respectively. However, they considered full state feedback, which may occur in an asynchronous manner. Furthermore, framework for

systems with parameter mismatch or plant disturbances is missing in literature. Multi-unit systems require inferential control of a process variable of one unit using the estimates obtained through the measurements from other interconnected subsystems. While this inferential control of a single process unit in presence of disturbance has been well studied (Padhiyar et al. 2006; Prasad et al. 2002), it is challenging to extend the notion to a distributed architecture.

The objective of this work is to implement a distributed control framework when primary variable of one or more units in the system is measured infrequently and measurements are available after some delay. The control variables are inferred from secondary measurements when the primary measurements are unavailable. When these measurements arrive with a delay, state augmentation formulation (Gopalakrishnan et al., 2011) is used to fuse them in centralized estimator framework. This method is more efficient than filter recalculation, and is appicable for varying time-delays and for nonlinear systems. The control structure based on linear models thus developed is illustrated using a reactor-separator system. The observability of the plant ensures the inferential control of the parameter till the arrival of the delayed measurements.

The rest of the paper is organized as follows. Next section introduces the Kalman filter based centralized estimator formulation using sampled state-augmentation for fusing delayed primary measurement (Gopalakrishnan et al. 2011). Subsequently, distributed MPC framework is described. The proposed approach is implemented on a reactor-separator system in the next section. Results are presented for distributed control with and without delayed primary measurements, followed by conclusions of this work.

2. State Estimator Formulation

2.1. Model Formulation

Most chemical process systems follow a nonlinear relationship between variables, which is represented as:

$$\dot{x} = f(x,u) \tag{1}$$

$$y = g(x) + \lambda \tag{2}$$

where $x \in \Re^n, u \in \Re^p, y \in \Re^m$; and λ represents the measurement noise associated with the sensor. The control formulation in this work is developed for a linear model, which is obtained by linearizing around the steady state of the plant. The discrete-time state-space model is given by:

$$x_{k+1} = Ax_k + Bu_k \tag{3}$$

$$y_{k+1} = Cx_{k+1} + \nu_{k+1} \tag{4}$$

The sensor noise, ν, is assumed to be additive Gaussian white noise with $N(0,\sigma^2)$ and uncorrelated with the states, and model-plant mismatch is considered. Specifically, we use state-space matrices (A^P, B^P, C^P) to distinguish the "plant" from the model used for computing the MPC control actions. The latter "model" is used in the state estimator as well as in the controller design.

2.2. Input Disturbance Model

Disturbances can enter the system as unmeasured disturbances or through errors in the model parameters. The latter case is considered in the subsequent case study. The

disturbance models are constructed by augmenting disturbance states and estimating the same through state estimation techniques. Muske & Badgwell 2002 provide a detailed account of various disturbance modelling techniques and necessary conditions in choosing the disturbance states. Load disturbances are commonly added to compensate for model plant mismatch. Based on the Hautus condition for detectability, the number of disturbance states augmented should be less than or equal to measurements (Muske & Badgwell 2002). Since the number of measured variables during minor and major instances are different in multi-rate estimation, considering a predefined number of disturbance states will create observability issues. Hence, to account for bias due to parameter mismatch, integrating white noise disturbance is assumed to propagate through the inputs, i.e., $u_k^* = u_k + d_k$ (Padhiyar et al. 2006). The disturbance model,

$$x_{k+1} = Ax_k + Bu_k + Bd_k \tag{5}$$

can be written in the standard form by defining augmented state $z_k \triangleq [x_k^T \quad d_k^T]^T$ and

$$z_{k+1} = \tilde{A}z_k + \tilde{B}u_k + \tilde{B}_{\mathrm{d}}\varepsilon_k \tag{6}$$

$$y_{k+1} = \tilde{C}z_{k+1} + \nu_{k+1} \tag{7}$$

$$\tilde{A} = \begin{bmatrix} A & B \\ 0 & I \end{bmatrix}, \qquad \tilde{B} = \begin{bmatrix} B \\ 0 \end{bmatrix}, \qquad \tilde{B}_{\mathrm{d}} = \begin{bmatrix} 0 \\ I \end{bmatrix}, \qquad \tilde{C} = [C \quad 0] \tag{8}$$

The number of augmented disturbance states will be equal to the number of inputs.

2.3. Measurement Delays

The primary measurements (composition or quality variables) are sampled at a slower rate and are available after a fixed time delay (n_D) from its sampling instance. At minor instances, when only secondary measurements are available,

$$y_k^s = C^{\mathrm{s}}x_k + \nu_k^{\mathrm{s}} \tag{9}$$

At major instances when delayed primary measurements also arrive:

$$\begin{bmatrix} y_k^{\mathrm{s}} \\ y_k^{\mathrm{p}} \end{bmatrix} = \begin{bmatrix} C^{\mathrm{s}}x_k \\ C^{\mathrm{p}}x_{k-n_D} \end{bmatrix} + \begin{bmatrix} \nu_k^{\mathrm{s}} \\ \nu_{k-n_D}^{\mathrm{p}} \end{bmatrix} \tag{10}$$

where, "s" and "p" represent secondary and primary measurements, respectively. As in Eq. (8), we define $\tilde{C}^{\mathrm{s}} = [C^{\mathrm{s}} \quad 0]$ and $\tilde{C}^{\mathrm{p}} = [C^{\mathrm{p}} \quad 0]$.

Multi-rate state estimation will be implemented using the sampled-state augmentation method to fuse the delayed measurements when they arrive. As we show later, this method offers some computational advantage over other available methods for state estimation with delayed measurements (Gopalakrishnan et al. 2011). Prior to sampling of the primary variable, Kalman filter is implemented with just the frequent secondary measurement. At the time instance when the primary variable is sampled, the state vector is augmented with the states at the same instance (sampling). The augmented state estimate and the covariance matrix is given by

$$\hat{Z}_{(s|s)} = \begin{bmatrix} \hat{z}_{(s|s)} \\ \hat{z}_{(s|s)} \end{bmatrix}, \qquad P_{(s|s)}^{\mathrm{a}} = \begin{bmatrix} P_{(s|s)} & P_{(s|s)} \\ P_{(s|s)} & P_{(s|s)} \end{bmatrix} \tag{11}$$

and the state space transition matrices are given by

$$\Phi = \begin{bmatrix} \tilde{A} & 0 \\ 0 & I \end{bmatrix}, \qquad \Gamma = \begin{bmatrix} \tilde{B} \\ 0 \end{bmatrix}, \qquad \Psi = \begin{bmatrix} \tilde{B}_d \\ I \end{bmatrix}, \qquad \Xi = \begin{bmatrix} \tilde{C}^{\mathrm{s}} & 0 \\ 0 & \tilde{C}^{\mathrm{p}} \end{bmatrix} \tag{12}$$

During minor instances, only the first row of Ξ is used since y^{P} have not arrived. Once the primary measurement arrives, the measurements are fused using Kalman filter and augmented state is un-augmented back to its original form. The improved state estimates thus obtained are considered for MPC predictions in the controller block.

3. Distributed MPC: Output Feedback

Distributed control system is a modification to the decentralized structure with added model information regarding interaction from the neighboring subunits. Figure 1(b) shows the architecture of distributed MPC with output feedback. Let x^i represent the state vectors for the i^{th} subunit. Thus, $x = [x^1 \quad x^2 \quad \ldots \quad x^n]^T$

Two subunits in the system interact with each other. This interaction is accounted for by sharing information between the controllers of various subunits. Let x^{ij} ($x^{ij} \subset x^j$) represent the states of subunit *j* that are required in prediction of subunit *i*. The general form of the interacting model (Venkat et al. 2005) is given by

$$x_{k+1}^i = A_i x_k^i + B_i u_k^i + \sum_{j \neq i} \left(A_{ij} x_k^{ij} + B_{ij} u_k^{ij} \right) \tag{13}$$

$$y_{k+1}^i = C^i x_{k+1}^i + v_{k+1}^i \tag{14}$$

We further assume that the interaction between sub-units is uni-directional. Thus, an upstream unit affects a downstream unit, but not vice versa. In our future work, we will relax this assumption. This assumption allows us to calculate the control actions of upstream units first, followed by downstream ones in a sequential manner.

3.1. Case study: Reactor-separator system

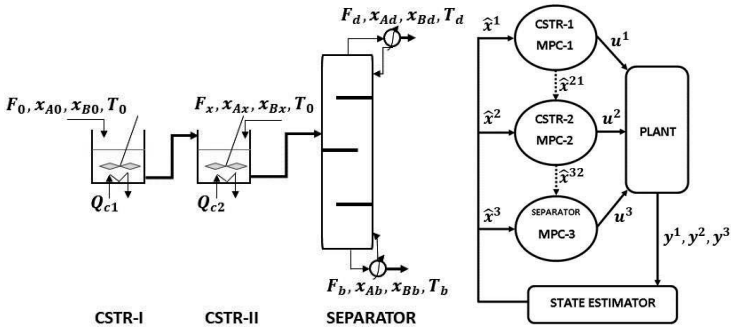


Figure 1: (a) Schematic of the reactor-separator system considered; (b) DMPC architecture

The case study of the two-reactor-separator system (Figure 1(a)) is considered for the implementation of the proposed distributed control formulation. Table 1 lists the key variables of the system. Since we have at least 3 measurements at each sampling instance, the number of disturbance states are always less than or equal to number of measurements thus satisfying the detectability condition for offset free control. Parameter mismatch is assumed between the linearized model and plant (in form of different values of heat transfer coefficients in CSTRs and relative volatility in separator). The sampling time is assumed to be 3 min; primary variables are sampled every 30 min and arrive with a delay of 15 min.

3.2. DMPC Algorithm

For the process in Figure 1, the distributed MPC formulation reduces to

$$x_{k+1}^{i} = \left(A_i x_k^i + B_i u_k^i\right) + A_{il} x_k^{il}, \qquad i = 1, 2, 3, \qquad 1 \le (l = i - 1) \le 2 \tag{15}$$

In other words, there is sequential transfer of information due to material transport across the three sub-units and the manipulated inputs of any sub-unit do not directly affect any other sub-unit. A future extension of this work is to compare the performance in presence of two-way interactions between the sub-units; for example, in presence of recycle from the separator to a reactor. Non-negativity constraints are used for mole fractions, and manipulated variables have both input magnitude and rate constraints. The proposed approach is implemented in MATLAB.

Table 1: List of variables of interest in the case-study

Variables	**CSTR-1**	**CSTR-2**	**Separator**
State Variables	x_A, x_B, T_1, T_{c1}	x_A, x_B, T_2, T_{c2}	x_{Ai}, x_{Bi} $(1 \le i \le 5)$
Controlled Variables	T_1	T_2	x_{B1}
Manipulated Variables	Q_{c1}	Q_{c2}	F_b
Measured (Frequent)	T_1	T_2	T_f
Measured (Infrequent)	–	–	x_{A1}, x_{B1}

4. Results and Discussion

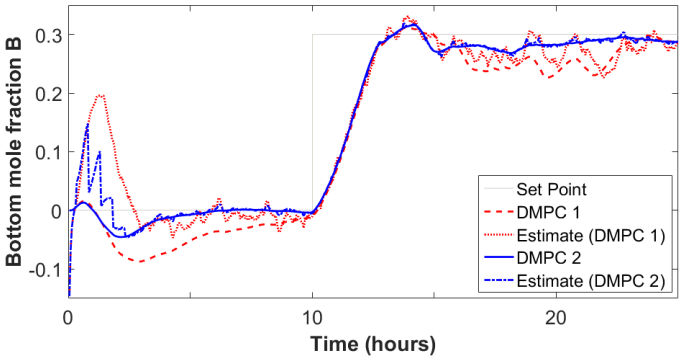


Figure 2: Comparison of DMPC with and without delayed primary measurements, DMPC 1 represents distributed MPC without primary measurements (secondary measurements only) and DMPC 2 includes delayed primary measurements.

Figure 2 and Figure 3 show the results of implementing the proposed approach on the reactor-separator system. For the sake of brevity, only the control of bottoms mole fraction of B (x_{Bb}) is discussed. This is chosen because x_{Bb} forms the delayed primary measurement in the third sub-unit; whereas the controlled variables T_1, T_2 for the CSTRs are available frequently and without delay. The thick solid line (blue color in online version of both figures) represents the values for distributed MPC with infrequent and delayed primary measurements. As can be seen from Figure 2, the performance of distributed MPC improves significantly when the primary measurements are used. A more meaningful comparison is that of distributed with centralized MPC in presence of delayed primary measurements. Figure 3 shows that the performance of distributed

MPC is close to that of the centralized control for controlling the bottoms mole fraction. Since measurements of the CSTRs are available without any delay, the distributed and centralized controllers match closely (results not shown). We compared this approach with distributed MPC using filter recalculation, and found our approach yields the same results, but with a slightly lower simulation time.

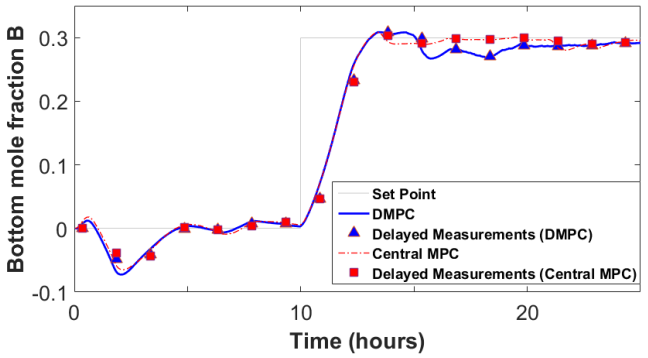


Figure 3: Comparison between centralized and distributed MPC (with primary variable sampled at 90 min). Markers represent the values of primary variables plotted at the arrival time.

5. Conclusions

The performance of distributed MPC with multi-rate estimation of primary variables was shown to be comparable to that of centralized control. Distributed MPC relies on seamless transfer of information across multiple controllers based on interactions in the physical plant. Fusing the infrequent and delayed primary measurements was important, as the improved state estimates provide better prediction of the control trajectory and hence improved control performance. Multi-rate estimation using augmentation of the sampled state is promising for distributed MPC and has potential scope for online implementation. This work will be extended to nonlinear system.

References

P.D. Christofides, R. Scattolini, D.M. de la Peña, J. Liu, 2013. Distributed model predictive control: A tutorial review and future research directions. *Comput. Chem. Engng.*, 51, 21–41

A. Gopalakrishnan, N.S. Kaisare, S. Narasimhan, 2011. Incorporating delayed and infrequent measurements in EKF- based nonlinear state estimation. *J. Process Control*, 21, 119–129.

J. Liu, X. Chen, D.M. de la Peña, P.D. Christofides, 2012. Iterative distributed model predictive control of nonlinear systems: Handling asynchronous, delayed measurements. *IEEE Transactions on Automatic Control*, 57, 528–534.

J. Liu, D.M. de la Peña, P.D. Christofides, 2010. Distributed model predictive control of nonlinear systems subject to asynchronous and delayed measurements. *Automatica*, 46, 52–61.

K.R. Muske, T.A. Badgwell, 2002. Disturbance modeling for offset-free linear model predictive control. *Journal of Process Control*, 12, 617–632.

N. Padhiyar, A. Gupta, A. Gautam, S. Bhartiya, F.J. Doyle III, S. Dash, S. Gaikwad, 2006. Nonlinear inferential multi-rate control of Kappa number at multiple locations in a continuous pulp digester. *Journal of Process Control*, 16, 1037–1053.

V. Prasad, M. Schley, L.P. Russo, B.W. Bequette, 2002. Product property and production rate control of styrene polymerization. *Journal of Process Control*, 12, 353–372.

R. Rocha, L.C. Oliviera-Lopes, 2016. A cooperative distributed model predictive control for nonlinear systems with automatic partitioning, *Comput. Aided Chem. Engng.*, 38, 2205-2210

A. Venkat, J.B. Rawlings, S. Wright, 2005. Stability and optimality of distributed model predictive control. *IEEE Conference on Decision and Control,* 6680–6685.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50082-3

On-line Bayesian-based Model-set Management Method with Case Study of Steam Reforming Prediction under Various Feed Compositions

Jia Li[a*], Mingheng Li[a]

[a]*Cal Poly Pomona, 3801 W Temple Ave, Pomona, CA 91768, USA*
jiali@cpp.edu

Abstract

Steam reforming is the most widespread process for the generation of hydrogen-rich synthesis gas from light carbohydrates. One of the operation and design challenges is that the feed materials are various from natural gas, shale gas, to liquid gas, or naphtha. The optimal operating conditions identified under one type of feed may not be functional well under another one. Even within the same type feed, the compositions may vary large enough in continuous operation that requires online operating conditions tuning. Currently, two types of control strategies are applied to address the issue: one is feedforward control that the feed composition is measured in advance so that the operating conditions could be set accordingly. However, the performance would rely on the accuracy of the process model, and may not be robust on feed composition disturbances. The other is simple PID feedback control that the conditions are adjusted by the product composition feedback. However, the complexity of the multi-input multi-output (MIMO) steam reforming may bring challenges to the PID controller design. In this paper, a Bayesian-based model-set management method is introduced for constructing a statistically superior model set for online model-based control application. First, a number of steam reforming models are developed on various feed types or compositions. Then the optimal operating conditions are identified on the weighted model predictions. During online application, the measured product compositions are feedback to adjust the weights by a Bayesian-based statistical method. A repeated use of the method keeps the weights updated constantly based on the newly available system data, which makes the model-set-based prediction more precise, robust, and adaptive to various feeds automatically. The efficiency of the method is demonstrated by studying a steam reforming system with different feeds.

Keywords: Bayesian method, online model-based control, steam reforming

1. Introduction

Steam reforming is the most widely used method for producing hydrogen, carbon monoxide, or other useful products from hydrocarbon fuels such as natural gas. The severe operating conditions and rapid system dynamics necessitate advanced model-based control or optimization. Great efforts have been made to develop reaction kinetics and heat/mass transfer models for various feeds and catalysts (Hou and Hughes, 2001; Abbas et al., 2017). However, during online applications, the process is typically under various fluctuations and disturbances such as feed composition change, which make it a great challenge to ensure the accuracy and robustness of model prediction.

To address the challenge, a Bayesian-based model-set management method is introduced for constructing a statistically superior model set for online model-based control application. The efficiency of the method is demonstrated by a steam reforming case study showing a superior prediction performance under various process feeds.

2. Multiple-Model-based Process Control with Bayesian-based Weight Adjustment

The presented system diagram with model online evaluation is shown in Figure 1. Instead of having only one model, the presented scheme has a model base including N models (m_1, m_2, …, m_N). Each model has the same type of inputs, $\boldsymbol{X}$, and outputs, $\widehat{\boldsymbol{Y}}$. However, each model may present a certain scenario of the process, or has different model structures. For example, the models could be developed under various feeds, catalyst types, or catalyst age. The model itself could be empirical, first-principle, neural network-based, or hybrid model. Compared to the one-model-based control system, it is expected that the model set could provide more robust predictions by covering the entire process scenarios. The model set could also minimize the necessity of online measurements (e.g. feed composition) by including those variables into modeling.

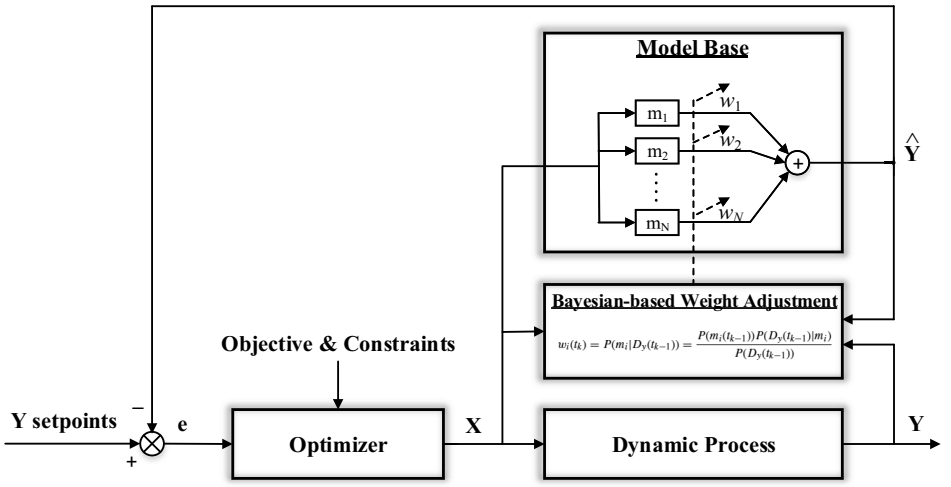


Figure 1 Bayesian-based model evaluation in the model based control diagram

The model prediction of model i at time t_k is presented as:

$$\hat{Y}_i(t_k) = m_i\big(X(t_k)\big) \qquad i = 1, 2, \dots, N \tag{1}$$

The model set prediction is the weighted average of all $\hat{Y}_i(t_k)$ with weights w_i:

$$\hat{Y}(t_k) = \sum_{i=1}^{N} w_i(t_k)\hat{Y}_i(t_k) \tag{2}$$

Having the weighted average model prediction, the control system could compare it to the output setpoint trajectory and identify the optimal manipulated variables in $\boldsymbol{X}$ to maximize (or minimize) the objective function under constraints.

When the process undergoes changes and the measured outputs deviate from the model prediction, a Bayesian-based model-set management method is introduced for constructing a statistically superior model set. In Bayesian equation, the posterior probability is constructed to evaluate the model accuracy by the product of prior probability and likelihood:

$$P(m|D) = \frac{P(m)P(D|m)}{P(D)} \tag{3}$$

Note that D is the newly available data set for evaluating model m. Prior probability $P(m)$ contains the prior knowledge about the applicability of model m. Posterior probability $P(m|D)$ is evaluated based on the past information about the model and the new process information. The model with a large posterior probability should be more desirable in model selection.

The presented weight adjustment system applies the posterior probability as the weights of each model. Assuming that the system information is available at time t_{k-1}, and the input and output data are contained, respectively. Then the Bayesian expression can be extended to time-variant as follows:

$$w_i(t_k) = P\big(m_i|Y(t_{k-1})\big) = \frac{P\big(m_i(t_{k-1})\big)P(Y(t_{k-1})|m_i)}{P\big(Y(t_{k-1})\big)} \tag{4}$$

The prior probability $P\big(m_i(t_{k-1})\big)$ presents the preference of model m_i before re-evaluation. Initially, all the models have the same preference. Then the posterior probability (or weight) of the previous time step can be set as the current prior probability. This should be valid based on such an analysis that a model showing a better prediction performance in the past may have a better chance to perform well in the current time step. Assuming the prediction error has a normal distribution with zero mean and deviation σ_j (j accounts for output variables), the likelihood $P(Y(t_{k-1})|m_i)$ is determined by the model prediction error probability. After equation derivation (Li and Huang, 2006), weight ω_i for model m_i at time t_k can be expressed of model predictions, measured outputs and preferred deviation:

$$w_i(t_k) = \begin{cases} \dfrac{\exp\left[-\frac{1}{2}\sum_j\left(\frac{y_j(t_{k-1})-\hat{y}_j(t_{k-1})}{\sigma_j}\right)^2\right]}{\sum_{i=1}^{N}\exp\left[-\frac{1}{2}\sum_j\left(\frac{y_j(t_{k-1})-\hat{y}_j(t_{k-1})}{\sigma_j}\right)^2\right]} & k=1 \\ \dfrac{w_i(t_{k-1})\exp\left[-\frac{1}{2}\sum_j\left(\frac{y_j(t_{k-1})-\hat{y}_j(t_{k-1})}{\sigma_j}\right)^2\right]}{\sum_{i=1}^{N}w_i(t_{k-1})\exp\left[-\frac{1}{2}\sum_j\left(\frac{y_j(t_{k-1})-\hat{y}_j(t_{k-1})}{\sigma_j}\right)^2\right]} & k>1 \end{cases} \tag{5}$$

where i is for model number, j is for output variables, k is for sampling time.

The Bayesian-based weight determination method can be used to evaluate periodically the degree of applicability of the models in the model set and to create a weighted average

prediction for system characterization. The method can be embedded in the following procedure for automatic, continuous management of the model set.

Step 1. Generate model set M that contains N process models. The models can be structurally or parametrically different.

Step 2. Initialize weights ω_i to $1/N$ and $k = 1$. Set deviation σ_j of the prediction error of each output variable. Calculate the initial model prediction sets $\hat{Y}_i(t_0)$.

Step 3. Obtain the real process outputs measurement values $Y(t_{k-1})$.

Step 4. Generate model prediction error $Y(t_{k-1}) - \hat{Y}_i(t_{k-1})$ for each model.

Step 5. Calculate the weights $w_i(t_k)$ using Eq. (5).

Step 6. Generate new model-based prediction $\hat{Y}(t_k)$ using Eq. (1) and (2).

Step 7. Return to Step 3 for the next time instant.

Note that after each update of the model weights, the model prediction $\hat{Y}(t_k)$ can be used for model-based control/optimization in the period before the next weight re-evaluation.

3. Steam Reforming Case Study

To show the efficiency of the presented method, the steam reforming process is applied which consists of three major reactions as below:

$$\text{Reaction 1: } CH_4 + H_2O \longleftrightarrow 3H_2 + CO$$

$$\text{Reaction 2: } CO + H_2O \longleftrightarrow H_2 + CO_2$$

$$\text{Reaction 3: } CH_4 + 2H_2O \longleftrightarrow 4H_2 + CO_2$$

According to Xu and Froment (1989a, 1989b), the kinetic models are developed as below:

$$r_1 = \frac{k_1}{P_{H_2}^{2.5}} \times \left(P_{CH_4} P_{H_2O} - \frac{P_{H_2}^3 P_{CO}}{K_1} \right) / \Omega^2 \tag{6}$$

$$r_2 = \frac{k_2}{P_{H_2}} \times \left(P_{CO} P_{H_2O} - \frac{P_{H_2} P_{CO_2}}{K_2} \right) / \Omega^2 \tag{7}$$

$$r_3 = \frac{k_3}{P_{H_2}^{3.5}} \times \left(P_{CH_4} P_{H_2O}^2 - \frac{P_{H_2}^4 P_{CO_2}}{K_3} \right) / \Omega^2 \tag{8}$$

$$\Omega = 1 + K_{CO} P_{CO} + K_{H_2} P_{H_2} + K_{CH_4} P_{CH_4} + K_{H_2O} P_{H_2O} / P_{H_2} \tag{9}$$

Where k_1, k_2, k_3, K_{CO}, K_{H2}, K_{CH4}, K_{H2O} are functions of temperature. The function parameters could be obtained by experimental data regression.

A model set with five steam reforming models are developed under different scenarios. The inputs are feed temperature and tube wall temperature, while outputs are product compositions and temperature. The model prediction error is defined as the sum of normalized outputs error to the real process. The set-point error is defined as the sum of normalized outputs error to the product composite set-points. Table 1 shows the model description. In the simulation test, the base case model 3 is applied as the pseudo real process for the first 30 sampling times and model 4 is set to the real process afterwards.

Table 1 Description of the Steam Reforming Models

Model #	**Description**
Model 1	Change the tube heat transfer coefficient from 50 to 70 [W/m2/K]
Model 2	Increase k_1 by 1.5%
Model 3	Base case model
Model 4	Change the steam to methane ratio from 1.75 to 1.96
Model 5	Increase K_{CH4} by 0.5%.

Figure 2 shows each model has different prediction error. Since the model 3 is assumed to be the real process at the first 30 sampling data, its prediction error keeps zero. With Bayesian-based method, the prediction error of the weighted model could approach to near zero quickly. After 30 data, the feed changes the steam to methane ratio and the model 4 becomes the pseudo real process. It can be seen that the prediction error of the weighted model could approach the minimal value as well in an automatic way.

A better model prediction performance could enhance the control performance as well. Figure 3 shows the error of the output compositions to the set-points. For the first 30 data, since the base case model 3 is also set as the real process, the model-based control could keep the setpoint error to the minimal value within the operation window. However, after the feed changes, the model 3 could not provide a good prediction and the setpoint error keeps high. As comparison, the Bayesian-based model could approach the same minimal setpoint error in several sampling times and adapt to the new scenario as well. It is shown that when the feed changes, the setpoint error does not change until ~60 sampling times when the weighted model gives new optimal inputs (the feed T and the tube wall T). To expedite the process, the deviation σ of the prediction error could be tuned as a trade-off between the quick response and the robust disturbance rejection.

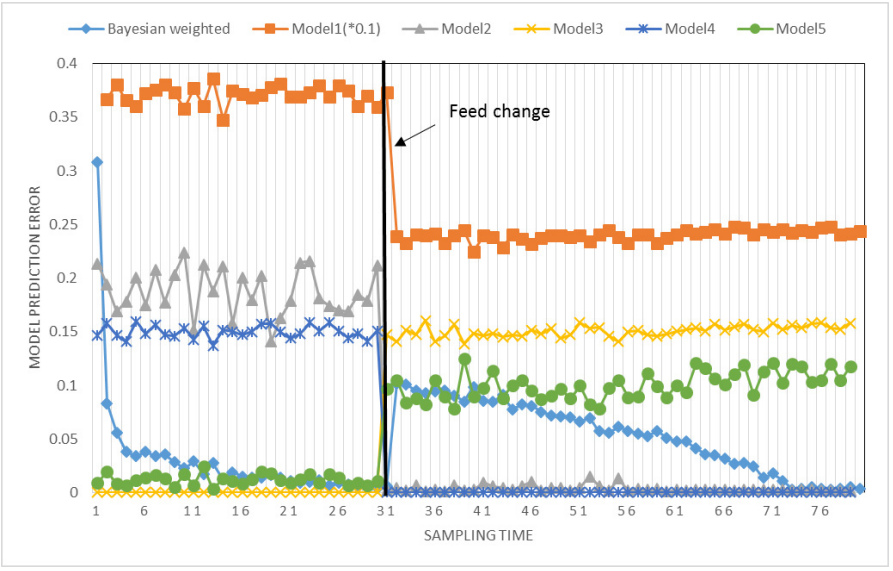


Figure 2. Model prediction error with feed change

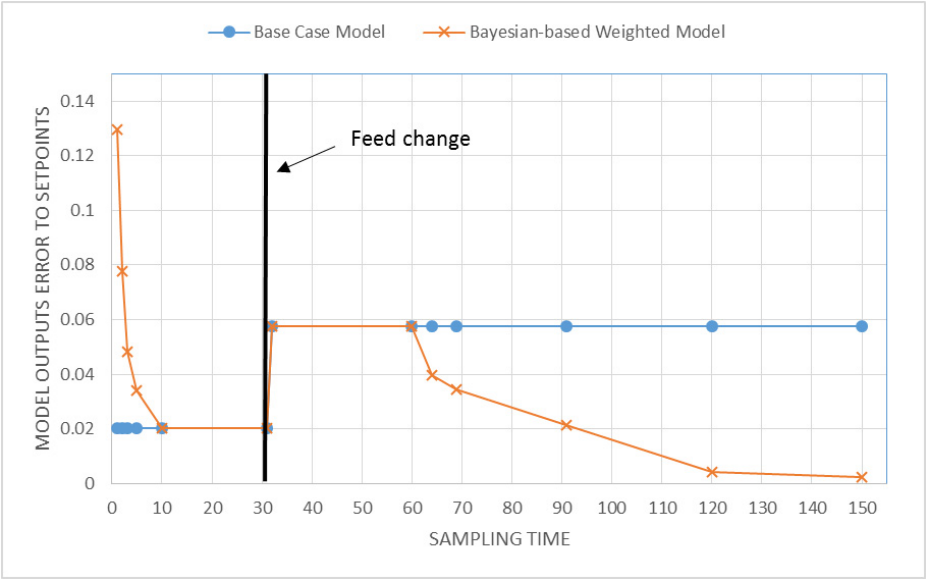


Figure 3. Model set-point error with feed change

4. Conclusions

In this paper, a Bayesian-based weight adjustment method is introduced in online applications. The periodic update of model-set weights by the model-set management procedure can be automated, which can ensure high-precision predictions, even if the process system undergoes disturbances and fluctuations. The steam reforming case study has demonstrated its attractiveness by shown a superior prediction performance under various process feeds. Also the presented method is generic and applicable for any model-based online applications.

References

S. Abbas, V. Dupont, T. Mahmud, 2017, Kinetics Study and Modelling of Steam Methane Reforming Process over a NiO/Al_2O_3 Catalyst in an Adiabatic Packed Bed Reactor, Journal of Hydrogen Energy, 42, 2889–2903.

K. Hou, R. Hughes, 2001, The Kinetics of Methane Steam Reforming over a Ni/α-Al_2O Catalyst, Chemical Engineering Journal, 82, 311–328.

J. Li, Y. Huang, 2006, Bayesian-based On-line Applicability Evaluation of Neural Network Models in Modeling Automotive Paint Spray Operations, Computers and Chemical Engineering, 30, 1392–1399.

J. Xu, G. Froment, 1989a, Methane Steam Reforming, Methanation and Water-Gas Shift: I. Intrinsic Kinetics, AIChE Journal, 35, 1, 88-96.

J. Xu, G. Froment, 1989b, Methane Steam Reforming: II. Diffusional Limitations and Reactor Simulation, AIChE Journal, 35, 1, 97-103.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50083-5

Dual-time scale based extending of the benchmark Tennessee Eastman process

Bo Chen[a], Feng-Wen Lan[a], Zhu Wang[a], Xiong-Lin Luo[a*]

[a]*Department of Automation, China University of Petroleum Beijing, 102249, China*
luoxl@cup.edu.cn

Abstract

The original Tennessee Eastman (TE) process which contains vapor-liquid phase is a single time scale simplified model. Dynamic equations associated with the vapor phase mass are usually ignored and treated as quasi-steady state, leading to the pressure completely depending on temperature. In this paper, to restore the true dynamic characteristics of the pressure, the inadequacy of single time scale and quasi-steady state model are firstly analyzed, and then a full dynamic modeling method for whole process of the TE process is proposed and full dynamic model is built. Finally, the simulation results of pressure and temperature between full dynamic model and quasi-steady state model only on the basis of mass balance control (pressure open-loop) are compared and analyzed to validate the effectiveness of dynamic modeling method.

Keywords: systems engineering; dual-time scale; Tennessee Eastman process

1. Introduction

Nowadays, more and more people are engaged in the study of Tennessee Eastman (TE) process, the TE process benchmark based on the FORTRAN language was proposed by Downs and Vogel (1993). It consists of several chemical devices of vapor-liquid coexistence, in these devices, pressure related to vapor phase and temperature related to liquid phase cannot be observed in the same time scale because of the large differences of vapor-liquid phase physical characteristics. The TE process is not only a multivariable large system with cycle stream, but also a typical dual-time scale system.

In order to make the TE process closer to the chemical engineering theory, since the day the TE process was put forward, the modeling and simulation are developing. The simple mechanism model only composed of mass balance equations was established (Ricker and Lee, 1995). Next, energy balance equations for reactor, flash tank and stripper were increased (Jockenhövel et al., 2003), but the equations of each device are still algebraic. Then, the TE process for study on distributed control strategy of large systems was identified on Matlab (Liu et al., 2013). In recent years, more and more experts put forward a variety of fault diagnosis methods (Bo et al., 2008; Chebel-Morello et al., 2016) and problems related to control (Antelo et al., 2008), and use the TE process as the test object.

However, above models are still single time scale simplified models, there is no independent dynamic change of pressure. Furthermore, mass balance control is essential in chemical process control, mass balance control is controlling the pressure and level, however, the pressure cannot be strictly controlled, and the mass balance control in the true sense cannot be achieved. To reflect the real dynamic characteristics of pressure, Luo et al. (2015) established the full order model of flash tank through increasing the

dynamic mass balance equation of pressure based on dynamic modeling principle of chemical process.

In this paper, for the original TE process, the inadequacy of single time scale and quasi-steady state model and the necessity of full dynamic modeling are firstly analyzed. Then, to make the TE process closer to chemical engineering theory, a full dynamic modeling method for the whole process of TE process is proposed and full dynamic model is built. Based on above analyses, finally, through simulation comparison to validate the effectiveness of dynamic modeling method.

2. Evaluation of the original TE process

According to the original model, dynamic equations associated with the vapor phase mass are ignored and treated as quasi-steady state. They are as follows.

The pressure equations of components are as Eq.(1). The temperature equation after unification of vapor-liquid phase is as Eq.(2).

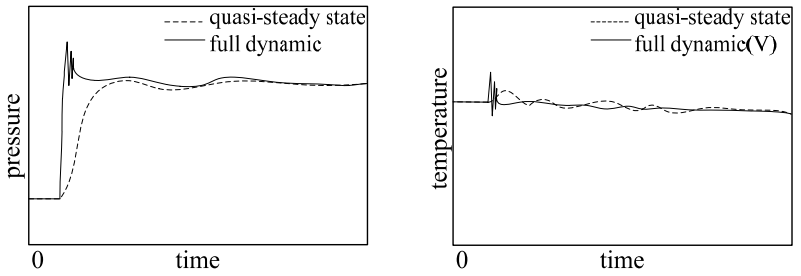
$$P_i = \frac{N_i \mathrm{R} T}{V^{\mathrm{V}}},\ i = \mathrm{A,B,C};\ P_i = x_i P_i^0,\ i = \mathrm{D,E,F,G,H};\ P = \sum_{\mathrm{i=A}}^{\mathrm{H}} P_i \tag{1}$$

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{F_{\mathrm{in}} h_{\mathrm{in}} - F_{\mathrm{vo}} h^{\mathrm{V}} - F_{\mathrm{lo}} h^{\mathrm{L}} + Q}{\rho^{\mathrm{L}} c_{\mathrm{p}} V^{\mathrm{L}}} \tag{2}$$

A-C – non-condensable components; D-H – condensable components; R – gas constant, V^V and V^L – vapor and liquid phase volume, m^3; P_i^0 – the saturated vapor pressure of component i, kPa; F_{in}, F_{vo}, F_{lo} – flow rate of inlet, vapor outlet and liquid outlet respectively, kmol·h^{-1}; h_{in}, h^V, h^L – enthalpy of inlet, vapor outlet and liquid outlet respectively, kJ·kmol^{-1}; Q – transfer heat, kJ·h^{-1}; ρ^L – liquid density, kmol·m^{-3}; c_p – liquid specific heat, kJ·kmol^{-1}·K^{-1}.

According to the above equations, the pressure is obtained by the temperature through the ideal gas equation of state directly in the modeling process, there is no independent dynamic change of pressure and the temperature within the whole device is represented by liquid temperature. When step signal is applied to the system ($\mathrm{d}T/\mathrm{d}t \neq 0$), the temperature is transferred from the original steady state value to a new steady state value, the value of each step in the transition is obtained by Eq.(2), and in every moment, the pressure is obtained by the temperature, as shown in Eq.(1). The transition process of temperature is over and pressure is end too until $\mathrm{d}T/\mathrm{d}t = 0$. In this case, process variables of vapor-liquid phase will change in the same time scale, as shown in Figure 1(a) and (b) by dotted line. As known to all, in the actual chemical engineering process, the changing rate of vapor pressure is much faster than that of liquid phase, if the difference between model and actual process is too large, then the real dynamic process will not be reflected and all simulation tests will lose their guiding significance to the actual process. The system analysis and synthesis results obtained by the simplified model are difficult to be applied to practice and this is true of the original TE process.

Therefore, restoring the dynamic characteristics of the TE process is especially important. Then, we assume that if the dynamic change of pressure and vapor temperature is restored, the pressure and vapor temperature are fast variables and liquid temperature is slow variables, at this moment, the dual-time scale characteristics are restored, as shown in Figure 1(a) and (b) by solid line.



(a) Dynamic response of pressure (b) Dynamic response of temperature

Figure 1. Dynamic response in quasi-steady state model and dynamic model

3. Fast dynamic pressure based full dynamic modeling

3.1. Principle of full dynamic modeling

In order to restore the true dynamic characteristics of vapor phase, this paper proposes a full dynamic modeling method. The full dynamic modeling method is shown in Figure 2, it represents the relationship between mass and energy variables after the vapor-liquid equilibrium. The vapor temperature is obtained from vapor phase mass balance equation and energy balance equation. The liquid temperature is obtained from the liquid phase mass balance equation and energy balance equation. The pressure is obtained from the vapor-liquid phase mass balance equation and vapor phase energy balance equation. The level is obtained from liquid phase mass balance equation, and the pressure will have an effect on the level, because the vapor-liquid phase coexists in the same device. The dynamic change of the pressure is associated with the change of vapor phase mass, vapor temperature and liquid level in full dynamic modeling. It shows that the pressure has the independent characteristic and is no longer completely dependent on the change of temperature.

3.2. Full dynamic modeling of operation units

The units of TE process are divided into four parts, namely, a mixing zone, a reactor, a stripper and a flash tank, wherein the flash tank portion includes a condenser, a separator and a circulating gas compressor. However, modeling schematic drawing is not fully in accordance with the TE process.

According to the proposed principle of full dynamic modeling, we take the reactor as an example to model and explain in detail. The modeling process of the other three parts is similar to that of reactor.

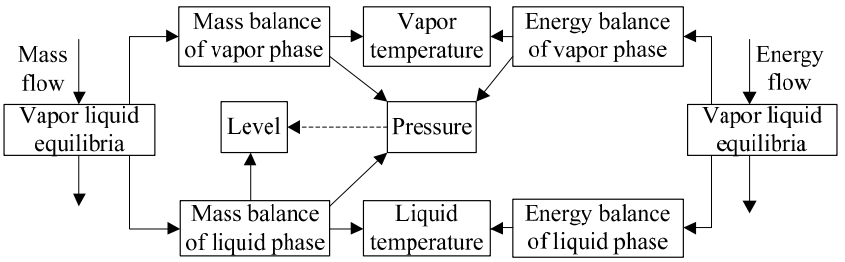
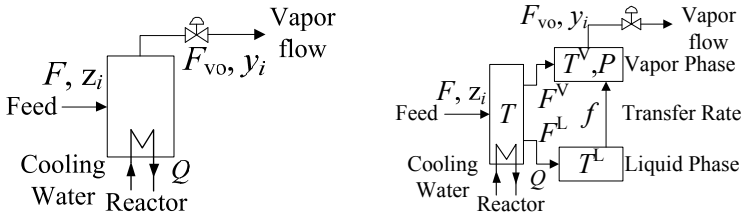


Figure 2. Schematic of full dynamic modeling



(a) Quasi-steady state modeling (b) Full dynamic modeling

Figure 3. Schematic of quasi-steady state and full dynamic modeling in reactor

F – feed flow rate; z_i – feed components; F_{vo} – vapor flow rate; y_i – vapor components; F^V, F^L – vapor-liquid flow rate of separation after reaction; f – transfer rate from liquid to vapor; P – pressure; T^V, T^L – vapor-liquid phase temperature; Q – heat exchange of cooling water.

The process of reactor with only vapor phase mass outlet: the vapor reactants are fed to the reactor and react. The products leave the reactor as vapors along with the unreacted feeds. The flow out of the vapor phase is determined by differential pressures and valve position of control valve. The reactor has an internal cooling bundle for removing the heat of the reaction. Quasi-steady state modeling in reactor is carried out according to the above, as shown in Figure 3(a). Full dynamic modeling is according to the schematic that transforms Figure 3(a) into Figure 3(b).

As shown in Figure 3(b), this paper considers the reactions are completing instantaneously and obtain vapor and liquid phase. Liquid flow comes into a virtual liquid tank, the liquid temperature and level are calculated in it. Vapor flow comes into a virtual vapor tank, the pressure and vapor temperature are calculated in it. The liquid phase enters the vapor phase in flash form. The reactor is satisfied the vapor-liquid equilibrium in whole process. And in whole process, in fact, there are three time scale, including reactions in reactor itself, liquid phase in virtual liquid tank and vapor phase in virtual vapor tank. Because the reactions are completing instantaneously, only considering the time scale of two virtual tanks.

After full dynamic modeling, the independent dynamic characteristic of pressure is restored, as shown in Eq.(3); and the temperature of vapor-liquid phase is calculated separately, as shown in Eq.(4).

$$V^{\mathrm{V}}\frac{\mathrm{d}P}{\mathrm{d}t}=\mathrm{R}T^{\mathrm{V}}\frac{\mathrm{d}N}{\mathrm{d}t}+N\mathrm{R}\frac{\mathrm{d}T^{\mathrm{V}}}{\mathrm{d}t}+P\mathrm{A}\frac{\mathrm{d}L}{\mathrm{d}t} \tag{3}$$

$$N\frac{\mathrm{d}T^{\mathrm{V}}}{\mathrm{d}t}=\frac{C_{\mathrm{p}}\left(F^{\mathrm{V}}+f\right)T-C_{\mathrm{pv}}F_7T^{\mathrm{V}}}{C_{\mathrm{pv}}}-T^{\mathrm{V}}\frac{\mathrm{d}N}{\mathrm{d}t};\ L\frac{\mathrm{d}T^{\mathrm{L}}}{\mathrm{d}t}=\frac{C_{\mathrm{p}}F^{\mathrm{L}}T-C_{\mathrm{pl}}fT^{\mathrm{L}}}{C_{\mathrm{pl}}\rho^{\mathrm{L}}\mathrm{A}}-T^{\mathrm{L}}\frac{\mathrm{d}L}{\mathrm{d}t} \tag{4}$$

V^V – vapor phase volume, m^3; P – pressure, kPa; R – gas constant, $J\cdot mol^{-1}\cdot K^{-1}$; T^V and T^L – vapor and liquid phase temperature, °C; N – amount of vapor phase substance, kmol; L – liquid level, m; A – cross-sectional area, m^2; C_p, C_{pv} and C_{pl} – molar heat capacity of mixed flow after reaction, vapor outlet flow and transfer flow from liquid phase to vapor phase respectively, $kJ\cdot kmol^{-1}\cdot K^{-1}$; F^V, F^L, F_{lv} and F_7 – vapor flow rate of separation, liquid flow rate of separation, the transfer flow rate from liquid phase to vapor phase and flow rate of outlet respectively, $kmol\cdot h^{-1}$.

Through the comparison between Eq.(1) and Eq.(3), Eq.(2) and Eq.(4), the main difference between the quasi-steady state model and the full dynamic model is whether

the pressure equation is dynamic or not, the pressure is not completely depending on temperature and it has an independent dynamic process. The other difference is temperature equation, the vapor and liquid temperature are calculated separately, and they both have their dynamic process.

Finally, full dynamic model of the whole process is connecting each unit device according to process to form a complete chemical engineering process. The connection between units is mainly through the pipeline, valves and pumps.

4. Simulation comparison

After full dynamic modeling, to verify our conjecture in section 1, the quasi-steady state model and the full dynamic model on the basis of mass balance control (level closed-loop and pressure open-loop) are simulated by gPROMS software, and the dynamic characteristics of the pressure and temperature are observed and compared.

4.1. Pressure

When the feed stream 1 changes, the simulation results of pressure are shown in Figure 4, it shows that when the independent dynamic change of pressure is restored, the dynamic characteristic of pressure under full dynamic model presents the fast change, and then the performance is consistent with that under quasi-steady state after the fast change is over. The dynamic response time of all variables will be influenced due to the control, but it will not change the dynamic characteristic of each variable.

4.2. Temperature

As shown in Figure 5, compared with quasi-steady state model, full dynamic model makes vapor and liquid phase temperature calculated separately, leading to two temperature curves. The change of vapor phase temperature is similar to that of pressure under full dynamic model; the change of vapor phase temperature is much faster than that of liquid phase temperature after applying step signal to the system, after that, the variation trend is consistent with that under quasi-steady state.

The section 4 indicates that full dynamic modeling method restores real dynamic characteristics of pressure and vapor phase temperature, the change of pressure is no longer completely dependent on temperature, vapor and liquid phase temperature also have their respective trends. Variables of vapor phase change fast and variables of liquid phase change slowly, they change in different time scale, namely, the dual-time scale characteristic of TE process is restored. It is consistent with our conjecture in section 1.

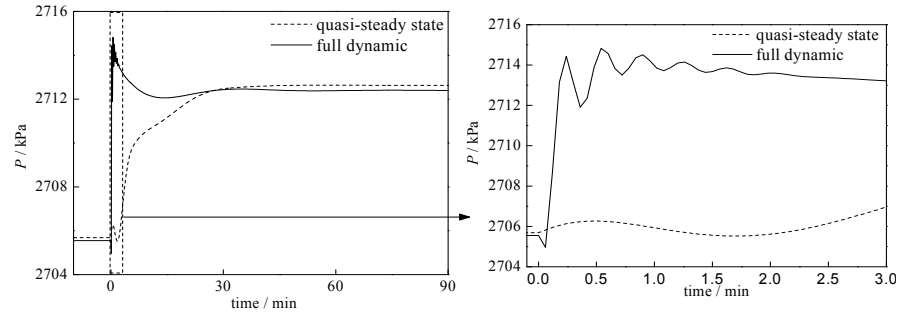


Figure 4. Dynamic response of reactor in quasi-steady state and full dynamic model

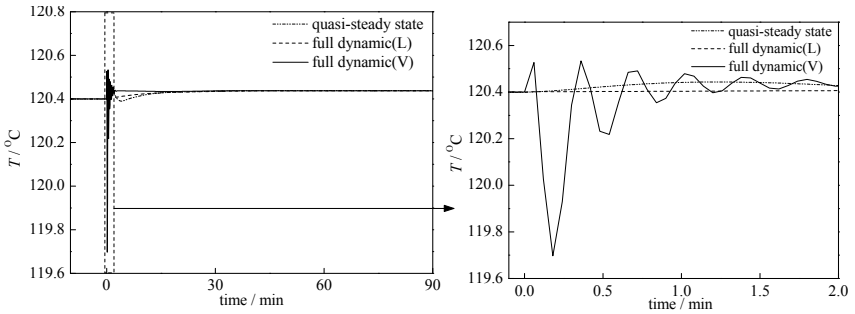


Figure 5. Dynamic response of reactor in quasi-steady state and full dynamic model

5. Conclusions

This paper aims at proposing a full dynamic modeling method and restoring the independent dynamic characteristic of pressure. Under quasi-steady state model, the pressure is completely controlled by temperature, and vapor and liquid phase temperature are unified, it is impossible to reflect the real dynamic process of vapor phase. The comparison of equations and simulation results indicate that full dynamic model can reflect the real dynamic characteristics of pressure and vapor phase temperature. Full dynamic model is closer to the chemical engineering theory and it will improve the quality of control algorithms and have guiding significance in optimization, nonlinear control, predictive control etc.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (21676295).

References

L. T. Antelo, J. R. Banga, and A. A. Alonso, 2008, Hierarchical design of decentralized control structures for the Tennessee Eastman Process, Comput. Chem. Eng., 32, 9, 1995-2015.

C. M. Bo, S. Zhang, G. M. Zhang, and Z. Zhang, 2008, Fault identification of Tennessee Eastman process based on FS-KPCA[J], CIESC Journal, 59, 7, 1783-1789.

B. Chebel-Morello, S. Malinowski, and H. Senoussi, 2016, Feature selection for fault detection systems: Application to the Tennessee Eastman Process, Applied Intelligence, 44, 1, 189-194.

J. J. Downs and E. F. Vogel, 1993, A plant-wide industrial process control problem, Comput. Chem. Eng., 17, 3, 245-255.

T. Jockenhövel, L. T. Biegler, A. Wachter, 2003, Dynamic optimization of the Tennessee Eastman process using the OptControlCentre, Comput. Chem. Eng., 27, 11, 1513-1531.

Y. B. Liu, X. L. Luo, F. Xu, 2013, Global coordination and stability analysis for distributed model predictive control system[J], CIESC Journal, 64, 4, 1318-1331.

X. L. Luo, H. Zhao, F. Xu, 2015, Two-time scale modeling and dynamic characteristics analysis of flash tank in TE process[J], CIESC Journal, 66, 1, 186-196.

N. L. Ricker and J. H. Lee, 1995, Nonlinear modeling and state estimation for the Tennessee Eastman challenge process, Comput. Chem. Eng., 19, 9, 983-1005.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50084-7

A Bilevel Programming Approach for the Dynamic Optimization of Cyanobacterial C-phycocyanin Production under Uncertainty

Dongda Zhang[a], Ehecatl A. del Rio-Chanona[a*]

[a]*Centre for Process Systems Engineering, Imperial College London, SW7 2AZ, UK*
a.del-rio-chanona@imperial.ac.uk

Abstract

C-phycocyanin is a high-value bioproduct synthesized by cyanobacterium *Arthrospira platensis* with a significant global market demand given its applications in the pharmaceutical, food and colorant industries. Unfortunately, its biosynthesis is currently characterized by low productivity and large uncertainty during the production process. High variability and unreliable expectations on product yields substantially hinder the industrialization of microorganism derived biochemicals as they present a risk to the profitability and safety of the underlying systems. Therefore, in this work, we propose a robust optimization approach to determine the lower and upper product yield expectations for the sustainable production of C-phycocyanin. Kinetic modeling is adopted in this study as a tool for fast prototyping, prediction and optimization of chemical and biochemical processes. On the upside, parameters in bioprocess kinetic models are used as a simplification of the complex metabolic networks to enable the simulation, design and control of the process. On the downside, this conglomeration of parameters may result in significant model uncertainty. To address this challenge, we formulate a bilevel max-min optimization problem to obtain the worst-case scenario of our system given the uncertainty on the model parameters. By constructing parameter confidence ellipsoids, we determined the feasible region along which the parameters can minimize the system's performance, while nutrient and light controls are used to maximize the biorenewable production. The inner minimization problem is embedded by means of the optimality conditions into the upper maximization problem and hence both are solved simultaneously. Through this approach, we determined pessimistic and optimistic scenarios for the bioproduction of C-phycocyanin and hence compute reliable expectations on the yield and profit of the process.

Keywords: uncertain dynamic systems, optimal control, bilevel programming, bioprocess optimization.

1. Introduction

C-phycocyanin is a blue antenna pigment used to enhance the photosynthetic efficiency of cyanobacteria and red algae (Eriksen 2008). It has been recognized as a high-value bioproduct with great potential in the pharmaceutical industry (Chen et al. 2013; Kuddus et al. 2013). Currently, extensive research has been conducted to identify the metabolic mechanisms of C-phycocyanin synthesis in cyanobacterium *A. platensis*. It has been demonstrated that nitrate concentration and illumination intensity are the most important factors determining both the content and the productivity of phycocyanin (Xie et al. 2015; del Rio-Chanona et al. 2015). Meanwhile, to accomplish the

industrialization of phycocyanin production, two requirements must be satisfied. First, it is necessary to identify the optimal operating conditions for phycocyanin synthesis and biomass growth so that the process efficiency can be maximized. Second, a final phycocyanin content higher than 10 % of cell dry weight must be guaranteed; otherwise the cost of the pigment downstream separation will be significantly increased.

To address the two requirements, mathematical models have become an indispensable tool to determine the process optimal operating conditions for long-term biomass cultivation and phycocyanin production. Recently, a kinetic model capable of simulating cyanobacterial biomass growth and phycocyanin production under different light intensities and nitrate concentrations has been proposed (del Rio-Chanona et al. 2015). However, as a kinetic model is a simplification of a highly complex metabolic network which involves a significant amount of metabolic reactions, the uncertainty of a kinetic model is in general large and the process optimization results heavily rely on the accuracy of the model. As a result, it is essential to consider the effects of model uncertainty on the process prediction and optimal control. Therefore, in this study a bilevel programming framework is designed to optimize C-phycocyanin production with uncertainty under a fed-batch long-term operation system.

2. Cyanobacterial C-phycocyanin production system

The *A. platensis* biomass growth and C-phycocyanin production model in (del Rio-Chanona et al. 2015) is adopted in the current study, and the reader is directed to the original source for details on the model. The dynamic equation system is:

$$\frac{dx}{dt} = r_M x - u_d x \quad (1), \quad \frac{dn}{dt} = -Y_{NO} r_M x + F_N \cdot N_{in} \quad (2), \quad \frac{dq}{dt} = kx - \frac{k_d q}{n + K_{Np}} \quad (3)$$

$$r_M = u_0 \frac{n}{n + K_n} \quad (4), \quad I(z) = I_0\left[\exp\left(-\left(\varpi x + K_a\right)z\right) + \exp\left(-\left(\varpi x + K_a\right)\left(L - z\right)\right)\right] \quad (5)$$

$$u_0 = u_m \frac{I(z)}{I(z) + k_s + I(z)^2 / k_i} \quad (6), \quad k = k_m \frac{I(z)}{I(z) + k_{sp} + I(z)^2 / k_{ip}} \quad (7)$$

Given the fact that nitrate inflow rate (F_N) is a control in this study, while the model was originally designed to simulate a batch process, a new term, $F_N \cdot N_{in}$ is added on the right-hand-side of Eq. (2) so that the model can be used to simulate the fixed-volume fed-batch process. The influent nitrate concentration (N_{in}) is assumed to be 500 mM. To maximize the production of C-phycocyanin, a 16-day fed-batch process was selected as the production mode for operation, and dynamic equations were modified accordingly. This would allow to treat the nitrate inlet flow and the light intensity as control variables, and hence we can formulate an optimal control problem. To maintain this scenario as close to reality as possible, controls were only allowed to vary once per day during a 16 day computational experiment.

3. Parameter estimation and confidence ellipsoids

3.1. Parameter estimation

Here we outline the procedure to estimate the parameters in Eq. (8) for Eqs. (1) - (7) using the data provided in the original article. To estimate these parameters, a nonlinear

least-squares optimization problem is formulated as Eq. (9). The output variables once the DAE system is solved are labelled y, while the experimental data is labelled $\hat{y}$, p_{lb} and p_{ub} are the upper and lower bounds of parameters, and Λ is the inverse of the covariance matrix estimated by the experimental measurement error. As the parameter estimation problem is nonconvex, and to avoid shallow local minima, a multi-start framework was implemented. One hundred initial starting points were generated by a Sobol sequence and the estimated parameters that produced the least objective function was chosen.

$$p = (u_m, K_N, u_d, Y_{NO_3}, k_d, K_{N_p}, k_m, k_s, k_i, k_{sp}, k_{ip}, \tau, K_a) \tag{8}$$

$$\min_p \sum_{i=1}^{N} \left(\hat{y}_i - y(t_i, p)\right)^T \Lambda_i \left(\hat{y}_i - y(t_i, p)\right)$$

$$\begin{aligned} s.t. \quad & \frac{dx}{dt} = f(x(t), p) \\ & p_{lb} \le p \le p_{lb}, \quad x(0) = x_0, \quad t_0 \le t \le t_f \end{aligned} \tag{9}$$

3.2 Confidence Ellipsoids

Joint confidence regions were determined by considering all simultaneous linear combinations of the parameters, computed by Eq. (10) where p^* is the optimal parameter solution determined by Eq. (9), and the right-hand side is the standard value for the Chi-square test given the number of parameters and a 95 % level of confidence.

$$\left(p - p^*\right)^T \Lambda \left(p - p^*\right) \le \chi^2_{1-\alpha, NP} \tag{10}$$

4. Optimization of process production

An optimal control problem was formulated to optimize the production of C-phycocyanin in a fed-batch operation mode. To maximize the process production, two control variables were used: light intensity and nitrogen inflow rate. The resulting dynamic optimization problem is the following:

$$\max_{Fin(t), L(t)} q(t_f)$$

$$\begin{aligned} s.t. \quad & \frac{dx}{dt} = f(x(t)) \\ & x(0) = x_0, \quad t_0 \le t \le t_f \\ & Fin^{\min} \le Fin(t) \le Fin^{\max}, \quad L^{\min} \le L(t) \le L^{\max} \end{aligned} \tag{11}$$

The solution of Eq. (11) would result in the optimal control sequences for *Fin*(*t*) and *L*(*t*). Problem (11) is discretized through orthogonal collocation over finite elements, where the collocation points are placed according to a fifth order Radau quadrature.

5. Worst-case scenario optimization

5.1. Bilevel programming approach

A bilevel max-min approach enables us to formulate a problem that would yield the worst-case scenario of the dynamic optimization in the system. In this case, given that we have no strict bound or inequality constraints that our system is likely to violate, the

worst-case is the scenario with the lowest product production even when the system is optimized. For this, we maximize the production of C-phycocyanin by using light and nitrate inlet as control parameters. Furthermore, we assume the system will select the worst possible set of parameters that are within the 95 % confidence intervals computed by (3). This bilevel optimization problem is the following:

$$
\begin{aligned}
&\max_{Fin(t),L(t)} \; q(t_f) \\
&s.t. \quad \frac{dx}{dt} = f(x(t)) \\
&\quad Fin^{\min} \le Fin(t) \le Fin^{\max}, \quad L^{\min} \le L(t) \le L^{\max} \\
&\quad x(0) = x_0, \quad t_0 \le t \le t_f \\
&\quad \min_{p} \; q(t_f) \\
&\quad s.t. \quad (p - p^*)^T \Lambda (p - p^*) \le \chi^2_{1-\alpha,NP}
\end{aligned}
\tag{12}
$$

Problem (12) can be reformulated by embedding the necessary and sufficient conditions for optimality of the inner minimization problem into the outer optimization problem. This formulation of a bilevel problem is referred as mathematical programs with complementarity constraints (MPCC). MPCC have a less complicated structure than the original Bilevel problems, particularly, the feasible sets are always closed (Allende & Still 2013). This reformulation yields the following single-level problem (13), where μ is the Lagrange multiplier for the constraint on the parameters feasible region, delimited by the confidence ellipsoid $G = (p - p^*)^T \Lambda (p - p^*) - \chi^2_{1-\alpha,NP}$ and $F = q(t_f)$.

$$
\begin{aligned}
&\max_{Fin(t),L(t)} \; q(t_f) \\
&s.t. \quad \frac{dx}{dt} = f(x(t)) \\
&\quad Fin^{\min} \le Fin(t) \le Fin^{\max}, \quad L^{\min} \le L(t) \le L^{\max} \\
&\quad x(0) = x_0, \quad t_0 \le t \le t_f \\
&\quad \nabla_p L(p^*) = \nabla_p F(p^*) - \mu \nabla_p G(p^*) = 0, \quad \mu G(p^*) = 0 \\
&\quad \mu \ge 0, \quad \nabla^2_{pp} L(p^*) \succeq 0
\end{aligned}
\tag{13}
$$

When embedding the lower bilevel problem into the upper problem by the conditions of optimality, it is advisable to use the Fritz-John (FJ) necessary conditions (Allende & Still 2013). However, as we do not have equality constraints for the lower problem, the Karush-Kuhn-Tucker (KKT) and the FJ necessary conditions are the same. Moreover, we consider the smoothing approach (Allende & Still 2013) where we replace the complementary equation $\mu G(p^*) = 0$ by $\mu G(p^*) = \varepsilon$, $\varepsilon > 0$ a small perturbation parameter. Finally, let us note that Eq. (3) is only twice differentiable for parameters k_{sp}, k_{ip}, K_a, T, and K_{Np}, so they are the only parameters considered as optimization variables.

6. Implementation

Optimization problems in this work were implemented in a Python programming environment, using Pyomo as an interface for the optimization solver IPOPT (Wächter & Biegler 2006). The value of ε vas set to be 10^{-6}. Furthermore, solving Eq. (13)

directly was found difficult due to inadequate starting points for the optimization problem. Thus, Eq. (11) was first solved to obtain preliminary initialization points. To ameliorate convergence issues, expressions which contain bilinear and trilinear terms were substituted by their linearized form (Adjiman et al. 1998). Subsequently, linearized expressions of (14) were introduced one-at-a-time for each parameter to update the initial guesses for the optimizer. Finally, (14) and (15) were re-introduced in the overall problem, and an optimal solution to (13) was attained.

$$\nabla_p L(p^*) = \nabla_p F(p^*) + \mu \nabla_p G(p^*) = 0 \tag{14}$$

$$\mu G(p^*) = \varepsilon \tag{15}$$

7. Results and discussion

To obtain an optimistic scenario of the process optimization, Eq. (11) is solved, however, parameters in Eq. (8) were also assigned as optimization variables, hence helping the maximization of the bioproduct. On the other hand, solutions obtained by Eq. (13) represent the worst-case scenario for the current process.

Table 1: Model parameters and optimization results.

Parameter	Pessimistic	Optimistic			Pessimistic	Optimistic
k_{sp}	37.328	11.323		Biomass	7.5 g/L	4.8 g/L
k_{ip}	197.78	800.0		conc., g/L		
K_a	3.799	0.0		Phycocyanin	533.9	773.2
τ	73.381	48.382		production,	mg/L	mg/L
K_{NP}	18.372	16.892		mg/L		

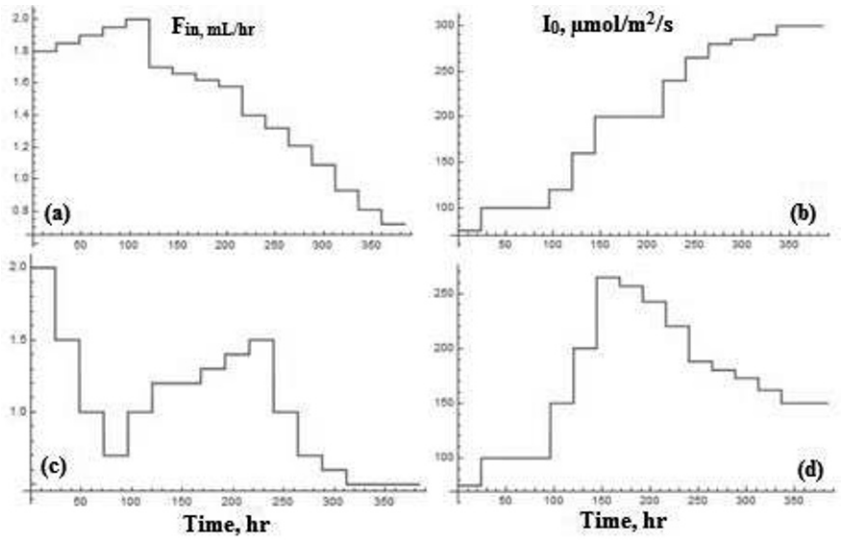


Figure 1: Optimal control actions under different scenarios. (a) and (b), (c) and (d): nitrate inflow rate and light intensity in the optimistic and pessimistic case, respectively.

The parameter values for both scenarios are listed in Table 1. By comparing the different results in Table 1, there is a significant change in the potential production on the process. This highlights the fact that an incorrect estimation of the best and worst production yields can severely endanger the safety and economics of the process. Furthermore, it is also worth noticing that there is a significant difference between the optimal control inputs for light and nutrient supply (Fig. 1), depending on the dynamic model parameters chosen. This means, that as expected, different parameter sets will result in different optimal control inputs to maximize the bioproduct.

In Table 1 we can see that k_{sp}, which is the light saturation term, is higher in the pessimistic scenario than it in the optimistic scenario. A system with a higher light saturation term is not able to use light to produce phycocyanin as efficiently as one that has a lower light saturation term. Moreover, the light inhibition term (k_{ip}) is higher in the optimistic scenario, which means that when there is a high light intensity, light will not inhibit the production of phycocyanin as much as that in a system with a lower value of k_{ip}, such as the pessimistic scenario. K_a and τ regulate how much the culture scatters and absorbs light. In both cases the pessimistic scenario has a higher value, which diminished the amount of light that can be utilized for phycocyanin production. Furthermore, K_{Np} can be seen to have practically the same value in both cases, reflecting its low sensitivity to phycocyanin production, and that other parameters can inhibit the systems performance more severely. Thus, it is important to identify the correct value of other parameters through model-based experimental design method in future study.

8. Conclusions

In this study, a bilevel max-min optimization framework was proposed to resolve the challenge on cyanobacterial C-phycocyanin production under uncertainty. Through this framework, phycocyanin production was predicted, and upper and lower bounds on productivity identified. Furthermore, given the wide confidence intervals, more experimental measurements are needed to narrow the varying range of the controls. This also highlights the need for an online identification and optimization strategy to be put in place (*e.g.* MPC, EMPC) so that the process optimization can be performed effectively.

References

C.S. Adjiman, et al., 1998. A global optimization method, αBB, for general twice-differentiable constrained NLPs — I. Theoretical advances. *Comput. Chem. Eng.*, 22(9), 1137–1158.

G.B. Allende and G. Still, 2013. Solving bilevel programs with the KKT-approach. *Mathematical Programming*, 138(1–2), pp.309–332.

C. Chen, et al., 2013. Engineering strategies for simultaneous enhancement of C-phycocyanin production and CO_2 fixation with Spirulina platensis. *Bioresour. Technol.*, 145, 307–312.

N.T. Eriksen, 2008. Production of phycocyanin--a pigment with applications in biology, biotechnology, foods and medicine. *Appl. Microbiol. Biotechnol.*, 80(1), 1–14.

M. Kuddus, et al., 2013. Recent developments in production and biotechnological applications of C-phycocyanin. *BioMed research international*, 2013, p.742859.

E.A. del Rio-Chanona, et al., 2015. Dynamic Simulation and Optimization for Arthrospira platensis Growth and C-Phycocyanin Production. *Ind. Eng. Chem. Res.*, 54(43), 10606–10614.

A. Wächter and L.T. Biegler, , 2006. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1), 25–57.

Y. Xie, et al., 2015. Fed-batch strategy for enhancing cell growth and C-phycocyanin production of Arthrospira (Spirulina) platensis under phototrophic cultivation. *Bioresour. Technol.*, 180, 281–287.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50085-9

Novel design of dynamic matrix control with enhanced decoupling control performance

Jyh-Cheng Jeng[a*], Yun-Ju Chang[a], Ming-Wei Lee[b]

[a]*National Taipei University of Technology, Taipei 10608, Taiwan*
[b]*China Steel Corporation, Kaohsiung, Taiwan*
jcjeng@ntut.edu.tw

Abstract

The decoupling control scheme is beneficial for controlling multivariable systems with strong interactions. This paper presents a novel control design approach where the concept of decoupling control is incorporated into the model predictive control design. We developed the decoupling dynamic matrix control that enables achieving effective decoupling control without adding any decouplers. Simulation study confirms the effectiveness of the proposed control design.

Keywords: model predictive control, dynamic matrix control, decoupling control.

1. Introduction

Most industrial and chemical processes are multivariable systems which often present undesirable couplings (interactions) between the controlled and manipulated variables. A multivariable system with strong interactions can be much more difficult to control. In this case, the decoupling control by using a decoupler is an effective control strategy to address the process interactions. However, the ideal decouplers are not always physically realizable, and the main problem of this decoupling approach is that the complexity of decoupler elements and apparent decoupled processes increases for high-dimensional multivariable systems (Garrido et al., 2012). Model predictive control (MPC) has become the standard method in several process industries to solve difficult multivariable control problems (Qin and Badgwell, 2003). MPC is a kind of multivariable control that can take process interactions into account. However, unlike the typical decoupler design, the conventional MPC design does not target explicitly on the ideal compensation of process interactions so that satisfactory decoupling control performance is not necessarily attained.

In this study, we propose a novel MPC design that enables achieving the condition of decoupling control without adding any decouplers. We term this control technique as decoupling dynamic matrix control (DDMC) as the step-response models are used for prediction. This novel design method conceptually decomposes each control signal into several constituent components for DDMC calculation. One of the components is used to drive a controlled variable to its reference signal, and the other components are used to eliminate the effects from other manipulated variables on this controlled variable. An additional term is then included in the DDMC cost function to penalize the incomplete compensation of process interactions. By tuning the constant weighting matrices, it is possible to deal with the trade-off among decoupling performance, tracking performance, and control efforts. Simulation study shows that the proposed DDMC design provides considerably improved decoupling control performance compared with the conventional DMC.

2. Dynamic matrix control

Dynamic matrix control (DMC) has achieved a great deal of success in the petroleum refining and petrochemicals industries. In this section, the conventional DMC algorithm, which uses the step-response model for output prediction, is briefly introduced.

Consider a general multivariable process involving r inputs, $\mathbf{u} = [u_1 \ u_2 \ ... \ u_r]^T$, and n outputs, $\mathbf{y} = [y_1 \ y_2 \ ... \ y_n]^T$. Let $\{s_{ij,1} \ s_{ij,2} \ ... s_{ij,N}\}$ be the N step-response coefficients for the model relating y_i and u_j. Assume that the initial condition is zero for simplicity. The p-step ahead prediction for the output y_i made at time instant t can be written as

$$\hat{y}_{i,t+p} = \sum_{j=1}^{r}\left(\sum_{h=1}^{N-1} s_{ij,h}\Delta u_{j,t+p-h} + s_{ij,N}u_{j,t+p-N}\right) = \sum_{j=1}^{r}\left(\sum_{h=1}^{p} s_{ij,h}\Delta u_{j,t+p-h}\right) + \hat{y}^{o}_{i,t+p} \tag{1}$$

where $\Delta u_{j,t} = u_{j,t} - u_{j,t-1}$ denotes the control move at time instant t, and $\hat{y}^{o}_{i,t+p}$ denotes the predicted unforced response obtained when $\Delta u_{j,t+h} = 0$ for $h \geq 0$.

Define vectors of predicted responses and predicted unforced response for the next P sampling instants, and control moves for the next M sampling instants as

$$\hat{\mathbf{Y}}_{t+1} = \begin{bmatrix} \hat{\mathbf{y}}_{t+1} \\ \hat{\mathbf{y}}_{t+2} \\ \vdots \\ \hat{\mathbf{y}}_{t+P} \end{bmatrix}; \qquad \hat{\mathbf{Y}}^{o}_{t+1} = \begin{bmatrix} \hat{\mathbf{y}}^{o}_{t+1} \\ \hat{\mathbf{y}}^{o}_{t+2} \\ \vdots \\ \hat{\mathbf{y}}^{o}_{t+P} \end{bmatrix}; \qquad \Delta\mathbf{U}_t = \begin{bmatrix} \Delta\mathbf{u}_t \\ \Delta\mathbf{u}_{t+1} \\ \vdots \\ \Delta\mathbf{u}_{t+M-1} \end{bmatrix} \tag{2}$$

where P is the prediction horizon and M is the control horizon, and we usually require $N \geq P \geq M$. The future output predictions of the model are written as

$$\hat{\mathbf{Y}}_{t+1} = \mathbf{S}\Delta\mathbf{U}_t + \hat{\mathbf{Y}}^{o}_{t+1} \tag{3}$$

where $\mathbf{S}$ is the dynamic matrix:

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{S}_2 & \mathbf{S}_1 & \mathbf{0} & \vdots \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{S}_M & \mathbf{S}_{M-1} & \cdots & \mathbf{S}_1 \\ \mathbf{S}_{M+1} & \mathbf{S}_M & \cdots & \mathbf{S}_2 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_P & \mathbf{S}_{P-1} & \cdots & \mathbf{S}_{P-M+1} \end{bmatrix}_{nP\times rM}; \qquad \mathbf{S}_h = \begin{bmatrix} s_{11,h} & s_{12,h} & \cdots & s_{1r,h} \\ s_{21,h} & s_{22,h} & \cdots & s_{2r,h} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1,h} & S_{n2,h} & \cdots & s_{nr,h} \end{bmatrix}_{n\times r} \tag{4}$$

Because there are uncertainties that cause output prediction error, we need to correct the future output predictions through output feedback. A typical approach is to add a bias correction to the predictions, which provides a vector of corrected predictions as

$$\tilde{\mathbf{Y}}_{t+1} = \hat{\mathbf{Y}}_{t+1} + \mathbf{\Phi}\left(\mathbf{y}_t - \hat{\mathbf{y}}_t\right) = \mathbf{S}\Delta\mathbf{U}_t + \hat{\mathbf{Y}}^{o}_{t+1} + \mathbf{\Phi}\left(\mathbf{y}_t - \hat{\mathbf{y}}_t\right) = \mathbf{S}\Delta\mathbf{U}_t + \tilde{\mathbf{Y}}^{o}_{t+1} \tag{5}$$

where $\tilde{\mathbf{Y}}^{o}_{t+1}$ is the corrected prediction for the unforced case, and $\mathbf{\Phi}$ is defined as

$$\mathbf{\Phi} = \begin{bmatrix} \mathbf{I}_n & \mathbf{I}_n & \cdots & \mathbf{I}_n \end{bmatrix}^T_{nP\times n}; \quad \mathbf{I}_n : n\times n \text{ identity matrix} \tag{6}$$

The general objective of the DMC control calculation is to determine $\Delta\mathbf{U}_t$ so that the following quadratic cost function is minimized:

$$\min_{\Delta \mathbf{U}_t} J_{\mathrm{DMC}} = \left(\mathbf{Y}_{t+1}^r - \tilde{\mathbf{Y}}_{t+1} \right)^T \mathbf{Q} \left(\mathbf{Y}_{t+1}^r - \tilde{\mathbf{Y}}_{t+1} \right) + \Delta \mathbf{U}_t^T \mathbf{R} \, \Delta \mathbf{U}_t \tag{7}$$

where $\mathbf{Y}_{t+1}^r$ is the reference trajectory over the prediction horizon, **Q** is the weighting matrix for predicted errors, and **R** is the weighting matrix for control moves. The DMC control law that minimizes the cost function in Eq.(7) can be calculated analytically. In the conventional DMC algorithm, the objective function is based on minimizing the predicted tracking errors and the control moves. However, the decoupling control performance is not *explicitly* considered in DMC design so that the conventional DMC does not necessarily provide an effective decoupling control. In the following, a novel DMC algorithm that enables the control design to achieve enhanced decoupling control performance is presented.

3. Decoupling design of dynamic matrix control

3.1. Condition for decoupling control

For a $n \times r$ multivariable process **G**(*s*), the typical decoupling control scheme is depicted in Figure 1, where **D**(*s*) is the decoupler and **C**(*s*) is the decentralized controller. To compensate the interactions among loops, the decoupler **D**(*s*) is designed such that the decoupled process matrix (**GD**) is a diagonal matrix. Namely, the process output y_i depends only on the controller output v_i, and the other controller outputs v_j ($j \neq i$) have no effect on y_i. In this case of ideal decoupling, a set-point change for one controlled variable has no effect on the other controlled variables, which is an important benefit of the decoupling control. In Figure 1, each decoupler output (process input) u_j comprises n components (u_{jk}, $k = 1, \cdots, n$) and is given by

$$u_j = \sum_{k=1}^{n} u_{jk} = \sum_{k=1}^{n} d_{jk} v_k \; ; \quad j = 1, 2, \cdots, r \tag{8}$$

where u_{jk} depends only on v_k. The process output can be written as

$$y_i = \sum_{j=1}^{r} g_{ij} u_j = \sum_{j=1}^{r} g_{ij} u_{ji} + \sum_{j=1}^{r} \left(g_{ij} \sum_{k=1, k \neq i}^{n} u_{jk} \right); \quad i = 1, 2, \cdots, n \tag{9}$$

As mentioned, the ideal decoupling is achieved when the process output y_i depends only on the controller output v_i, or equivalently depends only on u_{ji}. Therefore, according to Eq.(9), the condition for ideal decoupling is given by

$$\sum_{j=1}^{r} \left(g_{ij} \, u_{jk} \right) = 0; \quad \text{for all } \; k = 1, \cdots, i-1, i+1, \cdots, n \; \text{ and } \; i = 1, 2, \cdots, n \tag{10}$$

Development of the DDMC algorithm is based on the decoupling condition in Eq.(10).

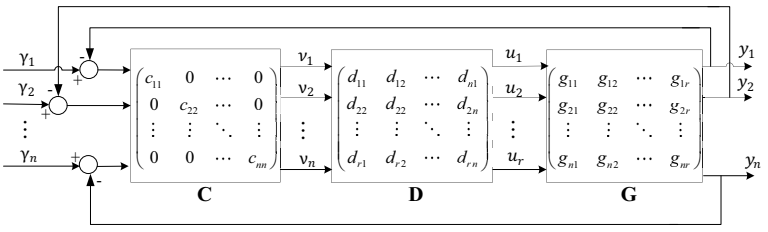


Figure 1. Decoupling control system for a $n \times r$ multivariable process.

3.2. Design of decoupling dynamic matrix control (DDMC)

Let $y_{ijk} = g_{ij}\, u_{jk}$ be the constituent components of y_i and $\hat{y}_{ijk,t+p}$ be the p-step ahead prediction for y_{ijk}. To achieve effective decoupling control, a new cost function is proposed for DMC control calculation as follows:

$$J_{\text{DDMC}} = \left(\mathbf{Y}^r_{t+1} - \tilde{\mathbf{Y}}_{t+1}\right)^T \mathbf{Q}\left(\mathbf{Y}^r_{t+1} - \tilde{\mathbf{Y}}_{t+1}\right) + \Delta\mathbf{U}_t{}^T \mathbf{R}\, \Delta\mathbf{U}_t + \sum_{i=1}^{n}\left[\sum_{k=1,k\neq i}^{n}\left[\left(\sum_{j=1}^{r}\hat{\mathbf{Y}}_{ijk,t+1}\right)^T \mathbf{W}_{ik}\left(\sum_{j=1}^{r}\hat{\mathbf{Y}}_{ijk,t+1}\right)\right]\right] \tag{11}$$

where $\hat{\mathbf{Y}}_{ijk,t+1} = [\hat{y}_{ijk,t+1}\ \ \hat{y}_{ijk,t+2}\ \cdots\ \hat{y}_{ijk,t+P_d}]^T$. The third term in J_{DDMC} is an additional penalty for the incomplete compensation of process interactions (the deviation of condition in Eq.(10)). Here, P_d is the decoupling horizon and $\mathbf{W}_{ik}$ is the weighting matrix for decoupling performance.

To calculate the predicted output component $\hat{y}_{ijk,t+p}$, each manipulated variable and control move is decomposed into n components in a similar manner given in Eq.(8):

$$u_j = \sum_{k=1}^{n} u_{jk}\ ; \quad \Delta u_j = \sum_{k=1}^{n} \Delta u_{jk}\ ; \quad j = 1,2,\cdots,r \tag{12}$$

Notably, such a decomposition is *conceptual* since a decoupler does not exist. The component Δu_{ji} is used to drive the output y_i to its reference signal, and the other components Δu_{jk} ($k \neq i$) are used to eliminate the effects from other manipulated variables on y_i. Define vectors of all decomposed control moves at time instant t and for the next M sampling instants as

$$\Delta\bar{\mathbf{u}}_t = \left[\Delta u_{11,t} \ \cdots\ \Delta u_{r1,t}\ \ \Delta u_{12,t}\ \cdots\ \Delta u_{r2,t}\ \cdots\ \Delta u_{1n,t}\ \cdots\ \Delta u_{rn,t}\right]^T \tag{13}$$

$$\Delta\bar{\mathbf{U}}_t = \left[\Delta\bar{\mathbf{u}}_t{}^T \ \ \Delta\bar{\mathbf{u}}_{t+1}{}^T \ \cdots\ \Delta\bar{\mathbf{u}}_{t+M-1}{}^T\right]^T_{rnM\times 1} \tag{14}$$

We have $\Delta u_{jk,t} = \mathbf{1}_{j+(k-1)n}\Delta\bar{\mathbf{u}}_t$, where $\mathbf{1}_{j+(k-1)n}$ denotes a $1\times rn$ row vector whose $[j+(k-1)n]$-th element is one and all other elements are zero. Also, a vector of decomposed control moves for the next M sampling instants can be expressed as

$$\Delta\mathbf{U}_{jk,t} = \left[\Delta u_{jk,t} \ \ \Delta u_{jk,t+1} \ \cdots\ \Delta u_{jk,t+M-1}\right]^T = \mathbf{\Pi}_{jk}\Delta\bar{\mathbf{U}}_t \tag{15}$$

where

$$\mathbf{\Pi}_{jk} = blockdiag\left[\mathbf{1}_{j+(k-1)n} \ \ \mathbf{1}_{j+(k-1)n} \ \cdots\ \mathbf{1}_{j+(k-1)n}\right]_{M\times rnM} \tag{16}$$

Therefore, the future predictions for the constituent components of output are written as

$$\hat{\mathbf{Y}}_{ijk,t+1} = \mathbf{S}_{ij}\Delta\mathbf{U}_{jk,t} + \hat{\mathbf{Y}}^o_{ijk,t+1} = \bar{\mathbf{S}}_{ijk}\Delta\bar{\mathbf{U}}_t + \hat{\mathbf{Y}}^o_{ijk,t+1}; \qquad \bar{\mathbf{S}}_{ijk} \triangleq \mathbf{S}_{ij}\mathbf{\Pi}_{jk} \tag{17}$$

where $\hat{\mathbf{Y}}^o_{ijk,t+1} = [\hat{y}^o_{ijk,t+1}\ \ \hat{y}^o_{ijk,t+2}\ \cdots\ \hat{y}^o_{ijk,t+P_d}]^T$ and $\mathbf{S}_{ij}$ is the $P_d \times m$ dynamic matrix for the SISO model between y_i and u_j.

The control moves for the next M sampling instants can be expressed as

$$\Delta\mathbf{U}_t = \mathbf{\Pi}\Delta\bar{\mathbf{U}}_t \tag{18}$$

where

$$\mathbf{\Pi} = blockdiag\left[\mathbf{\Omega} \ \ \mathbf{\Omega} \ \cdots\ \mathbf{\Omega}\right]_{rM\times rnM}; \quad \mathbf{\Omega} = \left[\mathbf{I}_r \ \cdots\ \mathbf{I}_r\right]_{r\times rn} \tag{19}$$

The corrected predictions of future outputs given in Eq.(5) can be written as

$$\tilde{\mathbf{Y}}_{t+1} = \bar{\mathbf{S}}\Delta\bar{\mathbf{U}}_t + \tilde{\mathbf{Y}}^o_{t+1}\ ; \qquad \bar{\mathbf{S}} = \mathbf{S}\boldsymbol{\Pi} \tag{20}$$

Substituting Eqs.(20), (18), and (17) into Eq.(11), the cost function for DDMC becomes

$$\begin{aligned} J_{\text{DDMC}} = &\left(\mathbf{Y}^r_{t+1} - \bar{\mathbf{S}}\Delta\bar{\mathbf{U}}_t - \tilde{\mathbf{Y}}^o_{t+1}\right)^T \mathbf{Q}\left(\mathbf{Y}^r_{t+1} - \bar{\mathbf{S}}\Delta\bar{\mathbf{U}}_t - \tilde{\mathbf{Y}}^o_{t+1}\right) + \Delta\bar{\mathbf{U}}_t{}^T\left(\boldsymbol{\Pi}^T\mathbf{R}\boldsymbol{\Pi}\right)\Delta\bar{\mathbf{U}}_t \\ &+ \sum_{i=1}^{n}\left[\sum_{k=1,k\neq i}^{n}\left[\left(\bar{\mathbf{S}}_{ik}\Delta\bar{\mathbf{U}}_t + \sum_{j=1}^{r}\hat{\mathbf{Y}}^o_{ijk,t+1}\right)^T \mathbf{W}_{ik}\left(\bar{\mathbf{S}}_{ik}\Delta\bar{\mathbf{U}}_t + \sum_{j=1}^{r}\hat{\mathbf{Y}}^o_{ijk,t+1}\right)\right]\right] \end{aligned} \tag{21}$$

where $\bar{\mathbf{S}}_{ik} = \sum_{j=1}^{r}\bar{\mathbf{S}}_{ijk}$. The cost function is now written in terms of $\Delta\bar{\mathbf{U}}_t$ so that the control design problem is transformed into the determination of $\Delta\bar{\mathbf{U}}_t$, instead of $\Delta\mathbf{U}_t$. Therefore, the DDMC control calculation is based on

$$\min_{\Delta\bar{\mathbf{U}}_t} J_{\text{DDMC}} \tag{22}$$

In the unconstrained case, the DDMC control law can be calculated analytically:

$$\Delta\bar{\mathbf{U}}_t = \mathbf{K}_T\left(\mathbf{Y}^r_{t+1} - \tilde{\mathbf{Y}}^o_{t+1}\right) - \sum_{i=1}^{n}\sum_{k=1,k\neq i}^{n}\left(\mathbf{K}_{D,ik}\sum_{j=1}^{r}\hat{\mathbf{Y}}^o_{ijk,t+1}\right) \tag{23}$$

where the gain matrices are given by

$$\begin{aligned} \mathbf{K}_T &= \left(\bar{\mathbf{S}}^T\mathbf{Q}\bar{\mathbf{S}} + \boldsymbol{\Pi}^T\mathbf{R}\boldsymbol{\Pi} + \sum_{i=1}^{n}\sum_{k=1,k\neq i}^{n}\left(\bar{\mathbf{S}}_{ik}{}^T\mathbf{W}_{ik}\bar{\mathbf{S}}_{ik}\right)\right)^{-1}\bar{\mathbf{S}}^T\mathbf{Q} \\ \mathbf{K}_{D,ik} &= \left(\bar{\mathbf{S}}^T\mathbf{Q}\bar{\mathbf{S}} + \boldsymbol{\Pi}^T\mathbf{R}\boldsymbol{\Pi} + \sum_{i=1}^{n}\sum_{k=1,k\neq i}^{n}\left(\bar{\mathbf{S}}_{ik}{}^T\mathbf{W}_{ik}\bar{\mathbf{S}}_{ik}\right)\right)^{-1}\bar{\mathbf{S}}_{ik}{}^T\mathbf{W}_{ik} \end{aligned} \tag{24}$$

Based on the receding horizon approach, only the first control move $\Delta\mathbf{u}_t$, which is obtained from $\Delta\mathbf{u}_t = \boldsymbol{\Omega}\Delta\bar{\mathbf{u}}_t$, is actually implemented and then a new set of control moves is calculated at the next sampling instant. Inequality constraints for process inputs and outputs can be included in the control calculation. In the constrained case, the control moves can be calculated by the programming techniques.

The extra design parameters in the DDMC design are the decoupling horizon P_d and the weighting matrix $\mathbf{W}_{ik}$, which provide more degrees for improved decoupling control performance. Increasing the value of $\mathbf{W}_{ik}$ generally improves the decoupling control performance, at the expense of the speed of tracking response. As P_d increases, the decoupling control performance tends to be improved. However, the decoupling control performance may become worse when a too large value of P_d is chosen. Our simulation study reveals that $P_d = P$ is a reasonable choice.

4. Illustrative example

Consider the Wood-Berry (Wood and Berry, 1973) binary distillation column process:

$$\begin{bmatrix} y_1(s) \\ y_2(s) \end{bmatrix} = \begin{bmatrix} \dfrac{-12.8e^{-s}}{16.7s+1} & \dfrac{-18.9e^{-3s}}{21s+1} \\ \dfrac{6.6e^{-7s}}{10.9s+1} & \dfrac{-19.4e^{-3s}}{14.4s+1} \end{bmatrix} \begin{bmatrix} u_1(s) \\ u_2(s) \end{bmatrix} \tag{25}$$

We compare the proposed DDMC design with the conventional DMC design. The tuning parameters used in the two designs are the same with $P = 15$, $M = 3$, and

$$\begin{aligned} \mathbf{Q} &= blockdiag\left[\mathbf{Q}_1, \mathbf{Q}_2, \cdots, \mathbf{Q}_P\right]; \quad \mathbf{Q}_i = diag\left[1 \quad 1\right]; \; i = 1, 2, \ldots, P \\ \mathbf{R} &= blockdiag\left[\mathbf{R}_1, \mathbf{R}_2, \cdots, \mathbf{R}_M\right]; \quad \mathbf{R}_i = diag\left[0.1 \quad 0.1\right]; \; i = 1, 2, \ldots, M \end{aligned} \tag{26}$$

For the DDMC design, the additional tuning parameters are set as $P_d = 15$ and

$$\mathbf{W}_{12} = \mathbf{W}_{21} = diag\left[W_1, W_2, \cdots, W_{P_d}\right]; \; W_i = 100; \; i = 1, 2, \ldots, P_d \tag{27}$$

Figure 2 shows the system responses for a unit step set-point change. The proposed DDMC gives superior decoupling performance because a set-point change for one controlled variable has negligible effect on the other controlled variables. The speed of tracking response for DDMC is slower compared to DMC. Nevertheless, when the speed of tracking response for DMC is slowed down (similar to that for DDMC) by setting R_i = 15, the DMC still cannot provide effective decoupling control. The result clearly demonstrates the effectiveness of the proposed DDMC design for improved decoupling control performance.

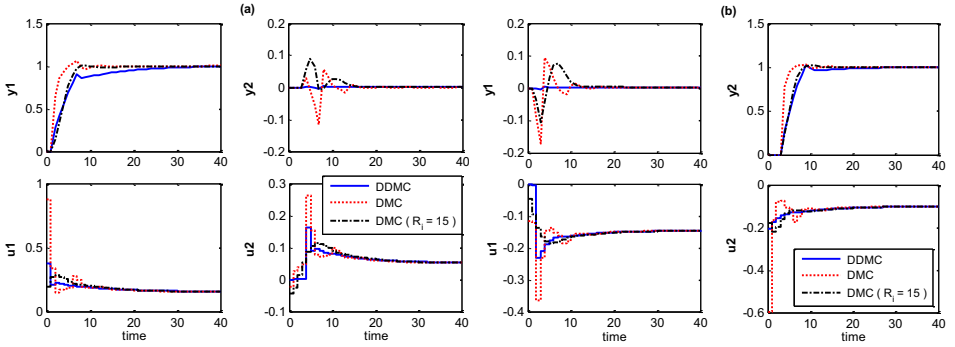


Figure 2. Closed-loop responses. (a) y_1 set-point change; (b) y_2 set-point change.

5. Conclusions

A novel design of DDMC controller for enhanced decoupling control has been proposed. The DDMC controller can provide an effective decoupling control without adding decouplers and without real-time adjustment of the design parameters. The effectiveness of the DDMC controller is validated through simulation studies. The DDMC design is expected to be a promising technology for the multivariable control systems where decoupling control performance is highly desirable.

Acknowledgement

The authors thank the support from MOST of Taiwan under grant 105-2221-E-027-128.

References

J. Garridoa, F. Vázqueza, F. Morilla, 2012, Centralized Multivariable Control by Simplified Decoupling, J. Process Control, 22, 1044–1062.

S. Qin, T. Badgwell, 2003, A Survey of Industrial Model Predictive Control Technology, Control Eng. Practice, 11, 733.

R. Wood, M. Berry, 1973, Terminal Composition Control of Binary Distillation Column, Chem. Eng. Sci., 28, 1707–1717.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50086-0

Nonlinear Dynamic Optimization for Improved Load-Shifting Agility of Cryogenic Air Separation Plants

Pascal Schäfer[a], Luise F. Bering[a], Adrian Caspari[a], Adel Mhamdi[a], Alexander Mitsos[a*]

[a]*Aachener Verfahrenstechnik – Process Systems Engineering, RWTH Aachen University, Forckenbeckstr. 51, 52074 Aachen, Germany*
amitsos@alum.mit.edu

Abstract

Time-variable energy prices motivate large energy consumers, such as the chemical industry, to adjust their electricity consumption to market signals, i.e., to perform demand side management (DSM). To maximize benefit, the processes must be able to perform large and fast load changes. In this paper, we apply dynamic optimization using a rigorous nonlinear model for offline computation of optimal load change trajectories. Considering characteristic load changes of an industrial-scale air separation unit (ASU) as a case study, we show that transition periods can be reduced by up to 50 % compared to solely using linear plant models to track load changes. Furthermore, we demonstrate that offline-computed optimal trajectories can be tracked efficiently by linearizing model predictive control in the presence of unknown temperature disturbances without violating the severe purity constraints. Hence, our results indicate a high potential of nonlinear model predictive control strategies for industrial DSM.

Keywords: Dynamic Optimization, Nonlinear Model Predictive Control, Cryogenic Air Separation

1. Introduction

Over the last decades, the share of power generation from intermittent renewable sources, in particular wind turbines and photovoltaics, has significantly increased. This leads to volatile electricity markets, where spot market prices differ substantially within few hours (EPEX SPOT SE, 2017). This price volatility motivates large energy consumers, e.g., the chemical/process industry, to perform demand side management (DSM), i.e., to adjust the electricity consumption and hence the production to market signals (e.g., Ghobeity and Mitsos, 2010).

In this context, the flexible operation of cryogenic air separation units (ASUs) has attracted interest due to their high energy consumption. Some ASU plants allow storing energy in times of low electricity prices in form of cryogenic liquid products that can be either sold to the market or re-vaporized to satisfy the demand for gaseous products in times of load reductions due to high energy prices. Most related works focus on the economic optimization of the production schedule by applying quasi-stationary models and identifying the optimal operation point as a function of current energy prices (e.g., Zhang et al., 2016 or Adamson et al., 2017). Recent work of Pattison et al. (2017) additionally includes process dynamics in economic optimization by using low-dimensional dynamic surrogate models.

However, previous works do not focus on the process control system. As industrial air separation comprises severe purity constraints, it is of paramount importance to satisfy these constraints while performing a load change and to reject unmeasured disturbances. For this purpose, we design a two-layer advanced process control system (APCS). The optimal trajectory for a desired load change is computed offline. Then, a linearizing online controller calculates base layer set-points for tracking the optimal trajectory. We compare the load-change performance of the proposed APCS to the industrial benchmark that uses only linear models. Furthermore, we evaluate disturbance rejection capacity for two load changes that enable energy market participation.

2. Methodology

2.1. Dynamic ASU Model

The flowsheet of the ASU considered in this work is depicted in Figure 1 (a). The ASU processes 135,000 Nm^3/h of fresh air. The vapor fraction of the incoming cooled air (HPF) enters the high-pressure column (HPC) in the bottom stage. On top of the HPC, a highly pure gaseous nitrogen stream is drawn off (HPGAN). Besides, after passing the condenser, a liquid nitrogen stream (LIN) is drawn off. The bottom product of the HPC is fed to the low-pressure column (LPC). In addition, the liquid fraction of the incoming air (LPF) and a fraction of the LIN-product are fed to the LPC. On top of the LPC, highly pure gaseous nitrogen is produced (LPGAN), while at the bottom, highly pure liquid (LOX) as well as gaseous oxygen (GOX) is produced. Besides, a gaseous stream is purged at an intermediate stage, drawing off the major part of the incoming argon. The condensing nitrogen of the HPC provides the required heat stream for the reboiler of the LPC.

A rigorous model of the described ASU is implemented in gPROMS version 4.2 (Process Systems Enterprise Ltd.) using the package Multiflash version 4.3 (KBC Advanced Technologies Ltd.) for physical property calculations. Values for pure compound as well as binary interaction parameters of the applied NRTL model are retrieved from the Aspen Plus database version 8.8 (Aspen Technology, Inc.). We assume that all streams entering the columns are either saturated liquid or vapor. The column stages are modelled with dynamic MESH-equations using the following assumptions (Huang et al., 2009 and Cao et al., 2016): (i) negligible vapor hold up, (ii) fast temperature dynamics, (iii) a constant pressure difference between neighboring stages and (iv) a linear relation for stage hydraulics. The HPC consists of 40 theoretical equilibrium stages, the LPC of 80. The respective pressures are 5.8 – 6.0 bar and 1.2 – 1.6 bar. A base layer control using PI controllers as depicted in Figure 1 (a) further extends the model. In total, the ASU process is represented by a differential-algebraic equation system (DAE) comprising 374 differential equations (in particular dynamic mass and component balances at each stage) and 1,885 algebraic equations, e.g., flash-calculations, pseudo steady-state energy balances to calculate the vapor streams within the column (c.f., Cao et al., 2016) and control laws of the PI controllers.

2.2. Advanced Process Control System

The APCS (Figure 1 (b)) is implemented using the in-house software OptoEcon Toolbox (Elixmann et al., 2014). Set-points of the base layer (PI) control are considered manipulated variables for the APCS that are used for realizing load changes, except for the liquid level in the reboiler, which is kept constant at its maximum value to ensure a complete covering of the integrated condenser. Product purities are considered controlled variables. Due to the high computational effort for nonlinear dynamic optimizations of

large-scale distillation models, the APCS is divided into a computationally intensive offline part for trajectory optimization and an online part for disturbance rejection, as suggested by Huang et al. (2009). First, an optimal trajectory is computed once for a pre-calculated load change (for instance, from quasi-stationary optimization of operating points) by solving the optimal control problem given below (see, White et al., 1996).

$$\min_{SP_i(t)} \sum_i \int_0^{t_f} \left(\frac{SP_i(t) - SP_i^*}{SP_i^*} \right)^2 dt \tag{1}$$

$$s.t.\ SP_i(t_f) = SP_i^* \qquad \forall\, i \tag{2}$$

$$\boldsymbol{h}(\boldsymbol{SP}(t)) \leq 0 \qquad \forall\, t \tag{3}$$

The objective is to minimize the quadratic relative difference between the current set-point $SP_i(t)$ of controller i and the set-point in the pre-calculated new operating point SP_i^* (Eq.(1)). End-point constraints in Eq.(2) ensure reaching the new operating point at final time t_f. The nonlinear function $\boldsymbol{h}(\boldsymbol{SP}(t))$ in Eq.(3), which is defined by the DAE model described above and relates controller set-points and purities, forces constraint satisfaction for all time points (10 ppm impurity for nitrogen and 0.5 % for oxygen). Dynamic optimizations are performed using gPROMS' built-in sequential NLPSQP solver (NL DO in Figure 1 (b)). A piecewise constant control vector parametrization is applied using a fixed grid with equally distributed control intervals. The set-point change between consecutive control intervals is further limited by an upper bound.

Second, a built-in block from the OptoEcon Toolbox for a linear time-invariant model predictive controller (LTI-MPC) calculates set-points for the base layer control to track the optimal trajectory while satisfying purity constraints. The LTI-MPC uses a linear state-space model that is obtained by linearizing the DAE once at nominal operating conditions. A prediction horizon of 16 sampling times is used, the control horizon is set to 12. Calculated set-points are provided to the plant surrogate via an OPC interface. We assume that all system states are estimated from measured outputs, i.e., state feedback, but add zero-mean white noise to simulate measurement errors. In the industrial benchmark APCS, the NL DO in Figure 1 (b) is dropped. Instead, ramps with maximum control action are applied as reference.

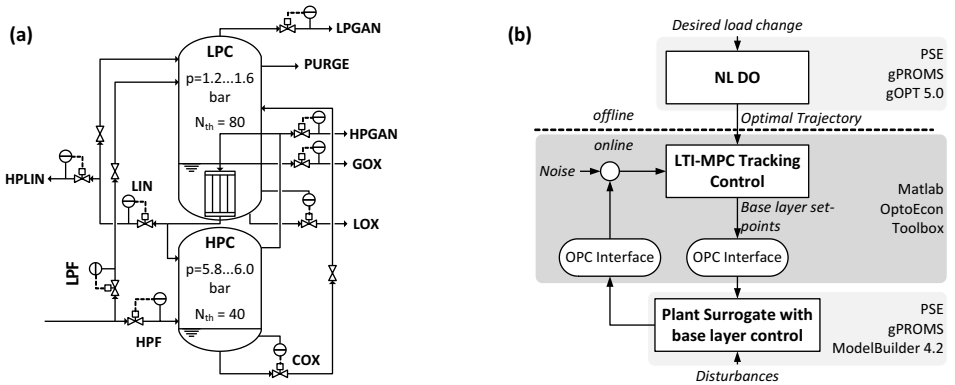


Figure 1: (a) Flowsheet of the ASU considered in this work. (b) Two-layer APCS incorporating offline nonlinear dynamic trajectory optimizations (NL DO) and online tracking control using a linear time-invariant model predictive controller (LTI-MPC).

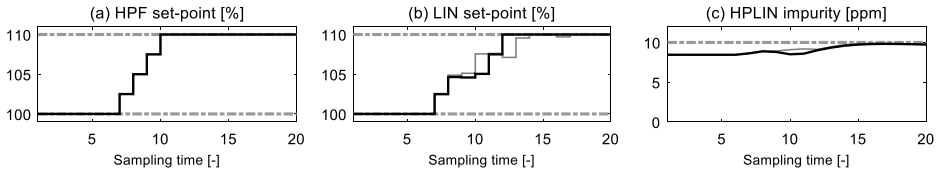


Figure 2: Temporal course of relevant process variables for a 10 % load increase of the ASU without disturbances. Solid black lines show optimal trajectories from NL DO. Thin grey lines show results of the industrial benchmark (without NL DO). Dashed-dotted grey lines show constraints.

3. Computational Case Studies and Discussion

3.1. Case Study A: Load Increase

First, we consider an operating scenario, where the ASU produces only liquids (HPLIN and LOX) that can be stored and discontinuously sold to off-site customers. The ASU can hence freely participate in the energy markets, i.e., if electricity is cheap, the load is increased in order to produce more. Likewise, if electricity is expensive, the load is decreased. For case study A, we consider a load increase by 10 % to be realized by the APCS. The comparison of the offline-optimized trajectories from NL DO solving Eq.(1) – (3) to the industrial LTI-MPC strategy (without NL DO) is given in Figure 2. In both cases, the HPF is increased with maximum allowed controller action per sampling time. If the LIN withdrawal was increased evenly, the limiting HPLIN purity constraint (Figure 2 (c)) would be violated, which requires corrective control actions. In both cases, constraint satisfaction is achieved by slower increasing the LIN withdrawal, which corresponds to a temporal reduction of the distillate-to-feed ratio that enhances the separation capacity of the column. By using the LIN trajectory from NL DO, the total transition period is reduced by 50% compared to the industrial benchmark.

In order to evaluate the disturbance rejection capacity of the proposed two-layer APCS, an unmeasured shift in HPF temperature (e.g., due to changing ambient temperature), that causes a partial condensation of the feed and that thus substantially disturbs the energy input to the column, is applied as disturbance scenario (Figure 3 (a)). We further assume zero-mean white noise on process states. Without using additional tracking control, the disturbances lead to a substantial violation of purity constraints during the load change (thin grey line in Figure 3 (c)), if the offline-optimized trajectories are applied. Using a LTI-MPC tracking control, this violation is circumvented by adjusting the distillate-to-feed ratio through temporal reductions of the LIN withdrawal compared to the offline-optimized trajectory (Figure 3 (b)) that keep the HPLIN impurity below its maximum allowed value (Figure 3 (c)).

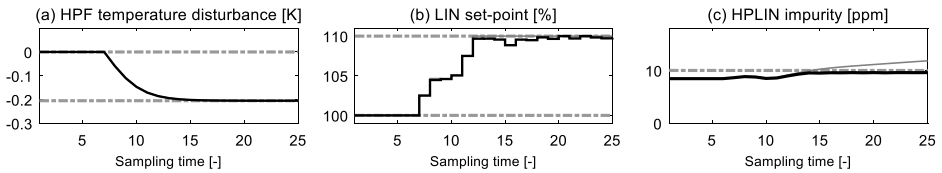


Figure 3: Temporal course of relevant process variables for a 10 % load increase of the ASU considering disturbances. Solid black lines show the simulation using the two-layer APCS. Thin grey lines show the results without tracking control. Dashed-dotted grey lines show constraints

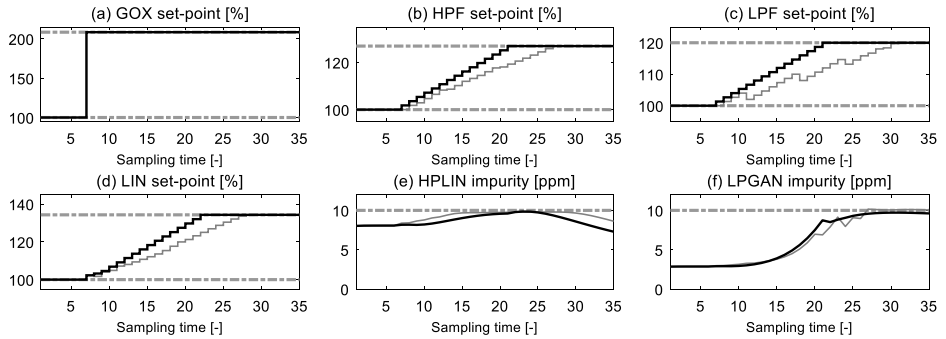


Figure 4: Temporal course of relevant process variables for a change of product mix of the ASU without disturbances. Solid black lines show optimal trajectories from NL DO. Thin grey lines show results of the industrial benchmark (without NL DO). Dashed-dotted grey lines show constraints.

3.2. Case Study B: Change of Product Mix

In the second operating scenario, an ASU with all possible products is considered. This kind of ASU has to ensure the supply of on-site customers with gaseous products. The fulfilment of supply agreements is the main objective. The ability of the ASU to participate in energy markets is thus limited to the liquid production. For case study B, we consider an instantaneous increase in GOX demand of an on-site customer (e.g., a steel mill). The increased oxygen production would lead to purity constraint violation without further measure. Thus, the operating point, in particular the liquid production, of the ASU is adjusted. We compare offline-optimized trajectories from NL DO to the industrial LTI-MPC strategy (Figure 4). Both the HPLIN impurity (Figure 4 (e)) and the LPGAN impurity (Figure 4 (f)) constraint are limiting and require corrective control actions. Similar to case study A, constraint satisfaction is ensured by temporal variations of the distillate-to-feed ratios now of both columns. We find that by applying input trajectories from the NL DO (Figure 4 (b) – (d)), transition periods are again substantially reduced by 37.5 % compared to the industrial benchmark.

Disturbance rejection capacity of the APCS is evaluated for this load change as well by introducing a shift in HPF temperature (Figure 5 (a)) and zero-mean white noise on process states. Similar to case study A, severe violations of both impurity constraints in case of missing tracking control can be avoided using a LTI-MPC (Figures 5 (e) and (f)), which in particular adjusts the LIN withdrawal from the HPC (Figure 5 (d)).

4. Conclusion

A two-layer APCS is proposed for an industrial-scale ASU. Load-change performance of the APCS is evaluated in two case studies using the in-house software OptoEcon-Toolbox. We show that by applying offline-computed trajectories from dynamic optimization of a nonlinear process model, transition periods for characteristic load changes can be significantly shortened, increasing the plant's agility. We further demonstrate that offline-optimized trajectories can be tracked in the presence of unknown disturbances without violating severe purity requirements by using industrial LTI-MPC controllers. Our results indicate a high potential of nonlinear model predictive control (NMPC) for industrial DSM, as quicker but still safe load changes are enabled. This is desirable in order to benefit from volatile energy markets, in particular from short-term spot market price peaks, such as negative electricity prices, that are expected to become more frequent. Future research should focus on the integration of economic objectives

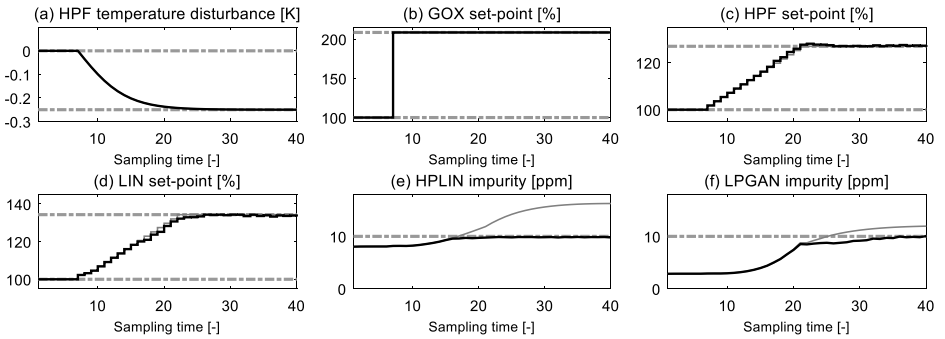


Figure 5: Temporal course of relevant process variables for a change of product mix of the ASU considering disturbances. Solid black lines show the simulation using the two-layer APCS. Thin grey lines show the results without tracking control. Dashed-dotted grey lines show constraints.

into the trajectory optimization to allow for a closed-loop economic NMPC. Particularly, this includes the incorporation and validation of reduced dynamic models for the distillation columns, such as, e.g., collocation-based models as suggested by Cao et al. (2016), in order to reduce the high computational demand of solving the NL DO problem.

Acknowledgements

The authors gratefully acknowledge the financial support of the Kopernikus-project "SynErgie" by the Federal Ministry of Education and Research (BMBF) and the project supervision by the project management organization Projektträger Jülich (PtJ). Further, the authors thank Dr. Andreas Peschel and Florian Schliebitz from Linde Engineering.

References

R. Adamson, M. Hobbs, A. Silcock, M. J. Willis, 2017, Integrated real-time production scheduling of a multiple cryogenic air separation unit and compressor plant, Comp. Chem. Eng., 104, 25–37.

Y. Cao, C. L. E. Swartz, J. Flores-Cerrillo, J. Ma, 2016, Dynamic Modeling and Collocation-Based Model Reduction of Cryogenic Air Separation Units, AIChE J., 62, 5, 1602–1615.

D. Elixmann, J. Puschke, H. Scheu, R. Schneider, I. J. Wolf, W. Marquardt, 2014, A software environment for economic NMPC and dynamic real-time optimization of chemical processes, at-Automatisierungstechnik, 62, 2, 150–161.

EPEX SPOT SE, www.epexspot.com, 2017.

A. Ghobeity, A. Mitsos, 2010, Optimal time-dependent operation of seawater reverse osmosis, Desalination, 263, 1–3, 76–88.

R. Huang, V. M. Zavala, L. T. Biegler, 2009, Advanced step nonlinear model predictive control for air separation units, J. Process Control, 19, 4, 678–685.

R. C. Pattison, C. R. Touretzky, I. Harjunkoski, M. Baldea, 2017, Moving horizon closed-loop production scheduling using dynamic process models, AIChE J., 63, 2, 639–651.

V. White, J. D. Perkins, D. M. Espie, 1996, Switchability analysis, Comp. Chem. Eng., 20, 4, 469–474.

Q. Zhang, A. Sundaramoorthy, I. E. Grossmann, J. M. Pinto, 2016, A discrete-time scheduling model for continuous power-intensive process networks with various power contracts, Comp. Chem. Eng., 84, 382–393.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50087-2

Reinforcement Learning Applied to Process Control: A Van der Vusse Reactor Case Study

G. O. Cassol[1], G. V. K. Campos[2], D. M. Thomaz[1], B. D. O. Capron[3*], A. R. Secchi[1]

[1]Universidade Federal do Rio de Janeiro, Programa de Engenharia Química, PEQ/COPPE/UFRJ, Caixa Postal 68.502, Rio de Janeiro, RJ 21941-914, Brazil
[2]University of California Davis, Department of Chemical Engineering, One Shields Avenue, Davis, CA 95616, United States
[3]Universidade Federal do Rio de Janeiro, Departamento de Engenharia Química, Escola de Química, Av. Horacio Macedo 30, CT-Bloco E, Rio de Janeiro, RJ 21949-900, Brazil
bruno@eq.ufrj.br

Abstract

With recent advances in industrial automation, data acquisition, and successful applications of Machine Learning methods to real-life problems, data-based methods can be expected to grow in use within the process control community in the near future. Model-based control methods rely on accurate models of the process to be effective. However, such models may be laborious to obtain and, even when available, the optimization problem underlying the online control problem may be too computationally demanding. Furthermore, the process degradation with time imposes that the model should be periodically updated to stay reliable. One way to address these drawbacks is through the merging of Reinforcement Learning (RL) techniques into the classical process control framework. In this work, a methodology to tackle the control of nonlinear chemical processes with RL techniques is proposed and tested on the well-known benchmark problem of the non-isothermal CSTR with the Van de Vusse reaction. The controller proposed herein is based on the implementation of a policy that associates each state of the process to a certain control action. This policy is directly deduced from a measure of the expected performance gain, given by a value function dependent on the states and actions. In other words, in a given state, the action that provides the highest expected performance gain is chosen and implemented. The value function is approximated by a neural network that can be trained with pre-simulated data and adapted online with the continuous inclusion of new process data through the implementation of an RL algorithm. The results show that the proposed adaptive RL-based controller successfully manages to control and optimize the Van de Vusse reactor against unmeasured disturbances.

Keywords: Reinforcement Learning; Model Predictive Control; Adaptive Control.

1. Introduction

In the past few decades the process industry has experienced a large increase in automation and data acquisition systems, and as acquiring good quality and reliable data becomes easier and cheaper, data-based methodologies become more susceptible to application. Machine Learning and Data Science methods have grown in importance

outside of the process community, driven by successful applications in real-life problems. It is important for the process industry to explore the application of data-based methods, and incorporate the most successful ideas to the traditional process control framework.

Another incentive for the application of data-based methods arises from the drawbacks of classical model-based control. Most model predictive control (MPC) methods still rely on the assumption of linearity, which is not true for the majority of chemical processes. Nonlinear MPC (NMPC) methods may be applied, but they require a nonlinear process model that may be laborious to obtain. In addition, solving the NMPC optimization problem may be too computationally demanding for online applications.

In order to tackle some of the aforementioned problems, a control approach based on model-free learning may be used. In the process control community model-free control strategies have been proposed since the early 1960s, while in the artificial intelligence community the field of reinforcement learning (RL) was developed during the 1980s, focusing on the design of control algorithms that work solely based on data (Busoniu *et al.*, 2010). There is not yet a vast literature on the application of RL techniques to the control of chemical processes, as evidenced by Syafiie *et al.* (2008), but a few authors proposed general purpose model-free controllers in the past. Stenman (1999) proposed a partially model-free approach combining MPC with adaptive techniques that estimate linear models online, called model-on-demand. Lee and Wong (2010) proposed a modification to the stochastic MPC framework based on approximate dynamic programming (ADP), and explored different aspects of the problem such as parametric vs. nonparametric value function approximators, and batch vs. continuous learning. Morinelly and Ydstie (2016) formulated an adaptive optimal control algorithm for systems with uncertain dynamics, called RL dual MPC (RLdMPC), in which an exploratory component is embedded in the objective function of the MPC and the competing results of the optimal solution and the cost function approximation provide a balance in exploration and exploitation.

Focusing on the application of RL techniques to chemical processes, Hoskins and Himmelblau (1992) proposed a neural network (NN) based control, in which two different NN models are used to predict the control performance measure and the control action/policy. Martinez (2000) proposed a sampling strategy (sequential experiment design) based on RL for optimization of batch processes, where the RL solution is used to systematically shrink the region of interest for the optimization. Syafiie *et al.* (2008) proposed a model-free learning control method based directly on RL techniques, more specifically on Q-learning with finite-discrete action space. Although easy to implement, the discrete version of Q-learning may not be the most appropriate for chemical processes, in which states and actions are continuous and must be discretized to obtain a good representation. Shah and Gopal (2016) proposed a model-free predictive control, employing a fuzzy inference system to approximate the value function. While the fuzzy model may account for uncertainties, it is still subject to the curse of dimensionality and requires tuning of many localized parameters.

In the present work, an adaptive controller based on an RL algorithm with value function approximation via NN is proposed and tested for the control of a nonlinear non-isothermal CSTR with the Van de Vusse reaction. We aim to assess the application of the RL algorithm to a chemical process, while addressing aspects such as continuous representation of states, combined optimization and control, and comparison with linear MPC.

2. Methodology

The problem addressed by RL is that of decision-making. An agent (a controller in our case) receives information about the state x of the system (process) and implements a policy, a mapping from the states into actions, which defines which action the controller has to take based on the actual state of the process. This policy is derived from the so-called action-value function $Q_{(x,u)}$, computed from data by an RL algorithm, that assigns an expected performance gain or return value G over time to every possible state-action pair. In our work, a multilayer perceptron (MLP) NN is employed to approximate $Q_{(x,u)}$ in order to use a continuous representation of the states.

A policy iteration algorithm is used to compute the policy used by our controller. Iteratively, the action-value function for the actual policy is evaluated using the Monte Carlo policy evaluation and the policy is then improved using the well-known *ε-greedy* algorithm (Busoniu *et al.*, 2010), in which the action that leads to the maximum value function is chosen with a $1-\varepsilon$ probability. Otherwise, a random action is taken to help guaranteeing that all the actions are considered in the optimization of the return value in a given state. The NN is trained taking into account the states and actions as inputs and the expected return value as output.

Different configurations of the controller are tested in this work:

(i) Pre-trained non-adaptive RL-based Controller (RL-C): the NN is only trained with pre-simulated data corresponding to different initial steady-state conditions randomly selected.

(ii) Untrained adaptive RL-C: the NN is only trained with the process data acquired online. In this case, a step control is used to take into account the difference e_{k-1} between the predicted reward from the NN model and the actual reward obtained in the last step. The control action is limited when the NN model is not capable to correctly predict the outcome, as it is shown in Eq. (1), where a_k^P is the action predicted by the NN and a_k^R the action really taken by the controller.

$$a_k^R = a_k^P\left(1 - 0.95 \cdot 1/(1 + e^{|e_{k-1}|})\right) \quad (1)$$

(iii) Pre-trained adaptive RL-C: the NN is pre-trained and continues to be trained online. Two hidden layers of seventy neurons are used by the NN model. The hidden layers use logistic functions and the output layer a linear one. The NN training is performed using MATLAB NN toolbox, which provides batch and online algorithms. 80% of the available data were used for training, 10% for validation and 10% for testing.

3. Results

3.1. Process description

The performance of the proposed controller is tested through the simulation of the control of a non-isothermal CSTR, with the Van de Vusse reaction, as described by Klatt and Engel (1998). The reaction consists in the synthesis of cyclopentenol (B) from cyclopentadiene (A) by acid-catalyzed electrophilic addition of water in a dilute solution. Due to the strong reactivity of both A and B, dicyclopentadiene (D) is produced by the Diels-Alder reaction as a side product, and cyclopentanediol (C) as a consecutive product by addition of another water molecule. It is usually desirable to maximize the production of component B, while minimizing the production of C and D. The reaction is presented in Eq. (2), and the mass and energy balances in Eq. (3) to (5), where $C_{i,in}$ and C_i are the inlet and outlet concentration of the i-th component, $k_{j(T)}$ is

the reaction rate parameter for reaction j (according to Arrhenius law), ΔH_{R_j} is the constant heat of reaction j, T_k is the jacket temperature, and other parameters are as defined in classical notation. The parameters values can be found in Klatt and Engel (1998).

$$A \xrightarrow{k_1} B \xrightarrow{k_2} C \qquad 2A \xrightarrow{k_3} D \tag{2}$$

$$\frac{dC_A}{dt} = \frac{F}{V}\left(C_{A,in} - C_A\right) - k_1(T)C_A - k_3(T)C_A^2 \tag{3}$$

$$\frac{dC_B}{dt} = \frac{F}{V}\left(C_{B,in} - C_B\right) + k_1(T)C_A - k_2(T)C_B \tag{4}$$

$$\frac{dT}{dt} = \frac{1}{\rho \cdot C_p}\left[k_1(T)C_A\left(-\Delta H_{R_1}\right) + k_2(T)C_B\left(-\Delta H_{R_2}\right) + k_3(T)C_A^2\left(-\Delta H_{R_3}\right)\right] + \frac{F}{V}(T_{in} - T) + \frac{K_w A_R}{\rho\, C_p\, V}(T_k - T) \tag{5}$$

This process was chosen due to its nonlinear characteristics, the most important of those being the gain inversion of the product concentration C_B with respect to inlet flow F. The location of the gain inversion depends on the reactor temperature T, which in turn is controlled by manipulating the cooling jacket temperature T_k.

3.2. Controller definition

The actions vector u includes the manipulated variables inlet flow F and cooling jacket temperature T_k increments. The state vector x includes the outlet concentrations C_A and C_B, the reactor temperature T, and the manipulated variables F and T_k. The control objective is to track a certain F_B ($F.C_B$) while maximizing C_B. The corresponding instantaneous reward R is defined in Eq. (6), where F_B^{SP} is the setpoint for F_B and the subscript k corresponds to the time instant. The proposed reward function prioritizes the set-point tracking of F_B (the sigmoidal function used goes to 1 when the difference between F_B and its set-point is large). Then, when F_B is close to the set-point, the reward starts to focus on the optimization of C_B. The weight w is used as a tuning factor. The return G, or accumulated reward, is simply defined as the discounted sum of rewards, as shown in Eq. (7), where $\gamma \in [0,1]$ is the discount factor, a measure of how "far-sighted" the controller is in considering its rewards.

$$R(k) = |F_B(k-1) - F_B^{SP}| - |F_B(k) - F_B^{SP}| + w \cdot [C_B(k) - C_B(k-1)]\left\{1 - 1\Big/\left(1 + e^{\left|\frac{F_B(k) - F_B^{SP}}{F_B^{SP}}\right|}\right)\right\} \tag{6}$$

$$G(\hat{k}) = \sum_{k=0}^{\hat{k}} \gamma^k R(k) \tag{7}$$

3.3. Initial Conditions

The initial conditions used for the NN model training in different episodes are random steady states for different values of F and T_k, with $C_{A,in} = 5.1$ mol/L, $C_{B,in} = 0$ mol/L and T_{in} = 110 °C. The initial steady state considered for the following results is defined by $F = 1000$ L/h and $T_k = 100$ °C, with the same conditions for $C_{A,in}$, $C_{B,in}$ and T_{in} as for training. The F_B set-point was set to 2000 mol/h. The time step used for all simulations was 0.01 h.

3.4. Process Control/Optimization without disturbances

In this section, the performance of the pre-trained non adaptive RL-C, the untrained adaptive RL-C, the linear MPC controller and the nonlinear MPC (NMPC) are compared when changes in the operating conditions take place and no disturbances enter the system. The model of the MPC is obtained from the linearization of the nonlinear model defined in Section 3.1 at the initial condition, when disturbance steps of ±10% are applied to the manipulated variables. The NMPC uses the complete model of the system.

The results in Figure 1 show that the linear MPC is not able to maintain F_B at the desired set-point. This system presents a gain inversion of C_B with respect to the manipulated variables, which means the available model given to the MPC is no longer accurate. The NMPC is the first to achieve the desired set-point of F_B, but takes more time than the pre-trained RL-C to reach the maximum of C_B. The pre-trained non adaptive RL-C, having been trained with data from different initial conditions, quickly manages to bring F_B to its set-point while maximizing C_B. Lastly, the untrained adaptive RL-C expresses a behavior that is initially similar to the linear MPC strategy, but then manages to change its course employing actions that are more similar to the pre-trained NN model. In addition, it is possible to verify that the untrained adaptive RL-C first takes actions that brings F_B to its set-point, and then starts to make changes in the system to optimize C_B (the control strategy "walks" in the F_B set-point curve).

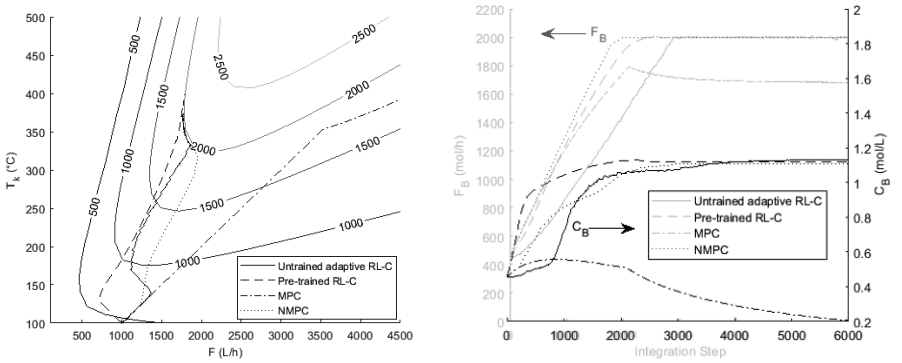


Figure 1. Simulation results without unmeasured disturbances

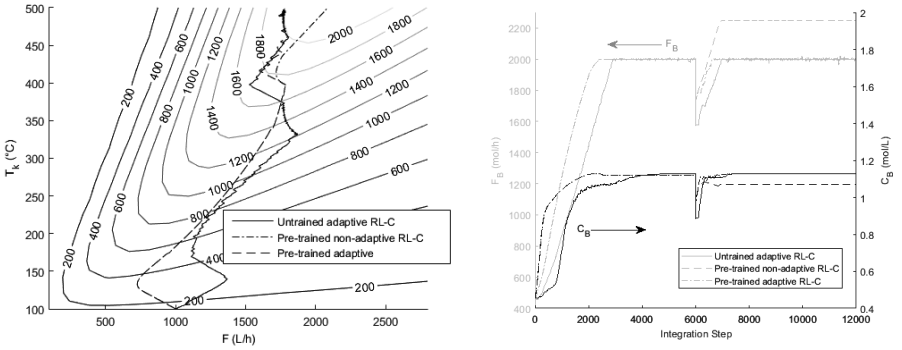


Figure 2. Simulation results with unmeasured disturbance

3.5. Process Control/Optimization with unmeasured disturbances

In this section, the performances of the three configurations of the RL-C defined in the methodology section are compared when an unmeasured disturbance occurs (the inlet temperature T_{in} changes from 110 ºC to 90 ºC at time step 6000). The results in Figure 2 show that both pre-trained strategies have the same behavior when the system is not disturbed. However, as expected, after the unmeasured disturbance occurs, only the pre-trained adaptive RL-C is able to bring the system back to the F_B set-point and to adapt to the new process conditions that are different from the ones used for training. In addition, the untrained adaptive RL-C also manages to bring F_B back to its set-point and to optimize C_B before and after the unmeasured occurs, although in a less efficient way, showing that the previous knowledge from the data can enhance its performance.

4. Conclusions

In this work a continuous RL algorithm was applied to the problem of adaptive control of chemical processes, in particular of chemical reactors with nonlinear response and gain inversion. A MLP neural network model was used to represent the action value function in terms of the continuous states and actions, which could be initially trained with a batch of pre-simulated process data. The results show that the adaptive RL-based controller successfully manages to control and optimize the Van de Vusse reactor against unmeasured disturbances and verify that the use of previous data from the system can help the NN model to adapt in a better way to changes in the system's behavior.

References

L. Busoniu, R. Babuska, B. De Schutter, D. Ernst, 2010, Reinforcement Learning and Dynamic Programming Using Function Approximators, CRC Press.

J. C. Hoskins, D. M. Himmelblau, 1992, Process Control via Artificial Neural Networks and Reinforcement Learning, Computers and Chemical Engineering, Vol. 16, No. 4, pp. 241-251.

K.-U. Klatt, S. Engell, 1998, Gain-scheduling trajectory control of a continuous stirred tank reactor, Computers & Chemical Engineering 22, Issues 4–5, pp. 491-502.

J. H. Lee, W. Wong, 2010, Approximate dynamic programming approach for process control, Journal of Process Control, Vol. 20, pp.1038– 1048.

E. C. Martinez, 2000, Batch Process Modeling for Optimization Using Reinforcement Learning, Computers and Chemical Engineering, Vol. 24, pp. 1187-1193.

J. E. Morinelly, B. E. Ydstie, 2016, Dual MPC with Reinforcement Learning, IFAC-PapersOnLine 49-7, pp. 266–271.

M. A. Mustafa, J. A. Wilson, 2012 Application of Reinforcement Learning to Batch Distillation, In: Proceedings of the Sixth Jordan International Chemical Engineering Conference, Amman, Jordan.

H. Shah, M Gopal, 2016, Model-free Predictive Control of Nonlinear Processes Based on Reinforcement Learning, IFAC-PapersOnLine 49-1 (2016) 089–094

A. Stenman, 1999, Model-free predictive control, In: Proceedings of the 38th IEEE Conference on Decision & Control, Phoenix, AZ, USA, pp.3712–3717.

S. Syafiie, F. Tadeo, E. Martinez, 2008, Model-Free Learning Control of Chemical Processes, Inside "Reinforcement Learning: Theory and Applications", Edited by C. Weber, M. Elshaw, N. M. Mayer, I-Tech Education and Publishing, pp.424.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50088-4

Dynamic Reduced Order Models for Polymerization Process Based on Molecular Weight Distribution

Jiayuan Kang[a], Zhijiang Shao[a], Xi Chen[a*]

[a]*State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou, Zhejiang 310027 P.R. China*
xi_chen@zju.edu.cn

Abstract

Key polymer properties are substantially based on the polymer molecular weight distribution (MWD). Modeling polymerization processes using kinetic-based approach involves the mass balances for all chain types and for all possible chain lengths. These models exhibit the nature of large-scale dynamic system. The numerical solution is computationally expensive and is challenging when applied for optimization problems. In this work, we proposed a reduced order model of MWD in dynamic systems. The basic idea is to reformulate the population balances to decompose the coupling in the chain lengths. We present a kind of model reformulation based on null-space projection method. The reformulation separates the fast equilibrium of living chains and the accumulative dead chains. Moreover, the coupling in the living chain lengths is decomposed by further derivation. Thus, the calculation of all possible chains is avoided. The large-scale MWD model is reduced to a much smaller scale one. Simulation results show that the computational cost is significantly reduced without essentially affecting the accuracy performance of the model. The model is also applied to a grade transition case study, which provides the prospection of dynamic optimization, state estimation and model predictive control based on MWD.

Keywords: Molecular weight distribution, Model order reduction, Dynamic optimization, Polymerization

1. Introduction

Research on polymer modeling is essential in precisely tuning the properties of polymers produced. The polymer model plays an important role in polymerization process monitoring, optimization and control. Modeling of a polymerization process for optimization depends on what kind of quality index is considered. Key polymer properties are substantially related to the polymer molecular weight distribution (MWD) (Soares et al. 2012). Modeling polymerization processes using kinetic-based approach involves the mass balances for all chain types and for all possible chain lengths. These balances, the so-called “population balances” can then be simultaneously solved to obtain the properties of each individual chain, which can derive the full MWD or average molecular weight (AMW). However, theoretically this requires simultaneously solving an infinite number of equations. For most macro quality indices, such as AMW, one can reformulate the population balances via moment method (Mastan et al. 2015). But for microstructural indices such as MWD, population balances are still necessary in many cases. Though the Flory distribution is applied in many steady state situations, it does not apply for dynamic processes. In general, it is not possible to find an analytical solution of the MWD in non-steady state. One can solve the population balances directly or cumulate the instantaneous

MWDs using pseudo steady state assumption (Soares et al. 2012; Weng et al. 2015). Either of the two approaches involves solving a large-scale dynamic system. Therefore, model reduction is required to improve the computational efficiency, particularly for the sake of process optimization and real-time applications, such as state estimation and model predictive control.

There are generally two types of methodologies of model reduction. One is the black-box approach, which generates surrogate models using data-driven techniques. The other is the physics-based approach, which manipulates the structure and equations of the rigorous model to find a lower dimensional approximation. In polymerization processes, the kinetic are highly nonlinear and the MWDs are high-dimensional output. It is difficult and expensive to surrogate the MWD or even the polymerization process (Kang et al. 2017). Research efforts have been made to reduce the computational expenses for MWD models. Saliakas et al. (2007) employed collation method and fixed pivot to calculate the dynamic MWD. Weng et al. (2015) speeded up dynamic simulation of MWD by multi-thread parallel computation. In this work, the model reformulation approaches are investigated to reduce the dynamic MWD model.

In this paper, we proposed a reduced order model of MWD in dynamic systems. The basic idea is to reformulate the population balances to decompose the coupling in the chain lengths. A reformulated MWD model based on null-space projection method (Nie et al. 2013) is presented. The reformulation separates the fast equilibrium of living chains and the accumulative dead chains. Moreover, it is worth to note that the coupling in the living chains are decomposed. Thus, the calculation of all possible chains are avoided. The MWD model is significantly reduced without essentially loss of model information. This benefits computational efficiency for on-line monitoring, dynamic optimization and control for polymerization processes based on MWD.

2. Dynamic Molecular Weight Distribution Models

Population balance is widely accepted to model the polymerization process. The mass balances for all chain types and for all possible chain lengths are derived in this method. These balances can be simultaneously solved to obtain the molecule fraction of each individual chain. MWD is derived based on the information of each chain calculated. Taking the high-density polyethylene (HDPE) process as an example, population balances for the living and dead chains in the reactor can be derived using the kinetic mechanism in Kang et al. (2017).

By assuming that the polymerization happens in a CSTR, the following population balances are applied to monomer-free active sites $P_0(j)$, living chains with length r $P_r(j)$, and dead chains with length r $D_r(j)$:

$$\frac{dP_0(j)}{dt} = k_{aA}(j)C_p(j)A - k_p(j)MP_0(j) + \left(k_{tH}(j)H_2 + k_{tA}(j)A + k_t(j)\right)\sum_{r=1}^{N_T} P_r(j) - k_d(j)P_0(j) \quad (1a)$$

$$\frac{dP_r(j)}{dt} = -k_p(j)M\left(P_r(j) - P_{r-1}(j)\right) - \left(k_{tH}(j)H_2 + k_{tA}(j)A + k_t(j)\right)P_r(j) \quad (1b)$$

$$\frac{dD_r(j)}{dt} = \left(k_{tH}(j)H_2 + k_{tA}(j)A + k_t(j)\right)P_r(j) \quad (1c)$$

where $j \in \{1, 2, \dots, N_s\}$ is the active site and N_T is the maximum length of chain considered, which could be 100000 or more.

The mass balances of monomer *M*, hydrogen H_2, catalyst $C_p(j)$ and cocatalyst *A* are:

$$\frac{dM}{dt} = -\sum_{j=1}^{N_s}\sum_{r=0}^{N_T}(k_p(j) + k_{tM}(j))MP_r(j) \quad (1d)$$

$$\frac{dH_2}{dt} = -\sum_{j=1}^{N_s}\sum_{r=1}^{N_T} k_{tH}(j)H_2P_r(j) \quad (1e)$$

$$\frac{dC_p(j)}{dt} = -k_{aA}(j)C_p(j)A \quad (1f)$$

$$\frac{dA}{dt} = -\sum_{j=1}^{N_s} k_{aA}(j)C_p(j)A \quad (1g)$$

Eqs. (1a)-(1g) is a 0-DOF system which contains $2N_TN_s + 2N_s + 3$ differential equations and $2N_TN_s + 2N_s + 3$ variables. The dynamic MWD is calculated by

$$w_r(t) = \frac{\sum_{j=1}^{N_s} r(P_r(j)+D_r(j))}{\sum_{r=1}^{N_T}\sum_{i}^{Ns} r(P_r(j)+D_r(j))} \quad (2)$$

The dynamic MWD model above is computational expensive due to the extremely large scale nature. The population balances can be reduced to a much smaller set of equations by applying moment method. However, the information about individual chains is sacrificed, thus it is not able to predict the full MWD. Though the Flory distribution can be derived at steady-state, it is only valid instantaneously. The conventional method for the dynamic MWD simulation accumulates the instantaneous distribution in short time intervals (see Soares et al. 2012). This integration process is also time consuming and some parallel computation techniques can help speed up (Weng et al. 2015). Moreover, this cumulative MWD method is not applied for dynamic optimization problems because of its nested calculation nature. To overcome these computational difficulties, we propose a reformulated population balance for dynamic MWD.

3. Dynamic Reduced Order Models

In this section, we present a kind of model reformulation based on null-space projection method. Among the polymerization reactions, the propagation rates $k_p$are significantly higher than those of the other rates. Thus, the living chains and dead chains showed quite different behaviors. The living polymer chains are fast consuming, and the life time is in the order of seconds. The dead polymer chains are slowly cumulative, and the life time can be long as the resistance time of the polymerization. The population model consequently leads to a two-time scale model and incurs numerical calculation issues. Given that the dynamic MWD is determine by the dead polymer chains, it is reasonable to separate the two-time scale behaviors.

The population balances in Eq (1) can be described by the following equation:

$$\dot{x} = \mathrm{A}r(x) \quad (3)$$

where $x \in R^{n_x}$ is the vector of component concentrations or populations, $r(x) \in R^{n_r}$ is the vector of reaction rates, A is the coefficient matrix. By partitioning the fast rate reactions, the following system is obtained:

$$\dot{x} = [A_1 \ A_2][r_1(x) \ \sigma r_2(x)]^T \quad (4)$$

where $\sigma r_2(x)$ represents the fast reaction rates, and σ is a large positive number that can approach infinity. A nullspace matrix Z is introduced such that $Z^TA_2 = 0$, and a corresponding matrix Y is also defined ensuring $[Y \ Z]$ to be nonsingular. Multiplying $[Y \ Z]^T$ to both sides of Eq. (4) gives

$$Y^T\dot{x} = Y^T A_1 r_1(x) + \sigma Y^T A_2 r_2(x) \tag{5a}$$

$$Z^T\dot{x} = Z^T A_1 r_1(x) \tag{5b}$$

Eq. (5b) does not include the fast reaction rates and is kept as part of the reformulation system. The matrices in Eq. (5a) is partitioned and arranged:

$$[Y_a^T \;\; Y_b^T]^T\dot{x} = [Y_a^T \;\; Y_b^T]^T A_1 r_1(x) + [0 \;\; \sigma f(x)]^T \tag{6}$$

where $f(x)$ is the non-zero elements in $Y^T A_2 r_2(x)$.

When $\sigma \to \infty$, $f(x) = 0$ is required to maintain the equation unaffected. This condition sketches the fast equilibrium manifold of the reaction system and contributes to part of the reformulated system. Finally, the reformulated system is derived:

$$Y_a^T\dot{x} = Y_a^T A_1 r_1(x) \tag{7a}$$

$$f(x) = 0 \tag{7b}$$

$$Z^T\dot{x} = Z^T A_1 r_1(x) \tag{7c}$$

By applying the nullspace projection method, the reformulated model of Eq. (1) can be derived after a sequence of matric operations. The reformulated mass balance that corresponds to Eq. (7a) are the monomer free sites:

$$\frac{dP_0(j)}{dt} = k_{aA}(j)C_p(j)A - k_d(j)P_0(j) \tag{8}$$

The fast equilibrium that corresponds to Eq (7b) are the living polymer chains, which can be

$$k_p(j)MP_0(j) - \big(k_{tH}(j)H_2 + k_{tA}(j)A + k_t(j)\big)Y(j) = 0 \tag{9}$$

$$k_p(j)M\big(P_r(j) - P_{r-1}(j)\big) - \big(k_{tH}(j)H_2 + k_{tA}(j)A + k_t(j)\big)P_r(j) = 0 \tag{10}$$

where $Y(j) = \sum_{r=1}^{N_T} P_r\,(j)$ is introduced as a new component.

The remained population balances that corresponds to Eq. (7c) are the slow cumulative dead polymer chains and the rest of the reactants. These balances remain the same by

substituting the $Y(j) = \sum_{r=1}^{N_T} P_r\,(j)$ definition into Eqs. (1c)-(1g).

The original population balance containing $2N_TN_s + 2N_s + 3$ equations in Eq. (1) is reduced to a DAE with $N_TN_s + 2N_s + 3$ differential equations and $N_TN_s + N_s$ algebraic equations. Moreover, it is worth to note that the decomposition of coupling in living chain lengths is accomplished by further substitution of Eq. (10):

$$P_r(j) = \left(\frac{k_p(j)M}{k_p(j)M + k_{tH}(j)H_2 + k_{tA}(j)A + k_t(j)}\right)^r P_0(j) \tag{11}$$

The dynamic MWD is calculated by adding another differential equation

$$\frac{dC}{dt} = \sum_{j=1}^{N_s}(k_p(j) + k_{tM}(j)MY \tag{12a}$$

$$w_r(t) = \frac{\sum_j^{N_s} rD_r(j)}{C} \tag{12b}$$

By combining the reformulated population balances Eqs. (1c)-(1g), (8), (9), (11) and (12), a chain-decoupling model is obtained. This model provides the property of discretized sampling in large chain intervals. One does not need to calculate all the chains to obtain a MWD profile. Thus, the reformulated model can figure out a MWD profile using $2\delta N_T N_s + \delta N_T + 3N_s + 4$ equations compared to the original model with $2N_T N_s + N_T + 2N_s + 3$ equations, where δN_T can be less than 100 but N_T can be 100000 or more.

4. Numerical Results

In this section, the proposed dynamic MWD model is validated by comparing with the conventional method in Weng et al. (2015). The conventional method accumulates the instantaneous Flory distribution in short intervals by applying pseudo steady state assumption. On the other hand, the proposed dynamic MWD model is transcript from DAEs to NLP and IPOPT is employed as the solver. The simulation for a grade transition process is presented. The number of active sites N_s is 5 in this case. The dynamic MWDs along time calculated by the two methods are compared in Figure 1.

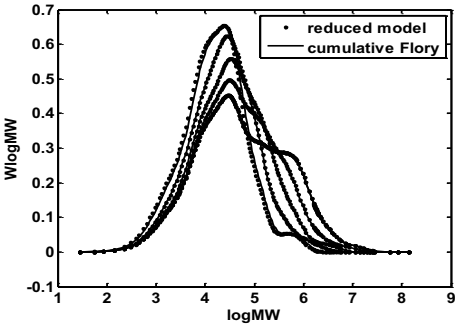


Figure 1. Dynamic MWDs along time with reduced model and cumulative Flory method

In the simulation case, the number of chains calculated in the reduced model were set as 40, i. e. $\delta N_T = 40$. These chains can be sampled very flexibly. In this case, the chain lengths are discretized evenly in logarithm scale. It indicates that the reduced model is predictive without essentially affects the performance. It achieved the same prediction performance as that of cumulative Flory method, which calculates over 50000 chains.
The model scale and computational efficiency of the reduced order model and cumulative Flory methods are further investigated. Despite the traditional serial computation of cumulative Flory method, parallel cumulative Flory method (Weng et al. 2015) is also tested for comparison.

Table 1. Model scale of the reduced order model and cumulative Flory method

Index	Chain calculated	Population balances	Moments	MWD equations
Reduced order model	40	213	-	41
Serial cumulative Flory method	50000	13	30	100000
Parallel cumulative method	50000	13	30	100000

Table 1. presents the equations that are solved in the reduce order model and cumulative Flory method. The reduced order model has the capability to solve a manageable number of population balances to derive the dynamic MWD. However, the cumulative Flory

method cannot deal with the large-scale population balances, and spends most of the computational efforts in MWD equations.

Table 2. presents the computational efficiency of reduced order model, serial cumulative Flory method, and parallel cumulative Flory method. Among the three methods, the reduced order model is the most efficient with a total cost of 12.1 s. It significantly reduced the time cost by more than 90% compared to the serial cumulative Flory method

Table 2. Computational efficiency of the three methods

Index	Thread	Moments (s)	MWD in sites (s)	Total (s)
Reduced order model	1	-	-	12.1
Serial cumulative Flory method	1	40.7	86.2	126.9
Parallel cumulative method	5	40.7	20.4	70.3

5. Conclusions

In this paper, a reduced order model is proposed for predicting MWD in dynamic systems. The original population balances are reformulated by applying null-space projection method. The fast equilibrium for living chains and slowly accumulative dead polymer chains are separated. Moreover, the coupling in chain lengths of the living polymers are decomposed. This decomposition significantly reduced the scale of necessary population balances to obtain the MWD. Numerical results indicate that the reduced model is predictive without essentially affects the performance. It significantly improves the computational efficiency and provides the prospective of real-time applications. Future work will focus on the on-line monitoring and optimal control based on MWD.

Acknowledgements

The authors gratefully acknowledge the financial support of National Key Research and Development Program of China (No. 2017YFE0106700) and NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization (No. U1509209). We would also like to thank Prof. Lorenz T. Biegler at Carnegie Mellon University for the constructive suggestions.

References

J. Kang, Z. Shao, X. Chen, X. Gu, L. Feng, 2017, Fast and reliable computational strategy for developing a rigorous model-driven soft sensor of dynamic molecular weight distribution, Journal of Process Control, 56, 79-99.

E. Mastan, S. Zhu, 2015, Method of moments: a versatile tool for deterministic modeling of polymerization kinetics, European Polymer Journal, 68, 139-160.

Y. Nie, L. T. Biegler, C. M. Villa, J. M. Wassick, 2013, Reactor modeling and recipe optimization of polyol processes: polypropylene glycol, AIChE Journal, 59, 2515-2529.

V. Saliakas, C. Chatzidoukas, A. Krallis, D. Meimaroglou, C. Kiparissides, 2007, Dynamic optimization of molecular weight distribution using orthogonal collocation on finite elements and fixed pivot methods: an experimental and theoretical investigation, Macromolecular reaction engineering, 1, 119-136.

J. B. P. Soares, T. F. L. McKenna, 2012, Polyolefin reaction engineering, Wiley-vch Verlag GmbH & Co. KGaA, Weinheim, Germany, 2012.

J. Weng, X. Chen, L. T. Biegler, 2015, A multi-thread parallel computational method for dynamic simulation of molecular weight distribution of multisite polymerization, Computers & Chemical Engineering, 82, 55-67.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50089-6

Implementation of a Radial Basis Function control strategy for the crystallization of Ibuprofen under uncertainty

Frederico Montes[a], Krist V. Gernaey[a], Gürkan Sin[a*]

[a]*PROSYS, Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts PladsBuilding 229, 2800 Kgs. Lyngby, Denmark*
gsi@kt.dtu.dk

Abstract

Batch crystallization is still one of the most commonly used processes in the pharmaceutical industry. Indeed, due to the flexibility of the process and the possibility of changing the product specification, the batch process is still the main choice. However, crystallization is a highly non-linear process and therefore difficult to model and simulate. Moreover, uncertainty in physical parameters and uncertainty propagation from previous process outputs has an important and critical influence on the process risk and therefore the control strategy needed to achieve robust process performance.

To this end, a two dimensional population balance model of ibuprofen batch crystallization is developed and used to analyse and study the operability space, where uncertainty is included. The uncertainty comes from previous synthesis steps and from reported uncertainty in the process parameters, described in the literature. A nonlinear control strategy is then applied, comparing different radial basis functions (quadratic, cubic, Gaussian…), in order to achieve the desired mean size by manipulating and updating the cooling profile of the same process.

The resulting three operation strategies are benchmarked with respect to key performance metrics (reference trajectory, deviation and input variation): open-loop, open-loop strategy under process uncertainties, and closed loop operation subject to process uncertainties. The final mean crystal size (MCS) is reported as a confidence interval, in order to be used for further downstream processing.

Keywords: modelling, Ibuprofen crystallization, process control, process uncertainty

1. Introduction

For traditional pharmaceutical industry, separation processes and strict time schedules are two common discussed topics. The former topic usually takes place in the form of batch units, mainly composed by distillations, filtrations and crystallizations amongst others. These operations are rather complex, expensive, and generate waste (such as solvents). Moreover, some of these steps might propagate deviations/disturbances for the next consecutive process unit, commonly referred as uncertainty in the process variables. New strategies are needed to the process understanding and drug development, such as the as the usage of principles of chemistry, engineering, material science, and quality control (Escotet-Espinoza et al. 2015). In addition, as encouraged by the U.S. Food and Drugs Administration (FDA), the use of Process Analytical Technology (PAT) tools to certify and control the end product quality and requirements

is of major importance for the whole process risk assessment. This leads to a necessity of implementing fast, reliable and easy to use control strategies and online monitoring, that takes into account or counters previous uncertainty. This is especially important when moving from lab-scale to mini-plant or pilot-plant, where time schedules are tight, and every day before a patent expiration is important, and valuable.

In this work, a control-strategy based on Radial Basis Functions Networks (Orr, 1996) is applied to a population balance Ibuprofen crystallization unit. This cooling crystallization operates in discontinuous mode, and the modelling parameters are adapted from available literature. The model construction and simulation is based in a Matlab/Simulink interface (The MathWorks®, Natick, MA), and the outputs are used to simulate the control strategy, that uses a reference trajectory of the mean size crystal to predict and actuate on the temperature profile, over the simulation time. The mean size is predicted using different Radial Basis Functions, evaluating them, and then re-simulate the system, with uncertainty in different parameters. We note that the objective of this work is not to provide the optimal control, but a fast and reliable method for upscaling processes when data is uncertain or missing.

2. Nomenclature

Symbol	Description	Units
C_{sat}	Saturation concentration	$g_{Sol}/g_{Solvent}$
k_{gx}	Growth rate constant in x direction	$\mu m/min.(g_{Sol}/g_{Solvent})$
G_x	Linear growth rate	μm/min
ρ_c	Crystal density	kg/m^3
ρ_w	Water density	kg/m^3
x_w	Water mass fraction	$g_{Water}/g_{Solvent}$
T_w	Serpentine water temperature	Celsius
$T_{w,in}$	Water Inlet temperature	Celsius
T_{cryst}	Crystallizer temperature	Celsius
C_{ibu}	Ibuprofen concentration	$g_{ibu}/g_{solvent}$
S_{sat}	Supersaturation	$g_{sol}/g_{solvent}$
$C_{p,i}$	Heat capacity	kJ/kg/K
ΔH_c	Heat of formation of crystal	kJ/kg
$V_{form,crystal}$	Volume of formed crystal	m^3
U	Heat transfer coefficient	$W/(m^2.K)$
A	Cooling area	m^2
$F_{w,in}$	Cooling water flow	m^3/s

3. Methodology

3.1. Model development

The population balance crystallization model is built using the method of classes, for two dimensions:

$$\frac{dN_{i,j}}{dt} + f_{i,j} - f_{i,j}^{"} = B_{nuc} \tag{1}$$

Where $f_{i,j}$ represents the inlet or outlet flux of crystals from nearby classes. For an inlet flux, for example:

$$f_{i,j} = \frac{G_{x,i-1}N_{i-1,j}}{2C_{x(i-1)}} + \frac{G_{y,j-1}N_{i,j-1}}{2C_{y(j-1)}} \tag{2}$$

Where C_x and C_y represent the discretization space, or the classes. Fort this population balance model (PBM), the linear growth rate described in the literature (Rashid et al. 2010). For instance, for the growth along the x characteristic length

$$G_x = k_x {S_{sat}}^n \tag{3}$$

$$S_{sat} = \max(0, \frac{C_{ibu} - C_{sat}}{C_{sat}}) \tag{4}$$

$$C_{sat} = 0.495 + 0.001026 {T_{cryst}}^2 \tag{5}$$

$$\frac{dT_{cryst}}{dt} = \frac{-\Delta H_c \rho_c V_{form,cryst} - UA(T_{cryst} - T_w)}{\rho_{magma} V c_{p,magma}} \tag{6}$$

$$\frac{dT_w}{dt} = \frac{F_{w,in}\rho_{water}C_{p,water}(T_{w,in} - T_w) + UA(T_{cryst} - T_w)}{\rho_{water}V_{water}c_{p,water}} \tag{7}$$

The birth rates were adapted from literature (Rashid et al. 2011), as well as the saturation profile for Ibuprofen in an ethanol/aqueous solution. Parameters for equation 6 were adapted from literature (Nayhouse et al. 2015)

3.2. Simulation

The assembled model is simulated with two cooling profiles: an initial linear decrease of temperature from 32 to 28 degrees (30 minutes), followed by an exponential decrease as stated in equation 9, until 12 degrees:

$$T_{w,in:t>30m} = 28 - (28 - 12)(\frac{(t - 30)}{Time_{simulation\ end} - 30})^{1.5} \tag{8}$$

This profile avoids secondary nucleation, while keeping the supersaturation levels high enough for a steady crystal growth. The resulting evolution of the mean crystal size is used for reference for the rest of this work.

3.3. RBF implementation

Radial Basis Functions (RBF) uses a series of basis functions that are symmetric and centered at each sampling point. The main feature of these functions is that their response decreases, or increases, monotonically with distance from a central point. The centre, the distance scale and the precise shape of the radial function are parameters of the model (Jin et al. 2002). Three different basis functions are tested for the RBF

network: Inverse Multiquadratic (IMQ), Thin Plate Spine (TPS) and Biharmonic (BH) (Jekabsons, 2009). Equation 9 represents the radial basis function node of TPS:

$$\Phi(dist) = dist\ \ln(dist) \tag{9}$$

3.4. Simulation with uncertainty

Selecting one of the different basis functions, the system is then simulated with different uncertainties (or disturbances in the initial conditions and parameters). The reference mean size of the crystal trajectory is the same as the undisturbed case study, in order to determine if the control system is able to detect and actuate on different conditions.

4. Results and discussion

The simulation of the crystallization of ibuprofen starts by providing the initial seed attributes, such as the mean size in both characteristic lengths, and the deviation The final obtained PSD has been used has reference trajectory for the RBF network (figure 1). The hidden layer of the RBF trains itself within the first half an hour of the process assuming a time delay of 3 minutes (enough time to sample and analyse the mean size of the crystals). Different RBF functions were tested and the results are shown in figure 2. From the observed results, the BH function is able to predict quite well the behaviour of the stem, and actuate on the temperature of the cooling profile in order to minimize the difference between the simulated and the reference mean size of the crystal. The early fluctuation is due to a small amount of training data, and the prediction is not as accurate as for the end of the simulation. By this time, the data available for training the RBF is vast (100 data points, assuming a 3 minute delay).

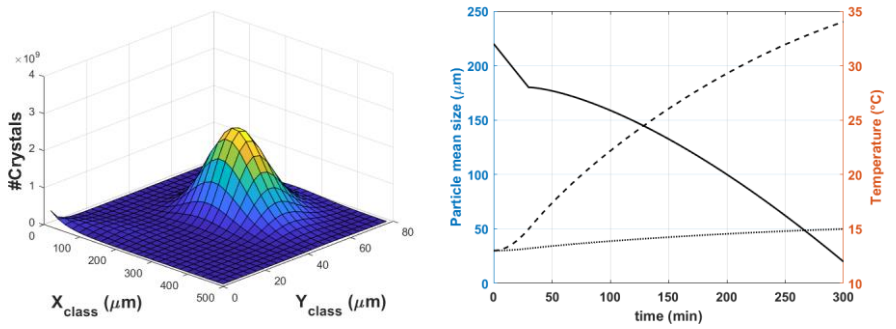


Figure 1- Final particle size distribution, at 5 hours of crystallization (left). Dynamic temperature profile over the process (full line, right), and the evolution of the mean sizes (dashed, right).

Using the BH RBF strategy, different simulations with deviations in process parameters (water content, initial seed mass) and uncertain parameters (k_g, k_b) were performed. Table 1 contains the results obtained for the 8 different simulations with uncertainty. Although the best approach to analyse the influence of this uncertainty would be to sample with Monte Carlo and propagate and quantify the uncertainty in the output (Sin et al 2009 et al), the results show that the RBH control is capable of reducing the off-spec MCS. For the example of the growth constant, figure 3 shows that by the end of the crystallization, the final mean size of the crystal is contained within the off-spec crystal in the open-loop simulation.

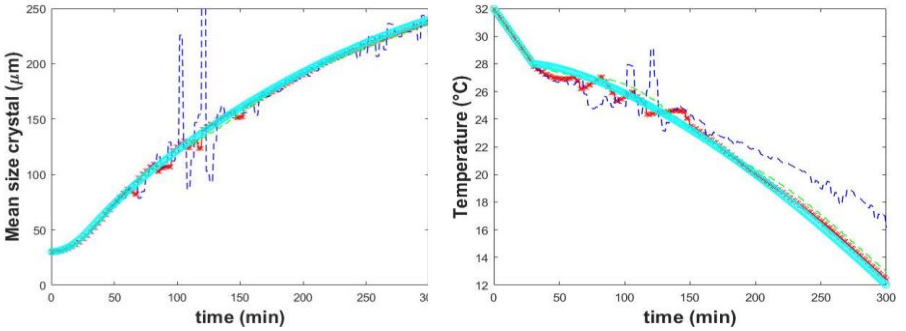


Figure 2- Mean particle size (left) and temperature profile predictions (right) of the different RBFs. (full: reference trajectory and temperature, dashed: IMQ RBF, remaining lines are similar to the reference trajectory)

Table1: Deviation from the mean size reference over time, due to uncertainty input

Input	Uncertainty	Mean size difference from ref. (μm)					Secondary nucleation
		t=1 h	t=2	t=3	t=4h	t=5h	
Water content	+50 %	0.80	-0.33	6.00	-2.86	1.25	no
	-50 %	-0.85	0.78	0.88	-0.85	0.21	no
Initial seed	+10%	-0.52	-2.93	-0.59	-1.61	-3.04	no
	-10%	0.58	4.51	8.01	1.43	-0.13	no
k_g	+20%	1.26	-5.25	1.49	-2.17	1.21	no
	-20%	-2.64	13.14	14.02	-6.14	0.34	yes
k_b	+65%	-0.21	-0.35	2.81	-1.47	<0.01	yes
	-65%	0.40	1.52	2.12	-0.40	0.39	no

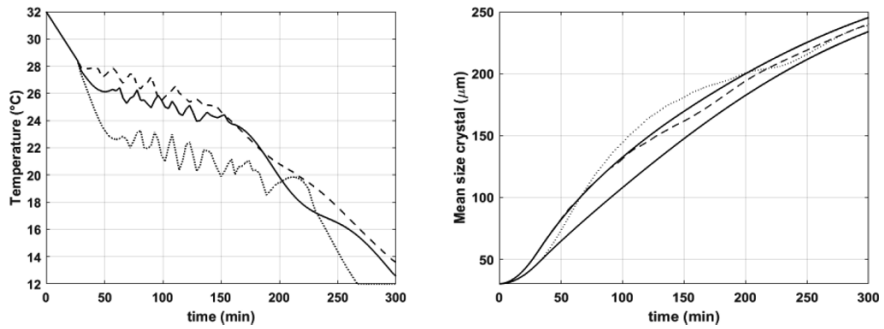


Figure 3- Temperature profiles (left) in closed loop for uncertainty in crystal growth (dashed: +20%, dotted: -20%, full: no disturbances) and the mean size evolution of crystals (right) under disturbance when in open-loop and closed loop (full: open loop for +20% and -20%, dashed:+20%, dotted: -20%,).

However, the problem was not posed to contain or avoid secondary nucleation. It is assumed that this data is unknown when moving from lab-scale to mini-plant. For the simulation, however, it was adapted from the literature (Rashid et al, 2010).

5. Conclusions

A control strategy using RBF network has been in an Ibuprofen crystallization model. It was shown that this is a reliable method to quickly move from smaller scales to mini-plant or micro-plant, when measurement (PAT) tools are available. The results show a good rejection of the disturbances made to the system, in the form of initial conditions of the batch and uncertain in critical parameters. The uncertainty contained in certain parameters replicates the case when available data from laboratory is not enough to have a good understanding of the process. Nevertheless, it is important to refer that this is not the optimal control strategy, as RBF is not trained on process input and output data generated from an optimal control (such as nonlinear model predictive control). The objective here is to show the ability of the RBF based control concept which can be trained using online measurements and which does not need a model to calculate control actions. Further work and development includes training of RBF to replace NMPC and laboratory validation of the control on a crystallisation unit.

6. Acknowledgments

This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement no.675251

References

M. S. Escotet-Espinoza, R. Singh, M. Sen, T. O'Connor, et al., 2015, Flowsheet Models Modernize pharmaceutical Manufacturing Design and Risk Assessment, Pharmaceutical Technology, 39, Issue 4.

G. Jekabsons, 2009, Radial Basis Function interpolation, available at http://www.cs.rtu.lv/jekabsons/regression.html

R. Jin, W. Chen, A. Sudjianto, 2002, On sequential sampling for global metamodeling in engineering design, Proceedings of ASME Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Montreal

M. Nayhouse, A. Tran, P. D. Christofides et al., 2015, Modeling and control of ibuprofen crystal growth and size distribution, Chemical Engineering Science 134

M. Orr, 1996, Introduction to Radial Basis Function Networks, Centre for Cognitive Science.

A. Rashid, E. White, I. Marziano et al., 2010, Growth rates of ibuprofen crystals grown from ethanol, Proc chemeca, Adelaide, paper 377

A. Rashid, E. White, I. Marziano et al., 2011, Nucleation Kinetics fo iburpofen crystals grown from aqueous ethanol, Proc chemeca, Sydney, paper 107

G. Sin, K. Gernaey,A. Lantz, 2009, Good modelling practise for PAT applications: Propagation of input uncertainty and sensitivity analysis, Biotechnology progress 25 (4).

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50090-2

Coordination of distributed MPC systems using a nonlinear dynamic plant model with closed-loop prediction

Hao Li[a], Christopher L.E. Swartz[a*]

[a]*Department of Chemical Engineering, McMaster University, 1280 Main Street West, Hamilton, L8S 4L7, Canada*
swartzc@mcmaster.ca

Abstract

This paper presents a novel coordination scheme for distributed MPCs using a dynamic real-time optimization (DRTO) formulation with closed-loop prediction and a nonlinear dynamic plant model. The DRTO formulation generates the predicted closed-loop response of a nonlinear plant under the action of constrained distributed MPC for computing optimal set-point trajectories based on an economic or target-tracking performance objective. The performance of the formulation is assessed through evaluation on a nonlinear case study, with comparisons made between distributed, centralized and decentralized configurations.

Keywords: dynamic real-time optimization, economic optimization, distributed MPCs, coordination, nonlinear.

1. Introduction

Model predictive control (MPC) utilizes a dynamic model to predict the future response of a plant, and computes the optimal control action as the solution of a suitably formulated optimization problem. The inputs corresponding to the first control interval are implemented, and the process repeated with a correction based on the most recent plant measurement. Due to its constraint-handling ability and flexibility, MPC has been widely adopted in the process industry as well as other industrial sectors. However, application of a single MPC system to a large industrial plant presents challenges that include multiple time scales, and reliability considerations of having all plant sections reliant on a single advanced controller. MPC is consequently typically implemented in a distributed manner in large industrial plants, which has motivated research studies on paradigms and algorithms for a distributed MPC architecture, recent reviews of which are given in Pannocchia (2013) and Christofides et al. (2013).

The approach in this paper utilizes a hierarchical dynamic real-time optimization (DRTO) approach in which the DRTO layer provides economically optimal set-point trajectories to an underlying MPC system, as proposed by Jamaludin and Swartz (2017b) for centralized MPC. A key feature of the approach is that the DRTO calculation is based on the predicted closed-loop response of the plant and its associated MPC system. This results in a multilevel optimization problem that is solved by replacing the MPC optimization subproblems by algebraic constraints corresponding to their first-order optimality conditions. More recently, the formulation has been extended to coordination of multiple distributed MPCs (Li and Swartz, 2017), using linear dynamic plant models for both the MPC and DRTO formulations. In this work, we extend our earlier approach

by using a nonlinear dynamic plant model at the DRTO level so that the plant response generated can more accurately reflect the true plant dynamics. In the DRTO formulation, the interaction between distributed MPCs and process plant is fully captured and predicted in a rigorous fashion, and set-point trajectories are generated for each MPC subsystem. The performance demonstrated by the nonlinear DRTO formulation shows significant improvement compared to that of the linearized counterpart. The formulation is also versatile in that it can be utilized for the purpose of both target tracking and economic optimization. The formulation of the DRTO problem is described, and the performance of the method is illustrated through application to a reactor-separator case study.

2. Problem Formulation

The general architecture of the two-layer coordination scheme for distributed MPCs is shown in Figure 1. A nonlinear dynamic plant model is embedded in the upper layer as a surrogate model to generate the plant response. Along the DRTO prediction horizon, a series of MPC optimization subproblems are embedded and generate the control actions applied to the nonlinear dynamic model at each time instance. This sequential closed-loop prediction is thus optimized under a typically economic objective function, and generates set-point trajectories for lower level distributed MPCs to perform subsequent control tracking.

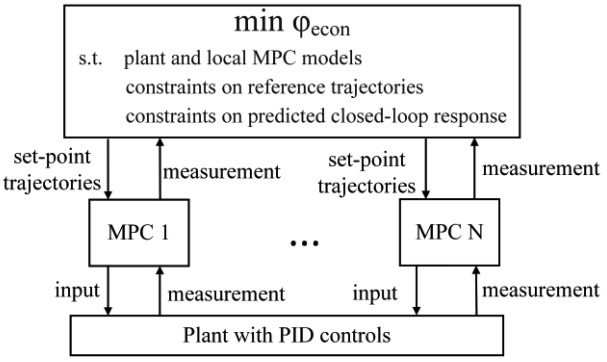


Figure 1: The control architecture for coordination of distributed MPCs with closed loop DRTO approach.

2.1. Model Discretization

Consider the nonlinear dynamic behaviour of the process plant to be accurately captured by a set of ODEs in the following form:

$$\begin{aligned} \dot{\mathbf{x}}^{\mathrm{DRTO}}(t) &= \mathbf{f}^{\mathrm{DRTO}}(\mathbf{x}^{\mathrm{DRTO}}(t), \mathbf{u}^{\mathrm{DRTO}}(t)) \\ \mathbf{y}^{\mathrm{DRTO}}(t) &= C\mathbf{x}^{\mathrm{DRTO}}(t) \\ \mathbf{x}^{\mathrm{DRTO}}(0) &= \mathbf{x}_0, \qquad \text{for } t \in [0, t_f] \end{aligned} \tag{1}$$

where $\mathbf{x}^{\mathrm{DRTO}}(t) \in R^{n_x}$ denotes the differential state vector, $\mathbf{x}^{\mathrm{DRTO}}(0) \in R^{n_x}$ the initial state vector, and $\mathbf{u}^{\mathrm{DRTO}}(t) \in R^{n_u}$ the input vector. t_f represents the duration of the DRTO prediction horizon. To solve the ODE system in the DRTO formulation, the system is discretized using an implicit Euler approximation. The controller sample time is assumed to be consistent with the finite element interval Δt_j and the piecewise constant inputs are applied to each finite element along the DRTO prediction horizon. There are various other discretization techniques available to achieve the same effect, such as orthogonal collocation on finite elements. However, implicit Euler integration is demonstrated to be sufficiently accurate to reflect the dynamic behaviour generated by the actual plant for

the case study considered. The plant is partitioned into subsystems $i \in O$, each of which is controlled by a local MPC based on a linear (or linearized) model of the form $\mathbf{x}_{k+1}^{(i)} = A_{ii}\mathbf{x}_k^{(i)} + B_{ii}\mathbf{u}_k^{(i)}$, $\mathbf{y}_k^{(i)} = C_{ii}\mathbf{x}_k^{(i)}$.

2.2. State and Disturbance Estimation for Nonlinear Plant Model

Prior to each DRTO execution, the current state estimate $\hat{\mathbf{x}}_j$ is first calculated based on the discretized model with previous state estimate $\hat{\mathbf{x}}_{j-1}$and the most recently implemented control action $\mathbf{u}_{j-1}^{\mathrm{MPC}}$. The predicted state estimate $\hat{\mathbf{x}}_j$ serves as the initial condition $\mathbf{x}_0$ for the DRTO model in the upper level. The disturbance added to the DRTO model is calculated by taking the difference between the current measured outputs and predicted outputs and assumed constant over the horizon:

$$\begin{aligned} \frac{\hat{\mathbf{x}}_j - \hat{\mathbf{x}}_{j-1}}{\Delta t_j} &= \mathbf{f}^{\mathrm{DRTO}}(\hat{\mathbf{x}}_j, \mathbf{u}_{j-1}^{\mathrm{MPC}}) \\ \mathbf{d}_j^{\mathrm{DRTO}} &= \mathbf{y}^{\mathrm{m}} - C\hat{\mathbf{x}}_j \\ \mathbf{x}_0 &= \hat{\mathbf{x}}_j \end{aligned} \tag{2}$$

2.3. Optimization Formulation

The dynamic models for the primary optimization problem and MPC optimization subproblems are embedded in the DRTO formulation as Eq.(3):

$$\begin{aligned} \min_{\mathbf{y}^{\mathrm{ref}},\mathbf{u}^{\mathrm{ref}}} \quad & \phi^{\mathrm{econ}}(\mathbf{x}^{\mathrm{DRTO}}, \mathbf{y}^{\mathrm{DRTO}}, \mathbf{u}^{\mathrm{DRTO}}) \\ \text{s.t.} \quad & \frac{\mathbf{x}_j^{\mathrm{DRTO}} - \mathbf{x}_{j-1}^{\mathrm{DRTO}}}{\Delta t_j} = \mathbf{f}^{\mathrm{DRTO}}(\mathbf{x}_j^{\mathrm{DRTO}}, \mathbf{u}_{j-1}^{\mathrm{DRTO}}), \quad j = 1, \ldots, N \\ & \mathbf{y}_j^{\mathrm{DRTO}} = C\mathbf{x}_j^{\mathrm{DRTO}} + \mathbf{d}_j^{\mathrm{DRTO}}, \quad j = 1, \ldots, N \\ & \mathbf{g}^{\mathrm{DRTO}}(\mathbf{x}^{\mathrm{DRTO}}, \mathbf{y}^{\mathrm{DRTO}}) \geq \mathbf{0} \\ & \mathbf{h}_i^{\mathrm{ref}}(\mathbf{y}^{\mathrm{ref}}, \mathbf{u}^{\mathrm{ref}}, \mathbf{y}^{\mathrm{sp}(i)}, \mathbf{u}^{\mathrm{sp}(i)}) = \mathbf{0}, \quad i \in O \\ & \mathbf{g}^{\mathrm{ref}}(\mathbf{y}^{\mathrm{ref}}, \mathbf{u}^{\mathrm{ref}}) \geq \mathbf{0} \\ & \mathbf{u}_j^{\mathrm{DRTO}(i)} = \mathbf{u}_{j,0}^{(i)}, \quad j = 0, \ldots, N-1 \\ & \left.\begin{aligned} & \mathbf{u}_{j,0}^{(i)} \in \arg \min_{\mathbf{u}_{j,k}^{(i)}} \phi_j^{\mathrm{MPC(i)}}(\mathbf{y}_{j,k}^{(i)}, \mathbf{y}_{j,k}^{\mathrm{sp}(i)}, \mathbf{u}_{j,k}^{(i)}, \mathbf{u}_{j,k}^{\mathrm{sp}(i)}) \\ & \text{s.t.} \quad \mathbf{x}_{j,k+1}^{(i)} = A_{ii}\mathbf{x}_{j,k}^{(i)} + B_{ii}\mathbf{u}_{j,k}^{(i)}, \quad k = 0, \ldots, m-1 \\ & \qquad \mathbf{x}_{j,k+1}^{(i)} = A_{ii}\mathbf{x}_{j,k}^{(i)} + B_{ii}\mathbf{u}_{j,m-1}^{(i)}, \quad k = m, \ldots, p-1 \\ & \qquad \mathbf{y}_{j,k}^{(i)} = C_{ii}\mathbf{x}_{j,k}^{(i)} + \mathbf{d}_{j,k}^{(i)}, \quad k = 1, \ldots, p \\ & \qquad \mathbf{d}_{0,k}^{(i)} = \mathbf{y}^{\mathrm{m}(i)} - C_{ii}\mathbf{x}_0^{(i)}, \quad k = 1, \ldots, p \\ & \qquad \mathbf{d}_{j,k}^{(i)} = \mathbf{y}_j^{\mathrm{DRTO}(i)} - C_{ii}\mathbf{x}_{j-1,1}^{(i)}, \quad k = 1, \ldots, p, \; j > 0 \\ & \qquad \mathbf{u}_{\min}^{(i)} \leq \mathbf{u}_{j,k}^{(i)} \leq \mathbf{u}_{\max}^{(i)}, \quad k = 0, \ldots, m-1 \end{aligned}\right\} \begin{array}{l} \text{MPC subproblems} \\ \text{subproblem } i \in O \\ j = 0, \ldots, N-1 \end{array} \end{aligned} \tag{3}$$

The closed-loop DRTO formulation consists of a primary optimization problem based on the nonlinear dynamic system to predict closed-loop response, and inner MPC optimization subproblems based on linear models of the plant subsystems to compute control actions. ϕ^{econ} denotes the objective function that captures the performance

criterion to be optimized. The main decision variables for the primary optimization problem are reference trajectories $\mathbf{y}^{\text{ref}}$ and $\mathbf{u}^{\text{ref}}$, from which the set-point trajectories for distributed MPCs $\mathbf{y}_{j,k}^{\text{sp}(i)}$and $\mathbf{u}_{j,k}^{\text{sp}(i)}$are extracted, while the decision variables in the inner MPC optimization subproblems are the control actions $\mathbf{u}_{j,k}^{(i)}$. $\mathbf{h}_i^{\text{ref}}$ represents the mapping of the DRTO reference trajectories to the MPC set-point trajectories, and can be used to enforce the trajectories to be constant within each DRTO interval over the DRTO prediction horizon, which can be explicitly represented as:

$$\begin{aligned} \mathbf{y}_{j,k}^{\text{sp(i)}} &= E_y^{(i)}\mathbf{y}_{j+k}^{\text{ref}}, \quad k=1,\ldots,p \\ \mathbf{u}_{j,k}^{\text{sp(i)}} &= E_u^{(i)}\mathbf{u}_{j+k}^{\text{ref}}, \quad k=0,\ldots,m-1 \end{aligned} \tag{4}$$

where $E_y^{(i)} \in R^{n_y^{(i)}\times n_y}$ and $E_u^{(i)} \in R^{n_u^{(i)}\times n_u}$ are matrices that map the full plant outputs and inputs to the subsystems. $\mathbf{g}^{\text{ref}}$ represents the upper and lower bounds of the trajectories. $\mathbf{x}^{\text{DRTO}}$, $\mathbf{y}^{\text{DRTO}}$, and $\mathbf{u}^{\text{DRTO}}$ are composite vectors of $\mathbf{x}_j^{\text{DRTO}}$, $\mathbf{y}_j^{\text{DRTO}}$, and $\mathbf{u}_j^{\text{DRTO}}$ respectively over the DRTO horizon. $\mathbf{u}_j^{\text{DRTO(i)}}$and $\mathbf{y}_j^{\text{DRTO(i)}}$ are subvectors of $\mathbf{u}_j^{\text{DRTO}}$and $\mathbf{y}_j^{\text{DRTO}}$ corresponding to the inputs and outputs of MPC subsystem i. A quadratic MPC objective function is used to yield a QDMC formulation as proposed in Garcia and Morshedi (1986) (see also Maciejowski (2002)). The disturbance estimate $\mathbf{d}_{j,k}^{(i)}$ is determined as proposed in Cutler and Ramaker (1980) and Garcia and Morshedi (1986), and utilizes the plant measurement $\mathbf{y}^{\text{m}(i)}$ for the first set of MPC subproblems, and the DRTO model output, $\mathbf{y}_j^{\text{DRTO(i)}}$, for subsequent time steps in the DRTO horizon, where $\mathbf{y}^{\text{m}(i)}$ is a subvector of the plant measurements corresponding to MPC subsystem i.

The multi-level optimization formulation of the closed-loop DRTO problem can be transformed into a single-level optimization problem by reformulating the MPC quadratic programming subproblems as a set of constraints using the Karush-Kuhn-Tucker (KKT) optimality conditions, as proposed by Baker and Swartz (2008). This correspondence for a general convex quadratic programming problem is:

$$\begin{array}{ll} \text{QP} & \text{KKT conditions} \\ \min_{\mathbf{z}} \ \frac{1}{2}\mathbf{z}^T H\mathbf{z} + g^T\mathbf{z} & H\mathbf{z}+\mathbf{g}-A^T\nu-\eta=\mathbf{0} \\ \text{s.t.} \quad A\mathbf{z}=\mathbf{b}, \quad \mathbf{z}\geq\mathbf{0} & A\mathbf{z}=\mathbf{b}, \quad z_i\eta_i=0, \quad (\mathbf{z},\eta)\geq\mathbf{0} \end{array} \tag{5}$$

where variable $\mathbf{z}$ denotes a generalized variable within the QP problem. The presence of the complementarity constraints in the reformulated problem may be effectively handled through an exact penalty formulation in which they are included as a penalty term in the objective function (Ralph and Wright, 2004; Jamaludin and Swartz, 2017a). The DRTO formulation are solved using AMPL with the IPOPT solver, and the lower-level control calculations and continuous plant simulation are implemented in MATLAB using the ode15s integrator.

3. Case Study

We consider a reactor-separator network studied by Baldea and Daoutidis (2007). A zero order reaction converts component A to component B in the reactor, and the product is then fed to the separator. The mole transfer rate between the vapour and liquid for component $i \in \{A,B,I\}$ in the vessel is modelled with the rate expression $N_j = K_j\alpha(y_j - P_j^s/P_c)x_j)(M_L/\rho_L)$. The differential equations used for this system are given below,

where M_R, M_V, M_L denote the molar holdup in the reactor, condenser vapour and liquid phase, and $N = N_A + N_B + N_I$. For simplicity, the molar hold-ups are assumed to be perfectly controlled.

$$\begin{aligned}
\dot{y}_{A,R} &= \tfrac{1}{M_R}[F_o(y_{A,o} - y_{A,R}) + R(y_A - y_{A,R}) - k_1 M_R y_{A,R}] & \dot{y}_I &= \tfrac{1}{M_V}[F(y_{I,R} - y_I) - N_I + y_I N] \\
\dot{y}_{I,R} &= \tfrac{1}{M_R}[F_o(y_{I,o} - y_{I,R}) + R(y_I - y_{I,R})] & \dot{x}_A &= \tfrac{1}{M_L}[N_A + x_A N] \\
\dot{y}_A &= \tfrac{1}{M_V}[F(y_{A,R} - y_A) - N_A + y_A N] & \dot{x}_I &= \tfrac{1}{M_L}[N_I + x_I N]
\end{aligned} \tag{6}$$

The nominal values for all process parameters are summarized in Baldea and Daoutidis (2007). The manipulated and control variables selected for the reactor are raw feed F_o and vapor composition of component A, $y_{A,R}$, respectively. The manipulated and controlled variables selected for the separator are condenser pressure P_c and liquid composition of component A, x_A, respectively. The tuning parameters for distributed MPCs and the DRTO formulation are summarized in Table 1. It is noted that the first move suppression weight value is small due to the small nominal value of the condenser pressure (0.338 MPa compared to feed flow rate 100 mol/min). In this case study, the system is given pre-determined targets for two outputs that are required to be tracked; the DRTO objective function is consequently formulated in a least-squares fashion. However, other forms of linear or nonlinear objective function can be readily used. The system is regulated separately for both centralized and decentralized MPC. For each case, the associated DRTO formulation is added to generate set-point trajectories for those MPCs. The performance is quantified using the sum-squared-errors (SSE) between the plant dynamics and final target values. The performance loss is defined as the percentange increase of SSEs with respect to the case with the lowest SSE value. Figure 2 shows the set-point trajectories and response of the continuous process dynamics for distributed cases using linear and nonlinear DRTO formulations. While the difference between the nonlinear and linear DRTO implementations appears slight, numerical results confirm the superior performance of the nonlinear approach. Table 2 summarizes the performance for several configurations: centralized and distributed DRTO with a nonlinear DRTO plant model, centralized and distributed DRTO with a linear DRTO plant model, and decentralized MPC with constant set-point trajectories corresponding to the set-point targets. From the results it can be seen that the nonlinear DRTO formulation outperforms the linear counterpart, which aligns with expectation since the DRTO formulation can generate a more accurate plant response based on the nonlinear model. The table also shows that the distributed cases experience only a slight performance drop when compared to the centralized counterparts due to the ability of DRTO formulation to predict the plant behavior based on a complete plant model that is not available to distributed MPCs. The performance drop can be justified by the additional flexibility and reliability offered by the distributed architecture over the centrazlied counterpart.

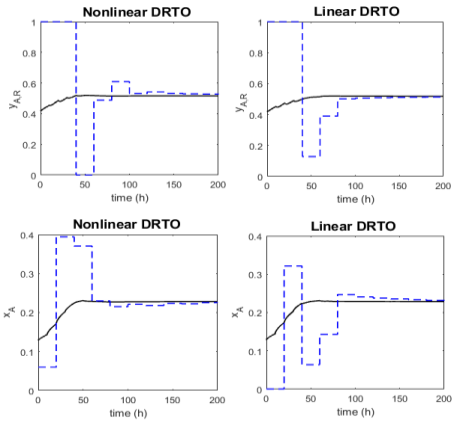


Figure 2: Set-points and output response for distributed MPC cases

Table 1: DRTO and MPC parameters for the process

Parameter	Description	Value (MPC1 & MPC2)
Δt_{DRTO}	DRTO sample time	20
Δt_{MPC}	MPC sample time	10
N	Optimization horizon	10
P	Prediction Horizon	5
M	Control Horizon	3
Q	Output tracking weight	10 & 2
R	Move suppression weight	10^{-4} & 5
y_{min}, y_{max}	Output and setpoint bounds	[0 1] & [0 1]
u_{min}, u_{max}	Input bounds	[0 250] & [0 1]

Table 2: The summary of performance under different configurations

Configuration	Control architecture	$\sum SSE$	performance loss
centralized	nonlinear DRTO + MPC	2.473	0%
distributed	nonlinear DRTO + MPC	2.556	3.36 %
centralized	linear DRTO + MPC	2.743	10.91%
distributed	linear DRTO + MPC	2.790	12.82%
decentralized	MPC	6.61	167.29%

4. Conclusion

The DRTO formulation is able to capture and predict the closed-loop behaviour of the plant response. Based on previous work of using a linear plant model in the DRTO formulation to generate plant responses, this paper demonstrates that the performance can be further improved when a nonlinear dynamic plant model is included in the DRTO formulation. The plant response is more accurate than that using a linear model, and consequently the set-point trajectories generated by the nonlinear DRTO strategy offer better target tracking capabilities for the lower-level distributed MPCs.

References

Baker, R., Swartz, C. L. E., 2008. Interior point solution of multilevel quadratic programming problems in constrained model predictive control applications. Ind. Eng. Chem. Res. 47, 1, 81.
Baldea, M., Daoutidis, P., 2007. Control of integrated process networks - A multi-time scale perspective. Comput. Chem. Eng. 31 (5), 426.
Christofides, P. D., Scattolini, R., Munoz de la Pena, D., Liu, J., 2013. Distributed model predictive control: A tutorial review and future research directions. Comput. Chem. Eng., 51, 21.
Cutler, C. R., Ramaker, B. L., 1980. Dynamic matrix control: a computer control algorithm. In: Joint automatic control conference., 72.
Garcia, C. E., Morshedi, A., 1986. Quadratic programming solution of dynamic matrix control (QDMC). Chem. Eng. Comm. 46 (1-3), 73.
Jamaludin, M. Z., Swartz, C. L. E., 2017a. Approximation of closed-loop prediction for dynamic real-time optimization calculations, Comput. and Chem. Eng., 103, 23.
Jamaludin, M. Z., Swartz, C. L. E., 2017b. Dynamic real-time optimization with closed-loop prediction. AIChE J., 63 (9), 3896.
Li, H., Swartz, C. L. E., 2017. Coordination of distributed mpc systems through dynamic real-time optimization with closed-loop prediction. In: 27th European Symposium on Computer Aided Process Engineering, 40,1603.
Maciejowski, J. M., 2002. Predictive control with constraints. Pearson education.
Pannocchia, G., 2013. Distributed Model Predictive Control. Springer London, London, p.1.
Ralph, D., Wright, S. J., 2004. Some properties of regularization and penalization schemes for MPECs. Optim. Methods Softw. 19, 5, 527.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50091-4

Optimization and Low-Level Control Design for Reactive Batch Distillation Columns including the Start-up

Alejandro Marquez-Ruiz[a*], J.H.A. Ludlage[a], Leyla Özkan[a]

[a] *Eindhoven University of Technology, 5612 AJ, Eindhoven, The Netherlands*
a.marquez.ruiz@tue.nl

Abstract

In this paper, the optimization and low-level control design for Reactive Batch Distillation Columns (RBD) including startup operation are investigated. This is a complex problem involving control of heat and mass transfer phenomena over a wide operating window. A new model for the startup operation of an RBD column is proposed. Using this model, an offline optimization problem is solved to minimize the start-up time. Additionally, the classical low-level controllers for RBD columns are reformulated, and a new cascade control strategy with decouplers is proposed. This control structure allows reducing the strong interaction among the input-output variables of the RBD. Finally, the model, the optimal startup, and the proposed controllers are tested on simulation on a polyesterification process.

Keywords: Reactive Batch Distillation Columns, Start-up, Low-Level Control, Optimal operation.

1. Introduction

Reactive distillation is a process in which a chemical reaction and distillation occur simultaneously in one single installation. A typical process operation of reactive distillation columns is batch since batch distillation is used in the chemical industry for the production of small amounts of products with high added value. Reactive batch distillation has gained a growing interest in both academic research and industrial applications since process limitations can be circumvented and investment and operational costs can be reduced by combining the reaction and separation processes.

There are basically three low-level control strategies for RBD: one-point bottom control, one-point column control and two-point control (Sørensen and Skogestad, 1994). In one-point bottom control strategy, the main objective is to maintain the reactor temperature close to the boiling point of the mixture (set-point). However, this set-point value may be difficult to know a priori, mostly in multi-component systems. In two-point control strategy (control of both reactor temperature and distillate composition), the strong interactions among the main variables (temperature and concentration) are the main drawback, and the reactor temperature has to be specified close to the boiling point. Last but not least, the one-point column control strategy (controlling the temperature on a tray in the column) gives a good performance for the RBD columns. However, this control strategy is controlling indirectly the most important variables of the process, and it is very sensitive to measurement delays and noise.

An important fact to be considered in the control design and optimization of the RBD columns is the start-up of the process. The start-up of batch distillation columns is a difficult operation in the chemical industry since it takes a long period leading to off-spec products and energy costs (Elgue et al., 2004). In general, modeling start-up phase where the thermodynamic equilibrium is not yet achieved is challenging. Few rigorous models for distillation columns start-up are available in the literature, and these models generally require a lot of parameters (Wozny and Li, 2004). Therefore, the inclusion of the startup in the classic control strategies is still a big challenge.

This paper is organized as follows: In Section 2 a new model for the startup operation for an RBD column is proposed. Using this model, in Section 3 the startup time is minimized by solving an offline optimization problem. In Section 4 the classical low-level controllers for RBD are reformulated, and a new decoupled-cascade control strategy is proposed. In Section 5, the model, the optimal startup, and the proposed controllers are tested on simulation on a polyesterification process. Finally, in Section 6 the conclusions are presented.

2. Modelling of RBD including the Start-up

2.1. Dynamic Modelling

Consider an RBD column with N_c species, R reactions and N_T stages where the reboiler stage (bottom) is the reactor. Three basic assumptions are applied in the formulation of the model: Perfect mixing and equilibrium between vapor and liquid on all stages, the accumulation of moles in the vapor phase is considered negligible, and chemical reactions in the vapor phase are neglected. A general mathematical model of the RBD is given by the following equations,

$$\frac{dn_{i,j}}{dt} = L_{i-1}x_{i-1,j} - L_i x_{i,j} - \xi_{i,j} + \xi_{i+1,j} + n_{Ti} r_{i,j},\ x_{i,j} = \frac{n_{i,j}}{n_{Ti}},\ n_{Ti} = \sum_{j=1}^{N_c} n_{i,j} \qquad (1)$$

$$\frac{d\left(n_{Ti}H_i\right)}{dt} = L_{i-1}H_{L_{i-1}} - L_i H_{L_i} - \xi_i H_{v_i} + \xi_{i+1}H_{v_{i+1}} + n_{Ti}\sum_{j=1}^{N_c}\left(-\Delta H_j\right)r_{i,j} + Q_{ext} \qquad (2)$$

$\forall i = 1,2,\cdots,N_T, \forall j = 1,2,\cdots,N_C$. In the model $y_{i,j}$, $\xi_{i,j}$, H_i, and H_{V_i} are given by:

$$\xi_{i,j} = V_i\, y_{i,j},\ y_{i,j} = \frac{\gamma_{i,j}P_{i,j}^{sat}(T_i)x_{i,j}}{\sum_{j=1}^{N_c}\gamma_{i,j}P_{i,j}^{sat}x_{i,j}},\ \xi_i = \sum_{j=1}^{N_c}\xi_{i,j},\ H_i = C_{p_i}T_i, H_{V_i} = C_{p_i}T_i + \lambda_{Vi}$$

Table 1: Boundary conditions of the RBD

	L_{i-1}	L_i	$\xi_{i,j}$	$\xi_{i+1,j}$	$r_{i,j}$	Q_{ext}	$x_{i-1,j}$
$i=1$	L_D	0	0	0	0	0	$x_{i+1,j}$
$i=N_T$	L_{i-1}	0	$V_i y_{i,j}$	0	$r_{i,j}$	Q_{ext}	$x_{i-1,j}$

The model given by (1)-(2) are subject to the boundary conditions in Table 1. To highlight that, the initial conditions of the model (1)-(2) assume an amount of vapor due to the equilibrium relationship $\xi_{i,j}$. This term depends on the concentration in the liquid phase. Therefore, this model is representative just for the case when the system has reached the VLE. Also, the boiling temperature of the mixture can be calculated solving the following

equation for the boiling temperature T_{B_i}, where $P_{i,j}^{sat}(T_{B_i})$ is given by the Antoine equation: $f_{T_{B_i}}(T_{B_i}) = \sum_{j=1}^{N_C} \gamma_{i,j} x_{i,j} P_{i,j}^{sat}(T_{B_i}) - P_i = 0$

2.2. Modelling of the start-up operation

In modelling of the start-up operation, the vapor flow behavior before and during the VLE must be understood. In general terms, the vapor flow has a kind of switch behavior, i.e., before the VLE it is zero, and after, it takes a value. First, the calculation of the vapor flow can be done assuming the temperature in steady state during the VLE condition. Hence the vapor flow can be calculated from the energy balance using the following equation,

$$V_i = H_{v_i}^{-1}\left(L_{i-1}H_{L_{i-1}} - L_i H_{L_i} + H_{v_{i+1}}V_{i+1} + n_{Ti}\sum_{j=1}^{N_c}\left(-\Delta H_j\right) r_{i,j} + Q_{ext} \right), \forall i = 1,\ldots,N_T \quad (3)$$

To include the switching behavior in the vapor flow, the next sigmoidal functions are used: $S(\alpha,\beta) = \left(1 + e^{-\alpha(x_i-\beta)}\right)^{-1}$. With such function in the Eq.(3), the vapor flow can be calculated using the following equation,

$$V_i = \omega_{V_i} H_{v_i}^{-1}\left(\omega_{L_i} L_{i-1}H_{L_{i-1}} - L_i H_{L_i} + H_{v_{i+1}}V_{i+1} + n_{Ti}\sum_{j=1}^{N_c}\left(-\Delta H_j\right) r_{i,j} + Q_{ext} \right), \forall i = 1,\ldots,N_T \quad (4)$$

Where $\omega_{V_i} = S(\alpha_{T_{B_i}}, T_{B_i})$, and $\omega_{L_i} = S(\alpha_{T_i}, T_{i-1})$ are switch functions with $\alpha_{T_{B_i}} > 0$ and $\alpha_{T_i} < 0$. The function ω_{V_i} activates the vapor flow when the temperature of the *i*-th tray reaches the boiling point and ω_{L_i} is not active when the temperature of the *i*-th tray reach the temperature of the (*i-1*)-th tray. Also, a switch function $\omega_{n_i} = S(\alpha_{n_i}, \varepsilon_i)$, $\alpha_{n_i} < 0$ is included in the calculation of the mole fraction $x_{i,j}$ because during the start-up $n_{T_i} = 0$, therefore, the switch is not active when $n_{T_i} \geq \varepsilon_i$. Finally, the model that is given by Eq. (1) and Eq. (2) can be written including the start-up as follows,

$$\frac{dn_{i,j}}{dt} = L_{i-1}x_{i-1,j} - L_i x_{i,j} - \xi_{i,j} + \xi_{i+1,j} + n_{Ti} r_{i,j},\ x_{i,j} = \frac{n_{i,j}}{\bar{n}_{Ti}},\ \bar{n}_{Ti} = n_{Ti} + \omega_{n_i} n^* \quad (5)$$

$$\frac{d\left(\bar{n}_{Ti} H_i\right)}{dt} = L_{i-1}H_{L_{i-1}} - L_i H_{L_i} - \xi_i H_{v_i} + \xi_{i+1} H_{v_{i+1}} + n_{Ti}\sum_{j=1}^{N_c}\left(-\Delta H_j\right) r_{i,j} + Q_{ext} \quad (6)$$

Remark 1: It is important to note that the inclusion of the switch variables does not affect the interpretation of the solution of the differential equations. These functions avoid singularities and eliminate to use if-then statements in modelling.

3. Optimal Operation and Start-up

Several cost functions have been proposed for the optimal operation of RBD columns (Sørensen et al. 1996). However, none of these objective functions have considered the start-up explicitly. To find the optimal operation trajectory of the process including the start-up, the following dynamic nonlinear optimization problem is proposed:

$$\min_u \phi(x,u) = \int_{t_0}^{t_f} \left(\underbrace{-X}_{Conversion} + \underbrace{\left\| T_{N_T} - T_{B_{N_T}} \right\|_S^2}_{start-up} \right) dt,\ \text{s.t: Eq(5), (6)},\ x \in \mathbf{X},\ u \in \mathbf{U} \quad (7)$$

In the selection of the cost function we have considered maximizing the conversion. This means more volatile are being produced and these components must be removed immediately to enhance the forward reactions. To this end, the temperature of the reactor (T) must be kept close to the boiling point of the mixture (T_b). Thus, by maximising the conversion and maintaining T close T_b, we reduce the start-up time. Finally, the optimisation problem (7) has been solved by parametrizing u and converting (7) into an NLP.

4. Low-Level Control Design including the Start-up

4.1. Classical low-level controllers for RBD including the start-up

We summarize the main characteristics of the classical control strategies in Table 2. Anyone of these control strategies can be found implemented in the industry. However, none of them have considered the start-up.

Table 2: Classical low-level controllers for RBD

	One-point bottom control	**One-point column control**	**Two-point Control**	
			LV-configuration	**DV-configuration**
Controlled variable	T_{N_T}	$T_i, \forall i \neq N_T$	$h_c, x_{1,j}, T_{N_T}$	$h_c, x_{1,j}, T_{N_T}$
Manipulated Variable	$L_2, Q_{ext} = Q_{\max}$	$L_2, Q_{ext} = Q_{\max}$	L_D, L_2, Q_{ext}	L_2, L_D, Q_{ext}

To include the start-up in the classical low-level controllers for RBD, the level of the condenser drum plays a key role. During the start-up, and before the level reaches the set-point $h_{c,ref}$, L_2 and L_D should be equal to zero. After that, the flows should satisfy $V_3 = L_2 + L_D$, reducing the manipulated variables to just one: the reflux ratio R. Therefore, to take into account this effect in the classic control strategies, the following equations are proposed for the liquid flows L_2 and L_D in terms of R : $L_2 = \omega_{tp} R V_3$, $L_D = \omega_{tp}(1-R)V_3$, with $\omega_{tp} = S(\alpha_{tp}, h_{c,ref}), \alpha_{tp} > 0$. In this case, a switch function is included to maintain the flows equal to zero before the start-up and manipulate just one variable after the condenser drum reaches the set-point. Based on this approach, the LV-configuration is the same DV-configuration. Finally, Table 3 summarizes the classic low-level controller for RBD including the start-up.

Table 3: Classical low-level controllers for RBD with the start-up

	One-point bottom control	**One-point column control**	**Two-point Control**
Controlled variable	T_{N_T}	$T_i, \forall i \neq N_T$	$h_c, x_{1,j}, T_{N_T}$
Manipulated Variable	$R, Q_{ext} = Q_{\max}, L_2 = \omega_{tp} R V_3$ $L_D = \omega_{tp}(1-R)V_3$	$R, Q_{ext} = Q_{\max}, L_2 = \omega_{tp} R V_3$ $L_D = \omega_{tp}(1-R)V_3$	$R, Q_{ext}, L_2 = \omega_{tp} R V_3$ $L_D = \omega_{tp}(1-R)V_3$

4.2. Cascade control strategy with decouplers

We have proposed a cascade control strategy with decouplers to improve the performance of the low-level controllers. This is motivated by the fact that the vapor flow generated in the reactor V_{N_T} is a function of the heat flow Q_{ext} (see Eq.4), and at the same time, the vapor flow can be considered a disturbance for the condenser drum level and the concentration of the accumulator. Based on these observations, the main characteristics of this control strategy are presented in Figure 1. To this end, we decouple the effect of the vapor flow on the concentration in the accumulator, and the effect of the reflux ratio on the reactor. Due to space limitations, the mathematical development of the decouplers has not been presented.

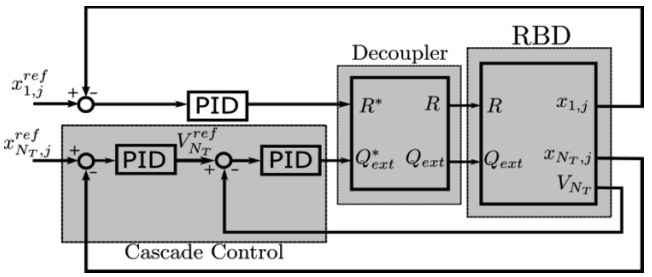


Figure 1: Cascade control structure with decouplers for the RBD

5. Case Study: Polyesterification Process

Consider a reactive batch distillation process of a polyesterification reaction with six stages ($N_T = 6$). The chemical process has 14 species ($N_c = 14$) with 17 reactions. Despite the multiple sub-reactions, a general reaction scheme is described by:

$$\text{Maleic Anhydride+Propylene Glycol} \rightleftarrows \text{Water+Polymer}$$

The control objective of this case study is to remove the water from the reactor to maximize the production of the polymer. To this end, fixed set-points $x_{1,\mathrm{H_2O}}^{ref} = 1$ and $T_{N_T}^{ref} = 450\,\mathrm{K}$ are selected for the local controllers. The simulations results of the control strategies and their comparison with the optimal trajectory are shown in Figure 2. In this Figure, it is possible to observe that, the one-point column control and the decoupled strategy have the better performance to remove the water. Besides, the concentration of water in the accumulator is close to the 80%

6. Conclusions

In this paper, a new model for the startup operation for an RBD column is proposed. Based on this model, an offline optimization problem is proposed in to minimize the start-up time. Additionally, the classical low-level controllers for RBD are reformulated, and a new cascade control strategy with decouplers is developed. Finally, the model, the optimal startup, and the proposed controllers are tested in simulation on a polyesterification process.

Acknowledgments: This work has been done within the project IMPROVISE in the Institute for Sustainable Process Technology (ISPT)

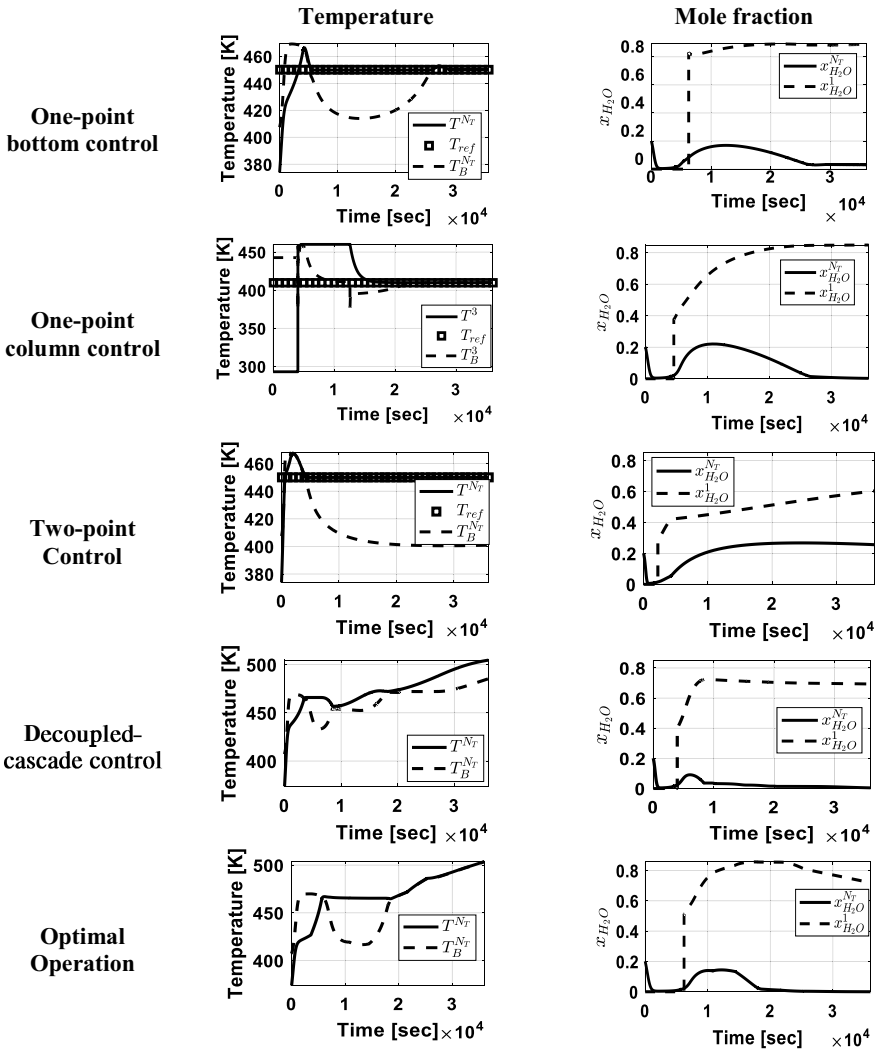


Figure 2: Closed loop trajectories of the main variables of the RBD. Temperature: T_i (—), T_{B_i} (--), and T_{ref_i} (□). Concentration: $x_{H_2O}^1$ (--), and $x_{H_2O}^{N_T}$ (—).

References

E. Sørensen, S. Skogestad, 1994, Control strategies for reactive batch distillation, Journal of Process Control, 4(4), 205-217.

S. Elgue, L. Prat, M. Cabassud, J.M. Le Lann, J. Cezerac, 2004, Dynamic models for start-up operations of batch distillation columns with experimental validation, Computers & chemical engineering, 28(12), 2735-2747

G. Wozny, P. Li, 2004, Optimization and experimental verification of start-up policies for distillation columns, Computers & chemical engineering, 28(1), 253-265.

E. Sørensen, S. Macchietto, G. Stuart, S. Skogestad, 1996, Optimal control and on-line operation of reactive batch distillation, Computers & chemical engineering, 20(12), 1491-1498.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50092-6

State and Parameter Estimation Based On Extent Transformations

Alejandro Marquez-Ruiz[a*], Carlos Mendez-Blanco[a], Marcella Porru[a], Leyla Özkan[a]

[a] *Eindhoven University of Technology, 5612 AJ, Eindhoven, The Netherlands*
a.marquez.ruiz@tue.nl

Abstract

This work presents a general sequential parameter-and-state estimation structure based on extents transformations for non-isothermal homogeneous reaction systems. The extents transformations allow for the development of a Linear Parameter-Varying (LPV) representation with a diagonal state matrix. The particular structure of the LVP system matrices and the physical interpretation of its parameters are used to propose an asymptotic estimator. The main advantage of this estimator is that its implementation does not depend on the chemical kinetic parameters. Based on the information given by the asymptotic estimators, two additional estimators are proposed: an adaptive estimator, and a Recursive Least Squares (RLS) estimator. The first one is used to estimate the chemical reaction rate, while the second to estimate the kinetic parameters. Convergence and tuning properties of the final structure are analyzed and tested on a CSTR example.

Keywords: State and parameter estimation, Extents transformations, Linear Parameter Varying (LPV) Systems, Non-isothermal Homogeneous Reaction Systems.

1. Introduction

The need for real-time information on the evolution of chemical reactions is critical for process control and fault detection. Additionally, the unavailability of specialized instruments motivates the development of state and parameter estimators for chemical processes. Estimators based on rigorous models are the first option for such purposes. However, most industrial processes are described by nonlinear differential algebraic equations, which are in general inconvenient for online parameter-and-state estimation. An alternative is to find different representations of rigorous models using state transformation such as projection methods, state space decomposition, or the inclusion of empirical linear models. Nevertheless, this type of approaches presents several disadvantages such as the loss of physical interpretation of the transformed variables. An attractive alternative is to use the concept of extent transformations (Amrhein et al. 2010) that allows the calculation of the contribution of every phenomenon in the process independently. Moreover, from the estimation perspective, the extents transformations allow for the development of a Linear Parameter-Varying (LPV) representation. The particular structure of the LVP system matrices and the physical interpretation of its parameters are used to propose a general structure for online estimation of non-isothermal homogeneous reaction systems.

There are many approaches for parameter and state estimation, however, due to the extent transformation, and the LPV representation, two well-known approaches for online parameter-and-state estimation are addressed in this paper: simultaneous and sequential. In the simultaneous approach, a set of differential equations that describe the

dynamics of the parameters are lumped to the original model of the system. The parameters are then assumed as a new set of extra states with no dynamics, and the parameter-and-state estimation problem is transformed into a single state estimation problem (Dochain, 2003). This approach has some notable drawbacks such as the selection of the models for the parameters, and the observability conditions of the system with the extended set of states. To tackle these issues, the sequential approach in a cascade structure has been proposed by Dochain, 2003. The main idea of sequential estimation is to solve two estimation problems: one for the states and one for the parameters. Typically, adaptive estimators (Bastin and Dochain, 1990) and least squares formulations have been suggested for parameter estimation. Nevertheless, specific characteristics such as convergence of the cascade structure, and models for the parameters (observer-based estimators) are still open issues. In this paper, the advantages of the LPV representation of non-isothermal homogeneous reaction systems are combined with the sequential estimation structure, to propose a general estimation structure with convergence guarantee.

This paper is organized as follows: In Section 2, the nonlinear model of the system is transformed into a Linear Parameter-Varying (LPV) system using the extent transformations. Based on the LPV model, In Section 3, an asymptotic estimator is proposed to estimate the states, and based on this information two additional estimators are proposed: an adaptive, and a Recursive Least Squares (RLS) estimators. In the same section, convergence and tuning properties of the final structure are analyzed. In section 4 the estimation structure is tested on a simulation of a CSTR. Finally, in Section 5 the conclusions are presented.

2. Extent-Based LPV Representations for Non-isothermal Homogeneous Reaction Systems

2.1. Extent of Reaction and Inlet Flow

Consider the mole balance equation for a homogeneous reaction system with *S* species, R_I independent reactions, *p* independent inlet flows and one outlet flow, given by:

$$\begin{aligned} \dot{n} &= VN^T r + W_{in}u_{in} - \theta n,\ n(0) = n_0 \\ \dot{T} &= -\theta T + \underbrace{\frac{C_{p_{in}}T_{in}}{mC_{p_{mix}}}}_{\alpha\in\mathbb{R}^{1\times p}} u_{in} - \underbrace{\frac{\Delta H_f N^T}{mC_{p_{mix}}}}_{\beta\in\mathbb{R}^{1\times R_I}} Vr + \underbrace{\frac{1}{mC_{p_{mix}}}}_{\gamma} Q_{in},\ T(0) = T_0 \end{aligned} \tag{1}$$

Where $n \in \mathbb{R}^S$ is the number of moles, V is the reaction mixture volume, $r \in \mathbb{R}^{R_I}$ is the reaction rate vector, and $u_{in} \in \mathbb{R}^p$ and $u_{out} \in \mathbb{R}^1$ are the inlet and outlet mass flows; $\theta = m^{-1}u_{out}$, where m is the reacting mixture mass; $N \in \mathbb{R}^{R_I \times S}$ is the stoichiometric coefficient matrix and $W_{in} \in \mathbb{R}^{S\times p}$ is the inlet composition matrix defined as $W_{in} = M_w^{-1}w_{in}$; $M_w \in \mathbb{R}^{S\times S}$ is the diagonal molecular weight matrix and $w_{in} \in \mathbb{R}^{S\times p}$ the matrix of weight fraction. In this work, the parameters α, β and γ are considered constant. To this end, a local controller for the total mass m is assumed. Based on the work of Amrhein et al. 2010, it is possible to find a linear transformation T_e such that the system (1) can be expressed in the following form:

$$\begin{bmatrix}\dot{x}_r\\ \dot{x}_{in}\\ \dot{\lambda}\\ \dot{T}\end{bmatrix}=\begin{bmatrix}-\theta I_{R_I} & 0 & 0 & 0\\ 0 & -\theta I_p & 0 & 0\\ 0 & 0 & -\theta & 0\\ 0 & 0 & 0 & -\theta\end{bmatrix}\begin{bmatrix}x_r\\ x_{in}\\ \lambda\\ T\end{bmatrix}+\begin{bmatrix}0 & 0\\ I_p & 0\\ 0 & 0\\ \alpha & \gamma\end{bmatrix}\begin{bmatrix}u_{in}\\ Q_{in}\end{bmatrix}+\begin{bmatrix}I_{R_I}\\ 0\\ 0\\ \beta\end{bmatrix}Vr \tag{2}$$

$$n = N^T x_r(t) + W_{in} x_{in}(t) + n_0 \lambda(t)$$

Where $x_r \in \mathbb{R}^{R_I}$ is the extent of reaction, $x_{in} \in \mathbb{R}^p$ is the extent of inlet flow, and $\lambda \in \mathbb{R}^1$ is the initial conditions discounting factor. The matrix T_e is calculated using a singular value decomposition of N^T and W_{in}. Details about this procedure can be found in Amrhein et al. 2010. With the representation Eq.(2) of the Eq.(1), the independent evolution of the reaction, the inlet flow, and outlet flow can be easily accounted.

2.2. LPV Representations for Homogeneous Reaction Systems

Let us assume the reaction rate r as an unmeasured disturbance, and the number of moles and the temperature as the outputs of the system, such that the equations (2) can be written as: $\dot{x} = A(\theta)x + Bu + Dd$, $y = Cx$, with $x \in \mathbb{R}^{R_I+p+2}$, $u \in \mathbb{R}^{p+1}$, $y \in \mathbb{R}^{S+1}$ and $d \in \mathbb{R}^{R_I}$. In this paper, it is assumed that the scheduling parameter θ is measured. Based on the LPV system (2), the state and parameter estimation based on extent transformations is developed.

2.2.1. Elimination of the reaction rate

To design asymptotic estimators, the reaction rate r must be eliminated from Eq. (2). Let us define a change of variable $z = \beta x_r + T$. Taking derivatives of both sides of the equation concerning time, and only considering time-varying the variables x_r and T, we obtain : $\dot{z} = -\theta z + \alpha u_{in} + \gamma Q_{in}$. The resulting overall dynamics become,

$$\begin{bmatrix}\dot{x}_{in}\\ \dot{\lambda}\\ \dot{z}\end{bmatrix}=\begin{bmatrix}-\theta I_p & 0 & 0\\ 0 & -\theta & 0\\ 0 & 0 & -\theta\end{bmatrix}\begin{bmatrix}x_{in}\\ \lambda\\ z\end{bmatrix}+\begin{bmatrix}I_p & 0\\ 0 & 0\\ \alpha & \gamma\end{bmatrix}\begin{bmatrix}u_{in}\\ Q_{in}\end{bmatrix}$$

$$y=\begin{bmatrix}W_{in}\left|C_N Z^{\dagger}\right.\\ 0\end{bmatrix}x+\begin{bmatrix}n_0\\ 0\end{bmatrix}\lambda + C_N q_z,\ C_N=\begin{bmatrix}N^T & 0\\ 0 & 1\end{bmatrix} \tag{3}$$

The system (6) can be seen as the following LPV system: $\dot{x} = A(\theta)x + Bu$ with $x \in \mathbb{R}^{p+1}$, $u \in \mathbb{R}^{p+1}$, $y \in \mathbb{R}^{S+1}$, where $q_z \in \mathrm{Ker(Z)}$ and $\mathrm{Z} = \begin{bmatrix}\beta & 1\end{bmatrix}$.

3. State and Parameter Estimation of Extent-Based LPV systems

In this section, the state and parameter estimation of Extent-Based LPV systems is addressed. First, let us define the measured and unmeasured variables of the system. In this paper, the temperature T and the number of moles of S_m components $n_m \in \mathbb{R}^{S_m}$, $\forall S_m \leq S$ are assumed as measured variables. Based on this, the number of moles are partitioned in measured $n_m \in \mathbb{R}^{S_m}$ and unmeasured $n_{um} \in \mathbb{R}^{S-S_m}$ variables. Also, the equation of the number of moles can be divided as:

$$\begin{bmatrix}n_{um}\\ n_m\end{bmatrix}=\begin{bmatrix}N_{um}^T\\ N_m^T\end{bmatrix}x_r+\begin{bmatrix}W_{in,um}\\ W_{in,m}\end{bmatrix}x_{in}+\begin{bmatrix}n_{0,um}\\ n_{0,m}\end{bmatrix}\lambda(t) \tag{4}$$

Where $N_{um} \in \mathbb{R}^{R_I \times (S-S_m)}$, $N_m \in \mathbb{R}^{R_I \times S_m}$, $W_{in,um}^T \in \mathbb{R}^{p \times (S-S_m)}$ and $W_{in,m}^T \in \mathbb{R}^{p \times S_m}$. Figure 1 shows the estimation structure proposed in this paper.

Theorem 1: *Consider the homogeneous reaction system given by the set of Eq. (2), where the reaction rate r is assumed as a bounded disturbance $r \in \Delta$, with Δ convex and compact. Then, if $rank([N_m^T \quad W_{in,m} \quad n_{0,m}]) = R_I + p + 1$, and $\theta \neq 0$ the system (2) is always observable.*

The proof of this theorem is omitted due to the space limitations.

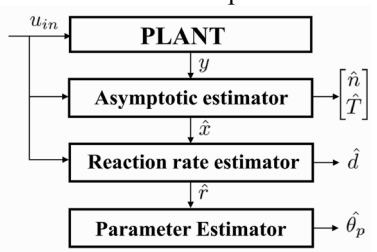


Figure 1. State and parameter estimation structure of Extent-Based LPV systems

3.1. Asymptotic estimator

In Table 1, the model, and the mathematical formulation of the asymptotic estimator are presented.

Table 1. Mathematical formulation of the asymptotic estimator

Model	Asymptotic Estimator
$\dot{x} = A(\theta)x + Bu$	Eq.(3)
$y_m = \begin{bmatrix} n_m^T & T \end{bmatrix}^T = C_m y$	$\hat{x}_r = \beta^{\dagger}(\hat{z} - T) + q_r,\ q_r \in \mathrm{Ker}(\beta)$
$y = C_z x + C_\lambda \lambda + C_N q_z$	$\hat{n} = N^{\intercal}\hat{x}_r(t) + W_{in}\hat{x}_{in}(t) + n_0\hat{\lambda}(t)$

The main advantage of this estimator is that its implementation does not depend on the chemical reaction rate r. The estimation of the extents of reaction depends on the variable $q_r \in \mathrm{Ker}(\beta)$. The calculation of this variable depends on the measured number of moles n_m, and this fact is stated in the following theorem.

Theorem 2. *Consider the asymptotic estimator presented in Table 2, where the estimation of the extents of reaction $\hat{x}_r$ is given by $\hat{x}_r = \beta^{\dagger}(\hat{z} - T) + q_r$ with $rank(\beta) = 1$ and $q_r \neq 0$. If $rank(N_m) = R_I$, then, the null space vectors q_r can be estimated using the following equation:*

$$q_r = \begin{bmatrix} (N_m)^{\dagger} & \beta^{\dagger} \end{bmatrix} \begin{bmatrix} n_m - W_{in,m}x_{in} - n_{0,m}\lambda \\ T - \hat{z} \end{bmatrix}$$

Also, the minimal number of species that need to be measured is $S_m \geq R_I$.

The proof of this theorem is omitted due to the space limitations. As it was stated in Bastin and Dochain (1990), the asymptotic observer is stable and convergent if $\theta > 0$. To satisfy such condition, in this paper, it is assumed that u_{out} does not remain equal to zero for too long.

3.2. Reaction rate estimation

In Table 2, the model, and the mathematical formulation of the reaction rate estimator is presented. For this purpose, the model given by Eq. (4) is used.

Table 2. Reaction rate estimation.

Model	Estimator
$\dot{x} = A(\theta)x + Bu + Dd$ $\dot{d} = -\theta d$ $y_m = y = Cx$	$\dot{\hat{x}}_e = A_e(\theta)\hat{x}_e + B_e u + L(\theta)(y - \hat{y}),\ \hat{y} = C_e\hat{x}_e$ $x_e = [x^T\ d^T]^T,\ A_e(\theta) = \begin{bmatrix} A(\theta) & D \\ 0 & -\theta \end{bmatrix},\ B_e = \begin{bmatrix} B \\ 0 \end{bmatrix},\ C_e = [C\ 0]$

This estimator is known in the literature as an adaptive estimator, and basically, it considers the unknown disturbances as unmeasured states with its dynamics equal to zero. However, in this paper, the disturbance is assumed with the dynamic of the extents of reaction because these variables are strongly related. Note that, this estimator uses as measured variables the temperature and the number of moles of all components. This information is taken directly from the asymptotic estimator. Due to the cascade nature of the states and parameter estimation structure, the convergence of the reaction rate estimator should be faster than the convergence of the asymptotic estimator. This can be achieved by the correct tuning of the matrix $L(\theta)$. However, the selection of this matrix has some limitation which are stated in the following theorem,

Theorem 3. *Consider the LPV model of Table 3. If C is a full row rank matrix, then, there exists a parametric observer gain $L(\theta)$ given by $L(\theta) = -(\theta + \vartheta)I_n C^{\dagger}$ where $\vartheta = \text{eig}\left(A_e(\theta) - L(\theta)C_e\right)$, such that $\theta \le \vartheta$.*
The proof of this theorem is omitted due to the space limitations.

3.3. Recursive Least Squares (RLS)

The final step in the estimation structure is the estimation of the kinetic parameters. To this end, a Recursive Least Squares method is used. First, the reaction rate r is typically parametrized as: $r_k = k_0 \prod_{i=1}^{S_r} C_{i,k}^{\mu_i}$. This equation can be written as $\overline{r}_k = P_k\theta_p + e_k$, where $\theta_p = \left[\ln(k_0), \mu_1, \cdots, \mu_n\right]^{\mathrm{T}}$, $\overline{r}_k = \ln(r)$ and $P_k = \left[1, \ln(C_{1,k}), \cdots, \ln(C_{n,k})\right]$. Finally, the estimation of the kinetic parameters contained in θ_p can be calculated using the following optimization problem: $\hat{\theta}_{p,k} = \arg\min_{\theta_p} \sum_{j=1}^{k} \left\| e_j \right\|_S^2$ .The solution of this problem is given by $\hat{\theta}_{p,k+1} = \hat{\theta}_{p,k} + L_{\theta_p}(\overline{r}_k - P_k\hat{\theta}_{p,k})$, where L_{θ_p} is the Kalman gain, and it can be calculated by the solution of the discrete-time Riccati equation.

4. Example

Let us consider a non-isothermal CSTR where a reversible reaction $A + B \leftrightarrows C + D$ takes place. The system has four species, one independent reaction, and two constant and independent inlets of *A* and *B*. The reaction rate is given by $r = k_f C_A C_B - k_r C_C C_D$, with $k_{f,r} = k_{o_{f,r}} \exp\left(\frac{-E_{a_{f,r}}}{RT}\right)$, where, $k_{o_f} = 6.06 \times 10^5\ \text{m}^3\text{kmol}^{-1}\text{h}^{-1}$, $k_{o_r} = 9.84 \times 10^6\ \text{m}^3\text{kmol}^{-1}\text{h}^{-1}$, $E_{a_f} = 63800\ \text{kJ kmol}^{-1}$, $E_{a_r} = 71710\ \text{kJ kmol}^{-1}$. The matrices N and W_{in} are given by: $N = \begin{bmatrix} -1 & -1 & 1 & 1 \end{bmatrix}$, $W_{in} = \begin{bmatrix} 0.01665 & 0 & 0 & 0 \\ 0 & 0.03121 & 0 & 0 \end{bmatrix}^{\mathrm{T}}$. The initial conditions are

$n_0 = [0.5,1,0.5,0]^T$ kmol, and $T_0 = 373$ K. Figure 2 shows the profiles of simulated and estimated variables including random measurement noise in the temperature T and outlet flow u_{out} (θ). These pictures are obtained in open loop performing step changes in the inlet flows and the initial condition of T. The performance of the estimation structure is quite good, even under noise in θ, however, at the beginning of the simulation, $\hat{n}_D$ is taking negative values, which is not desired. This problem can be corrected assuming perfect measurement of the outlet flow or by an appropriate selection of the desired poles.

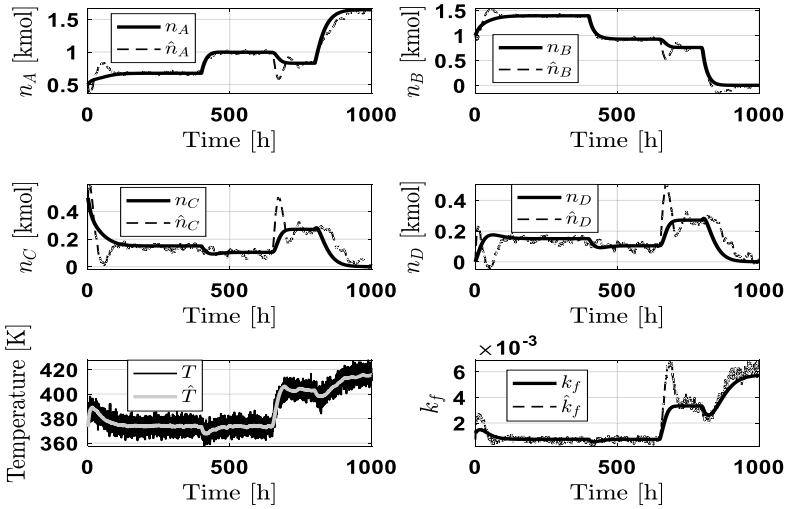


Figure 2. Trajectories of n_A, n_B, n_C, n_D, T and k_f : Simulated (—), Estimated (---)

5. Conclusions

In this paper, a general sequential parameter-and-state estimation structure based on extents-based LPV models for homogeneous and non-isothermal reaction systems is proposed. First, an asymptotic estimator is proposed based on an LPV model and without information of the chemical reaction rate. With the information given by the asymptotic estimator two estimators have been proposed: an adaptive estimator, and a Recursive Least Squares (RLS) estimator. Also, convergence and tuning properties of the final structure are analyzed. The estimation structure is tested on a simulation of a CSTR.

Acknowledgments: This work has been done within the project IMPROVISE in the Institute for Sustainable Process Technology (ISPT).

References

M. Amrhein, N. Bhatt, B. Srinivasan, D. Bonvin, 2010. Extents of reaction and flow for homogeneous reaction systems with inlet and outlet streams, AIChE Journal, 56(11), 2873-2886.

D. Dochain, 2003, State and parameter estimation in chemical and biochemical processes: a tutorial, Journal of process control, 13(8), 801-818.

G. Bastin, D. Dochain, 1990, On-line Estimation and Adaptive Control of Bioreactors. Elsevier, Amsterdam.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50093-8

Efficient Process Monitoring and Fault Isolation with the Integrated use of Markov Random Fields Learning and the Graphical Lasso

Changsoo Kim[a], Hodong Lee[a], Younggeun Lee[a], Kyeongsu Kim[a], Won Bo Lee[a*]

[a]*Seoul National University, Gwanakro 1, Gwanakgu, Seoul, 08826, Republic of Korea*
wblee@snu.ac.kr

Abstract

In this study, a framework for efficiently detecting faults, and isolating fault causes with the integrated use of Markov Random Fields (MRF) and the graphical lasso, is proposed. MRFs are capable of expressing any subtle relationship that appears within a process without any a priori knowledge, and can compute the joint probabilities of the variables, taking in the nonlinearity of chemical processes. The graphical lasso, a regularization algorithm for inducing the sparse precision matrix and learning the structure of a graphical model, mitigates the high computational complexity of learning and inferencing a MRF. Also, by splitting the process variables into inter-relevant groups, it enables quick identification of the fault variable, and fault propagation path detection.

The proposed algorithm was applied to the Tennessee Eastman benchmark process (TEP), to test its performance. The 28 fault cases within the TEP model were tested with the new monitoring method, and the results were compared to that of the conventional methods. The proposed method showed high accuracy and efficient fault identification performance, even with the tricky fault cases 4 and 9.

Keywords: Process Monitoring, Fault Isolation, Graphical Models, Graphical Lasso

1. Introduction

Process monitoring is an essential part of safe plant operation, and has been studied in various angles over the past few decades. Among the various methods, conventional data-driven monitoring methods, such as PCA, FDA, and ICA have been used and improved extensively. However, these methods are inherently limited in that they are not capable of taking in nonlinearity, and various data processing methods have been implemented to resolve this problem.

Yan et al. (2016) used the two variant autoencoders, the contractive autoencoder (CAE) and the denoising autoencoder (DAE) to develop a nonlinear process monitoring model. They developed the monitoring statistics H^2 and SPE using the outputs of the autoencoders, and applied their method to the 20 fault cases of the TEP. Their results clearly show that due to the nonlinear nature of the activation function, the autoencoder monitoring framework greatly increases the fault detection rate. The limitation of their work is that, monitoring accuracy decreases greatly when the fault is related to highly interacting group of variables. Also, since the hidden layer of nodes containing dimensionally altered form of variables cannot provide knowledge about the individual contribution of the variables to the fault, fault identification is almost impossible.

The work by Gonzalez et al. (2015) implement the use of Bayesian networks to process monitoring. Using the kernel density estimation (KDE) method to fit conditional and joint probabilities, they take advantage of the causal nature of Bayesian networks to dissect the process monitoring problem into a series of probability inference problems. However, these works are limited in that they require a priori knowledge of the variable causal structure to fully use the Bayesian network monitoring system.

In this study, a monitoring framework that resolves the limitations of the previous methods is proposed. Since the TEP has been used as the benchmark problem for testing out various monitoring algorithms, the proposed algorithm is tested out on the 28 faults provided within the TEP. Such work involving TEP include the papers by Wang and Melo (2016), where they used the Bayesian treed Gaussian process methods for process monitoring, and the paper by Rato and Reis (2017), where they apply the sensitivity enhancing transformation (SET) for data pre-processing prior to process monitoring for improved fault diagnosis.

2. Methodology

The proposed methodology consists of the integrated use of two theories: Markov random fields (MRF), and the graphical lasso.

MRFs serve as the basis for modelling the relationships of variables, often vaguely represented within process data. By modelling the chemical process as a MRF, subtle relationships as well as cyclic relationships can be expressed, unlike Bayesian networks. Since cyclic relationships are abundant within a process, such as in distillation columns, it is more expressive to model a process as a MRF. When modelling process variables into a MRF, it is assumed that the variables form a multivariate Gaussian distribution. Thus by obtaining a covariance matrix among the variables a MRF is defined. After running the graphical lasso and obtaining a structure of the grouped variables, the MRF is trained to obtain a covariance matrix for each group. MRFs can be trained by obtaining the maximum log-likelihood of the data:

$$\mathrm{l}(\Theta) = \log det\Theta - \mathrm{tr}(S\Theta) \tag{1}$$

$$\Theta^{-1} - \mathrm{S} - \Gamma = 0 \tag{2}$$

The term Θ represents the precision matrix, inverse of the covariance matrix, S represents the empirical covariane matrx, and Γ is a matrix of Lagrange parameters, having the value zero for existing edges. The detailed description of the equations are given in Hastie et al. (2010). The separate MRFs can be trained using the iterative proportional fitting (IPF) algorithm. After all MRFs are trained, the $(1-\alpha)$ trust region is used to obtain the monitoring limit of the normal data. A typical value of $\alpha = 0.95$ is used in this work.

The graphical lasso is the extension of the lasso regularization to graphical models, where the algorithm seeks to find the sparse form of the precision matrix, using L1 regularization. Lasso itself is often used as a variable selection technique, and implementing it to graphical models allows it to reduce the model to a more sparse form. Thus, it can be used to learn the structure of the MRF. The essence of graphical lasso is obtaining the maximum of equation (3), where ρ is the regularization parameter making the precision matrix sparse. The details are given in Friedman et al. (2007).

$$\mathrm{l}(\Theta) = \log det\Theta - \mathrm{tr}(S\Theta) - \rho\|\Theta\|_1 \quad (3)$$

This process is an essential step of the proposed methodology, serving two purposes. One is that it allows efficient process monitoring and quick fault identification. As a result of the graphical lasso, the entire set of variables are segmented into groups, each group consisting of variables that are highly interrelated with each other. This leads to natural scale-down of the process monitoring dimension, enabling the process monitoring method to more effectively monitor faults and isolate them. Another important outcome of the graphical lasso is that it greatly mitigates the computational complexity when dealing with MRFs. MRFs typically require the partition function when learning its parameters or performing inference, which is a factor that greatly increases the computational complexity. By dissecting the graph into groups and dismissing irrelevant edges, the computational complexity related to MRF can be greatly reduced, making the monitoring process tractable. Since a single run of the graphical lasso results in a small graph while the rest of the variables are singled out, an iterative approach is applied, so that all of the variables can be grouped together for monitoring.

The overall methodology is implemented in the order shown in Figure 1. First, all of the process variables subject to monitoring are considered to be in a form of fully connected graph. Then the entire graph is input to the graphical lasso algorithm, to find the initial group of nodes. Since the main purpose of graphical lasso is to divide the graph into different sections for monitoring, a large enough lasso parameter is used to achieve this

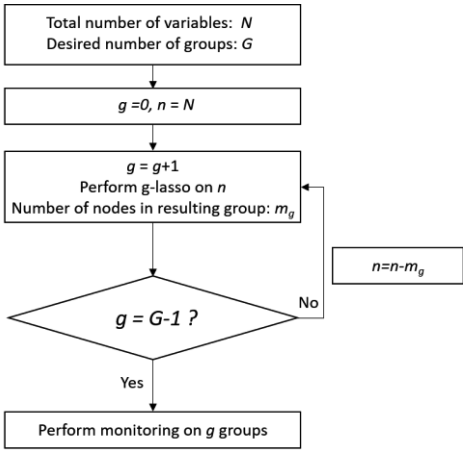


Figure 1. Proposed process monitoring framework

purpose. Usually this results in one small group of nodes, and all of the other nodes are isolated. For this reason, the graphical lasso is iteratively implemented on the rest of the results, to obtain multiple groups. An appropriate lasso parameter is selected empirically for each of the iterations. Afterwards, the grouped nodes are separately implemented into the MRF monitoring framework. Potential function and parameters of the MRFs are learned using IPF.

3. Results and Discussion

The proposed method is applied to the widely used Tennessee Eastman process (TEP), to verify its performance compared to that of conventional monitoring methods. For extensive analysis, rather than the original version by Downs and Vogel (1993), a revised version of the TEP, published by Bathelt et al. (2014) as a MATLAB Simulink model, was used to test out the monitoring framework. Compared to the original model, a more rigorous control structure was implemented in the Simulink model, and 8 extra faults were added, totalling up to 28 faults. The fault detection and isolation method was applied to each of the faults and the results were compared to that of the previous studies.

3.1. Grouping the process variables using graphical lasso

Prior to process monitoring, the 41 measurement variables of the TEP are divided into 6 groups, using the graphical lasso. After five dissections, the remaining nodes are gathered to form a final group. This process is shown in Figure 2.

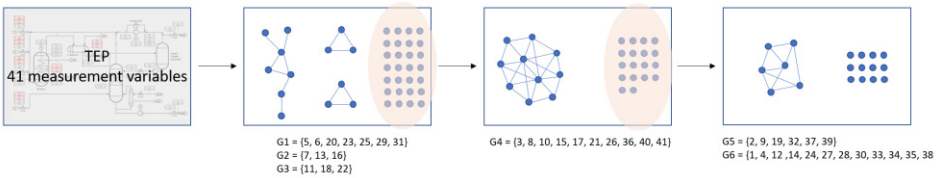


Figure 2. Performing iterative graphical lasso on 41 TEP variables

3.2. Monitoring: Improved fault detection performance

To emphasize the performance of the proposed method, the results of the faults 4 and 9 are presented in this paper, since they are usually deemed as the faults that are the most difficult to detect. Fault 1 is included as well as a representative case. Comparing with the monitoring results of fault 1 with the conventional PCA scheme, our framework shows an improvement in detection time by 1.5 hours. This is a significant improvement since 1.5 hours can be a critical time when dealing with process faults.

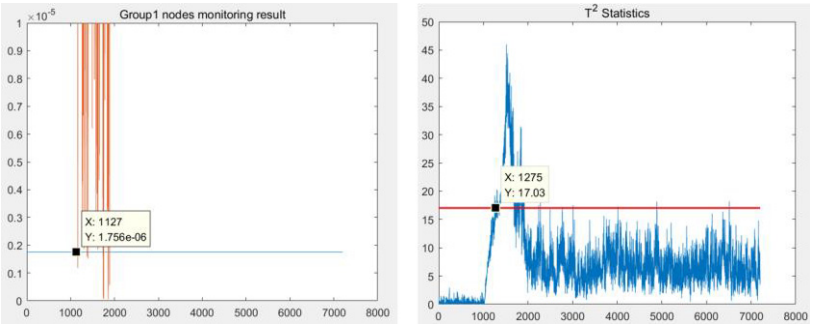


Figure 3. Monitoring results comparison of the proposed method and PCA

3.3. Fault propagation

To emphasize the effect of graphical lasso, it is tested out with the most recent, autoencoder monitoring framework. Due to the use of graphical lasso, process variables are monitored locally, and thus the fault propagation path can be detected. Considering

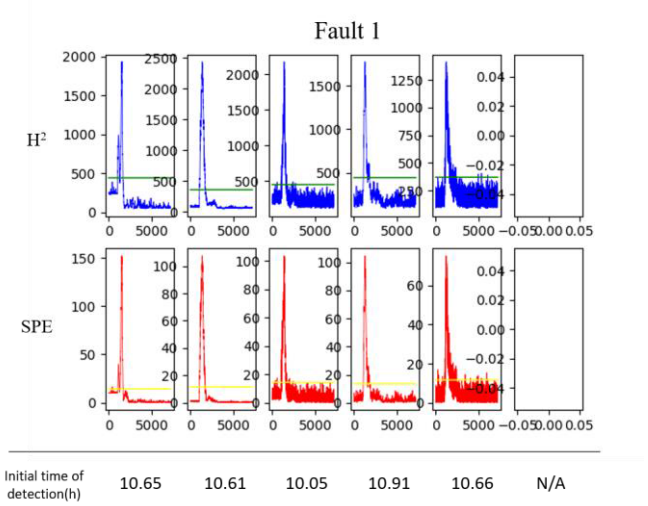


Figure 4. Fault propagation detection for fault 1

the fault detection results shown in Figure 4, the time in which fault is first detected within each group differs, suggesting a fault propagation path. The initial fault detection time of each group and the variables included in each group is given in Figure 4. The fault propagation among groups is evident from the initial detection time. Also, by analyzing the relevant variables within each group, it is possible to get a clear idea of the fault propagation path.

3.4. Fault isolation

A very important result of the proposed algorithm is the high efficiency in isolating a fault variable. For sensitive faults such as the first fault of the TEP, the process variable causing the fault can be isolated by investigating the group that is quickest in detecting the fault. However, the greatest strength of the proposed framework is revealed when dealing with complicated faults, such as the fourth and ninth fault of the TEP. Process

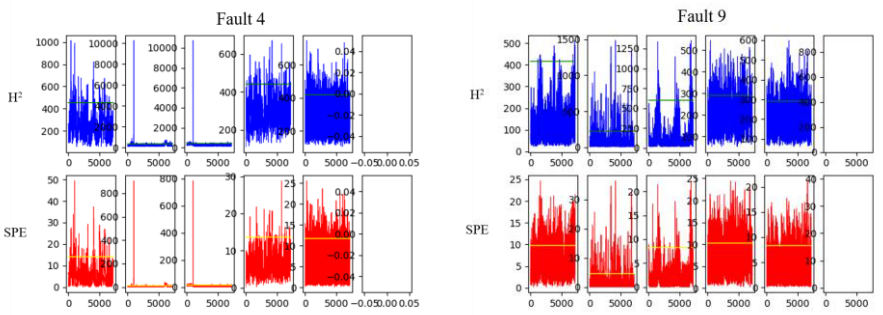


Figure 5. Fault isolation results for faults 4 and 9 of the TEP model

monitoring results of the two faults are given in Figure 5. Results from the recent process monitoring papers show that their methods detect only 30% of the faults, meaning it's almost impossible to get an idea of the location of the fault. However, as it can be seen from Figure 5, it is possible to restrict the region of fault variable and observe the trends for different groups when using this framework. The monitoring graph for fault 4 shows that groups number 2 and 3 are the only relevant regions for process monitoring. The benefit of the proposed framework is more evidently shown in the monitoring result for fault 9. When using only the autoencoder monitoring framework, one cannot even see whether a fault occurred at $t = 1000$, where in our case, group 3 clearly shows a change in value at that time, while all of the other groups are showing random noisy results. Considering the fact that fault number 9 is a random variation in D feed temperature, the results of our framework is consistent since group 3 contains variables 11, 18, and 22, all of them temperature variables. This suggests that, even without the knowledge of the fault such as in actual plants, it is possible to get evidence of the fault and start evaluating related variables.

4. Conclusions and Future Work

In this study, a framework for efficient process monitoring and fault isolation is proposed, via integrated use of the MRF learning and the graphical lasso. The proposed methodology was applied to the widely used TEP to test its performance. The results proved that the proposed method show great monitoring and fault isolation performance. These results show that with a simple application of the pre-processing method of graphical lasso to group the process variables, and using the remaining graph structure to perform MRF monitoring can not only enhance monitoring performance, but also provide process isolation and fault propagation results, allowing quick and accurate detection of process faults. This algorithm can be extended to industry-scale systems by simply adjusting the lasso parameter. This is to be included in future work.

References

W. Yan, P. Guo, L. Gong, Z. Li, 2016, Nonlinear and robust statistical process monitoring based on variant autoencoders, Chemometrics and Intelligent Laboratory Systems, 158, 31-40.

R. Gonzalez, B. Huang, E. Lau, 2015, Process monitoring using kernel density estimation and Bayesian networking with an industrial case study, ISA Transactions, 58, 330-347.

H. Wang and V. V. de Melo, 2016, Bayesian treed Gaussian process method for process monitoring, Computer Aided Chemical Engineering, 38, 1773-1778.

T. J. Rato and M. S. Reis, 2017, Improved Fault Diagnosis in Online Process Monitoring of Complex Networked Processes: a Data-Driven Approach, Computer Chemical Process Engineering, 40, 1681-1686.

T. Hastie, R. Tibshirani, J. Friedman, 2010, Undirected Graphical Models, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 625-643.

J. Friedman, T. Hastie, R. Tibshirani, 2008, Sparse inverse covariance estimation with the graphical lasso, Biostatistics, 9, 3, 432-441

J. J. Downs and E. F. Vogel, 1993, A Plant-wide Industrial Process Control Problem, Computers and Chemical Engineering, 17, 3, 245-255.

A. Bathelt, N. L. Ricker, M. Jelali, 2014, Revision of the Tennessee Eastman Process Model, 9th International Symposium on Advanced Control of Chemical Processes.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50094-X

Control of reactive distillation columns with top-bottom external recycle

Haisheng Chen[a*], Tengfei Wang[a], Kejin Huang[a]

[a]*Beijing University of Chemical Technology, North 3 ring road No. 15, Beijing, 100029, China*
chenhs@mail.buct.edu.cn

Abstract

Reactive distillation columns with external recycle between the top and bottom (RDCs-TBER) have a better steady-state performance as compared with conventional reactive distillation columns (CRDCs) for the separations of reacting mixtures featuring the most unfavourable ranking of relative volatilities (i.e., the reactants are the lightest and heaviest components with the generated products in between). In this article, their dynamics and control are studied in great detail with special attention given to the influences of external recycle on process dynamics and controllability. Because of the totally reboiled operation mode plus a large reboiler, the CRDC is generally characterized by severe under-dampness between the intermediate-product compositions and the reboiler heat duty and long transient time. With external recycle between the top and bottom of the RDC-TBER, the suppression of the under-dampness between the intermediate-product compositions and the reboiler heat duty and relatively short transient time is attained, thereby presenting favourable influences to process dynamics and controllability. In additional, the supplementary manipulated variable (i.e., the flow rate of the external recycle) can be used to improve the process operation. A reactive distillation system, executing a hypothetical reversible reaction, A + B ↔ C + D ($\alpha_A > \alpha_C > \alpha_D > \alpha_B$), is employed to inspect the dynamics and controllability of the RDC-TBER and CRDC. The CRDC figures rather sluggish tracking responses because the under-damped behaviour and slow inventory control. The RDC-TBER, on the other hand, shows greatly improved tracking performance with even enhanced disturbance rejection capabilities.

Keywords: reactive distillation column, external recycle, total reboiled operation, process dynamics, process controllability.

1. Introduction

For the separation of reacting mixtures featuring the most unfavorable ranking of relative volatilities (i.e., the reactants are the lightest and heaviest components with the generated products in between), it is difficult to achieved high product purities via conventional reactive distillation columns (termed as CRDCs, hereinafter) with side draw and operating at total reboiled and total refluxed modes simultaneously (Chen et al., 2013). To deal this issue, Tung and Yu (2007) proposed a kind of reactive distillation columns with two reactive sections at the top and bottom (RDC-TRS). Despite the RDC-TRS can acquire high product purity, huge energy consumption should be applied. In order to overcome this difficulty, we proposed to utilizing feed splitting technology and/or adding external recycle between the top and bottom and thus derived five kinds of process alternatives with feed splitting and/or external recycle

(Chen et al, 2016): (i) reactive distillation columns with two reactive sections and feed splitting (termed as RDC-TRS-FS, hereinafter); (ii) RDC-TRS-FS with external recycle (termed as RDC-TRS-FS-ER, hereinafter); (iii) reactive distillation columns with two reactive sections and external recycle (termed as RDC-TRS-ER, hereinafter); (iv) reactive distillation columns with a top-bottom external recycle (termed as RDC-TBER, hereinafter); (iii) RDC-TBER with feed splitting (termed as RDC-TBER-FS, hereinafter). The results showed that adding a top-bottom external recycle and/or utilizing feed splitting technology to the reactive distillation column can largely improve the steady-state system performance as compared with the RDC-TRS and the CRDC for the separations of reacting mixtures featuring the most unfavorable ranking of relative volatilities. Although the steady state researches on the two methods employed showed that they have great advantages on enhancing process thermodynamics efficiency, their dynamics and controllability are needed to be studied.

In terms of the dynamics and control of RDC-TRS-FS for separating a two-stage consecutive reacting mixture ($A + B \leftrightarrow C + D$, $C + B \leftrightarrow E + D$ with $\alpha_D > \alpha_B > \alpha_C > \alpha_A > \alpha_E$), Kaymak et al. (2017) provided four types of control structures including a temperature inferential control, a direct composition control, and two hybrid temperature and composition control and concluded that the last two control schemes can achieve effective control performance with little steady state deviations. Lately, Cao et al. (2017) found that deliberate arrangement and control of feed splitting can give great favorable effects on process dynamics and control. As for the dynamics and control of RDC-TBER (c.f. Figure 1b), few papers can be found to the best of our knowledge, even though it has comparable (or even better) steady state performance than RDC-TRS-FS in some reaction systems. The RDC-TBER have more manipulating valuable (i.e., the flow rate of external recycle) than the CRDC because of the process retrofitting, which means that the RDC-TBER might have improved dynamic performance than the CRDC with the operation of the flow rate of external recycle even though the former represents much stronger internal mass integration and internal heat integration than the latter. Therefore, researches on the effects of external recycle to process operation and control are necessary.

The main purpose of the current article is to gain insights into the dynamic behaviors of the RDC-TBER and develop effective control systems. Detailed comparisons are made between the three control systems for the RDC-TBER and that for the CRDC. The separations of a hypothetical ideal quaternary reaction is employed as illustrative example to evaluate the impact of external recycle on process dynamics and control. After a brief discussion of the obtained outcomes, the article ends with the conclusions that can be drawn.

2. Dynamics and control of RDC-TBER

As sketched in Figure 1a, the CRDC for the separation of the most unfavorable reacting mixtures is configured in such a way that the reactive section is arranged deliberately at the bottom (which include not only the reboiler but also some stages in the lower section divided by the stage for withdrawing the reaction products) for the separation of an exothermic reaction. The lightest reactant A and the heaviest reactant B are fed, respectively, into the reboiler and at the top of the reactive section with the light product C and the heavy product D withdrawn together as an intermediate product above the reactive section. It is noted that this arrangement leaves the CRDC in a totally refluxed and totally reboiled operation mode. The reflux flow rate and reboiler heat duty are used

to maintain the inventories of the condenser and reboiler, respectively, and the compositions of the intermediate product are controlled by side draw flow rate. The very unusual process configuration and operation mechanism actually lend the CRDC very unique process dynamics that can have profound influences on process controllability.

First, the regulation path is to be analyzed. Varying the side draw flow rate could not directly affect the compositions of the intermediate product. Its variation should do this job though the bottom inventory control loop. The variation of the side draw flow rate will affect the level of the bottom, then the reboiler heat duty would be adjusted, and the compositions of the intermediate product will change indirectly. Actually it has two routines to affect the C and D compositions of the intermediate product. Once a change occurs in the reboiler heat duty, it arouses an immediate variation in the vapor flow rate and has its first impact on the C and D compositions of the intermediate product. After the change moves upward and enters into the condenser, the totally refluxed operation mode transfers it completely into a perturbation in the liquid flow rate. After it reaches the stage for withdrawing the intermediate product, the perturbed liquid flow rate then presents the second impact on the C and D product compositions. It is noted that the first and second impacts are along opposite directions, and their net effect is the occurrence of an open-loop, under-damped behavior. Apparently, the unique phenomenon is closely related to the totally refluxed operation mode; more specifically, when the inventory in the condenser is larger, the degree of under-dampness becomes more severe, which presents definitely detrimental influences to process controllability. Apart from the intensification of interaction with the other control loop (i.e., the stoichiometric-balance control loop), the under-damped behavior of the CRDC also poses additional restrictions to tightening controller tuning. Second, the disturbance path is to be analyzed. The primary disturbance essentially originates from the A feed flow rate (and this assumes implicitly that it is taken as the production rate handle). Because of the totally reboiled operation mode, the open-loop under-damped behavior should still prevail despite that the reboiler heat duty is supposed to regulate the reboiler inventory here. Nonetheless, the severity of the under-dampness should be alleviated substantially. With the adoption of external recycle in process revamp, the exploited degrees of freedom can generally lead to a great improvement in the steady-state performance of the RDC-TBER. As sketched in Figure 1b, an external recycle is now added from top to bottom to enhance the possibility of contacting between the lightest reactant A and the heaviest reactant B. Relatively low vapor and liquid heat loads are now solicited by the RDC-TBER than CRDC because of improved reaction conversion. The including of external recycle can also improve the process controllability. In the regulation path, the impact to the intermediate product compositions of C and D is considerably suppressed because of total reflux operation mode is changed into partial one. Therefore, the severity of open-loop under-dampness should be greatly abated in the RDC-TBER. In the disturbance path, the alleviation of the open-loop under-dampness behavior can still be expected. The improvement of open loop process dynamics will certainly present favorable influences to the tight control of the C and D compositions in the intermediate product.

In the following, two reactive distillation systems, including the separations of a hypothetical reversible reaction, $A + B \leftrightarrow C + D$ ($\alpha_A > \alpha_C > \alpha_D > \alpha_B$) is used to examine the insights on the dynamics and controllability of the CRDC and RDC-TBER.

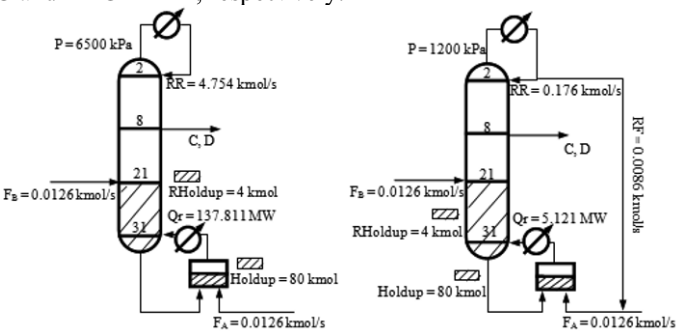
3. An illustrative example: A hypothetical ideal exothermic reaction

3.1. Process Studied

As shown in Figure 1, the two reactive distillation systems, i.e., the CRDC and RDC-TBER, are taken from Chen et al. (2013), which separate a hypothetical reversible reaction.

$$A + B \leftrightarrow C + D \quad (1)$$

To save the space, the main physicochemical properties, and other relevant information are omitted and one can found them in the relevant reference. The commercial software Mathematica is used in process simulation. In comparison with the CRDC, the RDC-TBER reduces the reboiler heat duty largely, highlighting the great potential of external recycle in process revamp. In this article, the solid and dashed lines stand for the outputs of the CRDC and RDC-TBER, respectively.



(a) (b)

Figure 1. Process synthesis and design. (a) CRDC, (b) RDC-TBER

3.2. Open-Loop Dynamic Characters

Figure 2 depicts the open-loop responses of the CRDC and RDC-TBER after they encounter ±5% impulse disturbances in the reboiler heat duty, separately. For both the positive and the negative perturbations, the CRDC shows seriously under-damped responses in the C and D compositions of the intermediate product. It is noted that the C composition of the intermediate product competes intensively with the D composition of the intermediate product, giving rise to quite similar responses but with opposite changing directions. For the RDC-TBER, external recycle makes it exhibit a degree of under-dampness that is considerably reduced as compared with that in the CRDC, implying a favorable influence on process dynamics by such a process revamp.

Figure 3 depicts the open-loop responses of the CRDC and RDC-TBER after ±5% impulse changes are introduced, separately, into the A feed flow rate. With the adoption of external recycle, the under-damped behaviors are subdued in the RDC-TBER, implying again a favorable influence on process dynamics by such a process revamp.

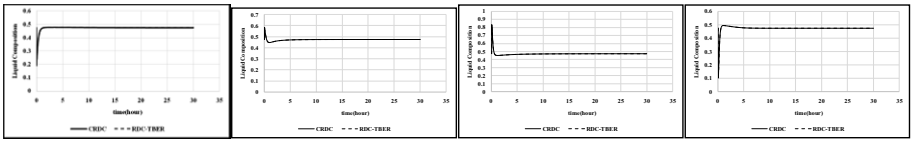


Figure 2. The open-loop responses of the CRDC and RDC-TBER after ±5% impulse disturbances in the reboiler heat duty

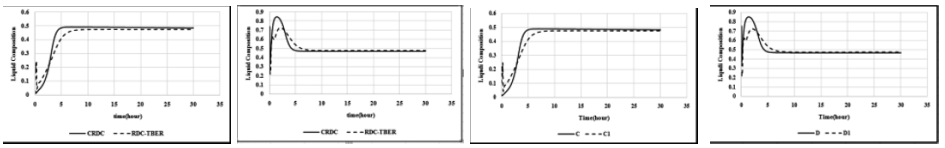
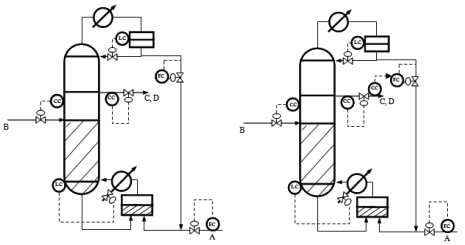


Figure 3. The open-loop responses of the CRDC and RDC-TBER after ±5% impulse disturbances in the A feed flow rate

3.3. Closed-Loop Operation



(a) (b)

Figure 4. Control systems of RDC-TBER, (a) CS1, (b) CS2

Direct composition control schemes are employed here as shown in Figure 4. Column pressure is regulated with the condenser heat removal, and the inventories of the reflux drum and reboiler are regulated with the reflux and reboiler heat duty, respectively, via a P-only controller. The product compositions are controlled with two methods. One is that the C composition of the intermediate product is regulated by the intermediate product flow rate via a PI controller, leaving the D composition of the intermediate product uncontrolled (c.f. Figure 4a); and the other is that the C composition of the intermediate product is regulated by the intermediate product flow rate and the D composition of the intermediate product is regulated by external recycle flow rate (c.f. Figure 4b). The B feed flow rate is employed to control the composition of B on stage 5 in the CRDC and on stage 10 in the RDC-TBER since the composition variation is the largest, serving to keep the stoichiometric ratio between the lightest reactant A and the heaviest reactant B. Although a P-only composition controller or a PI composition controller can be used here, for the sake of simplicity, the former option is chosen. The A feed flow rate is flow-controlled and works as the production rate handle. The composition controllers, designed in-line with the Tyreus− Luyben tuning rule. Composition measurement devices are assumed to act like a first order process with a 3 min time constant, and control valves are all set at the half-open position in the nominal steady state.

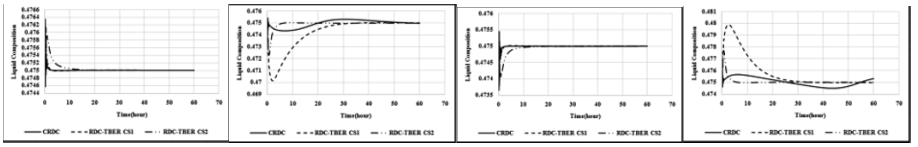


Figure 5. The regulatory responses of the CRDC and RDC-TBER after ±5% step changes are introduced into the A feed flow rate

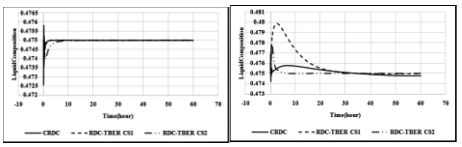


Figure 6. The regulatory responses of the CRDC and RDC-TBER after ±5% step changes are introduced into the A feed composition

The regulatory responses of the CRDC and RDC-TBER are illustrated in Figure 5 and 6, after ±5% step changes are introduced into the A feed flow rate and the A feed composition, respectively. Note that with the manipulating of external recycle, RDC-TBER (CS2) is found to be advantageous over the CRDC in this situation.

4. Conclusions

The unusual process configuration and operation mechanism, i.e., the totally refluxed and totally reboiled operation mode plus an intermediate-product withdrawal, lend the CRDC very unique process dynamics. In the regulation and disturbance path, seriously under-damped behaviours could occur between the light or heavy product compositions of the intermediate product and the side draw flow rate. These dynamic characteristics are likely to pose great difficulties for the tight quality control of the intermediate product. With the adoption of external recycle in the RDC-TBER, the reaction conversion is carefully coordinated, and this leads to not only the enhancement of steady-state performance but also the suppression of those unfavourable dynamic characteristics aroused by the totally refluxed and the totally reboiled operation mode plus an intermediate-product withdrawal. The reactive distillation system carrying out a hypothetical reversible reaction has been employed to scrutinize the dynamics and control of the CRDC and RDC-TBER. Two control systems of the RDC-TBER are developed. All of the obtained results have confirmed that with the adoption of external recycle the RDC-TBER helps to substantially alleviate the inherent drawbacks and therefore presents favourable influences to process dynamics and controllability.

References

H. Chen, K. Huang, W. Liu, L. Zhang, S. Wang, S. J. Wang, 2013, Enhancing mass and energy integration by external recycle in reactive distillation columns, AIChE Journal, 59, 6, 2015.

H. Chen, L. Zhang, K. Huang, Y. Yuan, X. Zong, S. F. Wang, L. Liu, 2016, Reactive distillation columns with two reactive sections: feed splitting plus external recycle, Chemical Engineering and Processing: Process Intensification, 108, 189.

D. B. Kaymak, H. Unlu, T. Ofkeli, 2017, Control of a reactive distillation column with double reactive sections for two-stage consecutive reactions, Chemical Engineering and Processing: Process Intensification, 113, 86.

S. T. Tung, C. C. Yu, 2007, Effects of relative volatility ranking to the design of reactive distillation, AIChE Journal, 53, 5, 1278.

Y. Cao, K. Huang, Y. Yuan, H. Chen, L. Zhang, S. Wang, 2017, Dynamics and control of reactive distillation columns with double reactive sections: Feed-Splitting Influences, Industrial and Engineering Chemistry Research, 56, 8029.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50095-1

Multivariable Control Structure Design of Heat Exchange Networks Based on Mixed-Integer Quadratic Programming

Lautaro Braccia[a,*], Pablo A. Marchetti[b], Patricio Luppi[a] and David Zumoffen[a]

[a] *Grupo de Ingeniería de Sistemas de Procesos (GISP); Centro Franco-Argentino de Ciencias de la Información y de Sistemas (CIFASIS); CONICET-UNR, 27 de Febrero 210 bis, Rosario (S2000EZP), Argentina*
[b] *Instituto de Desarrollo Tecnológico para la IndustriaQuímica (INTEC); UNL-CONICET, Güemes 3450, Santa Fe (3000), Argentina*
braccia@cifasis-conicet.gov.ar

Abstract

In this work, a new systematic methodology to address multivariable control structure (MCS) design for heat exchange networks (HENs) is proposed. This methodology is based on recent developments presented by Braccia et al. (2017) where the MCS design for medium/large-scale processes is performed based on a mixed-integer quadratic programming (MIQP) formulation. This methodology is based on an equation-oriented superstructure which addresses simultaneously: the selection of controlled variables (CVs) and manipulated variables (MVs), the input-output pairing definition, the controller topology, the initial heuristics considerations, and the stability test by using steady-state information only. Usually, in the HENs cases, all the outputs variables need to be controlled, i.e. there are no uncontrolled variables. In this context, the MCS design procedure presented in Braccia et al. (2017) requires a reformulation to quantify the manipulated variables instead of the uncontrolled variables. Furthermore, some modifications of the original MIQP algorithm related to the augmented functional cost are analyzed. The performance of different controller topologies, i.e. decentralized, sparse, and full, are evaluated. Two examples are given in this work to illustrate the potential of the proposed methodology and to support the final conclusions.

Keywords: Heat exchanger network synthesis, Multivariable controller design, Mixed-integer quadratic programming.

1. Introduction

The control structure design of heat exchange networks (HENs) is one of the most studied problems in the last decades. The aim of the control structure is to accommodate set-point changes and reject load disturbances in the network. Several approaches are proposed in the literature to ensure an appropriate control structure. Mathisen, et al. (1994) provided a heuristic method for bypass placement. Yang et al. (2001) and Escobar and Trierweiler (2009) proposed a retrofit HEN strategy with optimal bypass placement based on a simplified model for disturbance propagation and control. Morever, the pairing is based on the partial relative gain array (RGA), which is defined similar to the regular RGA.

In this work, a new MCS design formulation for heat exchange networks is proposed. The methodology of Braccia et al. (2017), which is based on the sum of squared deviations (SSD) and the net load evaluation (NLE) indexes, is considered as a starting point. A

reformulation of this methodology is made to quantify the manipulated variables instead of the uncontrolled variables, since control structure design problems for HENs require all the output variables to be controlled. Different controller topologies, i.e. decentralized, sparse, and full, are designed and analyzed.

2. Control Structure Design.

The proposed approach is based on an equation-oriented superstructure which addresses simultaneously: the selection of controlled variables (CVs) and manipulated variables (MVs), the input-output pairing definition, the controller design, the initial heuristic considerations, and the stability test by using only steady-state information. Given the transfer function matrix representation of a stable or stabilized plant, it can be partitioned as:

$$\begin{bmatrix} \boldsymbol{y}_s(s) \\ \boldsymbol{y}_r(s) \end{bmatrix} = \begin{bmatrix} \boldsymbol{G}_s(s) & \boldsymbol{G}_s^*(s) \\ \boldsymbol{G}_r(s) & \boldsymbol{G}_r^*(s) \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{u}_s(s) \\ \boldsymbol{u}_r(s) \end{bmatrix} + \begin{bmatrix} \boldsymbol{D}_s(s) \\ \boldsymbol{D}_r(s) \end{bmatrix} \cdot \boldsymbol{d}^*(s) \tag{1}$$

where $\boldsymbol{y}_s(s)$ and $\boldsymbol{y}_r(s)$ are the selected controlled variables (CVs) and the uncontrolled variables, respectively, $\boldsymbol{u}_s(s)$ are the selected manipulated variables (MVs) to control the subprocess $\boldsymbol{G}_s(s)$, and $\boldsymbol{u}_r(s) = 0$. Besides, $\boldsymbol{d}^*(s)$ are the disturbance variables (DVs). In Braccia et al. (2017), the process partitioning is done in the mixed-integer quadratic programming (MIQP) formulation by constraining specific entries in the inputs ($\boldsymbol{u}$) and the outputs ($\boldsymbol{y}$). The main idea is to minimize two indexes, i.e. SSD and NLE, which quantify the deviation of some specific process variables from their nominal operating point when set-point changes and disturbances occur individually. Obviously, the best solution depends on the CVs, MVs, and controller structure selection. The major contribution of this formulation is to translate the selection criteria based on the process matrices (G and D) to the input-output process variables (u and y). For this purpose, binary variable vectors to select the input and output variables ($\mathrm{z^I}$ and $\mathrm{z^O}$, respectively) and matrices to select the input-output pairing definition and controller design ($\mathrm{z^n}$ and $\mathrm{z^{nd}}$, respectively) are used. Using these binary variables, for each set-point and disturbance change the output and input variables are quantified by solving a system of equations, and the squared deviations are added up into the cost function.

In the proposed work, the control structure of the HEN is designed by only selecting the MVs, since all the process output variables (output temperatures) need to be controlled. In this sense, for a given HEN design ($\boldsymbol{G} = [g_{k,l}], \boldsymbol{D} = [\mathrm{d}_{k,j}]$), where $\mathrm{N_H}$ hot streams and $\mathrm{N_C}$ cold streams are integrated, the MIQP formulation presented in Braccia et al. (2017) is reformulated as follows:

$$\min_{\boldsymbol{u},\boldsymbol{y},\boldsymbol{z}} \mathrm{SSD} + \mathrm{NLE} = \sum_{i=1}^{m}\sum_{l=1}^{n} (u_{l,i}^{c})^2 + \sum_{j=1}^{d}\sum_{l=1}^{n} (u_{l,j}^{d})^2 + \sum_{i=1}^{m}\sum_{k=1}^{m} (y_{i,k}^{nc})^2 + \sum_{j=1}^{d}\sum_{k=1}^{m} (y_{j,k}^{nd})^2 \tag{2}$$

$$\text{subject to } \sum_{l=1}^{n} g_{k,l}\, u_{l,i}^{\mathrm{c}} - \phi_{k,i} = 0, \forall i,k \tag{3}$$

$$\sum_{l=1}^{n} g_{k,l}\, u_{l,j}^{\mathrm{d}} - \mathrm{d}_{k,j} = 0, \forall j,k \tag{4}$$

$$-\mathrm{M}\, z_l^{\mathrm{I}} \le u_{l,i}^{\mathrm{c}} \le \mathrm{M}\, z_l^{\mathrm{I}}, \forall i, l \tag{5}$$

$$-\mathrm{M}\, z_l^{\mathrm{I}} \le u_{l,j}^{\mathrm{d}} \le \mathrm{M}\, z_l^{\mathrm{I}}, \forall j, l \tag{6}$$

$$\sum_{l=1}^{n} z_l^{\mathrm{I}} = \mathrm{N_H} + \mathrm{N_C} \tag{7}$$

$$-\mathrm{M}\left(1 - z_{k,l}^{\mathrm{n}}\right) \le u_{i,k,l}^{\mathrm{nc}} - u_{l,i}^{\mathrm{c}} \le \mathrm{M}\left(1 - z_{k,l}^{\mathrm{n}}\right), \forall i, k, l \tag{8}$$

$$-\mathrm{M}\left(1 - z_{k,l}^{\mathrm{n}}\right) \le u_{j,k,l}^{\mathrm{nd}} - u_{l,j}^{\mathrm{d}} \le \mathrm{M}\left(1 - z_{k,l}^{\mathrm{n}}\right), \forall j, k, l \tag{9}$$

$$-\mathrm{M}\, z_{k,l}^{\mathrm{n}} \le u_{i,k,l}^{\mathrm{nc}} \le \mathrm{M}\, z_{k,l}^{\mathrm{n}}, \forall i, k, l \tag{10}$$

$$-\mathrm{M}\, z_{k,l}^{\mathrm{n}} \le u_{j,k,l}^{\mathrm{nd}} \le \mathrm{M}\, z_{k,l}^{\mathrm{n}}, \forall j, k, l \tag{11}$$

$$y_{i,k}^{\mathrm{nc}} + \sum_{l=1}^{n} g_{k,l}\, u_{i,k,l}^{\mathrm{nc}} - \phi_{k,i} = 0, \forall i, k \tag{12}$$

$$y_{j,k}^{\mathrm{nd}} + \sum_{l=1}^{n} g_{k,l}\, u_{j,k,l}^{\mathrm{nd}} = 0, \forall j, k \tag{13}$$

$$z_{k,l}^{\mathrm{n}} \le z_l^{\mathrm{I}}, \forall l, k, \quad z_l^{\mathrm{I}} \le \sum_{k=1}^{m} z_{k,l}^{\mathrm{n}}, \forall l, \quad z_{k,l}^{\mathrm{nd}} \le z_{k,l}^{\mathrm{n}}, \forall k, l \tag{14}$$

$$z_l^{\mathrm{I}} \le \sum_{k=1}^{m} z_{k,l}^{\mathrm{nd}}, \forall l, \quad 1 = \sum_{l=1}^{n} z_{k,l}^{\mathrm{nd}}, \forall k \tag{15}$$

$$-\mathrm{M}\left(1 - z_{i,l}^{\mathrm{nd}}\right) \le y_{i,l}^{\mathrm{r}} - g_{i,l}\, u_{l,i}^{\mathrm{c}} \le \mathrm{M}\left(1 - z_{i,l}^{\mathrm{nd}}\right), \forall i, l \tag{16}$$

$$\delta_1 z_{i,l}^{\mathrm{nd}} \le y_{i,l}^{\mathrm{r}} \le \delta_2 z_{i,l}^{\mathrm{nd}}, \forall i, l \tag{17}$$

$$\sum_{i=1}^{m} (1 - y_{i,i}^{\mathrm{nc}}) \ge \delta \tag{18}$$

In the proposed reformulation model, the SSD subproblem is addressed by eqs. (3) to (6). The variable $\boldsymbol{u}_i^c = \left[u_{l,i}^{\mathrm{c}}\right]$ and $\boldsymbol{u}_j^d = \left[u_{l,j}^{\mathrm{d}}\right]$ are the complete input vectors (including selected and not selected MVs) for the *i*-th set-point change and *j*-th disturbance effect, respectively. The unitary vectors $\phi_{\mathrm{i}} = \left[\phi_{\mathrm{k,i}}\right]$ generate the (normalized) change of set point i and the binary vector variable $\mathrm{z^I} = \left[\mathrm{z_l^I}\right]$ is used to select the MVs in the big-M formulation. The eq. (7) guarantees that the number of CVs and MVs are the same (square control structure). Note that, in this equation the number of CVs is equal to the total

number of currents in the network $(N_H + N_C)$ since all the process output variables (output temperatures) need to be controlled.

Moreover, the NLE subproblem is addressed by eqs. (8) to (13). The big-M inequalities given in eqs. (8) to (11) define the input response ($u_i^{nc} = [u_{i,k,l}^{nc}]$ and $u_j^{nd} = [u_{j,k,l}^{nd}]$) associated to the controller structure selected in z^n for each type of change. In this subproblem $y_i^{nc} = [y_{i,k}^{nc}]$ and $y_j^{nd} = [y_{i,k}^{nd}]$ are the complete net load effect vectors associated with control structure selection for ith set-point change and jth disturbance effect. The constraints in eqs. (14) and (15) are structural constraints. On the one hand, they guarantee that the controller structure z^n has only entries in allowed positions defined by the selection of CVs and MCs. In the other hand, they ensure that at least a decentralized policy is used. The selection of the decentralized control structure is represented with z^{nd}. This binary vector variable is used into a big-M constraint (eq. 16) to perform the RGA ($y^r = [y_{i,l}^r]$) matrix. The eq. (17) is used to guarantee a feasible and useful input-output pairing according to scalar parameters δ_1 and δ_2 fixed by the user. Finally an alternative stability/robustness test is include in the formulation in eq. (18), where a small positive lower bound δ is considered. A detailed description of SSD and NLE indexes, the perform of the RGA matrix or the stability criterion is not included here on account of the available space, but can be found in Braccia et al. (2017).

3. Case Study

In this section, a structure HEN02 studied in Escobar et al. (2009) are used to illustrate the application of the proposed MCS design procedure. In table 1 the design data are presented.

Table 1. Case study problem data

Stream	Tin (°C)	Tout (°C)	Flow Capacity ($kW°C^{-1}$)	Heat transfer coefficient ($kW°C^{-1}m^{-2}$)
H1	270	160	18	1
H2	220	60	22	1
C1	50	210	20	1
C2	160	210	50	1
CU	15	20		1
HU	250	250		1

On the one hand, the system input variables are the exchanger bypass on hot (and cold) side and the utility flow rate. On the other hand, the inlet temperature and flow rate of stream are the disturbance variables. The HEN02 configurations are composed by 11 inputs (where u_9, u_{10}, and u_{11} are associated with the flow rate of utilities), 4 outputs, and 8 disturbances. The disturbance propagation and control model presented in Yang et al. (2001) is used to obtain the steady-state gain of the process matrices (***G*** and ***D***). Since this model does not consider the hot and cold utilities, it is modified to: (a) incorporate these utilities as new heat exchangers and (b) include new process streams. In this sense, the hot and cold utilities are also characterized by the resulting steady-state gain matrices. Finally, the implementation utilizes an interface between MATLAB and GAMS, in which MATLAB generates the necessary data for each problem while GAMS solves the MIQP formulations. The state-of-the-art solver CPLEX 12.6 has been used to solve the MIQP problem.

3.1. RGA-based Input-Output Pairing.

We first analyse the results obtained by the proposed MIQP approach considering several alternatives for the interaction range $[\delta_1, \delta_2]$, which are presented in Table 2. It is important to recall that all results are associated to different decentralized control structures for HEN02. According to the discussion in Braccia et al. (2017), the parameter δ_1 is the most important limit, since it represents the minimum acceptable interaction.

In table 2, it can be observed that the performance index (SSD+NLE) increases if the interaction range is tighter. It is clear that a smaller lower bound for the interaction, i.e. $\delta_1 = 0.1$, allows the RGA matrix to include smaller elements. In this sense, the value of the RGA number (RGAn) increases.

Table 2. HEN02 decentralized results – Alternative interaction ranges $[\delta_1, \delta_2]$

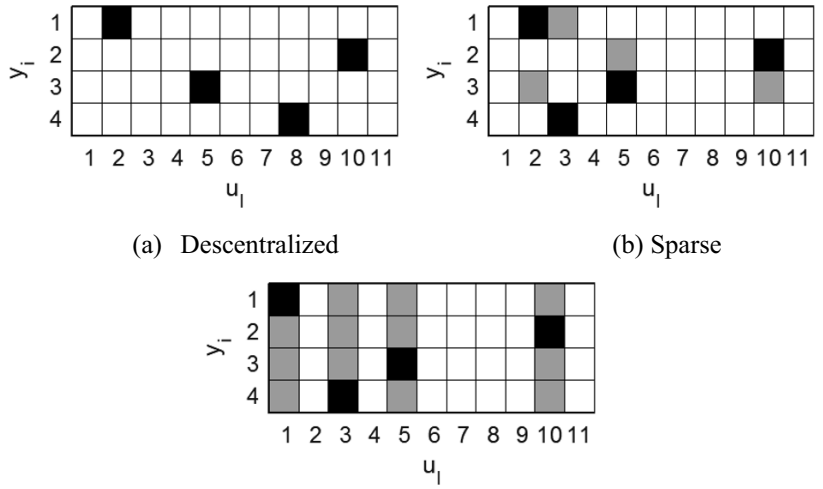
$[\delta_1, \delta_2]$	SSD+NLE	Optimization Time [s]	RGA Number
$[0.1 , 1\times10^{5}]$	66574.60	16.02	2.18
$[0.2 , 1\times10^{4}]$	66574.60	14.57	2.18
$[0.5 , 1\times10^{3}]$	69171.23	32.64	1.82
$[0.9 , 1\times10^{2}]$	81975.06	11.12	0
$[1 , 1\times10^{1}]$	81975.06	20.90	0

3.2. Multivariable Control Structure Design

In this section different controller structures, i.e. descentralized, sparse and full, for HEN02 configurations are analyzed. Figures 1 show the optimal MCS designs obtained using the MIQP formulation for each configuration and controller type. These figures display the input-output pairing and plant-model mismatch obtained (which are marked using black and gray background, respectively).
Furthermore, in order to quantify and compare the performance of the solutions, three indexes are calculated and reported in Table 3: the performance index (SSD+NLE), the RGAn, and condition number (ρ). In this sense, this table shows that for HEN02 configuration the sparse structure has the best SSD+NLE index.

4. Conclusions

In this work, a new systematic methodology to address the multivariable control structure (MCS) design of heat exchanger networks based on a mixed-integer quadratic programming (MIQP) formulation is proposed. The optimal solutions of the proposed formulation can be obtained using state-of-the-art solvers such as CPLEX in GAMS. On the one hand, several alternatives for the interaction range were considered to evaluate their impacts on MCS design. On the other hand, the proposed approach has been successfully tested on HEN02 configuration. Three types of MCS based on three different controller topologies were designed. The results show that the optimal sparse controllers have the best performance. However, the improvement in the SSD+NLE indexes produced by this controller topology requires adding several control loops. In future work the integration between process control and process synthesis and design will be analyzed. In particular, an appropriate MCS design to guarantee a suitable HEN operation over a specified range of expected variations in the inlet temperatures of the process streams will be developed.



(a) Descentralized (b) Sparse

(c) Full

Figure 1: HEN02 Results – Selected paring and process-model mismatch

Table 3. Case Study Results - Utility wt.: 500 - Bypass wt.: 0. $[\delta_1, \delta_2] = [0.5, 100]$

	Control Structure	SSD+NLE	RGA Number	$\rho = \frac{\bar{\sigma}(\boldsymbol{G}_s)}{\underline{\sigma}(\boldsymbol{G}_s)}$
HEN02	Decentralized	299527.61	1.81	241.61
	Sparse	250143.21	1.81	264.79
	Full	280389.50	1.81	189.32

References

L. Braccia, L., Marchetti, P.A., Luppi, P., Zumoffen, D., 2017. Multivariable control structure design based on mixed-integer quadratic programming. Ind. Eng. Chem. Res., 56 (39), 11228–11244. DOI: 10.1021/acs.iecr.7b02270

Z. Escobar, M., Trierweiler, J.O., 2009. Bypass design for control and optimization of heat exchanger networks Computer Aided Chemical Engineering. 27, 1665-1670.

Y. Mathisen, K. W., 1994. Integrated design and control of heat exchanger networks. Ph.D. Thesis. Trondheim University. NTH Norway.

X. Yang, Q. Z., Yang, Y. H., Huang, Y. L., 2001. Cost-effective bypass design of highly controllable heat-exchanger networks. AIChE Journal 47 (10), 2253–2276.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50096-3

Control Allocation Based Plantwide Control Structure Design for Heat Exchange Networks

Patricio Luppi[a,b*], Lautaro Braccia[a], Pablo A. Marchetti[c], David Zumoffen[a]

[a]*Centro Internacional Franco-Argentino de Ciencias de la Información y de Sistemas (CIFASIS, CONICET-UNR), 27 de Febrero 210 bis, Rosario (S2000EZP), Argentina*
[b]*Departamento Control – Escuela Ing. Electrónica – FCEIA – UNR, Pellegrini 250, Rosario (S2000BTP), Argentina*
[c]*Instituto de Desarrollo Tecnológico para la Industria Química (INTEC, UNL-CONICET), Guemes 3450, Santa Fe (3000), Argentina*
luppi@cifasis-conicet.gov.ar

Abstract

In this work, a novel design methodology based on the control allocation (CA) approach is presented. The CA structure allows to explicitly handle the input constraints (e.g. actuator position limits) as well as the secondary control objectives (e.g. minimization of the control energy). In addition, it can manage actuator saturations (and faults) without performing a reconfiguration of the control structure. The proposed architecture involves two levels: (i) a high-level decentralized structure implemented with conventional PI controllers and (ii) a CA module based on a weighted least-squares formulation which utilizes an active set method algorithm. The complete procedure only requires a steady-state model of the process for the structure design. Several dynamic simulations demonstrate the performance of the suggested strategy.

Keywords: plantwide control design, control allocation, heat exchange networks

1. Introduction

The control of heat exchanger networks (HEN) represents an important problem in chemical engineering. On the one hand, HEN are commonly responsible of managing much of the process energy recovery, which significantly impacts on the global operating costs. On the other hand, HEN typically introduce important interactions among the process variables, complicating the control structure design and the controllers tuning.

In this work, the control allocation (CA) philosophy (Johansen and Fossen, 2013) is considered in order to design control structures (CS) for HEN. Commonly, CS can be distinguished depending on their degree of centralization, where decentralized and centralized are the common control policies. Traditional decentralized architectures based on multi-loop PID controllers present a significant drawback: they are not able to handle actuator constraints. While centralized strategies, such as MPC-based control systems, present several advantages over multi-loop structures, most contributions in the literature propose decentralized CS for HEN. The main reasons lie in the costs associated with the model identification task, the commissioning/maintenance, and the typical computational burden of the MPC systems.

This work proposes CS for HEN based on a novel architecture. This alternative design presents two hierarchical levels: (i) a high-level control (HLC) strategy consisting of

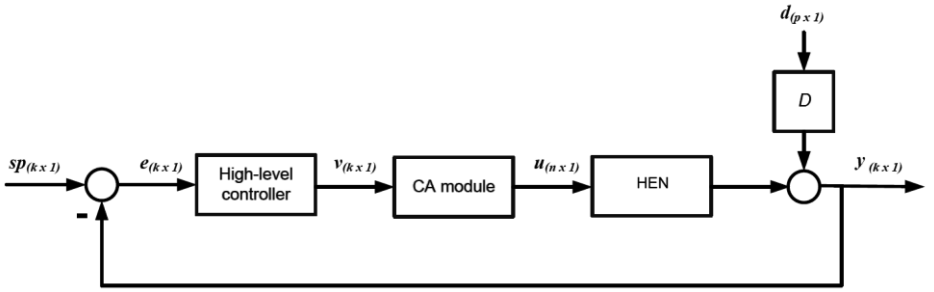


Figure 1: CA-based control structure

multi-loop PID controllers for computing a vector of virtual inputs and (ii) a control allocation module for mapping the vector of virtual inputs on the manipulated variables vector. This modular architecture allows to design the HLC by taking into account minimum information about the CA block. In addition, the CA module can handle the actuator constraints by itself. Another advantage is that it has the ability to accommodate actuator saturations and/or faults due to the automatic reutilization of the manipulated variables (redistribution of the manipulated energy).

The CS design procedure begins with the partition of the HEN model. The objective is to isolate the HLC design task from the CA module configuration. The next stage focuses on the HLC design, which is based on the well-known methodology presented by Braccia et al. (2017). In addition, a weighted least-squares formulation based on an active set method (Harkegard, 2002) is adopted for solving the CA problem due to its reduced computational load. The performance of the proposed CA-based CS design is evaluated through simulation of a HEN dynamic model, and it is compared against a conventional decentralized control structure.

2. Control structure design methodology

Consider a stabilized process with *n* inputs, *m* outputs and *p* disturbances expressed as:

$$y(s) = G(s)u(s) + D(s)d(s) \tag{1}$$

where *y(s)*, *u(s)*, and *d(s)* correspond to column vectors of dimension *(m x 1)*, *(n x 1)*, and *(p x 1)*, respectively. *G(s)* and *D(s)* are transfer function matrices which relate the inputs-outputs and disturbances-outputs, respectively, with dimension *(m x n)* and *(m x p)*. As illustrated in Fig. 1, a control allocation architecture is based on a high-level control (HLC) strategy and a control allocation (CA) module (Johansen and Fossen, 2013). The HLC strategy must compute a (column) vector of *k* virtual variables *v(s)*, which are usually defined based on the main process objectives. In addition, the CA module deals with the optimal management of the manipulated variables *u(s)* in order to satisfy:

$$v = M\,u, \quad u \in C \tag{2}$$

where *M* is a *(k x n)* matrix (commonly called control effectiveness matrix) and *C* represents the control constraints. Typically, the number of manipulated variables is greater than the number of outputs to be controlled (i.e. $n > k$). This allows to formulate

the CA problem as an optimization problem, where secondary control objectives can be easily included.
For the *(m x n)* matrix *G*, the singular value decomposition is defined as:

$$G = USV^T = [U_1 \quad U_2]\begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix}\begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \tag{3}$$

where *U* and *V* are *(m x m)* and *(n x n)* unitary matrices and the *(m x n)* matrix *S* contains real non-negative singular values arranged in descending order along its main diagonal. In Eq.(3), S_1 contains the *k* largest singular values. If the remaining (relatively small) singular values contained in S_2 are forced to zero, then an approximate representation of *G* is produced:

$$G \cong U_1 S_1 V_1^T \tag{4}$$

where U_1, S_1 and V_1 have dimension *(m x k)*, *(k x k)* and *(n x k)*. By discarding the smallest singular values of *S*, the least effective control directions of the plant are not taken into account. Note that Eq.(4) can be expressed as:

$$G \cong G_1 G_2 \quad with \quad G_1 = U_1 \quad and \quad G_2 = S_1 V_1^T \tag{5}$$

Assuming that the problem of Eq.(2) (unconstrained case) is based on $M = G_2$ and that G_2 has full rank, then it is simple to demonstrate that the transfer function which links *y* with *v* is U_1. Inversely, assuming that *y* and *v* are related via $G_1=U_1$ and that the pseudo-inverse of *G* from Eq.(4) is $G^+=V_1(S_1)^{-1}(U_1)^T$, then it is easy to show that:

$$u = V_1 S_1^{-1} v \tag{6}$$

i.e. the CA problem is based on $M = G_2$. In this framework, the proposal is to utilize the matrix $G_1=U_1$ for the HLC structure design and $G_2=S_1(V_1)^T$ for configuring the CA module. Note that the *(m x k)* matrix U_1 directly defines *k* virtual variables which are linked to the process outputs. Thus, it is possible to apply any know design methodology such as Braccia et al. (2017) for developing a decentralized (diagonal) HLC structure from $G_1=U_1$ and *D*.
Concerning the CA problem, the weighted least-squares (WLS) formulation proposed by Harkegard (2002) is considered here:

$$\min_u \left\| W_u(u - u_p) \right\| + \gamma \| W_v(G_2 u - v) \| \quad s.t. \quad \underline{u} \le u \le \overline{u} \tag{7}$$

where u_p corresponds to the nominal control input, W_u and W_v are weighting matrices which affect the control distribution among the actuators and the virtual variables, respectively, and gamma (γ) is a weighting parameter for relatively prioritizing the minimization of the control allocation error and the actuators movement. In addition, the constraints of Eq.(7) represent the position and/or the rate actuator limits. In this work, the active set based solver proposed by Harkegard (2002) is utilized to obtain the optimal solution of the CA problem. Such algorithm is very efficient for CA because a good estimate of the optimal active set is available from the previous sampling period. Moreover, active set methods are simple to implement with low computational burden.

3. Application example: HEN system

In this section it is considered the HEN studied in Braccia et al. (2015). It consists of two cold streams C_1 and C_2, one hot stream H_1, three heat exchangers (E_{111}, E_{121}, E_{122}), and one cooling utility (CU_1). The objective is to keep the outlet temperatures of hot and cold streams at their reference values (i.e. $Thout_1$ = 318 K, $Tcout_1$ = 393 K, $Tcout_2$ = 393 K) through the manipulation of the heat exchangers, rejecting flow and temperature disturbances at the inlet streams. In the following, uh_{ijk} (and uc_{ijk}) represent the exchanger bypass of hot (and cold) side between hot stream i and cold stream j at stage k, and fcu_1 represents the cooling utility flow, which constitute 7 system input variables (uh_{111}, uh_{121}, uh_{122}, uc_{111}, uc_{121}, uc_{122}, fcu_1). In addition, $Thin_i$ (and $Tcin_i$) correspond to the inlet temperature of hot (and cold) stream i, and fh_i (and fc_i) represent the flow capacity of hot (and cold) stream i, which constitute 6 system disturbance variables ($Thin_1$, $fhin_1$, $Tcin_1$, $fcin_1$, $Tcin_2$, $fcin_2$). The normalized steady-state gain matrices G and D corresponding to the nominal operating point presented in Braccia et al. (2015) are considered here for the CS design. The associated inputs, outputs and disturbances are: $u = [uh_{111}, uh_{121}, uh_{122}, uc_{111}, uc_{121}, uc_{122}, fcu_1]^T$, $y = [Thout_1, Tcout_1, Tcout_2]^T$ and $d = [Thin_1, fhin_1, Tcin_1, fcin_1, Tcin_2, fcin_2]^T$.

3.1. High-level control

In order to satisfy the HEN requirements, all outputs $Thout_1$, $Tcout_1$ and $Tcout_2$ must be controlled. Thus from Eq.(3), k = 3 must be configured. Taking into account the RGA-number minimization procedure proposed by Kariwala and Cao (2010), the obtained pairings between the virtual and controlled variables are: v_1-$Thout_1$, v_2-$Tcout_2$ and v_3-$Tcout_1$, which correspond to a decentralized (diagonal) CS implemented with 3 PI loops. The corresponding tuning parameters are: Kc_1 = 70, Ti_1 = 3.4, Kc_2 = 110, Ti_2 = 0.45, and Kc_3 = -43.3, Ti_3 = 1.25, which were obtained through the IMC method proposed by Garcia and Morari (1985).

3.2. Control allocation module

The proposed CA algorithm is based on the active set method for the weighted least-squares (WLS) formulation presented in section 2. In the following, the dynamic performance and timing properties of the CA are evaluated taking into account different simulation scenarios. On the one hand, the tuning of the gamma (γ) parameter as well as the control weighting matrix W_u are considered to show the CA ability for handling secondary control objectives (i.e. the control energy). On the other hand, an adjusted version of the nominal constraint set is proposed to demonstrate the advantages of the WLS algorithm.

Table 1: simulation scenarios

Scenario	Time [s]	Event	Variable	Type	Magnitude
Sc1	10	SP change	$Thout_1$	Step	+1.0 K
	500	"	$Tcout_1$	"	-1.0 K
	1000	"	$Tcout_2$	"	-1.0 K
Sc2	10	DV change	$Thin_1$	"	+2.5 K
	500	"	$Tcin_1$	"	+2.5 K
	1000	"	$Tcin_2$	"	+2.5 K
Sc3	10	DV change	$fhin_1$	"	+0.3 $kWC^{-1}m^{-2}$
	500	"	$fcin_1$	"	+0.3 $kWC^{-1}m^{-2}$
	1000	"	$fcin_2$	"	+0.3 $kWC^{-1}m^{-2}$

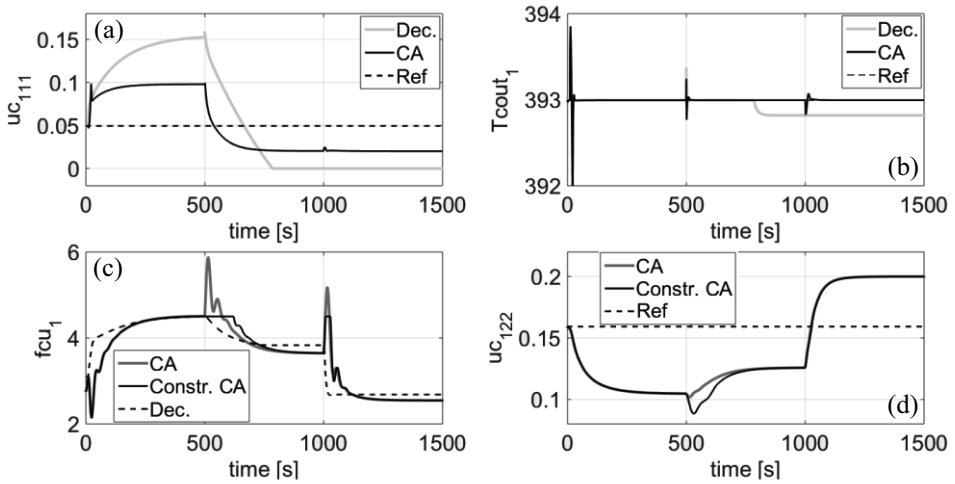


Figure 2: Dynamic responses

3.3. Results

In this section, several simulations of the complete system involving the dynamic model of the HEN controlled with the high-level controller plus the CA module are presented. In this context, different scenarios characterized by alternative CA configurations, constraints sets, setpoint (SP) changes and disturbances (DV) are considered, see Table 1. Also, a conventional decentralized CS for the HEN system which is composed of 3 control loops was developed for the sake of comparison. It was obtained taking into account the methodology proposed by Braccia et al. (2017). The selected manipulated variables and the corresponding pairings are: fcu_1-$Thout_1$, uc_{111}-$Tcout_1$ and uc_{122}- $Tcout_2$. The associated tuning parameters are: Kc_1 = -0.28, Ti_1 = 0.5, Kc_2 = -0.07, Ti_2 = 0.0097, and Kc_3 = -0.058, Ti_3 = 0.0089, respectively. In order to quantify and compare the performance of the solutions, three well-know indexes are utilized: the integral absolute error (IAE), the error improvement percent (EIP) and the control energy (E_u).

The first analysis corresponds to a performance comparison considering alternative gamma (γ) values for the WLS algorithm, taking into account the scenario *Sc2* superimposed with the scenario *Sc3* (i.e. a new scenario named here as *Sc4*). A gamma value of γ = 1000 was considered as a reasonable tuning because it provides an acceptable trade-off between control energy and dynamic performance, see Eq.(7). In this regard, for γ = 1000 all outputs reduce the IAE (up to a 60 %) with respect to a configuration with γ = 100, subject to a higher control energy requirement (which is less than 10% for all manipulated variables).

Unlike the obtained conventional decentralized CS, the CA-based system (with γ = 1000) meets all the control requirements for all the simulation scenarios, namely: *Sc1*, *Sc2*, *Sc3* and *Sc4*, and also for a critical version of *Sc4* where the disturbance $fhin_1$ changes its direction (i.e. a step of -0.3 $kWC^{-1}m^{-2}$ with respect to the nominal value). This is illustrated in Fig. 2 (a) and (b), where the evolution of the manipulated variable uc_{111} and the output $Tcout_1$ are compared taking into account both control strategies, subject to the commented critical scenario. As can be seen, the conventional decentralized CS is not able to regulate the output temperature due to the saturation of uc_{111}. Contrarily, the CA-based CS takes optimal advantage of all available degrees of

freedom, with a remarkable reduction of the control energy (up to a 81 %) for uc_{111}, uc_{122} and fcu_1.
The second analysis corresponds to the tuning of the control weighting matrix W_u in order to reduce the cooling utility flow (fcu_1). The considered test scenario was *Sc3*. For this purpose, the default W_u (i.e. an identity matrix) was replaced by W_u = *diag*([1,1,1,1,1,1,5]). With this configuration, the cooling utility flow was reduced 1.2% with respect to the nominal tuning, and 0.8% when compared against the decentralized CS, without affecting the dynamic performance of the system outputs.
Finally, the ability of the CA-based CS design for constraint handling was evaluated. To this end, a position high limit for the cooling utility flow (fcu_1) was defined (equal to 4.5) in order to remove some fcu_1 consumption peaks, see Fig. 2 (c). It can be noted that during the activation of the fcu_1 position high limit, the WLS algorithm further manipulates the remaining degrees of freedom so as to enhance the system performance. This effect is illustrated in Fig. 2 (d) where the control energy requirement is a 7% higher for uc_{122}. In addition, Fig. 2 (c) shows how the CA-based approach reduces 1.5 % the cooling utility flow (fcu_1) with respect to the decentralized CS, which does not consider explicitly the constraint.

4. Conclusions

The proposed CA-based CS design presents several interesting characteristics such as the management of secondary control objectives and constraint handling, which are usually provided by more complex controllers like MPC. Note that the design of the overall structure can be addressed considering only a steady-state model of the HEN process. In addition, the weighted least-squares (WLS) formulation was selected due to the good trade-off between the obtained dynamic performance and computational burden. In fact, the maximum computation time per sample was less than 10 milliseconds taking into account a compiled version (C-language) of the WLS solver. In this context, the WLS algorithm can be considered as an interesting option for real-time industrial implementations.

References

T. Johansen and T. Fossen, 2013, Control allocation - A survey, Automatica, 49, 1087-1103.

L. Braccia, P. Marchetti, P. Luppi and D. Zumoffen, 2017, Multivariable Control Structure Design Based on Mixed-Integer Quadratic Programming, Ind. Eng. Chem. Res., 56, 11228-11244.

O. Harkegard, 2002, Efficient active set algorithms for solving constrained least squares problems in aircraft control allocation, Proceedings of the 41st IEEE Conference on Decision and Control.

L. Braccia, P. Luppi, M. García and M. Basualdo, 2015, Synthesis of flexible heat exchanger networks integrated with reconfigurable control design, 12th International Symposium on Process Systems Engineering.

V. Kariwala and Y. Cao, 2010, Branch and Bound method for multiobjective pairing selection, Automatica, 46, 932-936.

C. Garcia and M. Morari, 1985, Internal model control - Design procedure for multivariable systems, Industrial and Engineering Chemistry Product Research and Development, 24, 472-484.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50097-5

Modeling and Simulation of Autothermal Reforming Reactor of Diesel over Ni-based Catalyst in Solid Oxide Fuel Cell based Auxiliary Power Unit System

Sohyun Jeong, Daewook Kim, Jay H. Lee*

Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon, 34141, Republic of Korea
jayhlee@kaist.ac.kr

Abstract

The catalytic reforming process in the fuel cell based APU system is effective in supplying the fuel cell with a hydrogen-rich gas stream. Of several reforming technologies, autothermal reforming (ATR) is regarded as the most reliable for transportation applications, due to its durability during long-lasting operation and independence of heat sources. As autothermal reforming is a thermally neutral process involving exothermic and endothermic reactions, the analysis of the reaction coupled with transport phenomena inside the reformer is needed for obtaining operating conditions with respect to a specific fuel source. Such analysis ensures that the reformate gas enter the fuel cell at suitable composition and temperature. Thus, this study proposes a dynamic model of ATR of diesel, while considering the heat transfer phenomena involved with the kinetics of the Ni-based catalyst. For rate parameters for the Ni-based catalyst, model-based parameter estimation is performed in gPROMS Process Builder with micro-reactor experimental data. The established kinetics is then incorporated into the one-dimensional homogeneous model that accounts for heat transfer phenomena inside the monolithic reformer. The developed model shows good agreement with the corresponding experiment of the 1kW-class diesel ATR reactor conducted with step changes in the operating conditions.

Keywords: autothermal reforming, diesel reforming, dynamic modelling, fuel cell based auxiliary power unit system

1. Introduction

In efforts to reduce the greenhouse gas emissions, particularly carbon dioxide, a global movement toward more efficient and cleaner energy resource utilization has expanded and developed. For a clean and sustainable energy, a fuel reforming process has received much attention as it converts fossil fuel including diesel, coal, and natural gas to hydrogen-rich gas, providing a source for the fuel cell based auxiliary power unit (FC-APU) to generate electric power. APUs are commonly installed on heavy-duty vehicles, aircrafts, and other mobile systems for environmental purposes. Specifically, SOFC-APU on-board heavy-duty trucks enable to reduce unfavorable gas emission, noise during idling, and fuel consumption and idling time of primary engine (Xu et al., 2013).

As a primary part of the FC-APU system, reforming process has been investigated by a number of studies on the effects of fuel source, catalyst type, catalyst material, and reformer type on the reforming efficiency. A commercial FC-APU application widely uses autothermal reforming among several reforming technologies. ATR is a combination of exothermic partial oxidation and endothermic steam reforming reactions. The ATR occurs independently of external heat sources while achieving a high hydrogen yield. Since the reformer efficiency determines the SOFC performance and subsequently the overall output power, the reforming performance is of high importance and therefore is controlled in operation. The fuel source of the reformer also largely contributes in determining the efficiency and operation conditions. When operated with methane or light hydrocarbons as the fuel source, the reformer may show a complete conversion and stable operation (Hoang, 2004). However, the kinetic studies on diesel or heavy hydrocarbon reforming are hardly reported for its complex characteristics. The analysis of heavy hydrocarbon fuel reforming is challenging mainly due to its large carbon contents, which may cause coke formation and potential catalyst failure (Xu et al., 2013). The conventional ATR reactions may be inadequate to represent the complexity of diesel reformation such as decomposition of heavy hydrocarbons.

Thus, this study proposes a comprehensive mathematical model that describes ATR of diesel over the Ni-based catalyst. The ATR reaction kinetics regarding diesel reforming is refined to reflect the dynamic behaviour of an actual conversion process. We examine the kinetic model developed with model-based parameter estimation, followed by the dynamic model construction and simulation.

2. Model Development

2.1. Reaction kinetics

The autothermal reformer in the SOFC-APU system is implemented with a monolithic type catalyst. To examine the Ni-based catalyst performance in the monolith, a preliminary experiment has been performed at a micro-reactor scale. Experiments are carried out with varying input conditions such as steam-to-carbon ratio (SCR) of 1.5, 2.0, 2.5, 3.0, oxygen-to-carbon ratio (OCR) of 0.6, 0.8, 1.0, and temperature of 800, 850, 900 ^{o}C. The commercial-grade diesel is injected with an ultrasonic fuel injector. The composition of reformate gases is analyzed with a gas chromatograph. When compared to the thermodynamic equilibrium simulation based on adiabatic condition in Aspen Plus, a significant difference in product composition is noticed: two main components entering the SOFC stack, hydrogen and carbon monoxide, are produced 26 % and 33 % less than ideal. These differences suggest that an accurate prediction of the product composition requires an appropriate kinetic model of the reactios with the Ni-based catalyst.

Conventional autothermal reforming process comprises three main reactions: partial oxidation (POX), steam reforming (SR), and water-gas shift (WGS) reactions. Many of previous studies report that it is difficult to obtain the kinetics of POX for the heavy hydrocarbons due to the operation at high temperature and short contact time (Parmar et al., 2010). This difficulty may be solved by replacing POX with total oxidation (TOX) reaction in the ATR process. In the ATR operation condition where steam is supplied to the reactor in excess, Parmar et al. has suggested that the TOX may be regarded as a combination of POX, WGS, and hydrogen oxidation reactions.

The overall ATR reaction including TOX instead of POX is then compared with a stoichiometric balance of the experimental data, to confirm our assumption about

components and reactions. It is shown that when methane production is neglected as in the conventional ATR reactions, the overall reaction is not able to represent the practical behavior inside the reformer. Therefore, the reaction set has been refined based on the least squares method, minimizing the production and consumption difference between theoretical predictions and experimental data. Four key reactions are selected among the theoretically possible reactions (Xu et al. 2013): TOX, pre-reforming (PR), SR and WGS reactions (Table 1).

For the specific rate parameters for the Ni-based catalyst, the model-based parameter estimation is performed in gPROMS Process Builder with the data obtained with the commercial diesel. For each reaction, eight rate parameters of pre-exponential factors and activation energies are optimally determined by using the method of Maximum Likelihood Estimation. The assumptions used in the estimation are:

1. Ideal gas behavior
2. Fully developed laminar flow
3. Channels in the monolith assumed as one single channel
4. Reactor assumed as an isothermal plug flow reactor with mass balance of

$$\frac{dF_i}{dV} = \sum_j v_{ij} r_j \tag{10}$$

where F_i is the molar flow rate of species, V is the control volume, v_{ij} is the stoichiometry of species i in j^{th} reaction and r_j is the j^{th} reaction rate.

5. Rate law expression with powers of one
6. Reformate gas: H_2, CO, CO_2, O2, N_2, H_2O, CH_4 and light hydrocarbons (C_{2+}).

2.2. Mass and heat balances

To enhance the understanding of process dynamic behavior of the reformer, heat transfer phenomena inside the monolithic reformer are associated with the kinetics. As the main components pass through the monolith channels, they react on the catalyst-coated walls. Based on the assumptions used in the kinetic modeling as well as the following, the governing equations and accompanying boundary conditions for our reformer model are defined in Table 2.

Table 1. List of refined autothermal reactions with corresponding reaction rate expressions

Refined ATR reaction set selected (C_nH_m: diesel surrogate)		ΔH_{298} (kJ/mol)	
TOX	$C_nH_m + \left(n + \frac{m}{4}\right) O_2 \rightarrow nCO_2 + \frac{m}{2} H_2O$	-7513	(1)
	$r_{TOX} = \boldsymbol{k_{TOX}} y_{fuel} y_{O_2}$		(2)
PR	$C_nH_m + \left(\mathrm{n} - \frac{m}{4}\right) H_2O \rightarrow \left(\frac{n}{2} - \frac{m}{8}\right) CO_2 + \left(\frac{n}{2} + \frac{m}{8}\right) CH_4$	-92.67	(3)
	$r_{PR} = \boldsymbol{k_{PR}} y_{fuel} y_{H_2O}$		(4)
SR	$C_nH_m + nH_2O \rightarrow nCO + \left(n + \frac{m}{2}\right) H_2$	1928	(5)
	$r_{SR} = \boldsymbol{k_{SR}} y_{fuel} y_{H_2O}$		(6)
WGS	$CO + H_2O \leftrightarrow CO_2 + H_2$	-41.20	(7)
	$r_{WGS} = \boldsymbol{k_{WGS}} \left(y_{CO} y_{H_2O} - y_{CO_2} y_{H_2} / K_{eq}\right)$		(8)
	where $\boldsymbol{k_i} = k_{0,i}\, exp\left(-\frac{E_{A,i}}{RT}\right)$, and $K_{eq} = \exp\left(\frac{4577.8}{T} - 4.33\right)$ (Creaser et al., 2011)		(9)

Table 2. Governing equations for ATR reactor with corresponding boundary conditions

Mass balance for gas phase	$\frac{\partial \rho_g x_i}{\partial t} + \frac{\partial (u x_i)}{\partial z} = MW_i \sum_j \nu_{ij} r_j$	(11)
Energy balance for gas phase	$\rho_g C_{p,g} \frac{\partial T_g}{\partial t} + \rho_g C_{p,g} u \frac{\partial T_g}{\partial z} + \hat{a}_s u_{Overall}(T_g - T_s) = \sum_j (-\Delta H_j) r_j$	(12)
Energy balance for solid phase	$\rho_s C_{p,s} \frac{\partial (T_s)}{\partial t} = \hat{a}_s u_{Overall}(T_g - T_s) - \hat{a}_{out} h_{out}(T_s - 298.15)$	(13)
	where $\frac{1}{u_{Overall}} = \frac{1}{h_g} + \frac{w_s}{k_s}$	(14)
Boundary conditions	At reformer inlet ($z = 0$), $x_k = x_{k0}$, $T_g = T_{g0}$	

Table 3. Heat transfer correlations applied in the dynamic model

Heat transfer between gas and solid (Hoang, 2004)

$$j_{H,ln} = \frac{Nu_{ln}}{RePr^{1/3}} = \frac{h_{ln}}{\langle \rho v \rangle \hat{C}_p} \left(\frac{C_p \mu}{k}\right)^{2/3} = \frac{h_{ln} S}{\omega \hat{C}_p} \left(\frac{C_p \mu}{k}\right)^{2/3} \quad (15)$$

$$j_D = j_H = 0.91 Re_j^{-0.51} S_f \quad for\ Re_j < 50 \quad (16)$$

$$j_D = j_H = 0.61 Re_j^{-0.41} S_f \quad for\ Re_j > 50 \quad (17)$$

Natural convection from the solid (Bird, 1960)

$$Ra_L = Gr_L Pr = \frac{g\beta(T_s - T_\infty)L^3}{\nu\alpha} \quad (18)$$

$$\overline{Nu}_D = \left\{0.60 + \frac{0.387 Ra_D^{1/6}}{[1 + (0.559/Pr)^{9/16}]^{8/27}}\right\}^2 \quad for\ Ra_D \le 10^{12} \quad (19)$$

1. One-dimensional homogeneous system
2. Negligible pressure drop
3. Neglected heat transfer among channels
4. Neglected radiation and diffusion

This mathematical model is simulated in gPROMS Process Builder with thermodynamic physical properties obtained from Multiflash, database provided in gPROMS Process Builder. The partial differential equations in the balance equations are solved with the backward finite difference method. The heat transfer coefficient between the gas and solid phases are determined from the Chilton-Colburn factor (Bird, 1960). The heat loss to the air is considered as natural convection from the wall and the corresponding correlations are employed in the model (Table 3).

3. Results and Discussion

3.1. Kinetic model

The estimated rate parameters for the reactions are presented in Table 4. The activation energies for SR and WGS are roughly compared to the literature, despite variations in reaction mechanism for different catalysts (Creaser et al., 2011). The order of magnitude for each reaction is consistent with the reported value. For validation, the outlet molar flowrates of H_2 and CO in the micro-reactor experiment are compared with the kinetic model (Figure 1).

Table 4. Estimated rate parameters for refined reaction set

	Pre-exponential factor (mol/kg$_{cat}$/s)	Activation Energy (kJ/mol)	
	Estimated	Estimated	Literature
TOX	2.212×10^{7}	31.54	
PR	2.339×10^{11}	196.6	
SR	5.847×10^{12}	201.6	116.9
WGS	6.791×10^{2}	43.56	54.07

The kinetic model describes the measured data with a significantly improved accuracy as all data points fall within 15% error range. From Figure 2, we could observe the predicted behavior inside the reformer. The total oxidation reaction immediately occurs at the initial stages of the reformer, consuming the entire oxygen source. Steam reforming gradually takes place with the remaining fuel and generates hydrogen throughout the reformer.

3.2. Dynamic model

The simulation presents the dynamic response of the reforming process of diesel over the Ni-based catalyst. The predictions are validated with the experimental data at the reformer exit (Figure 3). To examine the effect of input conditions, step changes of SCR and OCR are imposed on the simulation that has been initially set with OCR of 0.7 and SCR of 2.5. We have varied SCR in ±0.3 and OCR in ±0.05.

One critical aspect to be concerned in the dynamic simulation is heat transfer. The total oxidation releases a large amount of heat near the initial stage of the reformer. As it is strongly exothermic, the mode to control the temperature inside the channel is necessary. Without considering any heat transfer phenomena, case 1 simulation response shows only the heat generated from the reactions and the outlet temperature is approximately 100 K higher than the experimental data. This temperature gap can be controlled by heat transfer correlations between the gas in the channel and the wall, as well as between the wall and ambient air. Case 2 simulation accounts for the heat transfer between the gas and the wall. Its response exhibits a reasonable time delay due to the heat transfer in every step sequence since the wall begins to hold the heat transferred from the gas. In the simulation considering all the reactions and the heat transfer correlations, the dynamic response follows the experiment trend closely.

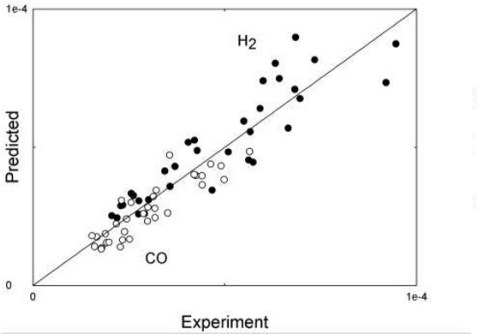


Figure 1. Parity plot for outlet molar flowrates (mol/s) of H_2 (black) and CO (white) in all input conditions.

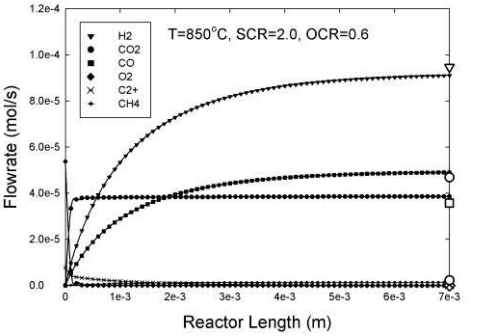
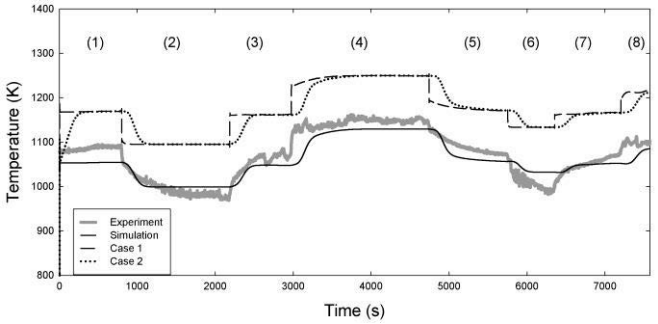


Figure 2. Flowrate profile of main components at different operating conditions. Simulation (closed symbol) compared with the micro-reactor experimental data (open symbol).



Interval	SCR	OCR
1	2.5	0.7
2		0.65
3		0.7
4		0.75
5		0.7
6	2.8	
7	2.5	
8	2.2	

Figure 3. Temperature profile at reformer exit compared with experimental results in the sequence of step changes. Adiabatic simulation results (dotted line) and non-adiabatic simulation results (solid line). Step change details are listed in the table.

4. Conclusions

A comprehensive model for ATR of diesel over Ni-based catalyst has been suggested in this paper. The characteristics of Ni-based catalyst in the ATR has been studied and defined by model-based parameter estimation by using the method of MLE, which determined the optimal rate parameter values that minimize the differences with the given experimental data. The validation of the micro-reactor experimental data confirmed that the developed kinetics predicts the actual diesel reforming conversion quite well. The dynamic model has been developed considering the established kinetics, as well as heat transfer phenomena between the channels and the surroundings. Given varying operating conditions in step sequences, the dynamic simulation showed a good agreement with the corresponding experimental data.

The developed model ensures the prediction of the diesel reforming performance by including the hydrocarbon decomposition in the kinetics. The diesel ATR should be carefully modelled and analysed given the complexity in diesel compositions and the uncertainty in the inputs during reforming. It is found from the simulation that the efficiency of the diesel reforming is sensitive to the input variables and temperature. The optimal operation and control strategy will be developed with the aid of sensitivity analysis with respect to operating conditions in our future study.

References

R.B. Bird, W.E. Stewart, E.N. Lightfoot, 1960, Transport Phenomena, John Wiley and Sons, New York, NY.

D. Creaser, X. Karatzas, B. Lundberg, L.J. Pettersson, J. Dawody, 2011, Modeling study of 5kW e-scale autothermal diesel fuel reformer, Applied Catalysis A: General, 404, 1, 129-140

D.L. Hoang, S.H. Chan, 2004, Modeling of a catalytic autothermal methane reformer for fuel cell applications. Applied Catalysis A: General, 268, 1, 207-216

R.D. Parmar, A. Kundu, C. Thurgood, B.A. Peppley, K. Karan, 2010, Kinetic studies of the autothermal reforming of tetradecane over Pt/Al2O3 catalyst in a fixed-bed reactor. Fuel, 89, 1212-1220

X. Xu, P. Li, Y. Shen, 2013, Small-scale reforming of diesel and jet fuels to make hydrogen and syngas for fuel cells: A review. Applied Energy, 108, 202-217

A. Sharma, M. Jelemensky, R. Paulen, M. Fikar, 2016, Estimation of membrane fouling parameters for concentrating lactose using nanofiltration. Computer Aided Process Engineering, 38, 151-156

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50098-7

Modeling and Simulation of LNG Fuel Vessel (LFV) Fuel Tank Reflecting CFD Result

Shin Hyuk Kim[a], Sungyoon Choi[b], Jay H. Lee[a]*

[a]*Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon, 34141, Republic of Korea*
[b]*Samsung Heavy Industries, 23, Pangyo-ro 227beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do, Republic of Korea*
jayhlee@kaist.ac.kr

Abstract

This study develops a mathematical model that predicts the dynamics of the temperature and pressure distributions, and phase changes in a Liquefied Natural Gas (LNG) fuel tank during the fuelling process. Using the Computational Fluid Dynamics (CFD) simulation results, it is shown that the LNG tank exhibits a vertical temperature distribution over the vapor phase during fuelling. Reflecting this insight from the CFD simulation, the developed model is a one-dimensional model for the vapor phase and a well-mixed model for the liquid phase. The developed model uses a moving boundary method to track the vapor-liquid interphase, reflecting the changing volume in a fixed discretised domain. Therefore, conservation equations are re-derived in order to account for the volume of the vapor phase that changes with the LNG injection. In addition, the model includes the interphase mass transfer, which reflects the phase change due to the pressure drop and heat transfer. The developed mathematical model was simulated using *gPROMS*^{TM} and the results were validated with the CFD results.

Keywords: Moving boundary, Bunkering, LNG, LNG fuel vessel, CFD

1. Introduction

With tightening of environmental regulations on ship fuel, LNG is expected to replace Heavy Fuel Oil (HFO) as the main fuel. However, the current state of vessel design for LNG bunkering lacks established standards and needs significant research efforts.

Most LNG storage tanks operate at ambient pressure. Consequently, as fuel is injected into the Liquid Fuel Vessel (LFV), the tank is depressurized to the ambient state. The pressure drop in the LFV tank causes the formation of Boil-Off Gas (BOG) from the remaining LNG. The LFV tanks are relatively small in volume compared to the tanks used in the past to store LNG. Also, the fuelling in this case is a much faster process due to the requirement of a short charge time. These factors lead to a sudden pressure drop and the BOG generation inside the tank. Temperature of the fuel entering the LFV tank during the bunkering process is lower than the temperature of the fuel remaining inside. The inflow of cold LNG can act to suppress the phase change in the tank. The system can be further stabilized by controlling the LNG feed for a gradual pressure drop.

It is not possible to predict and analyse the above phenomenon by modeling it as a flash drum, as often done in the past. Since LNG tanks in vessels are relatively large in size and cannot be assumed to be in equilibrium, a full dynamic model should be employed. It should be possible to use the computational fluid dynamics (CFD) to simulate the

situation where the liquid and vapor phases are clearly separated and external influences cause a phase change. The simulation results can be used to analyse the flow patterns inside the tank, which can be useful in developing a reduced-order mathematical model.

To this end, in this study, we propose a new LNG tank simulation model. The structure of the proposed model is based on CFD results: The liquid phase is assumed to be well mixed and the vapor phase is described by a one-dimensional model, to express vertical distributions of temperature and pressure. In order to reflect volume changes in the phases due to the fuel injection, it is necessary to implement a moving vapor-liquid interphase. The moving boundary method helps to track the changing domain in an efficient manner [Robalo, et al, 2005]. The multi-phase model enables the prediction of detailed heat and mass transfer based on the two-film resistance model. The designed mathematical model is simulated using the commercial software of *gPROMS*^{TM}, where thermal physical properties are calculated using the *Multiflash*^{TM} package.

2. CFD simulation

2.1. CFD Model description

This study involves an independent cylindrical tank of type C to be used in small to medium sized vessels (in Figure 1). The NG outlet is located in the dome at the top of the tank, and the inlet is designed to allow a horizontal injection from the bottom side of the dome. The volume is about 900 m^3 and the mesh number is about 700,000 in total.

LNG and NG coexist in the tank. Fuel is separated as liquid and vapor phase. We have simulated a situation where cooled LNG fuel (stored at lower pressure) is injected and the NG is discharged to ambient pressure. At this time, the incoming LNG and the leaving NG may constitute a multiphase fluid with a distinct phase boundary. Therefore, we simulated the flow and heat transfer of LNG and NG by using the Volume of Fluid (VOF) model. The *k-epsilon* model is used to predict the turbulence flow caused by the fast LNG injection and NG emissions. The NG outlet is modelled using a user defined function (UDF) to allow the fuel to be released by a choked valve, as in equation (1), where $\dot{m}$ is the mass flowrate in, C_d is the dimensionless discharge coefficient, A is the discharge hole cross-sectional area in m^2, γ is c_p/c_v, ρ is the gas density in kg m^{-3}, P is the gas pressure in Pa (Perry, et al, 2007). The CFD simulation is conducted to focus only on the convective phenomena without considering the phase change, due to the computational difficulty associated with introducing a phase change into the CFD simulation code. The detailed operating conditions for the simulation are shown in table 1.

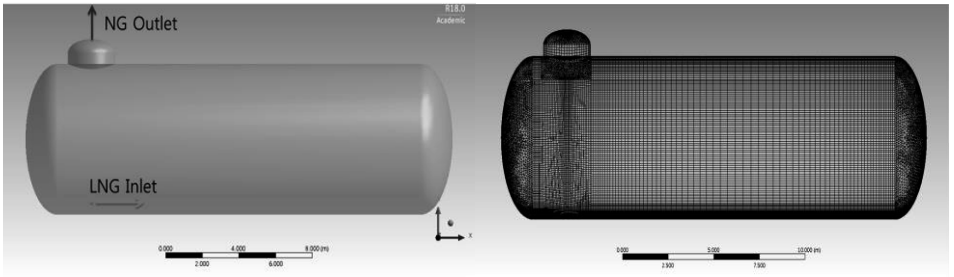


Figure 1. Geometry and mesh of a LNG tank

Table 1. Initial, boundary, and operating conditions for a simulation.

Initial condition		Gauge Pressure 3 bar Gas Temperature 183 K Liquid Temperature 131 K LNG Volume Fraction 0.1	Wall condition	Thermal	Temperature : 298.15 K U : 0.067 $Wm^{-2}K^{-1}$
Boundary condition	Inlet	Mass flow inlet T : 111.6 K LNG : 130 kg/s		Material	ρ: 1200 kg/m3 Cp : 1800 J/kg K k : 0.025 W/m K
	Outlet	Velocity inlet Using UDF (choked flow equation [Perry, et al, 2007]) $\dot{m} = C_d A \sqrt{\gamma \rho P \left(\frac{2}{\gamma+1}\right)^{\frac{\gamma+1}{\gamma-1}}}$ (1)	Materials	NG	CH_4 vapor (Single component)
				LNG	CH_4 liquid (Single component)

2.2. CFD simulation result

From the CFD analysis, we can expect that the vapor phase will be dominated by the vertical flow and the liquid phase will be well mixed. Figure 2 a) shows the vertical temperature distribution, due to the vertical pressure drop, external heat input, and heat exchange in the interphase. The liquid temperature appears to be constant throughout the liquid phase. Figure 2 b) suggests that the liquid phase exhibit turbulent flow due to the fuel injection and the tank structure. CFD results can provide physical properties which cannot be computed in a simple process unit model. For example, variables using dimensionless numbers that include slip velocity can be predicted only if two or more dimensions of the multiphase fluid are simulated in a same domain. Therefore, the heat transfer coefficients of the interphase and the heat flux of the wall required for the heat transfer simulation in a process unit model are calculated from the CFD results as shown in Eq.(2).

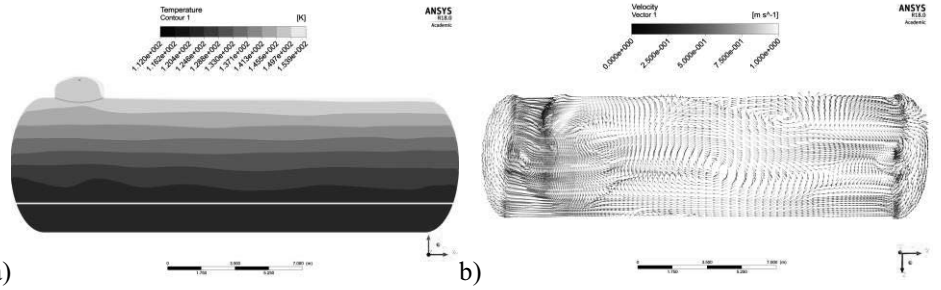
$$q_w \cong 11\ Wm^{-2},\ \ h_l \cong 99.5\ Wm^{-2}K^{-1},\ \ h_v \cong 21.4\ Wm^{-2}K^{-1} \quad (2)$$

, where q_w is heat flux from ambient, h_l and h_v are heat transfer coefficient of liquid and vapor at interphase.

3. LNG tank Model

3.1. Conceptual design for modeling (moving boundary method)

The model structure derived from the CFD simulation result is shown in Figure 3. Liquid phase is modeled as a well-mixed lumped parameter system. The vapor phase is modeled as a distributed system in one dimension to reflect the distinct vertical temperature



a) b)

Figure 2. CFD result of a LNG tank in a pseudo-steady state, a) temperature, b) velocity.

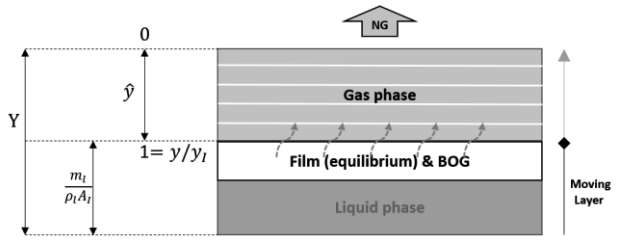


Figure 3. Conceptual design of a process unit model

distribution. It is assumed that there is a virtual film at the interphase between the vapor phase and the liquid phase. The film is always in vapor-liquid equilibrium, and the vapor phase and liquid phase exchange mass and heat with the film. The film follows the liquid level change tracked by the moving boundary method and the domain of the vapor phase is changed as the film is moving, as shown in Figure 3.

3.2. Model description

The moving boundary method uses a dimensionless domain scaled to reflect a moving boundary system. The boundary of a domain is usually calculated by a physical equation. Tracking the boundary point and reflecting it in the first principles equations using a dimensionless domain can make them appear as if a fixed domain is moving. The Eq.(3) represents the dimensionless domain, and Eq.(4) represents the induction formulas for reconstructing the first principles equations in the dimensionless domain.

$$\hat{y}(t) = \frac{y}{\mathrm{y}_I(t)} \tag{3}$$

$$\frac{\partial \phi}{\partial y} = \frac{1}{\mathrm{y}_I}\frac{\partial \phi}{\partial \hat{y}}, \qquad \frac{\partial^2 \phi}{\partial y^2} = \frac{1}{\mathrm{y}_I^2}\frac{\partial^2 \phi}{\partial \hat{y}^2}, \qquad \frac{\partial \phi}{\partial t} = \frac{\partial \phi}{\partial t} + \frac{\partial \phi}{\partial \hat{y}}\left(\frac{-\hat{y}}{\mathrm{y}_I}\frac{\partial \mathrm{y}_I}{\partial t}\right) \tag{4}$$

, where $\hat{y}$ is dimensionless domain, y is real domain, y_I is interface for tracking domain.

Equations (5) to (7) are conservation equations for the liquid phase. Equation (5) is the mass balance, Equation (7) is the energy balance, and Equation (6) calculates the interface location from the liquid level, y_I.

$$\frac{\partial m_l}{\partial t} = \dot{\mathrm{m}}_{in} - \mathrm{F} \cdot A_I \tag{5}$$

$$\mathrm{y}_I = \mathrm{Y} - \frac{m_l}{\rho_l A_I} \tag{6}$$

$$\frac{\partial m_l H_l}{\partial t} = \dot{\mathrm{m}}_{in} \cdot H_{l,in} - \mathrm{Q}_l \cdot A_I + q_w \cdot A_{l,w} \tag{7}$$

, where m_l is the mass of liquid in kg, F is the mass flux from liquid to vapour due to phase change in kg m^{-2} s^{-1}, A_I is the interphase area in m^2, Y is the tank height in m, ρ_l is the liquid density in kg m^{-3}, H_l is the liquid enthalpy in J kg^{-1}, Q_l is the heat flux from liquid to vapour in J m^{-2} s^{-1}, and $A_{l,w}$ is the heat transfer area of the liquid space wall in m^2.

Equations (8) to (10) are the newly derived conservation equations in the dimensionless domain, representing mass, momentum, and energy balances, respectively.

$$\frac{\partial \rho_v}{\partial \hat{y}}\left(\frac{-\hat{y}}{y_I}\cdot\frac{\partial y_I}{\partial t}\right) + \frac{\partial \rho_v}{\partial t} = -\frac{1}{y_I}\cdot\frac{\partial u_v \cdot \rho_v}{\partial \hat{y}} \tag{8}$$

$$\rho_v \frac{\partial u_v}{\partial \hat{y}}\left(\frac{-\hat{y}}{y_I}\cdot\frac{\partial y_I}{\partial t}\right) + \rho_v \frac{\partial u_v}{\partial t} = -\rho_v \cdot u_v \frac{1}{y_I}\cdot\frac{\partial u_v}{\partial \hat{y}} - \frac{1}{y_I}\cdot\frac{\partial P_v}{\partial \hat{y}} + \mu \frac{1}{y_I^2}\frac{\partial^2 u_v}{\partial \hat{y}^2} \tag{9}$$

$$\frac{\partial \rho_v E_v}{\partial \hat{y}}\left(\frac{-\hat{y}}{y_l}\cdot\frac{\partial y_l}{\partial t}\right)+\frac{\partial \rho_v E_v}{\partial t}=-\frac{1}{y_l}\cdot\frac{\partial u_v(\rho_v E_v+P_v)}{\partial \hat{y}}+\mathrm{k}_v\frac{1}{y_l^2}\frac{\partial^2 T_v}{\partial \hat{y}^2}+\mathrm{q}_w\cdot A_{v,w} \tag{10}$$

,where ρ_v is the vapor density in kg m^{-3}, u_v is the vapor velocity in m s^{-1}, E_v is the total vapor energy in J kg^{-1}, P_v is the vapor pressure in Pa, T_v is the vapor temperature in K, $A_{v,w}$ is the heat transfer area of the vapor space wall in m^2, and μ is the vapor viscosity in Pa.s.

Equations (11) and (12) express the heat flux the phase change by heat transfer and pressure change as a two-resistance model. As a general method of expressing heat transfer, the two-resistance model considers a separate heat transfer process with different heat transfer coefficients in the interphase.

$$\mathrm{Q}_l=h_l\cdot(T_{sat}(\mathrm{P}_v)-T_l)-F\cdot H_{l,s}=h_v\cdot(T_v-T_{sat}(\mathrm{P}_v))-F\cdot H_{v,s}=\mathrm{Q}_v \tag{11}$$

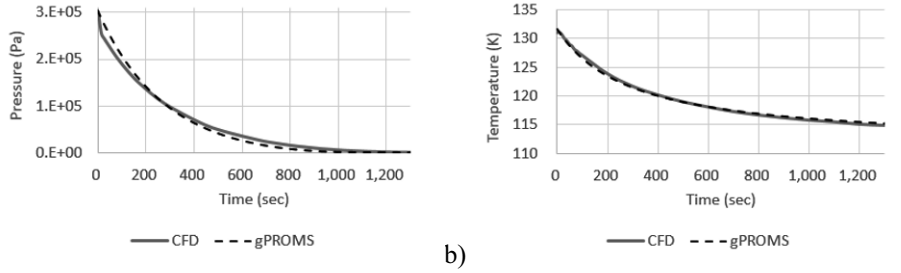
$$\begin{aligned} &IF\ F\geq 0\ (\text{'}evaporation\text{'}),\ H_{l,s}=H_l(T_l),\ H_{v,s}=H_v(T_{sat}(\mathrm{P}_v)) \\ &\text{Else}\quad (\text{'}condensation\text{'}),\ H_{l,s}=H_l(T_{sat}(\mathrm{P}_v)),\ H_{v,s}=H_v(T_v) \end{aligned} \tag{12}$$

, where, $T_{sat}(\mathrm{P}_v)$ is the saturated temperature for pressure P_v and T_l is the liquid temperature in K, $H_{l,s}$ and $H_{v,s}$ are liquid and vapor saturation enthalpies for phase change in J kg^{-1}, and Q_v heat flux for vapor from phase change in J m^{-2} s^{-1}.

3.3. Simulation result

To verify the results, we simulated the above one-dimensional mathematical model under the same condition as the CFD simulation. As shown in Figure 4 below, the pressure profile for the outlet and the temperature profile of the liquid phase were quite similar to those of the CFD simulation. The error range was within 0.03 Pa for the pressure, and 0.1 K for the liquid temperature. It took about 1200 seconds to reach the pseudo steady state after pressure and temperature drops. We confirmed that this transient behavior was consistent with the CFD result.

We simulated a control situation in which the choked valve and fuel injection were simultaneously considered. Unlike in the CFD simulation, the process unit model accounted for the phase change. The model reflects geometrical characteristics of a cylindrical tank: An increase in water level decreases the interfacial area representing the interphase. The height of the tank is 7 m, and we confirmed that the interface area showed a concave trend with a maximum at 3.5 m (as shown in Figure 5 a)). Figure 5 b) examines the mass flux by phase change. Evaporation occurs briefly during the initial phase of simulation. Immediately, the mass flux is reversed by condensation. After that, the evaporation starts again at about 756 seconds, and the inside of the tank starts to stabilize. Evaporation continues until the fuelling is completed. Temperatures of the vapor, liquid, and saturated temperature (film) in contact with the interphase can help explain the phenomenon more clearly (as shown in Figure 5 c)). The liquid temperature



a) b)

Figure 4. Comparison of CFD and gPROMS simulation results, a) outlet pressure, b) liquid temperature

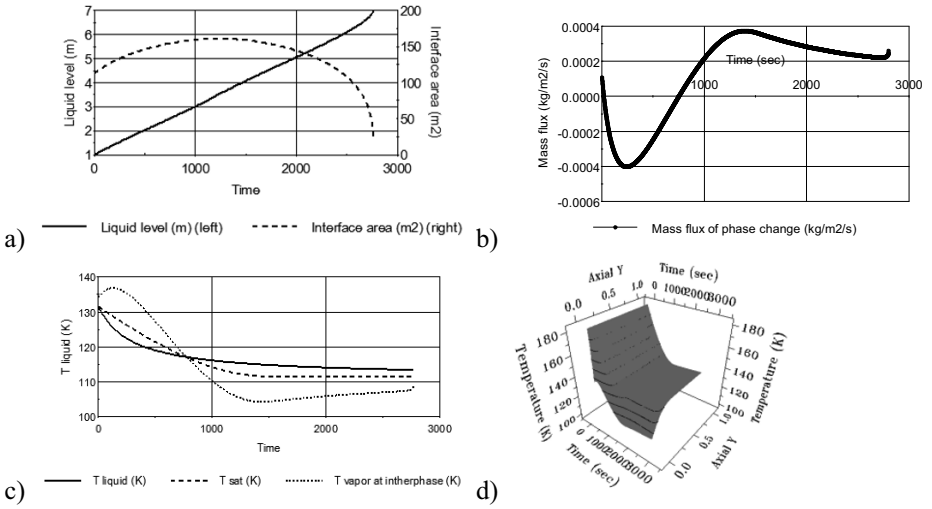


Figure 5. Dynamic simulation results of developed LNG tank model, a) interphase tracking, b) mass flux of phase change, c) interphase temperatures, d) vapor temperature profile in a tank

decreases depending on LNG fuelling, and the saturated temperature decreases depending on the pressure drop due to the NG emission. The vapor temperature is sensitive to the heat generated by the phase change. The condensation occurrs due to heat transfer by the liquid phase rather than energy released by pressure drop until 756 seconds. At about 240 seconds, the vapor temperature reaches a maximum. This is a consequence of the convective motion, which is predicted to occur at about the time it takes for the pressure drop to affect the vapor at interphase. After 756 seconds, the pressure drop causes the bubble point to be lower than the liquid phase, and evaporation occurs. The temperature change inside the tank reflecting these phenomena can be observed with time transient and vertical dimension (as shown in figure 5 d)).

4. Conclusions

The proposed LNG fuel tank model uses a dimensionless domain that can track the moving boundary based on total mass change of the liquid. All conservation equations have been rewritten in this domain. The proposed model includes a virtual film to enable the prediction of phase change and heat transfer between the phases. To demonstrate the use of the developed LNG tank model, a simple choked valve is implemented to predict the closed-loop dynamic response. The result shows the potential of using the model to develop an effective operation strategy. Our further study aims to improve the developed model to consider a multi-component fuel. The model also is expected to contribute to the establishment of a control strategy that optimizes the overall bunkering process with detailed control devices.

References

R. Robalo, C. A. Sereno, M.C. Coimbra and A. E. Rodrigues, (2005), The numerical solution of moving boundary problems using the moving finite element method, Computer Aided Chemical Engineering, 20, 79-84

R. H. Perry, and D. W. Green, (2007), Perry's Chemical Engineerings' Handbook, Eighth Edition, McGraw Hill Professional.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50099-9

Stochastic NMPC of Batch Processes Using Parameterized Control Policies

Eric Bradford[a*], Lars Imsland[a]

[a]*Department of Engineering Cybernetics, Norwegian University of Science and Technology, O. S. Bragstads plass 2D, 7034 Trondheim, Norway*
eric.bradford@ntnu.no

Abstract

Nonlinear model predictive control (NMPC) is an effective method for optimal operation of batch processes. Most dynamic models however contain significant uncertainties. It is therefore important to take these uncertainties into account in the formulation of the open-loop MPC problem to prevent infeasibilities or worse performance. An issue of such formulations is the disregard of feedback in the predictions, which leads to overly conservative control actions. The introduction of feedback through parametrized control policies is one way to solve this issue. In this work we compare the performance of affine feedback policies against more complex policies given by radial basis function networks. We incorporate these feedback policies into a polynomial chaos based stochastic NMPC algorithm to gauge their efficiency. The parameters of the feedback policies are either determined online by the NMPC algorithm or are pre-computed offline.

Keywords: Nonlinear model-based control, Polynomial chaos, Closed-loop policies, Chemical process control, Uncertain dynamic system

1. Introduction

Batch processes are used in many chemical sectors due to their inherent flexibility. These are operated at unsteady state and are often highly nonlinear, which motivates the application of nonlinear model predictive control (NMPC) (Nagy and Braatz, 2003). Many dynamic models however have limited accuracy due to various uncertainties. This can lead to constraint violations and worse control performance, which can be circumvented by incorporating these uncertainties in the NMPC algorithm (Mesbah, 2016). If we assume the uncertainties to be described by known probability density functions (pdfs), then the inclusion of the uncertainties in the NMPC algorithm leads to stochastic NMPC (SNMPC) formulations. In SNMPC constraints are addressed probabilistically, which allows for the systematic trade-off of constraint violation in probability with the conservativeness of the MPC solution (Mesbah, 2016). SNMPC methods include successive linearization (Cannon et al., 2009), sample-average NMPC (Bradford and Imsland, 2017), unscented sampling NMPC (Bradford and Imsland, 2017) and polynomial chaos expansion (PCE) NMPC (Fagiano and Khammash, 2012). A well-known problem of MPC under uncertainty is the fact that open-loop control actions are exceedingly conservative. To ensure reasonable predictions of the uncertainty, feedback needs to be considered. One way to achieve this is to optimize over parametrized feedback policies (Goulart et al., 2006). For linear stochastic MPC it is common to either evaluate a feedback matrix offline for pre-stabilization (Cannon et al., 2011) or by determining the parameters of the feedback control law online as decision variables (Hokayem et al., 2012). For SNMPC this problem is often ignored as in the PCE based SNMPC paper by

(Mesbah et al., 2014) or heuristics are applied as in Bradford and Imsland (2017). In this paper we focus on parametrized feedback policies for batch processes. For the SNMPC algorithm we use PCE due to its accuracy. We compare two different approaches to determine the parameters of the feedback policies: online optimization in the SNMPC algorithm or pre-computation offline. The first approach allows for revaluation of the feedback policy with knowledge of the new measurements, while the second allows for more complex feedback policies to be used. In addition, we consider three different parametrizations: affine time invariant, affine time varying and lastly radial basis function networks. In particular, radial basis function networks allow us to consider arbitrarily complex parametrizations. The various constellations were compared on a semi batch reactor case study via closed-loop simulations.

2. SNMPC with feedback policies

The aim of SNMPC is to control a discrete-time stochastic nonlinear system:

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{\theta}) \tag{1}$$

where k represents the discrete time, $\mathbf{x} \in \mathbb{R}^{n_\mathbf{x}}$ are the system states, $\mathbf{u} \in \mathbb{R}^{n_\mathbf{u}}$ are the control inputs, $\boldsymbol{\theta} \in \mathbb{R}^{n_\boldsymbol{\theta}}$ are uncertain parameters and $\mathbf{f}(\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{\theta})$ denotes the system dynamics. We assume $\boldsymbol{\theta}$ to be independent Gaussian distributed random variables. The components have known mean μ_i and standard deviation σ_i. We can then formulate a general SNMPC problem with parametric feedback policies as follows:

$$\underset{\boldsymbol{\mathcal{u}}_N}{\text{minimize}}\ \mathbb{E}\big(J(N, \mathbf{x}(\mathrm{n}), \boldsymbol{\mathcal{u}}_N, \boldsymbol{\theta})\big)$$

subject to:

$$\begin{aligned}
&\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{\theta}) && \forall k \in \{0, \dots, N-1\} \\
&\mathbb{P}\big(g_j^{(k)}(\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{\theta}) \le 0\big) \ge 1 - p_j^{(k)} && \forall k \in \{1, \dots, N\} \times \left\{1, \dots, n_g^{(k)}\right\} \\
&\mathbf{u}_k = \boldsymbol{\varphi}\big(\boldsymbol{\nu}_k + \boldsymbol{\kappa}(\mathbf{x}_k, \boldsymbol{\gamma}_k)\big) && \forall k \in \{0, \dots, N-1\} \\
&\mathbf{x}_0 = \mathbf{x}(\mathrm{n})
\end{aligned} \tag{2}$$

where the objective function is the expectation of $J(N, \mathbf{x}(n), \boldsymbol{\mathcal{u}}_N, \boldsymbol{\theta})$, $g_j^{(k)}(\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{\theta})$ are individual chance constraints, $p_j^{(k)} \in (0,1) \subset \mathbb{R}$ is the probability of constraint violation, $\boldsymbol{\mathcal{u}}_N = \{\boldsymbol{\nu}_0, \dots, \boldsymbol{\nu}_{N-1}, \boldsymbol{\gamma}_0, \dots, \boldsymbol{\gamma}_{N-1}\}$ is the set of decision variables, $\boldsymbol{\varphi}(\cdot)$ is a saturation function, $\boldsymbol{\kappa}(\cdot)$ is the feedback policy employed, $\boldsymbol{\nu}_k$ are the feed-forward control inputs, $\mathbf{x}(n)$ is the known initial state at time step n and N is the time horizon.

The saturation function in this report was defined to be individual sigmoid functions for each control input to introduce implicit constraints on $\mathbf{u}_k$: $\boldsymbol{\varphi}\big(\boldsymbol{\nu}_k + \boldsymbol{\kappa}(\mathbf{x}_k, \boldsymbol{\gamma}_k)\big) = \left[\frac{u_{max,1} - u_{min,1}}{1+\exp(-[\boldsymbol{\nu}_k + \boldsymbol{\kappa}(\mathbf{x}_k, \boldsymbol{\gamma}_k)]_1)} + u_{min,1}, \dots, \frac{u_{max,n_\mathbf{u}} - u_{min,n_\mathbf{u}}}{1+\exp(-[\boldsymbol{\nu}_k + \boldsymbol{\kappa}(\mathbf{x}_k, \boldsymbol{\gamma}_k)]_{n_\mathbf{u}})} + u_{min,n_\mathbf{u}}\right]^T$, i.e. $u_{min,i} \le u_{k,i} \le u_{max,i}\ \forall i \in \{1, \dots, n_\mathbf{u}\}$.

Three different parametrizations of feedback control laws were considered:

1) Affine time invariant (ATI)

$$\boldsymbol{\kappa}(\mathbf{x}_k, \boldsymbol{\gamma}) = \mathbf{K}\mathbf{x} \tag{3}$$

where $\mathbf{K} = \begin{bmatrix} \gamma_1 & \cdots & \gamma_{n_\mathbf{x}} \\ \vdots & \ddots & \vdots \\ \gamma_{n_\mathbf{u}} & \cdots & \gamma_{n_\mathbf{u} \times n_\mathbf{x}} \end{bmatrix}$, Number of parameters: $n_\mathbf{u} \times n_\mathbf{x}$

2) Affine time varying (ATV)

$$\boldsymbol{\kappa}(\mathbf{x}_k, \boldsymbol{\gamma}_k) = \mathbf{K}_k \mathbf{x} \tag{4}$$

where $\mathbf{K}_k = \begin{bmatrix} \gamma_{k,1} & \cdots & \gamma_{k,n_\mathbf{x}} \\ \vdots & \ddots & \vdots \\ \gamma_{k,n_\mathbf{u}} & \cdots & \gamma_{k,n_\mathbf{u} \times n_\mathbf{x}} \end{bmatrix}$, Number of parameters: $n_\mathbf{u} \times n_\mathbf{x} \times (n_k - 1)$

3) Radial basis function network time invariant (RBFN-n_{RBF})

$$\kappa_i(\mathbf{x}_k, \boldsymbol{\gamma}) = \sum_{j=1}^{n_{RBF}} \omega_j \phi_j(\mathbf{x}_k) \tag{5}$$

$$\phi_j(\mathbf{x}_k) = \exp\left(-0.5\left(\mathbf{x}_k - \boldsymbol{\mu}_j\right)^T \mathrm{diag}\left(\boldsymbol{\lambda}_j\right)\left(\mathbf{x}_k - \boldsymbol{\mu}_j\right)\right) \tag{6}$$

where n_{RBF} is the number of radial basis functions with parameters $\left[\omega_j, \boldsymbol{\mu}_j, \boldsymbol{\lambda}_j\right]^T = \boldsymbol{\gamma}_{(j-1)(1+2n_\mathbf{x})+1:j(1+2n_\mathbf{x})}$, Number of parameters: $n_\mathbf{u} \times n_{RBF}(1 + 2n_\mathbf{x})$

The open-loop case is given by $\boldsymbol{\kappa}(\cdot) = \mathbf{0}$. The nonlinear radial basis function network parametrization was taken from Deisenroth and Rasmussen (2011) and allows for arbitrarily complex feedback policies by adjusting the number of radial basis functions.

3. PCE based SNMPC

The problem defined in Eq.(2) is intractable, since it involves probability and expectations on nonlinear functions of $\boldsymbol{\theta}$. PCE can be used to solve this efficiently. Assume we are given a nonlinear transformation of independent standard normal variables $\zeta(\mathbf{z})$ with finite second order moments. According to PCE theory $\zeta(\mathbf{z})$ can be approximated by a truncated orthogonal polynomial basis (Owen et al., 2017):

$$\zeta(\mathbf{z}) \approx \sum_{0 \le |\boldsymbol{\alpha}| \le p} a_{\boldsymbol{\alpha}} He_{\boldsymbol{\alpha}}(\mathbf{z}) = \mathbf{a}^T \boldsymbol{\Phi}(\mathbf{z}) \tag{7}$$

where $\mathbf{z} \in \mathbb{R}^{n_\boldsymbol{\theta}}$ is a vector of standard normal variables with components $z_i \sim \mathcal{N}(0,1)$, $a_{\boldsymbol{\alpha}}$ are unknown expansion coefficients, $|\boldsymbol{\alpha}| = \sum_{j=1}^{n_\boldsymbol{\theta}} \alpha_j$, p is the order of truncation, $\mathbf{a}$ is a vector of coefficients $a_{\boldsymbol{\alpha}}$ and $\boldsymbol{\Phi}(\mathbf{z})$ contains $L = \frac{(n_\boldsymbol{\theta}+p)!}{n_\boldsymbol{\theta}! p!}$ polynomial elements $He_{\boldsymbol{\alpha}}(\mathbf{z})$ of the truncated expansion. $He_{\boldsymbol{\alpha}}(\mathbf{z})$ are known multivariate Hermite polynomials defined by a tensor product of univariate Hermite polynomials: $He_{\boldsymbol{\alpha}}(\mathbf{z}) = He_{\alpha_1}^{(1)}(z_1) \times \ldots \times He_{\alpha_{n_\boldsymbol{\theta}}}^{(n_\boldsymbol{\theta})}\left(z_{n_\boldsymbol{\theta}}\right)$, where $He_j^{(i)}$ is the univariate Hermite polynomial of order j in terms of the i[th] component of $\mathbf{z}$. Univariate Hermite polynomial can be found in look-up tables.

We now need to find values for the coefficients $a_{\boldsymbol{\alpha}}$ to obtain a good approximation to $\zeta(\mathbf{z})$, which is done by using observations of $\zeta(\mathbf{z})$ at different values of $\mathbf{z}$. In this report we used a Sobol quasi random design to determine the sample points and transform it to follow a Gaussian distribution by using the inverse cumulative normal distribution function. We will refer to this design as $\mathcal{Z} = \left\{\mathbf{z}^{(0)}, \ldots, \mathbf{z}^{(n_s-1)}\right\}$, which contains n_s sampling points. We then obtain a response for each sample, which gives us $\mathbf{Y} = \left[\zeta\left(\mathbf{z}^{(0)}\right), \ldots, \zeta\left(\mathbf{z}^{(n_s-1)}\right)\right]^T$. The coefficients can be found by least-squares estimation:

$$\hat{\mathbf{a}} = (\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \mathbf{Y} \tag{8}$$

where $\Psi_{ij} = He_{\boldsymbol{\alpha}_j}\left(\mathbf{z}^{(i)}\right)\ i = 0, \dots, n_s - 1, j = 1, \dots, L$ is a data matrix containing the polynomial terms in Eq.(7) and $He_{\boldsymbol{\alpha}_j}$ is the j[th] Hermite polynomial in the series expansion.

The variance and mean of $\zeta(\mathbf{z})$ are approximately given by (Mesbah et al., 2014):

$$\mathbb{E}\left(\zeta(\mathbf{z})\right) \approx \hat{a}_1, \quad \mathrm{Var}\left(\zeta(\mathbf{z})\right) \approx \sum_{j=2}^{L} \hat{a}_j^2\, \mathbb{E}\left(He_{\boldsymbol{\alpha}_j}^{\ 2}(\mathbf{z})\right) \tag{9}$$

Using PCEs the problem in Eq.(2) can be simplified as follows:

$$\underset{\boldsymbol{u}_N}{\text{minimize}}\ \mathbb{E}\left(\xi_J\right)$$

subject to:

$$\mathbf{x}_{k+1}^{(i)} = \mathbf{f}\left(\mathbf{x}_k^{(i)}, \mathbf{u}_k, \boldsymbol{\theta}^{(i)}\right) \quad \forall (k,i)\ \in\ \{0, \dots, N-1\}\ \times \{0, \dots, n_s - 1\} \tag{10}$$

$$\sqrt{\mathrm{Var}\left(\zeta_{g_{jk}}\right)\left(1 - p_j^{(k)}\right)/p_j^{(k)}} + \mathbb{E}\left(\zeta_{g_{jk}}\right) \le 0 \quad \forall k\ \in\ \{1, \dots, N\} \times \left\{1, \dots, n_g^{(k)}\right\}$$

$$\hat{\mathbf{a}}_J = (\boldsymbol{\Psi}^T\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^T\mathbf{Y}_J, \qquad \hat{\mathbf{a}}_{g_{jk}} = (\boldsymbol{\Psi}^T\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^T\mathbf{Y}_{g_{jk}}$$

$$\mathbb{E}\left(\xi_J\right) = \hat{a}_{J,1}, \qquad \mathbb{E}\left(\zeta_{g_{jk}}\right) = \hat{a}_{g_{jk},1}$$

$$\mathrm{Var}\left(\zeta_{g_{jk}}\right) = \sum_{m=2}^{L} \hat{a}_{g_{jk},m}^2\, \mathbb{E}\left(He_{\boldsymbol{\alpha}_m}^{\ 2}(\mathbf{z})\right)$$

$$\mathbf{u}_k^{(i)} = \boldsymbol{\varphi}\left(\boldsymbol{\nu}_k + \boldsymbol{\kappa}\left(\mathbf{x}_k^{(i)}, \boldsymbol{\gamma}_k\right)\right) \qquad \forall k\ \in\ \{0, \dots, N-1\}$$

$$\mathbf{x}_0^{(i)} = \mathbf{x}(\mathrm{n}) \qquad \forall i\ \in\ \{0, \dots, n_s - 1\}$$

where $\theta_l^{(i)} = \sigma_l z_l^{(i)} + \mu_l$, $\mathbf{x}^{(i)}$ represents the state vector of scenario i with uncertain parameter $\boldsymbol{\theta}^{(i)}$, $\mathbf{Y}_J = \left[J\left(N, \mathbf{x}(\mathrm{n}), \boldsymbol{u}_N, \boldsymbol{\theta}^{(0)}\right), \dots, J\left(N, \mathbf{x}(\mathrm{n}), \boldsymbol{u}_N, \boldsymbol{\theta}^{(n_s-1)}\right)\right]^T$ *and* $\mathbf{Y}_{g_{jk}} = \left[g_j^{(k)}\left(\mathbf{x}_k^{(i)}, \mathbf{u}_k^{(i)}, \boldsymbol{\theta}^{(0)}\right), \dots, g_j^{(k)}\left(\mathbf{x}_k^{(i)}, \mathbf{u}_k^{(i)}, \boldsymbol{\theta}^{(n_s-1)}\right)\right]$ are data matrices to determine the required coefficients**.** It should be pointed out that while $\mathbf{Y}_J$ and $\mathbf{Y}_{g_{jk}}$ change each iteration of the optimization algorithm**,** the sample design $\mathcal{Z}$ remains constant, such that $(\boldsymbol{\Psi}^T\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^T$ and the $\mathbb{E}\left(He_{\boldsymbol{\alpha}_m}^{\ 2}(\mathbf{z})\right)$ can be calculated offline. The probability constraints were approximated robustly using Chebyshev inequality, for more information see (Mesbah et al., 2014).

4. Case study

The case study involves an isothermal semi batch reactor with a second-order exothermic reaction taking place; A + B → C. The exact differential equation for the case study can be found in Lucia and Paulen (2014). The control input is the feed rate u and the state vector is $\mathbf{x} = [C_A, C_B, C_C, V]^T$, where C_i are the concentrations of species i and V is the reactor volume. The flow rate u can be adjusted between 0 and 0.4. The initial conditions were set to $[C_{A0}, C_{B0}, C_{C0}, V_0] = [3,0,0,0.7]$ and the temperature is kept at 70°C. The semi batch reactor is controlled by a shrinking horizon implementation of Eq.(10). The objective of the SNMPC problem was set to maximize the average amount of C at the end of the batch, while keeping the adiabatic temperature below 85°C to prevent uncontrollable behaviour and the concentration of B at the final time below $0.5\mathrm{mol dm}^{-3}$.

The probability of constraint violation for both was set to 0.05. The resulting OCP was solved employing direct collocation with 10 control intervals and a final time of 1h utilizing CasADi in Python (Andersson et al., 2012).

To gauge the performance of the various feedback policy methods, we ran 100 simulations based on a Sobol design for 5 different constellations with 4 computed online: Open-loop without feedback, ATI, ATV and RBFN with a single radial basis function referred to as RBFN-1. A more complex RBFN was computed offline with 10 radial basis functions referred to as RBFN-10. In Fig.1 a box-plot is shown of the attained amount of product C and hence the higher the amount of C, the better the closed-loop performance. In Tab.1 the corresponding average computational times are shown with the number of decision variables of the optimization problem. We can see that the open-loop approach has the lowest median and the fastest computational time, which is expected since it is the most conservative and has the least number of decision variables. ATI and RBFN-1 perform slightly better than open-loop implementation but, due their simplicity, they are worse than the others. In particular, RBFN-1 is a poor choice, since it leads to the same performance as ATI with larger computational times. ATV on the other hand has the highest median production rate, yet it requires also the highest computational time due to its large number of decision variables. RBFN-10 is only marginally worse, however has a computational time that is significantly lower than the other feedback policy approaches. It does however take twice as long as the open-loop approach.

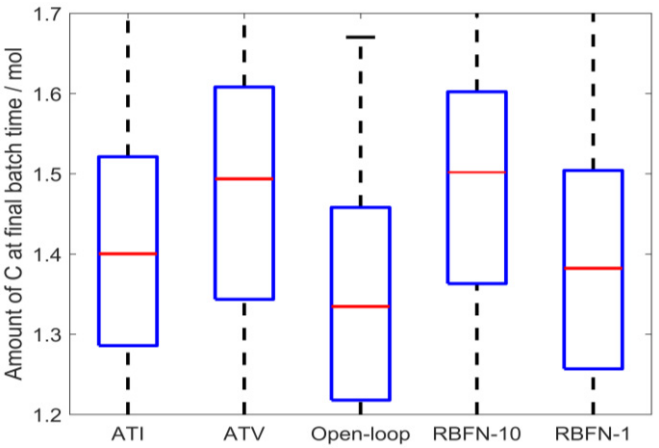


Figure 1 Box-plot of amount of product for 100 Sobol design points using different feedback policy approaches

Table 1 Comparison between number of decision variables evaluated online and average evaluation time of a single NMPC control step of the feedback policies

Feedback policy	No. decision variables	Evaluation time (s)
ATI	14	5.9
ATV	46	11.6
Open-loop	10	0.87
RBFN-10	10	1.87
RBFN-1	19	9.58

5. Conclusions

In conclusion, we investigated different approaches to introduce feedback into a SNMPC algorithm. In general, feedback always improves the performance of the otherwise open-loop algorithm. For the feedback policies evaluated online it could be shown that ATI and RBFN-1 were outperformed by the more complex ATV, however ATV carries the largest computational time due to the number of decision variables it introduces. On the other hand, the pre-computed RBFN-10 managed to perform only marginally worse while

being significantly faster in terms of computational time. We therefore conclude that the use of pre-computed complex feedback control parametrizations is an interesting approach to obtain good closed-loop performance without introducing a too heavy computational load.

6. Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska- Curie grant agreement No 675215.

References

J. Andersson, J. Åkesson, and M. Diehl, 2012, CasADi: A symbolic package for automatic differentiation and optimal control, Recent advances in algorithmic differentiation, Lecture Notes in Computational Science and Engineering, Volume 87, p. 297-307.

E. Bradford, and L. Imsland, 2017, Expectation constrained stochastic nonlinear model predictive control of a batch bioreactor, Computer Aided Chemical Engineering, Volume 40, p. 1621-1626.

E. Bradford, and L. Imsland, 2017, Stochastic Nonlinear Model Predictive Control with State Estimation by Incorporation of the Unscented Kalman Filter, arXiv preprint arXiv:1709.01201.

M. Cannon, B. Kouvaritakis, S. V. Rakovic, and Q. Cheng, 2011, Stochastic tubes in model predictive control with probabilistic constraints, IEEE Transactions on Automatic Control, Volume 56, Issue 1, p. 194-200.

M. Cannon, D. Ng, and B. Kouvaritakis, 2009, Successive linearization NMPC for a class of stochastic nonlinear systems, Nonlinear Model Predictive Control, Lecture Notes in Control and Information Sciences, Volume 384, p. 249-262.

M. Deisenroth, and C. E. Rasmussen, 2011, PILCO: A model-based and data-efficient approach to policy search, *in* Proceedings of the 28th International Conference on machine learning (ICML-11), p. 465-472.

L. Fagiano, and M. Khammash, 2012, Nonlinear stochastic model predictive control via regularized polynomial chaos expansions, *in* Proceedings of the 51st IEEE Conference on Decision and Control (CDC), p. 142-147.

P. J. Goulart, E. C. Kerrigan, and J. M. Maciejowski, 2006, Optimization over state feedback policies for robust control with constraints, Automatica, Volume 42, Issue 4, p. 523-533.

P. Hokayem, E. Cinquemani, D. Chatterjee, F. Ramponi, and J. Lygeros, 2012, Stochastic receding horizon control with output feedback and bounded controls, Automatica, Volume 48, Issue 1, p. 77-88.

S. Lucia, and R. Paulen, 2014, Robust nonlinear model predictive control with reduction of uncertainty via robust optimal experiment design, IFAC Proceedings Volumes, Volume 47, Issue 3, p. 1904-1909.

A. Mesbah, 2016, Stochastic model predictive control: An overview and perspectives for future research, IEEE Control Systems, Volume 36, Issue 6, p. 30-44.

A. Mesbah, S. Streif, R. Findeisen, and R. D. Braatz, 2014, Stochastic nonlinear model predictive control with probabilistic constraints, *in* Proceedings of the IEEE 2014 American Control Conference, p. 2413-2419.

Z. K. Nagy, and R. D. Braatz, 2003, Robust nonlinear model predictive control of batch processes, AIChE Journal, Volume 49, Issue 7, p. 1776-1786.

N. Owen, P. Challenor, P. Menon, and S. Bennani, 2017, Comparison of surrogate-based uncertainty quantification methods for computationally expensive simulators, SIAM/ASA Journal on Uncertainty Quantification, Volume 5, Issue 1, p. 403-435.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50100-2

Forecast of persistent disturbances using *k*-nearest neighbour methods

Francesco Borghesan[a*], Moncef Chioua[b], Nina F. Thornhill[a]

[a]*Imperial College London, South Kensington Campus, London SW7 2AZ, UK*
[b]*ABB Corporate Research Center, Wallstadter Str. 59, Ladenburg D-68526, Germany.*
** f.borghesan@imperial.ac.uk*

Abstract

This paper focuses on the prediction of persistent disturbances based on their past measurements using two versions of the *k*-nearest neighbours method: an unweighted and a weighted version. Results of tests on data from a refinery show that the two methods can predict the future trend of a disturbance. They also show that the weighted version is more robust against the choice of the number of nearest neighbours used. The method opens up the possibility of model-free feedforward control without the constraint of causality based on the whole history of a measurement.

Keywords: Plantwide disturbance, time series prediction, nearest neighbours, process control

1. Introduction

In process plants, a disturbance is an undesired, transitory deviation of a process variable from its desired set point. Such disturbances can affect the quality of the product, or they can cause a malfunction of the site machinery and accelerate its wear (Yuan & Qin, 2014). Further, a disturbance originating from one unit can propagate to other units because of material, energy or information and control interconnections.

Control systems have a role in propagating a disturbance. A controller compensates the effect of the disturbance on the controlled variable by adjusting the manipulated variable. However, the manipulated variable can be an external disturbance for another plant unit. In this way, the plantwide disturbance continues to propagate. Various disturbances, each characterised by its own time trend, can affect a plant over time. Therefore, there should be an optimal control action for each disturbance. For this, it would be useful for the control algorithm to be able to detect that a disturbance is occurring and to predict the future evolution of the disturbance.

This paper focuses on disturbance prediction. Being able to predict the future evolution of a disturbance can improve the performance of a controller. In process systems, feedforward control is used routinely to compensate for measured disturbances when they are detected (Liptak 2005; Camacho and Bordons 2008), but the standard implementation of feedforward control does not include a prediction of the future evolution of the disturbance. Koerber & King (2013), and Mehra *et al.* (1997) have described how previewing a disturbance improved the control action with applications in wind turbines and automotive industries. However, modelling and predicting a disturbance remains an open question. Literature provides many methods such as ARIMA models and Artificial Neural Networks (De Gooijer and Hyndman 2006). The methods are classified into first principle models, statistical methods and data-driven methods. Difficulties in choosing the method arise mainly when the disturbances are

caused by non-linear effects such as limit cycles, or when they arise because of random events. Cairano et al. (2014) improved the control of the battery management of a hybrid electric vehicle by predicting the disturbance, the driver driving style, with a data-driven Markov chain. Klenske et al. (2016) improved the control of an electrically actuated telescope proposing a model based on Gaussian processes that learns to fit nearly-periodic disturbances. However, these methods adapt their models based on the last measured samples and do not exploit the whole history of the measured disturbance. This paper presents two versions of a *k*-nearest neighbours method. Advantages of the *k*-nearest neighbours method include the ability to capture the repeating pattern typical of persistent disturbances, and its easy implementation. The result give an advance over previous uses of *k*-nearest neighbours for time series prediction which have typically predicted just one sample a few steps ahead. This paper shows it is possible to do the more challenging task of predicting the future evolution of the time series.
Section 2 describes the methods, and Section 3 presents the results of applying the method to real data coming from a refinery. The paper ends with a discussion and conclusions.

2. Method description

2.1. Introduction

The *k*-nearest neighbours method supposes that the current time series segment will evolve in future like a past time series segment (not necessarily a recent one) evolved previously (Kantz and Schreiber 2004). The task therefore is to identify past segments of the time series which are similar to the present one according to a certain norm.
The algorithm considers a measurement signal made of N samples $\boldsymbol{y}_M(N) = [y(1)\ y(2)\ \dots\ y(N)]$, termed 'memory'. The memory size increases as more samples $y(\cdot)$ are recorded. The segment made of the last m samples of $\boldsymbol{y}_M(N)$ is $\boldsymbol{y}_E(N) = [y(N-m+1)\ \ y(N-m+2)\ \dots y(N)]$. This segment represents the current evolving disturbance pattern and is called $\boldsymbol{y}_E(N)$ 'evolution', while m is known as the 'embedding dimension' (Kantz and Schreiber 2004). The algorithm searches within the memory $\boldsymbol{y}_M(N)$ for the k time series intervals of length m that are most similar to $\boldsymbol{y}_E(N)$. The similarity is measured in this paper with the Euclidean distance (Eq 1). These k time series intervals $\boldsymbol{y}_j,\ j = 1, \dots, k$ are the *k*-nearest neighbours and are those k time series intervals within $\boldsymbol{y}_M(N)$ with smallest Euclidean distances:

$$d\left(\boldsymbol{y}_E(N), \boldsymbol{y}_j\right) = \sqrt{\sum_{i=1}^{m}\left(y_E(i) - y_j(i)\right)^2} \tag{1}$$

As depicted in Figure 1, each nearest neighbour $\boldsymbol{y}_j$ is followed by a time series of length h, where h is the desired prediction horizon. These time series of length *h* are called the 'prediction contributions' $\boldsymbol{y}_{P\,j}, j = 1, \dots, k$. They show how their corresponding nearest neighbours evolved over an interval of length h. Therefore, they are the basis for building the prediction vector, as will be described in Section 2.2. Since all prediction contributions must stay within the memory $\boldsymbol{y}_M(N)$, there is a constraint for the starting points r_j of the nearest neighbours $\boldsymbol{y}_j$

$$\boldsymbol{y}_j = \left[y(r_j)\ \ y(r_j+1) \dots\ \ y(r_j+m)\right], \quad j = 1, \dots, k, \qquad r_j \in [1; N-h-m+1] \tag{2}$$

Consequently $\boldsymbol{y}_{P\,j}$

$$\boldsymbol{y}_{P\,j} = \left[y(r_j+m+1)\ \ \dots\ \ y(r_j+m+h)\right] \qquad\qquad j = 1, \dots, k \tag{3}$$

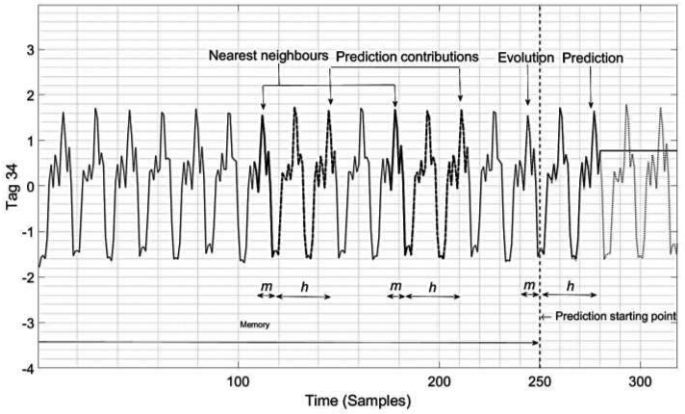


Figure 1 An example of a persistent disturbance and representation of terminology

2.2. The two versions of the algorithm

Unweighted and weighted versions of the algorithm differentiate themselves at this step. For the unweighted version, the prediction is an average of the prediction contributions. The prediction is a vector $\hat{\boldsymbol{y}}(N) = [\hat{y}(N+1) \dots \hat{y}(N+h)]$ where:

$$\hat{y}(N+i) = \frac{1}{k}\sum_{j=1}^{k} y_{P\,j}(i),\ i = 1, \dots, h \tag{4}$$

In the weighted version of the algorithm, the nearest neighbours contribute to the prediction proportionally to their distance from the evolution. First a weight w_j for each nearest neighbours is calculated. The weight $w_j = 1$ if the nearest neighbour $\boldsymbol{y}_j$ has the smallest distance with $\boldsymbol{y}_E(N)$. Conversely, $w_j = 0$ if the corresponding nearest neighbour $\boldsymbol{y}_j$ is the most distant from $\boldsymbol{y}_E(N)$.

$$w_j = \frac{\max_{\ell=1\dots k}(d(\boldsymbol{y}_E(N),\boldsymbol{y}_\ell)) - d(\boldsymbol{y}_E(N),\boldsymbol{y}_j)}{\max_{\ell=1\dots k}(d(\boldsymbol{y}_E(N),\boldsymbol{y}_\ell)) - \min_{\ell=1\dots k}(d(\boldsymbol{y}_E(N),\boldsymbol{y}_\ell))} \qquad k \geq 2 \qquad j = 1, \dots, k \tag{5}$$

$$w_1 = 1 \qquad k = 1$$

The prediction vector $\hat{\boldsymbol{y}}(N) = [\hat{y}(N+1) \dots \hat{y}(N+h)]$ is a weighted average of the prediction contributions, where:

$$\hat{y}(N+i) = \frac{1}{\sum_{j=1}^{k} w_j}\sum_{j=1}^{k} w_j\, y_{P\,j}(i) \qquad i = 1, \dots, h \tag{6}$$

2.3. Measuring the accuracy of the prediction for varying values of k and m

The tuning parameters of the methods are m and k. Varying them causes the average prediction distance to change. The prediction distance E_{av} is the Euclidean distance between the prediction and the corresponding segment of the actual future time series (not known in advance to the algorithm) divided by the chosen prediction horizon h so that the value does not depend on the length of the prediction horizon.

$$E_{av}(N) = \frac{1}{h}d(\hat{\boldsymbol{y}}(N), [y(N+1) \dots y(N+h)]) \tag{7}$$

The smaller $E_{av}(N)$ is, the more accurate the prediction is.

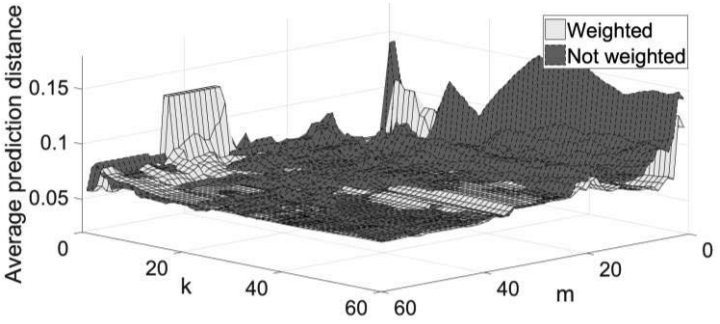


Figure 2 Effect of parameters *k* and *m* on prediction of Tag 33

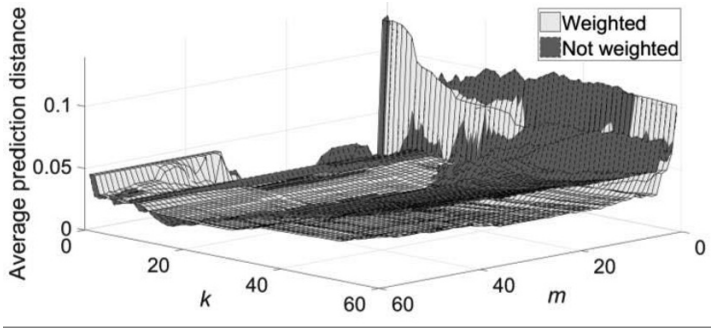


Figure 3 Effect of parameters *k* and *m* on prediction of Tag 34

3. Results

3.1. Description of the data

Figure 2 and Figure 3 show the performance of the prediction methods on two time trends called Tag 33 and Tag 34 which came from a SE Asian refinery (Thornhill 2005) and are composed of 512 samples. The signals have been mean-centred and scaled to unit standard deviation. The prediction starting point in these examples is $N = 450$ (the 450th over 512 samples) to give the algorithm enough samples to look for the nearest neighbours. The prediction horizon has been chosen to be 30 because this requires the algorithm to predict two cycles of the oscillation, constituting a stringent test.

3.2. Effect of k *and* m *on* E_{av}

Figure 2 and Figure 3 show the values of E_{av} as a function of $k \in [1,60]$ and $m \in [1,60]$ for the two methods. The figures show that for very small values of m, the prediction error can be high. This is because small embedding dimension might not identify the shape of the disturbance. For example, two points might have similar values in the ascending or descending phase of an oscillation but the prediction contributions $\boldsymbol{y}_P$ could be very different, thereby worsening the prediction. After a threshold, increasing m does not influence E_{av} for a periodic (or nearly periodic) signal like Tag 34. Instead m influences E_{av} for Tag 33, which is made of spikes different from each other. Each pattern lasts between 8 and 17 samples and indeed there is a decrease in values of E_{av} for $8 \leq m \leq 17$. A good choice is $m = \frac{1}{f_0}$, where f_0 is the main frequency of the disturbance. Therefore $m = 17$ in this case.

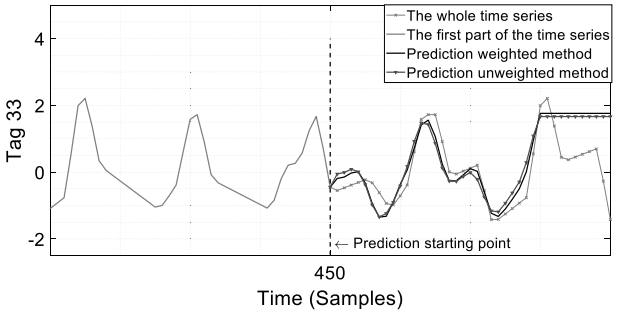


Figure 4 Prediction of Tag 33 with $N = 450, h = 30, k = 26, m = 27$ using the weighted ($E_{av} = 0.0679$) and unweighted ($E_{av} = 0.0855$) k-nearest neighbor methods.

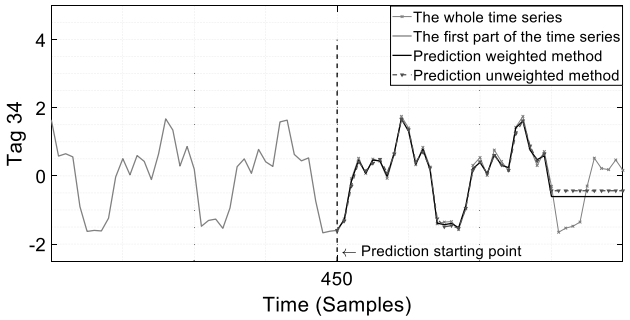


Figure 5 Prediction of Tag 34 with $N = 450, h = 30, k = 26, m = 27$ using the weighted ($E_{av} = 0.0178$) and unweighted ($E_{av} = 0.0202$) k-nearest neighbor methods.

The prediction error E_{av} rises for high values of k. Increasing k too much results in time series intervals being considered as nearest neighbours even though they are not very similar to $\boldsymbol{y}_E(N)$ and do not develop in the same way. Figures 2 and 3 show the weighted method reduces the impact of a large value of k compared to the unweighted nearest neighbour method and gives smaller values for E_{av}.

For small values of k, the surface of the plots is more irregular, showing E_{av} is sensitive to small changes of parameters k and m. The reason is that even persistent disturbances have some randomness. By averaging over more prediction contributions, the resulting prediction is closer to the true value. Therefore, the best value of k depends on the number of nearest neighbours sufficiently close to $\boldsymbol{y}_E$ available in the memory. A satisfactory choice is $k = \frac{N}{m}$ and therefore $k = 26$ in this case. A question for future work is to make the selection of k adaptable, because the memory vector $\boldsymbol{y}_M(N)$ grows as time goes by.

3.3. Quality of the prediction

Predictions for the Tag 33 and Tag 34 using the suggested parameters $m = 17, k = 26$ with both methods are shown in Figure 4 and Figure 5. The predictions are based on the history up to sample number 450. The figures also show what actually happened to the time trends. The accuracy is good for both methods, although slightly better for the weighted method. The figure captions give the E_{av} values. As mentioned earlier, Tag 34 is inherently more predictable than Tag 33 because Tag 33 is characterized by peaks with some variability from one to the next. The lower E_{av} values for Tag 34 reflect this qualitative observation.

4. Discussion

Figures 4 and 5 suggest the value of having a prediction of the disturbance. The ability to predict is of great value for a control system. It provides a mechanism for non-causal feedforward control, i.e. when the disturbance is predicted with good accuracy before it happens. Moreover, it is not necessary for a similar disturbance to have happened recently. Provided one or more similar deviations have been recorded within the history then the *k*-nearest neighbours method will give prediction based on those similar deviations. This is in contrast to time series modelling methods which use recent history with a forgetting mechanism to discard historical data. The next step in the work is to build predictions based on *k*-nearest neighbours into a control algorithm.

5. Conclusions

The *k*-nearest method has been proposed to predict the future values of process disturbances. The method has been able to provide good predictions of two signals coming from a refinery. The impacts of the parameters k and m have been described, and default values have been provided. The paper gave a comparison between the unweighted and weighted *k*-nearest method and concluded the weighted method is more robust and that it provides better performance.

Acknowledgement

Financial support is gratefully acknowledged from the Marie Skłodowska Curie Horizon 2020 EID-ITN project “PROcess NeTwork Optimization for efficient and sustainable operation of Europe’s process industries taking machinery condition and process performance into account – PRONTO", Grant agreement No 675215.

References

S.D. Cairano, D. Bernardini, A. Bemporad, and I.V. Kolmanovsky. 2014. Stochastic MPC with learning for driver-predictive vehicle control and its application to HEV energy management. *IEEE Transactions on Control Systems Technology,* 22 (3), 1018–1031.

E. Camacho and C. Bordons. 2008. *Model Predictive Control*. Springer Verlag.

J.G.De Gooijer and R.J. Hyndman. 2006. 25 Years of time series forecasting. *International Journal of Forecasting,* 22 (3), 443–473.

H. Kantz and T. Schreiber. 2004. *Nonlinear Time Series Analysis*. Second Edition. Cambridge: Cambridge University Press.

E.D. Klenske, M.N. Zeilinger, B. Schölkopf, and P. Hennig. 2016. Gaussian process-based predictive control for periodic error correction. *IEEE Transactions on Control Systems Technology*, 24 (1), 110–121

A. Koerber and R. King. 2013. Combined feedback-feedforward control of wind turbines using state-constrained model predictive control. *IEEE Transactions on Control Systems Technology* 21 (4), 1117–1128.

B.G. Liptak. 2005. Feedback and feedforward control. In *Instrument Engineers’ Handbook, Fourth Edition, Volume Two : Process Control and Optimization*, Fourth Ed. CRC Press.

R.K. Mehra, J.N. Amin, K.J. Hedrick, C. Osorio, and S. Gopalasamy. 1997. Active suspension using preview information and model predictive control. *Proceedings of the 1997 IEEE International Conference on Control Applications,* 860–865, doi:10.1109/CCA.1997.627769.

N.F. Thornhill. 2005. Finding the source of nonlinearity in a process with plant-wide oscillation. *IEEE Transactions on Control Systems Technology,* 13 (3), 434–443.

T. Yuan and S.J. Qin. 2014. Root cause diagnosis of plant-wide oscillations using Granger causality, Journal of Process Control, 24 (2), 450-459.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50101-4

Input Designs to Obtain Uncorrelated Outputs in MIMO System Identification

Kurt E. Häggblom

Faculty of Science and Engineering, Åbo Akademi University, Biskopsgatan 8, Åbo (Turku) FI-20500, Finland
khaggblo@abo.fi

Abstract

A problem in open-loop identification of multiple-input multiple-output (MIMO) systems is that standard designs using uncorrelated inputs tend to produce correlated outputs. If the system is ill-conditioned, this correlation may be very strong. Such a correlation reduces identifiability and may result in a model with different controllability properties than the true system. If the model is used for control system design, the result may be poor closed-loop performance and even instability.

The author has recently presented an experiment design method for MIMO system identification that solves the main problem by producing uncorrelated outputs for the model used in the design. The solution to the design problem is not unique, however. For a given type of input, the same output covariance can be obtained by different input designs. In this paper, various design options are studied. One of them is to minimize the input amplitudes. The considered signal types are pseudo random binary sequences (PRBS) and multi-sinusoidal signals with optimized phase shifts. Two systems having different directionality properties and number of inputs/outputs are used for illustration.

Keywords: System identification, Multivariable systems, Experiment design, Identification for control, Convex optimization.

1. Introduction

In the control of industrial MIMO systems, integral controllability (IC) is usually a required property. This means that the closed-loop system must be stable when controllers with integral action are used, and remain stable if the controllers are detuned (Garcia and Morari, 1985). Because IC is determined by the static gain matrix of the system, an experiment design procedure based on an estimate of the gain matrix was proposed by Koung and MacGregor (1992). In this design, the static gain directions of the system are explicitly excited. Although dynamics are not considered in this experiment design, it tends to work well also for identification of dynamics when the dynamics are aligned with the gain directions (Häggblom and Böling, 1998, 2013). Overviews of various developments of this procedure is given by Häggblom (2014).

Dynamics are explicitly taken into account in experiment design methods proposed by Rivera et al. (2007), Darby and Nikolaou (2014), and Kumar and Narasimhan (2016). These methods are rather complicated. Although not used as an optimization criterion, the designs yield almost uncorrelated outputs. A design procedure that directly addresses the output distribution has been proposed by Häggblom (2017). The aim is to produce uncorrelated outputs to maximize identifiability. The problem can be formulated as a convex optimization with linear matrix inequalities (LMIs) as constraints,

which is easy to solve. The user can select the type of input perturbation, e.g. random binary signal (RBS), pseudo random binary signal (PRBS), or multi-sinusoidal signal.

The solution to the optimization problem is not unique, however. For a given type of input, the same output covariance can be obtained by different input designs. In this paper, various design options are studied. One of them is to minimize the input amplitudes, which is equivalent to minimizing the crest factor. Two systems having different directionality properties and number of inputs/outputs are used for illustration.

2. Design to obtain uncorrelated outputs

It is assumed that an initial model is available for the dynamic system to be identified. The experiment design is based on a discrete-time state-space model,

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) \\ y(k) &= Cx(k) + Du(k) \end{aligned}, \tag{1}$$

where k denotes sample number; $u(k)$, $x(k)$, and $y(k)$ are vectors of inputs, states, and outputs, respectively, at sample instant k; A, B, C, and D are matrices of appropriate dimensions in the state-space description. An initial model of any standard form can be converted to this form. Time delays can be modeled exactly by inclusion of additional states (Häggblom, 2017).

2.1. Covariance model

Equation (1) yields the covariance model (Häggblom, 2017)

$$\begin{aligned} P_x(k+1) &= AP_x(k)A^{\mathrm{T}} + AP_{xu}(k)B^{\mathrm{T}} + BP_{xu}^{\mathrm{T}}(k)A^{\mathrm{T}} + BP_u(k)B^{\mathrm{T}} \\ P_y(k) &= CP_x(k)C^{\mathrm{T}} + CP_{xu}(k)D^{\mathrm{T}} + DP_{xu}^{\mathrm{T}}(k)C^{\mathrm{T}} + DP_u(k)D^{\mathrm{T}} \end{aligned}, \tag{2}$$

where $P_u(k)$, $P_x(k)$ and $P_y(k)$ are (sample) covariance matrices and $P_{xu}(k)$ is the (sample) cross-covariance matrix between states and inputs, all at sample instant k.

Here, $P_{xu}(k)$ is problematic because it makes Eq. (2) underdetermined. In Häggblom (2017) it is shown that $P_{xu}(k)$ depends on the auto-covariance properties of the input. Including this dependency in the model would introduce undesired complications.

To overcome this problem, $P_{xu}(k) = 0$ is used in Eq. (2), and a scaled version of $P_y(k)$ is introduced to compensate for the exclusion of $P_{xu}(k)$. With the approximation $P_x(k+1) \approx P_x(k)$, and omission of the sample index k, Eq. (2) is thus replaced by

$$\begin{aligned} P_x &= AP_xA^{\mathrm{T}} + BP_uB^{\mathrm{T}} \\ S_yP_yS_y^{\mathrm{T}} &= CP_xC^{\mathrm{T}} + DP_uD^{\mathrm{T}} \end{aligned}, \tag{3}$$

where the scaling matrix S_y is determined as part of the optimization.

2.2. Optimization

The outputs are uncorrelated if P_y is diagonal. Thus, it is desired to have an input covariance P_u that produces a diagonal P_y. According to Hadamard's inequality (Horn and Johnson, 2013), $\det P_y \le \Pi P_{y,ii}$, where $P_{y,ii}$, $i = 1,\ldots,n$, are the diagonal elements

of P_y, with equality holding if and only if P_y is diagonal. This means that uncorrelated outputs can be obtained by maximizing $\det P_y$ subject to upper bounds on the diagonal elements of $P_{y,ii}$. The upper bounds $\text{var}\, y_i$ are the maximum desired output variances.

If there are no (active) constraints on input variances, the same solution is obtained by maximizing the minimum eigenvalue of P_y, i.e., $\max \underline{\lambda}(P_y)$ (Häggblom, 2017). The optimization can be formulated as $\max \lambda$ subject to $P_y - \lambda I \succcurlyeq 0$, which requires $P_y - \lambda I$ to be a positive semidefinite matrix.

The optimization is done subject to Eq. (3), which is linear with respect to the decision variables P_u, P_x and P_y. The scaling matrix S_y is updated iteratively as

$$S_y \leftarrow S_y P_y^{1/2} R_y^{-1/2}, \tag{4}$$

where R_y is the sample covariance matrix of the outputs resulting from a simulation of the system using the input determined from the current P_u in the iteration. The simulation can be done using Eq. (1), or a model from which Eq. (1) was determined. With $S_y = I$ as starting value, this makes R_y converge to P_y (if all matrices have full rank).

Thus, the input covariance matrix P_u is obtained as the solution to

$$\max_{P_u, P_x, P_y} \lambda \text{ s.t. } P_y - \lambda I \succcurlyeq 0, \ \text{var}\, y_i \geq P_{y,ii}, \text{ and Eq. (3)}, \tag{5}$$

with S_y updated iteratively by Eq. (4).

2.3. The input signal

The solution to Eq. (5) does not yield the type of input signal $u(k)$, only its covariance P_u. The user is free to choose any type of (persistently exciting) input signal normally used in system identification. It can, e.g., be a random binary signal (RBS), a pseudo random binary sequence (PRBS), or some version of multi-sinusoidal signals.

Let $\xi(k)$ denote the k^{th} sample of such a design signal. The sample covariance matrix of $\xi(k)$, P_ξ, should ideally be close to a diagonal matrix. This can usually be achieved by using time-shifted copies of a base sequence as components $\xi_i(k)$, $i = 1, \ldots, n$. The idea is to determine a constant matrix T such that the transformation

$$u(k) = T\xi(k) \tag{6}$$

yields an input $u(k)$ with the covariance P_u. From Eq. (6) it follows that

$$P_u = T P_\xi T^{\mathrm{T}}, \tag{7}$$

where P_u and P_ξ are known. Singular value decompositions (SVD) of these positive definite matrices yield

$$P_u = V_u \Lambda_u V_u^{\mathrm{T}}, \quad P_\xi = V_\xi \Lambda_\xi V_\xi^{\mathrm{T}}, \tag{8}$$

where V_u and V_ξ are orthogonal matrices and Λ_u and Λ_ξ are diagonal matrices of

singular values. From Eqs (7) and (8) it follows that

$$T = V_u \Lambda_u^{1/2} Q^{\mathrm{T}} \Lambda_\xi^{-1/2} V_\xi^{\mathrm{T}}, \tag{9}$$

where Q is an arbitrary orthogonal matrix of appropriate dimension. Different choices of Q yield different input and output distributions with the covariances P_u and P_y.

In the following examples, the effect of some choices of Q that follow naturally from the problem description are investigated. These choices are $Q = I$, $Q = V_\xi^{\mathrm{T}}$, $Q = V_u$, and $Q = V_\xi^{\mathrm{T}} V_u$. In addition, a solution minimizing $\| T \|_\infty$ is considered. Since $\| T \|_\infty$ is the largest row sum of absolute values $| T_{ij} |$, this minimizes the peak value of $u(k)$.

3. Examples

The optimizations in the following examples are done in the MATLAB environment using the YALMIP toolbox (Löfberg, 2004) and Mosek as SDP solver. For minimization of $\| T \|_\infty$, the MATLAB command fmincon is used with Eq. (7) as nonlinear constraint. The design signals are generated by the idinput command in the MATLAB System Identification Toolbox (Ljung, 2017).

3.1. A moderately ill-conditioned 2×2 system

The first example is a distillation column model presented by Wood and Berry (1973). It has been used as example in numerous studies, also in Häggblom (2017), where the model is given. The condition number of the gain matrix is 7.45. The transfer function model is sampled with sample time $T_s = 0.2$ and converted into the form of Eq. (1).

The design signal in the first study of the Wood and Berry column is a PRBS with minimum switching time $T_{\mathrm{sw}} = 1$ and sequence length $N = 127$, yielding the period length $P = T_{\mathrm{sw}} N = 127.$ The two signals are made uncorrelated by time-shifting one of them by half a period. Two periods of opposite signs are used. The PRBS is effective in the frequency range $2\pi / P \ldots 0.8\pi / T_{\mathrm{sw}}$.

In the second study of this model, a multi-sinusoidal design signal is applied. To cover the same frequency range as in the PRBS design, using the same period P, the frequencies $\omega_k = 2\pi k / P$, $k = 1, \ldots, n_{\mathrm{f}}$, with $n_{\mathrm{f}} \approx 0.4N \approx 51$, are used. To choose phase shifts, $100 n_{\mathrm{f}}$ randomly distributed phase shifts are generated, and a combination resulting in the smallest overall amplitude is selected by the idinput command. This is an alternative to the minimization of the crest factor by an intractable optimization (see Rivera et al., 2007). As in the PRBS design, the two signals are made uncorrelated by time-shifting one of them by half a period. Two periods of opposite signs are used.

In all design cases, the output specification $R_y = P_y = I$ was reached. The results of the designs are summarized in Table 1 and Table 2. Minimization of $\| T \|_\infty$, denoted $\min T_\infty$ in the tables, results in the smallest input peak values $| u_i |_{\max}$, as expected. The choice $Q = V_\xi^{\mathrm{T}}$, which is easy to apply, comes close to this result. The multi-sinusoidal designs result in considerably higher input peak values that the PRBS designs. Figures 1 and 2 show scatter plots of the output distributions for the choice $Q = V_\xi^{\mathrm{T}}$.

Table 1. PRBS designs (2x2).

Q	T	var u	$\|u\|_{\max}$	$\|y\|_{\max}$
I	$\begin{bmatrix} 0.48 & -0.57 \\ 0.32 & -0.11 \end{bmatrix}$	$\begin{bmatrix} 0.56 \\ 0.11 \end{bmatrix}$	$\begin{bmatrix} 1.05 \\ 0.43 \end{bmatrix}$	$\begin{bmatrix} 2.43 \\ 3.20 \end{bmatrix}$
V_ξ^{T}	$\begin{bmatrix} 0.71 & 0.08 \\ 0.34 & -0.16 \end{bmatrix}$	$\begin{bmatrix} 0.51 \\ 0.15 \end{bmatrix}$	$\begin{bmatrix} 0.79 \\ 0.51 \end{bmatrix}$	$\begin{bmatrix} 2.82 \\ 2.56 \end{bmatrix}$
V_u	$\begin{bmatrix} 0.27 & -0.71 \\ 0.01 & -0.43 \end{bmatrix}$	$\begin{bmatrix} 0.58 \\ 0.18 \end{bmatrix}$	$\begin{bmatrix} 0.97 \\ 0.44 \end{bmatrix}$	$\begin{bmatrix} 2.91 \\ 2.89 \end{bmatrix}$
$V_\xi^{\mathrm{T}} V_u$	$\begin{bmatrix} 0.64 & 0.28 \\ 0.28 & 0.32 \end{bmatrix}$	$\begin{bmatrix} 0.49 \\ 0.18 \end{bmatrix}$	$\begin{bmatrix} 0.92 \\ 0.60 \end{bmatrix}$	$\begin{bmatrix} 2.78 \\ 2.96 \end{bmatrix}$
$\min T_\infty$	$\begin{bmatrix} 0.60 & 0.00 \\ 0.34 & 0.20 \end{bmatrix}$	$\begin{bmatrix} 0.47 \\ 0.16 \end{bmatrix}$	$\begin{bmatrix} 0.69 \\ 0.55 \end{bmatrix}$	$\begin{bmatrix} 2.37 \\ 2.66 \end{bmatrix}$

Table 2. Multi-sinusoidal designs (2×2).

Q	T	var u	$\|u\|_{\max}$	$\|y\|_{\max}$
I	$\begin{bmatrix} 0.97 & -1.18 \\ 0.74 & -0.31 \end{bmatrix}$	$\begin{bmatrix} 0.52 \\ 0.14 \end{bmatrix}$	$\begin{bmatrix} 1.82 \\ 0.92 \end{bmatrix}$	$\begin{bmatrix} 2.41 \\ 2.98 \end{bmatrix}$
V_ξ^{T}	$\begin{bmatrix} 1.62 & 0.14 \\ 0.69 & -0.29 \end{bmatrix}$	$\begin{bmatrix} 0.57 \\ 0.12 \end{bmatrix}$	$\begin{bmatrix} 1.71 \\ 0.85 \end{bmatrix}$	$\begin{bmatrix} 2.42 \\ 2.68 \end{bmatrix}$
V_u	$\begin{bmatrix} 0.60 & -1.38 \\ -0.08 & -0.85 \end{bmatrix}$	$\begin{bmatrix} 0.50 \\ 0.16 \end{bmatrix}$	$\begin{bmatrix} 1.75 \\ 0.91 \end{bmatrix}$	$\begin{bmatrix} 2.27 \\ 2.58 \end{bmatrix}$
$V_\xi^{\mathrm{T}} V_u$	$\begin{bmatrix} 1.52 & 0.56 \\ 0.55 & 0.52 \end{bmatrix}$	$\begin{bmatrix} 0.56 \\ 0.12 \end{bmatrix}$	$\begin{bmatrix} 1.86 \\ 0.92 \end{bmatrix}$	$\begin{bmatrix} 3.20 \\ 2.67 \end{bmatrix}$
$\min T_\infty$	$\begin{bmatrix} 1.59 & 0.00 \\ 0.72 & 0.30 \end{bmatrix}$	$\begin{bmatrix} 0.55 \\ 0.13 \end{bmatrix}$	$\begin{bmatrix} 1.59 \\ 0.91 \end{bmatrix}$	$\begin{bmatrix} 2.93 \\ 2.66 \end{bmatrix}$

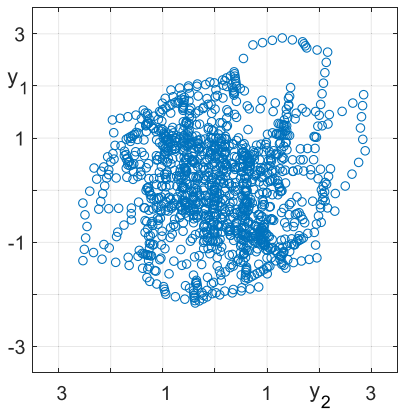


Figure 1. Output distribution of PRBS design for $Q = V_\xi^{\mathrm{T}}$.

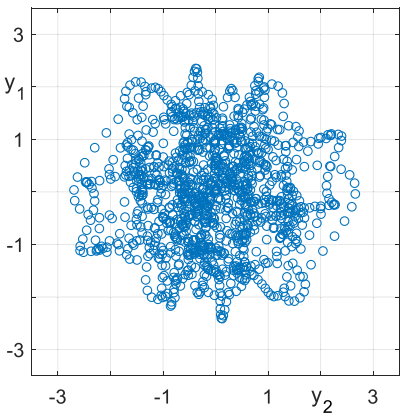


Figure 2. Output distribution of multi-sinusoidal design for $Q = V_\xi^{\mathrm{T}}$.

3.2. An ill-conditioned 3×3 system

The second example is a distillation column model presented by Vasnani (1994). In Häggblom (2016), a rescaled version of the model was introduced to make it slightly more ill-conditioned, resulting in the condition number 30. Here, this model is used. The transfer function model of size 3×3 is sampled with sample time $T_s = 1$ and converted into the form of Eq. (1).

A minimum switching time $T_{\mathrm{sw}} = 5$ and sequence length $N = 255$ are used in the PRBS design. The signals are made uncorrelated by time-shifting consecutive components by one third of the period. Two periods of opposite signs are used. The multi-sinusoidal design is done similarly as above with $n_{\mathrm{f}} = 102$.

The output specification $R_y = P_y = I$ was reached in all design cases. The results for $Q = V_\xi^{\mathrm{T}}$ and $\min \| T \|_\infty$ are summarized in Tables 3 and 4 with T omitted. The conclusions are similar to those for the 2×2 case.

Table 3. PRBS designs (3×3).

Q	$\operatorname{var} u$	$\lvert u\rvert_{\max}$	$\lvert y\rvert_{\max}$
V_ξ^{T}	$\begin{bmatrix}0.2957\\0.0032\\0.0016\end{bmatrix}$	$\begin{bmatrix}0.5500\\0.0799\\0.0584\end{bmatrix}$	$\begin{bmatrix}2.6655\\2.7126\\2.6002\end{bmatrix}$
$\min T_\infty$	$\begin{bmatrix}0.2957\\0.0032\\0.0017\end{bmatrix}$	$\begin{bmatrix}0.5437\\0.0825\\0.0595\end{bmatrix}$	$\begin{bmatrix}2.6543\\2.6938\\2.5708\end{bmatrix}$

Table 4. Multi-sinusoidal designs (3×3).

Q	$\operatorname{var} u$	$\lvert u\rvert_{\max}$	$\lvert y\rvert_{\max}$
V_ξ^{T}	$\begin{bmatrix}0.2268\\0.0025\\0.0013\end{bmatrix}$	$\begin{bmatrix}1.1930\\0.1424\\0.1034\end{bmatrix}$	$\begin{bmatrix}2.8995\\2.9893\\3.0450\end{bmatrix}$
$\min T_\infty$	$\begin{bmatrix}0.2267\\0.0025\\0.0013\end{bmatrix}$	$\begin{bmatrix}1.1910\\0.1391\\0.1122\end{bmatrix}$	$\begin{bmatrix}2.8931\\2.9325\\2.9903\end{bmatrix}$

4. Conclusions

Input designs to obtain uncorrelated outputs in MIMO system identification was studied. Inputs with peak values close to the optimized minimum can be obtained by the simple choice $Q = V_\xi^{\mathrm{T}}$ in the calculation of the transformation matrix T used for generating inputs from a design signal. The same output covariance is obtained with PRBS inputs having significantly smaller input peak values than multi-sinusoidal inputs.

References

M.L. Darby and M. Nikolaou, 2014, Identification test design for multivariable model-based control: An industrial perspective, *Control Eng. Pract.,* 22, 165–180.

S.E. Garcia and M. Morari, 1985, Internal model control. 2. Design procedure for multivariable systems, Ind. Eng. Chem. Process Des. Dev., 24, 2, 472–484.

K.E. Häggblom, 2014, On experiment design for identification of ill-conditioned systems, IFAC Proceedings Volumes, 47, 3, 1428–1433.

K.E. Häggblom, 2016, Evaluation of experiment designs for MIMO identification by cross-validation, IFAC-PapersOnLine, 49, 7, 308–313.

K.E. Häggblom, 2017, A new optimization-based approach to experiment design for dynamic MIMO identification, IFAC-PapersOnLine, 50, 1, 7321–7326.

K.E. Häggblom and J.M. Böling, 1998, Multimodel identification for control of an ill-conditioned distillation column, J. Process Control, 8, 3, 209–218.

K.E. Häggblom and J.M. Böling, 2013, Experimental evaluation of input designs for multiple-input multiple-output open-loop system identification, Proc. IASTED Int. Conf. Control and Applications, 157–163, Honolulu, HI, USA.

R. A. Horn and C. R. Johnson, 2013, Matrix Analysis, Cambridge UP, New York, NY, USA.

C.-W. Koung and J.F. MacGregor, 1992, Design of identification experiments for robust control. A geometric approach for bivariate processes, Ind. Eng. Chem. Res., 32, 8, 1658–1666.

A. Kumar and S. Narasimhan, 2016, Optimal input signal design for identification of interactive and ill-conditioned systems, Ind. Eng. Chem. Res., 55, 14, 4000–4010.

L. Ljung, 2017, System Identification Toolbox User's Guide, The MathWorks, Inc., Natick, MA.

J. Löfberg, 2004, YALMIP: A Toolbox for Modeling and Optimization in MATLAB, Proc. IEEE Int. Symp. Computer Aided Control Systems Design (CACSD), 284–289, Taipei, Taiwan.

D.E. Rivera, H. Lee, H.D. Mittelmann, and M.W. Braun, 2007, High-purity distillation: using plant-friendly multisine signals to identify a strongly interactive process, IEEE Control Syst. Mag., 27, 5, 72–89.

V.U. Vasnani, 1994, Towards relay feedback auto-tuning of multi-loop systems, PhD Thesis, National University of Singapore.

R.K. Wood and M.W. Berry, 1973, Terminal composition control of a binary distillation column, Chem. Eng. Sci., 28, 9, 1707–1717.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50102-6

Control Strategy Scheme for the Prehydrolysis Kraft Process

Antton Lahnalammi*, Herbert Sixta, Sirkka-Liisa Jämsä-Jounela

Aalto University, Kemistintie 1, Espoo 02150, Finland
antton.lahnalammi@aalto.fi

Abstract

This paper presents a control strategy scheme for the prehydrolysis stage in the prehydrolysis kraft process. The key element of the proposed control strategy is a soft-sensor, which estimates the rate of the unobservable reactions based on a physico-chemical dynamic model of the process. The rate of hydrolysis depends on the wood temperature and hydrogen ion concentration, which can be calculated by estimating the reaction heat due to deacetylation of wood and subsequent wood component degradation reactions. The estimated hydrogen ion concentration and reaction enthalpies are combined with dynamic temperature profile and kinetic reaction models for estimation of the main quality parameters for dissolving pulp, namely the lignin content and the content and degree of polymerization of the alpha-cellulose fraction. The developed model is tested based on parameters approximated from the literature and the feasibility of the model for testing the suggested control strategy scheme is discussed.

Keywords: prehydrolysis kraft process, dissolving pulp, dynamic model, soft-sensor, model-based control

1. Introduction

The prehydrolysis kraft (PHK) process is an extension of kraft cooking, where a separate prehydrolysis stage is conducted prior to the cooking stage, in order to selectively dissolve the major part of the wood hemicelluloses. The resulting almost pure (> 90 %) α-cellulose, denoted as dissolving pulp, can be used for the production of cellulose derivatives and regenerated cellulose fibers. The dissolving pulp demand is growing steadily, largely due to increased textile demand, which can no longer be met by increasing cotton production (Hämmerle, 2011).

The PHK-process is mainly conducted in batch digesters, where the prehydrolysis stage is accomplished by heating the wood chips with direct steam or pressurized water to 160 – 180 °C for up to two hours. The acetyl groups attached to the hemicelluloses are released as acetic acid, which gradually creates an acidic environment in the aqueous phase. The generated proton activity catalyzes the hydrolysis of the wood biopolymers into smaller, eventually water-soluble fractions. The addition of mineral acid is avoided due to adverse effects related to the formation of sulfate or chloride containing salts. The removal of wood hemicelluloses is accompanied by unavoidable partial hydrolysis of cellulose, which causes enhanced depolymerization and increased yield loss during the subsequent cooking stage due to alkaline hydrolysis and secondary peeling reactions, respectively. The degree of polymerization, determined as intrinsic viscosity, is the main dissolving pulp control parameter. Furthermore, the formation of delignification resistant condensed lignin from the released highly reactive acid soluble lignin hinders the following cooking stage and limits the recovery of sugars from the prehydrolyzate. On-line monitoring of

these aforementioned reactions is missing. Thus, the process has to be optimized off-line and only minor adjustments are made based on long run deviations.

Reaction kinetic models allow insight into the unobservable reactions. However, the extremely inhomogeneous and partially indeterminate morphology of hardwood on cell wall and macromolecular levels prevents comprehensive modelling of the reaction rates and the interlinked rate limiting mass transfer phenomena. Consequently, generalizations based on experimental results are unavoidable. Simmonds et al. (1955) showed early on that the removal of xylan during prehydrolysis consists of an initial fast phase, followed by a slower secondary phase. Based on this result, Conner (1984) developed a prehydrolysis model where xylan is divided into fast and slow reacting fractions, both of which are degraded following pseudo first order kinetics. Following the same logic, Borrega et al. (2011a, 2011b) developed kinetic models for the degradation of the main carbohydrates and the removal and condensation of lignin in hardwood. More recently Ahmad et al. (2016) developed an extensive physico-chemical model, which estimates the acid equilibrium and mass transfer based on the Nerst-Planck equation and the major wood component degradation pathways and molecular weight distribution of xylo-oligosaccharides by second order kinetics. Reaction kinetics of the following kraft cooking stage are outside the scope of this study, and have been reviewed elsewhere (Nieminen and Sixta, 2012).

Control of the PHK-process is limited by shared resources and recirculation of residual heat and chemicals, which impose scheduling limits to displacement batch cooking facilities. As a result, the temperature profile is the only available control variable during the prehydrolysis stage. In industrial settings, the extent of hemicellulose removal is traditionally estimated by the P-factor (Brasch and Free, 1965; Lin, 1979) which is analogously to Vroom's (1957) H-factor used in kraft cooking: time integral computed over Arrhenius type equation using experimentally determined average activation energy and frequency factor for overall delignification. The plant control system's optimization tool solves optimal temperature profile leading to the predetermined P-factor, subject to limitations imposed by available time and resources.

The earliest applications of model based control strategies for batch digesters were single corrections to target H-factor based on effective alkali conduction measurements after initial fast reactions, which were correlated with simplified kinetic models (Kerr, 1976; Noreus et al., 1974). Later, Paulonis and Krishnagopalan (1991) introduced an adaptive inferential control system, which utilized additional UV and refraction index liquor analysis methods for real time estimation of kappa number based on material balances. Lee and Datta (1994) further developed the approach into a state observer based model predictive controller for batch kraft cooking. More recently, de Vaal and Sandrock (2007) introduced a model based control strategy for acid sulphite pulping in a batch digester, which utilizes a fundamental model (Kilian and De Vaal, 2000) for estimating the degree of polymerization. However, no model based control strategy for the PHK-process has been published to date.

The temperature profile remains the only available signal from the digester during the prehydrolysis stage, which suggests the utilization of energy balance for real time estimation of the reaction rates. Courchene et al., (2005) successfully determined heat of reaction of kraft cooking from overall energy balance of a batch laboratory digester. This paper presents a fundamental dynamic model for the prehydrolysis kraft process, for the estimation of the energy and component balances. Subsequently a control strategy design based on the model is outlined and the validity of the model is tested with approximated parameters.

2. Process description

Batch pulp digester consists of a steel reactor vessel and external heat exchanger. According to academic inclination, the model system described here utilizes water phase prehydrolysis, where liquid fraction is continuously circulated through the heat exchanger, while the solid wood fraction remains immobilized inside the reactor. The simplest case is assumed here, where the circulation liquid flows in from the top and out from the bottom of the reactor (Figure 1 Right). Sieves are installed into the top and bottom of the reactor vessel: the flow is spread evenly on vertical dimensions, allowing simplification to plug flow. The heat exchanger is simulated as a multi-tube system.

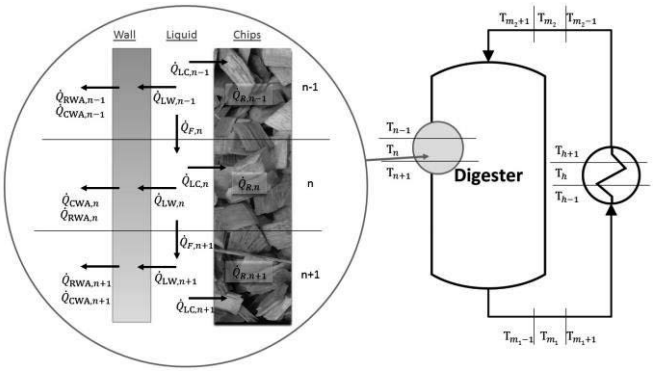


Figure 1. Right: Digester system liquid circulation. Left: Direction of the simulated thermal energy fluxes inside the reactor vessel. The wood chip fraction is absent in the pipes and the heat exchanger.

3. Dynamic model

The dynamic model of the digester is formulated through energy and material balances. The former are defined for the circulating liquid ($\dot{Q}_L$), the wood chips including entrapped liquid ($\dot{Q}_C$) and the digester system walls ($\dot{Q}_W$) and the latter for α-cellulose [C], xylan [X], lignin [L], glucomannan [G], acetyl groups [OAc] and acetic acid [AcOH]. The reactor, the outflow pipe, the heat exchanger and the inflow pipe are divided into N, M_1, H and M_2 sections respectively. Inside each section uniform temperature and component concentration is assumed.

3.1. Energy balance

Heat is carried from one section on to the other with the flow of the liquid fraction ($\dot{Q}_F$) and convected form the liquid into the chips ($\dot{Q}_{LC}$) and the walls ($\dot{Q}_{LW}$). Inside the heat exchanger, the flux from the walls is positive. Heat is lost from the wall sections to ambient by convective ($\dot{Q}_{CWA}$) and radiative heat transfer ($\dot{Q}_{RWA}$). Conduction between separate sections is assumed to be insignificant. The general energy equations for the chip, liquid and wall sections are described below (Eq. 1-3, Figure 1).

$$\dot{Q}_{L,n} = \dot{Q}_{F,n} - \dot{Q}_{LC,n} - \dot{Q}_{LW,n} \tag{1}$$

$$\dot{Q}_{C,n} = \dot{Q}_{LC,n} + \dot{Q}_{R,n} \tag{2}$$

$$\dot{Q}_{W,n} = \dot{Q}_{LW,n} - \dot{Q}_{CWA,n} - \dot{Q}_{RWA,n} \tag{3}$$

3.2. Component balances

Wood degradation reactions take place inside the wood chips. Main wood components [X], [C], [L] and [G] are divided into fast and slow reacting fractions. Additionally, the solubilized and condensed lignin are included. Degradation and condensation reactions and the acetyl group cleavage are modelled using second order kinetics as described in Eq. 4, where $[R_i]$, $[H^+]$ and k stand for the reactant concentration, hydrogen ion concentration and reaction rate coefficient respectively. The water autoionization and acetic acid dissociation are modelled as equilibrium reactions. The diffusion of acetic acid, acetate anions and hydrogen cations form the entrapped water to the free liquid fraction is modelled according to Fick's law of diffusion. The development of cellulose degree of polymerization (DP)—which relates to the pulp viscosity—can be approximated by the concept of degradation increase (DI) and degree of degradation (DD) described in (Kilian and De Vaal, 2000).

$$\left[\dot{R}_{i,n}\right] = -k_{R_i}\left[R_{i,n}\right][H_n^+] \tag{4}$$

4. The control strategy scheme

The aim of the proposed control strategy is to improve pulp quality control by estimating the unobservable reactions during the prehydrolysis stage. The hypothesis is that variations in raw material quality lead to variable hydrolysis rate, which can be explained by the variation in the development of the temperature and hydrogen ion profiles inside the wood chips. By estimating these profiles from the measured data, the developed model can adapt to the variations in raw material quality and calculate the optimal T-profile for achieving the target pulp quality. Consequently, measured liquid temperature ($T_{L,n,m}$) and measured rate of change in liquid energy balance ($\dot{Q}_{L,n,m}$) are combined with the model energy balance for development of soft sensor ($[]_{ss}$) for estimation of the wood chip fraction temperature (Eq. 5) which can in turn be used for estimation of hydrogen ion concentrations (Eq. 6). The Equations 5 and 6 are derived by rearranging Equations 1 and 2 respectively and by observing that the $\dot{Q}_{LC}$ term consists of average heat transfer coefficient between liquid and chip ($\bar{h}_{LC,n}$) and the temperature gradient, while the reaction enthalpy term ($\dot{Q}_R$) is summation over all N_R reaction rates (Eq. 4) multiplied by the corresponding heat of reaction (λ_i). The measurements from the soft sensor are combined with the estimates from the dynamic model component balance using state-estimator, the model states are updated, and finally the optimal temperature set point trajectory is computed using nonlinear model predictive controller (Figure 2).

$$T_{C,n,ss} = \left(\left(\dot{Q}_{L,n,m} + \dot{Q}_{LW,n} - \dot{Q}_{F,n}\right)\bar{h}_{LC,n}^{-1} + T_{L,n,m}\right) \tag{5}$$

$$[H_n^+]_{ss} = \left(\dot{Q}_{C,n} - \dot{Q}_{LC,n}\right)\left(\sum_{i=1}^{N_R} k_i\left[R_{i,n}\right]\lambda_i\right)^{-1}\Big|_{T_{C,n,ss},T_{L,n,m}} \tag{6}$$

For model validation and parameter estimation with industrial data, a correlation between fully cooked and prehydrolysis treated PHK-pulp needs to be established, as the dissolving pulp quality parameters can be measured only after the entire PHK-cycle. This can be accomplished by a simple regression model or a full cooking model and will

furthermore allow the periodic model parameter estimation for adaption to long term deviations (Figure 2). In case a full cooking model is used, the degradation increase computed in estimation of DP allows also the computation of the increase in reducing end groups, which has major effect on the rate of the yield and viscosity reducing peeling reactions.

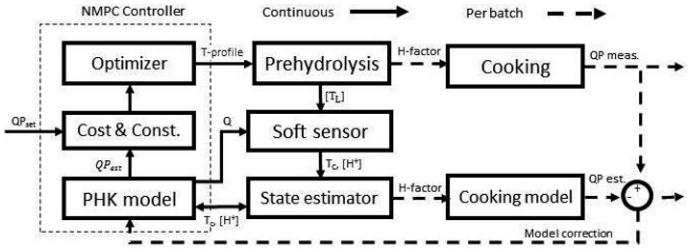


Figure 2. Overall scheme of the proposed control strategy.

5. Preliminary testing and discussion

All the kinetic parameters and initial concentrations, with the exception of xylan, were obtained from Ahmad et al. (2016), where the cellulose degradation is modelled as one fraction and the proportion of fast reacting lignin was 50.06%. The xylan degradation was modelled using the parameters from (Borrega et al., 2011b), scaled to second order kinetics by dividing with the average hydrogen ion concentration. The diffusion rate constants for acetic acid and lignin were obtained from (Haynes et al., 2017), where the solubilized lignin was approximated as ethylbenzene. Temperature relation was modelled by Stokes-Einstein relationship, and multiplied by effective capillary cross sectional area of 0.1 (Inalbon et al., 2017). The heat transfer coefficients used for liquid-wood, liquid-wall and wall-ambient interfaces were 200, 500 and 37 W/(m^{2}K) respectively.

With the used parameters, the temperature of the solid fraction lags behind the liquid fraction by 10 – 15 ºC during the heating period, resulting in a smoother curve and attenuation of the overshoot in temperature control. The hydrogen ion and solubilized lignin diffusion and the mixing through the circulation flow results in almost homogenous pH and solubilized lignin concentration. Lignin and xylan concentrations deviate only slightly between the reactor sections, regardless of the temperature gradient of almost 0.5 ºC. The trends are similar in shape to those presented in the original kinetic model and the model seems to offer a realistic framework for development and testing of the suggested control scheme. Furthermore, the model has potential to provide information about the in batch variance of pulp quality parameters.

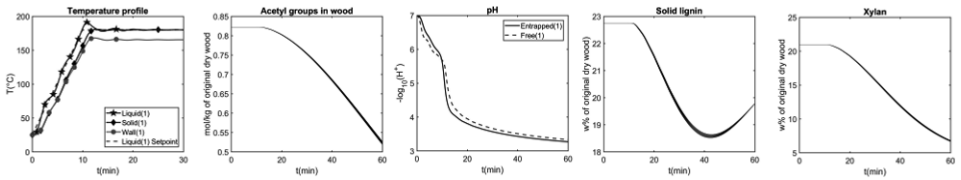


Figure 3. Model output for temperature profile and component balances for acetyl groups, hydrogen ions, lignin and cellulose during autohydrolysis

6. Conclusions

A dynamic mechanistic model for the prehydrolysis stage was presented and a model based control strategy was outlined. The validity of the control strategy relies on the exactness of the computed enthalpy signal ($\dot{Q}_R$), which in turn requires a precise modelling of the energy balance, mass transfer phenomena, reaction enthalpies and reaction rates. However, as the pulping processes are highly non-ideal, the convective heat transfer coefficients and the reaction rate constants are system dependent. In addition, no model for loss of viscosity or heats of reaction during prehydrolysis have been published. The implementation of the model for a real system consequently calls for additional experimental research for determining of these key parameters. Furthermore, the prehydrolysis model needs to be combined with a full cooking model for control of the final PHK-pulp quality. However, initially a simple regression model can be used for model testing and validation.

References

Ahmad, W., Kuitunen, S., Borrega, M., Alopaeus, V., 2016. Physico-chemical modeling for hot water extraction of birch wood. Ind. Eng. Chem. Res. 55, 11062–11073.

Borrega, M., Nieminen, K., Sixta, H., 2011a. Effects of hot water extraction in a batch reactor on the delignification of birch wood. BioResources 6, 1890–1903.

Borrega, M., Nieminen, K., Sixta, H., 2011b. Degradation kinetics of the main carbohydrates in birch wood during hot water extraction in a batch reactor at elevated temperatures. Bioresour. Technol. 102, 10724–10732.

Brasch, D.J., Free, K.W., 1965. Prehydrolysis-kraft pulping of Pinus radiata grown in New Zealand. Tappi 48, 245–248.

Conner, A., 1984. Kinetic Modeling of Hardwood Prehydrolysis. Part I. Xylan Removal by Water Prehydrolysis. Wood Fiber Sci. 16, 268–277.

Courchene, C.E., McDonough, T.M., Hart, P.W., Malcolm, E.W., Carter, B.R., 2005. Determining the heat of reaction of kraft pulping. Tappi J. 4, 9–13.

de Vaal, P., Sandrock, C., 2007. Control of a batch pulp digester using a simplified mechanistic model to predict degree of polymerisation. Comput. Chem. Eng. 31, 1222–1230.

Haynes, W.M., Lide, D.R., Bruno, T.J., 2017. CRC Handbook of Chemistry and Physics. CRC Press, Florida.

Hämmerle, F.M., 2011. The Cellulose Gap (the future of cellulose fibres). Lenzinger Berichte 89, 12–21.

Inalbon, M.C., Solier, Y.N., Zanuttini, M.Á., 2017. Hydrothermal treatment of Eucalyptus wood: Effects on Ion permeability and material removing. Ind. Crops Prod. 104, 195–200.

Kerr, A.J., 1976. The kinetics of kraft pulping batch--digester control. Tappi 59, 89–91.

Kilian, A., De Vaal, P.L., 2000. Potential for improved control of an acid sulfite batch digester using a fundamental model. Tappi J. 83, 68.

Lee, J.H., Datta, A.K., 1994. Nonlinear inferential control of pulp digesters. AIChE J. 40, 50–64.

Lin, C.K., 1979. Prehydrolysis-Alkaline Pulping of Sweetgum Wood (Liquidambar Styra L.). North Carolina State University at Raleigh.

Nieminen, K., Sixta, H., 2012. Comparative evaluation of different kinetic models for batch cooking: A review. Holzforschung 66, 791–799.

Noreus, S., Wallin, G., Olsson, T., Sandblom, H., 1974. Computer Control of Batch Digesters. Pulp Pap. Mag. Canada 75, 102–106.

Paulonis, M.A., Krishnagopalan, A., 1991. Adaptive Inferential Kraft Batch Digester Control Based on Cooking Liquor Analysis. Tappi J. 74, 169–175.

Simmonds, F.A., Kingsbury, R.M., Martin, J.S., 1955. Purified Hardwood Pulps for Chemical Conversion. Tappi 38, 178–186.

Vroom, K.E., 1957. The “H” factor: - A Means of Expressing Cooking Times and Temperatures as a Single Variable. Pulp Pap. Mag. Canada 58, 228–231.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50103-8

Using Cognitive Computing for the Control Room of the Future

Sambit Ghosh[a], B. Wayne Bequette[a*]

[a]*Department of Chemical and Biological Engineering, Rensselaer Polytechnic Institute, 110 Eighth Street, Troy, NY 12180, USA*
bequette@rpi.edu

Abstract

Chemical manufacturing processes are often complex and large-scale, with many variables to monitor and control. The startup and shutdown of process units can be particularly challenging, with specific and coordinated sequences of events to assure safety while completing operation goals in a reasonable amount of time. Recent advances in cognitive computing, "big data" and process data analytics have great potential to improving the transient operation of chemical processes. A framework to study process startups, shutdowns and abnormal events using graph theory, cognitive computing and machine learning is presented. Preliminary results based on a simple two-tank problem show the usefulness of graph-based analyses. The procedure detects a sensor fault and startup protocol deviations. Ongoing research tests the proposed strategy on more realistic larger scale chemical processes.

Keywords: Startup/shutdown, cognitive computing, process operators, smart manufacturing

1. Introduction

Startups and shutdowns (SU/SD) are the most hazardous times during process operations, yet most process systems engineering research has been focused on continuous, steady-state operation. Startups and shutdowns have an incident rate 10 times that of steady-state operation (CSB, 2007). While chemical processes are highly automated, it is during these times that the human-in-the-loop plays a critical role, and several disasters have occurred because proper protocols were not followed.

SU/SD involves a combination of manual and automatic operations. Sometimes loops are put into manual because the response is faster than a closed-loop controller Li et al. (2016) use a fuzzy logic based approach to analyze the interactions between operators in a nuclear power plant. Sand and Terwiesch (2013) note that process plant operators have the tasks of SU/SD, supervising the automated operation, and diagnosing and managing disturbances. They stress the need to increase the performance of human-machine interfaces.

Our research effort uses several mathematical and computational tools to analyse the complex non-linearities during SU/SD and provides a framework for the "control room of the future,". This paper focuses on the use of graph theory to detect process deviations or faults. Network and Graph theory has been applied to a range of problems, with important results by Fiedler (1973). Spectral graph theory, especially using the normalized Kirchoff (or Laplacian matrix in graph theory) matrix was discussed in Chung (1994). In process dynamics and control, Jogwar and Daoutidis (2017) use the concept of

communities to design controllers, while Heo and Daoutidis (2015) use graph representations to simplify the analysis of a complex plant.

Much of the focus on fault detection has been steady-state or small dynamic perturbations from steady-state. Isermann (2005) provide reviews of fault detection, with selected applications, of quantitative model-based fault detection and diagnosis techniques. Data-driven approaches for detecting multiple faults are detailed by He et al. (2014). Severson et al. (2016) provide an overview of recent work in process monitoring. Most notable for our application, Wang et al. (2012) develop a hybrid fault diagnosis strategy for chemical process start-up. In the following sections, a novel method is proposed that involves human-machine interactions during a dynamic process

2. Methodology

The goal of this paper is to define a framework to analyse complex dynamic processes and include the human elements in the analysis to understand the temporal patterns over which such a process operates. Fig 1 outlines our complete research goal and vision. The work highlighted in blue in Fig 1 is presented by using a simple proof-of-concept example.

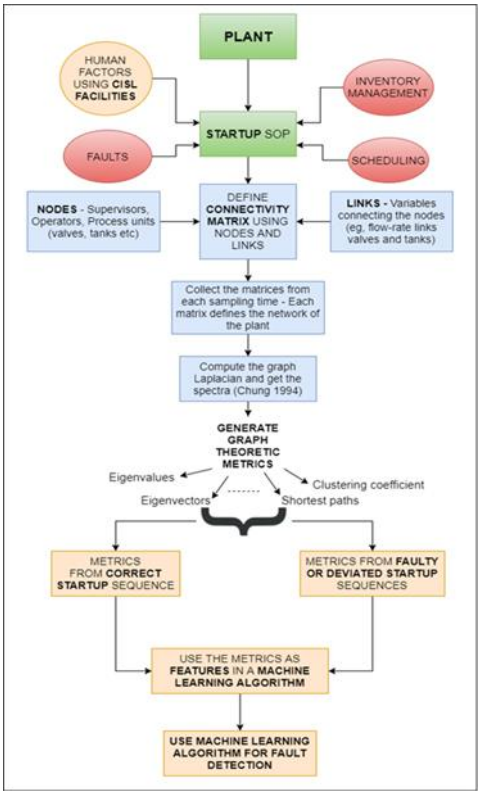


Figure 1. Flowchart of proposed fault detection methodology using process dynamics, graph theory, machine learning and CISL facilities. Color code – Green (available information for any process), Red (external disturbances affecting the startup sequence), Blue (Ongoing work in the lab), Yellow (Future work)

Consider a system of two tanks in series along with the individuals responsible for managing the process (Fig 2) Process model equations are shown in Table 1. The inlet flow to T_1 is controlled by control valve 1 (CV_1). CV_2 controls the flow between T_1 and T_2, and CV_3 controls the flow out of T_2. There are two shifts of operators with their respective supervisors. Shift 1 has Supervisor 1 (S_1) with Operators 1 and 2 (O_1 and O_2). Shift 2 has Supervisor 2 (S_2) with Operators 3 and 4 (O_3 and O_4). Here we assume only shift 1 is active. O_1 controls CV_1 and O_2 controls CV_2 and CV_3. Finally, Sensor 1 and 2 (Sen_1 and Sen_2) give the tank height readings to the shift operators.

2.1 Spectral analysis

Using the above description, the following steps are executed:

1. The total number of physical units = 13 (6 personnel, 3 CVs, 2 tanks, 2 sensors).
2. Define a symmetric matrix called the adjacency matrix ***Ad*** of size 13X13 where each node is a physical unit. This is described in Table 2.
3. Define the links (connection between each node) as follows:

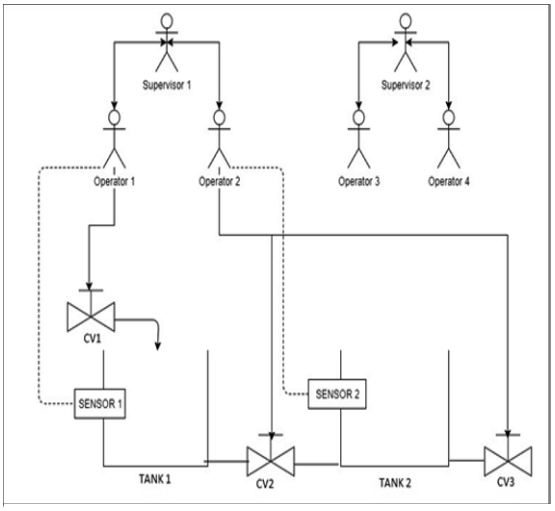


Figure 2. Two tank interacting system hierarchy

S_1 with O_1, O_2 (and S_2 with O_3, O_4) – weights vary from 0 to 1 depending on the level of interaction between the supervisor and the operators. A higher weight indicates more interaction during a process. In the present example, the weights are assigned based on assumptions. In future studies these will be determined from human-interaction based models (Riley et al, 2016) and cognitive computing. O_1 with CV_1 - fraction of the valve opened (K_1, ranges from 0-1). O_2 with CV_2 and CV_3 - fraction of the valve opened (K_2, K_3 ranges from 0-1). CV_1 with T_1 – Flow in (F_1). CV_2 with T_1, T_2 – Flow out of T_1 or flow in to T_2 (F_2). CV_3 with T_2 – Flow out of T_3 (F_3). Sen_1 with T_1 and Sen_2 with T_2 – 1 in case of no fault and 1.3 (i.e a sensor bias) in the faulty scenario. Sen_1 with O_1 and Sen_2 with O_2 – Tank height readings. 4. From these definitions, at each time step, the matrix ***Ad*** can be created using data from the plant. ***Ad*** thus forms a connectivity matrix (or adjacency matrix) for the entire plant. The connectivity changes with respect to the SU sequences with time. As a result, analysing changes in the matrix ***Ad*** connectivity gives us consolidated information about the entire plant.

Table 1: Two-tank interacting system model

Equations	Definitions
$dh_1/dt = (F_1 – F_2)/A_1$	h_1, h_2 fluid heights of T_1 and T_2 (m). A_1, A_2 are cross-sectional area
$dh_2/dt = (F_2 – F_3)/A_2$	F_{in} is $1m^3$ F_1 is flow in to tank 1 in m^3/min
$F_1 = K_1*F_{in}$	, F_2 is flow in to tank 2 in m^3/min, F_3 is flow out of tank 2 in m^3/min
$F_2 = K_2*\beta_1\sqrt{(h_1-h_2)}$	β_1 (=1) and β_2 (=1) are flow coefficients in $m^3/min/m^{0.5}$
$F_3 = K_3* \beta_2\sqrt{h_2}$	K_1-K_3 are fraction open of CV_1-CV_3

An example of the ***Ad*** matrix is given below

$$Ad = \begin{bmatrix} 0 & 0 & [0,1] & [0,1] & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & [0,1] & [0,1] & 0 & 0 & 0 & 0 & 0 & h1 & 0 \\ [0,1] & 0 & 0 & 0 & 0 & 0 & K1 & 0 & 0 & 0 & 0 & 0 & h2 \\ [0,1] & 0 & 0 & 0 & 0 & 0 & 0 & K3 & K3 & 0 & 0 & 0 & 0 \\ 0 & [0,1] & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & [0,1] & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & K1 & 0 & 0 & 0 & 0 & 0 & 0 & F1 & 0 & 0 & 0 \\ 0 & 0 & 0 & K3 & 0 & 0 & 0 & 0 & 0 & F2 & F2 & 0 & 0 \\ 0 & 0 & 0 & K3 & 0 & 0 & 0 & 0 & 0 & 0 & F3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & F1 & F2 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & F2 & F3 & 0 & 0 & 0 & 1 \\ 0 & h1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & h2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Table 2: Defined nodes for the process

Nodes	Description
1	Supervisor 1 (S_1)
2	Supervisor 2 (S_2)
3	Operator 1 (O_1)
4	Operator 2 (O_2)
5	Operator 3 (O_3)
6	Operator 4 (O_4)
7	Control valve 1 (CV_1)
8	Control valve 2 (CV_2)
9	Control valve 3 (CV_3)
10	Tank 1 (T_1)
11	Tank 2 (T_2)
12	Sensor 1 (Sen_1)
13	Sensor 2 (Sen_2)

5. We now borrow concepts from network/graph theory to analyse ***Ad***. Let $\mathbf{L} = \mathbf{D}-\boldsymbol{Ad}$ where ***Ad*** is the adjacency matrix, **D** is the degree matrix and **L** is the Kirchoff (also called Laplacian in graph theory) matrix. The degree matrix is defined as a diagonal matrix whose elements are the sum of all edges for a particular node in ***Ad***.
6. Define $\mathbf{L^*} = \mathbf{D^{-1/2}LD^{-1/2}}$ where **L*** gives us the normalized Kirchoff matrix and the eigen-decomposition of $\mathbf{L^*} = \boldsymbol{\Theta\Lambda\Theta^T}$ where $\boldsymbol{\Theta}$ are the eigenvectors and $\boldsymbol{\Lambda}$ the eigenvalues. For a nXn matrix, there will be n eigenvalues and n eigenvectors (Chung, 1994). Some key features from this decomposition are as follows:

• The set of eigenvectors $\boldsymbol{\Theta}$ form an orthonormal basis of the network and eigenvalues of **L*** are scaled from 0-2 for any symmetric matrix
• The multiplicity of zero eigenvalues give the number of connected components in the graph/network and number of eigenvalues equal to 1 gives information about similar nodes (especially isolated nodes)
• The eigenvector corresponding to the first non-zero (and thus second smallest) eigenvalue is called the Fiedler vector - gives information about the clustering or grouping of nodes in a network and is often used to compare networks. In this example, the Fiedler vector will be used to show the differences in the SU sequences.

2.2 Shortest paths
Shortest path analyses are very useful to understand the information flow in a complex network. In this work, we use MATLAB's shortest path module on the ***Ad*** matrices and Dijktra's algorithm (Dijktra, 1959) to find the information flow between the components. An analysis of three different SU procedures is presented below.

3. Results

The SU/SD of a plant is defined by a Standard Operating Procedure (SOP) containing a series of sequential steps. As a result, SU/SDs are temporal in nature and any deviations in it should be studied in a simple manner to detect process deviations. Since SU/SD disasters occur due to deviations from the SOPs, it is pertinent to have a method to detect these deviations. In this study, the use of network parameters like Fiedler vectors (FVs) and shortest paths are used to show how different kinds of information can be extracted from a dynamic process as defined before.

For the process in Fig 2, an SOP is defined as given in Table 3.

Table 3: SOP of the two-tank system

Startup sequence	Fraction of valve opened in the startup protocol
Correct	0-50mins – K_1=0.5, K_2=1, K_3=0
	51-100mins – K_1 = 0.1, K_2=0, K_3=0
	101-150mins – K_1=1, K_2=1, K_3=1
Deviated	0-30mins – K_1=1, K_2=1, K_3=0
(to fill the tank faster compared to the Correct startup sequences)	31-150mins – K_1=1, K_2=1, K_3=1
Deviated+Faulty	0-30mins – K_1=1, K_2=1, K_3=0
(Starts with the same protocol as Deviated but changes to a more ad-hoc scenario due to the faulty sensor reading)	31-60mins – K_1=1, K_2=1, K_3=1
	61-90mins – K_1=0.8, K_2=1, K_3=1
	91-120mins – K_1=0.92, K_2=1, K_3=1
	121-150mins – K_1=0.92, K_2=0.9, K_3=1

Simulating the process for the three different SU sequences, the ***Ad*** matrices for the three SU: Correct, Deviated and Deviated+Faulty are found at each time step using steps 3-4

in Methodology. Using the three sets of ***Ad*** matrices, the spectral decomposition is computed and the FV are obtained using steps 5-6 in Methodology. Fig 3 plots the FV for the three SU at 3 different time steps. At t=15 and 60 mins, both the Deviated and Deviated+Faulty SU show differences with the Correct SU as this is the transient region. However, at t=135 mins, the Correct and Deviated cases have the exact same plant state (all weights are same) as they reach steady state. As a result, their FV match exactly, indicating that the two processes are the same at all hierarchy. In Fig 4, it is shown that when we change S_1-O_1 and S_1-O_2 to 0.5 (indicating occasional interaction of operators with supervisors) only for the Deviated scenario, even at t=135 when all the process variables are having the same values (steady state), due to a change in S_1-O_1 and S_1-O_2 weights from the Correct scenario (=0.1), the FV do not match.

The shortest path can provide insights about the flow of information in the network. Using the graphshortestpath() command in MATLAB the following paths were found:

1. S_1-O_2-Sen_2-T_2, S_1-O_2-CV_2-T_2, S_1-O_2-CV_2-T_1, S_1-O_2-CV_3-T_2
2. S_1-O_1-Sen_1-T_1, S_1-O_1-CV_1-T_1
3. S_2-O_3, S_2-O_4

It is easy to see that if Shift 2 was in charge, then the paths would have been different and hence we end up with a technique to see changes in shift with time. Interestingly, this change is also captured in the eigenvalues reported in Table 4. The multiplicities of eigenvalues (=1) equals the number of human personnel not connected to the process due to a different shift. This can be generalized to any process equipment as well.

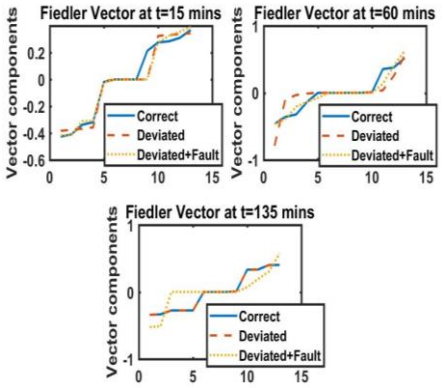


Figure 3. Fiedler vector comparison of the three SU sequences at three time snapshots.

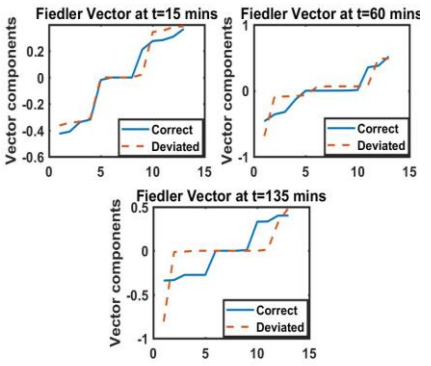


Figure 4. Fiedler Vector of Correct and Deviated. S1-O1 and S1-O2 are 0.1 and 0.5 for Correct and Deviated respectively.

Table 4: Eigenvalues of the normalized Laplacian at t=135 minutes.

Correct	0	0	0.009	0.813	0.921	1	1	1	1.078	1.186	1.99	2	2
Deviated	0	0	0.009	0.812	0.921	1	1	1	1.078	1.186	1.99	2	2
Deviated+Faulty	0	0	0.068	0.617	0.991	1	1	1	1.080	1.382	1.931	2	2

4. Conclusion and Future work

Graph theory was used to analyse SU of a simple process. The results from the FV show that sets of sequential processes will show differences if the links in the individual networks change. It is also useful to analyse the sequences with time; a user-defined time window can be defined to get average ***Ad*** matrices and FV. Similarly, deviations in shortest paths of SU sequences and eigenvalues of a process indicate a change in information flow paths between the physical units and the connectivity of a node in a network. It is evident from these results that for a complex process, such an information transformation helps to identify the changes that are taking place dynamically. These are preliminary results and that we have a vision of the control room of the future that includes the use of IBM Watson based facilities at RPI and machine learning approaches to improve operations.

References

CSB report on Texas City explosion. U.S. Chemical Safety and Hazard Investigation Board. Investigation report: refinery explosion and fire, BP Texas City, Texas. Report no. 2005-04-I-TX. March 2007. Available at: http://www.csb.gov/ assets/1/19/csbfinalreportbp.pdf.

Chung, F. R. K. (1994) 'Spectral Graph Theory', Conference Board of the Mathematical Sciences, (92), p. 214.

Dijkstra, E.W., 1959. A note on two problems in connexion with graphs. Numerische mathematik, 1(1), pp.269-271.

Fiedler, M., 1973. Algebraic connectivity of graphs. Czechoslovak mathematical journal, 23(2), pp.298-305.

He Y-L, Wang R, Kwong S, Wang X-Z. Bayesian classifiers based on probability density estimation and their applications to simultaneous fault diagnosis. Information Sciences 2014;259:252-268.

Heo, S. and Daoutidis, P. (2015) 'Graph-Theoretic Analysis of Multitime Scale Dynamics in Complex Material Integrated Plants', Ind. Eng. Chem. Res., 54(42), pp. 10322–10333.

Isermann R. Model-based fault-detection and diagnosis—status and applications. Annu Rev Control. 2005;29:71-85.

Jogwar, S. S. and Daoutidis, P. (2017) 'Community-based synthesis of distributed control architectures for integrated process networks', Chem. Eng Sci., 172, pp. 434–443.

Li, P.C., Chen, G.H., Dai, L.C., Zhang, L., Zhao, M. and Chen, W., 2016. Methodology for analyzing the dependencies between human operators in digital control systems. Fuzzy Sets and Systems, 293, pp.127-143.

Riley, W.T., Martin, C.A., Rivera, D.E., Hekler, E.B., Adams, M.A., Buman, M.P., Pavel, M. and King, A.C., 2016. Development of a dynamic computational model of social cognitive theory. Translational behavioral medicine, 6(4), pp.483-495.

Sand G, Terwiesch P. Closing the loops: An industrial perspective on the present and future impact of control. European J. Control. 2013;19:341-350.

Severson K, Chaiwatanodom P, Braatz RD, Perspectives on process monitoring of industrial systems, In Annual Reviews in Control, 2016; 42:190-200.

Wang Z, Zhao J, Shang H. A hybrid fault diagnosis strategy for chemical process startups. J. Proc. Cont. 2012;22:1287-1297.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50104-X

Model Approximation in Multiparametric Optimization and Control – A Computational Study

Justin Katz[a,b], Nikolaos A. Diangelakis[a,b], Efstratios N. Pistikopoulos[a,b*]

[a]*Artie McFerrin Department of Chemical Engineering, Texas A&M University 100 Spence St., College Station, TX 77843, United States*
[b]*Texas A&M Energy Institute, Texas A&M University 1617 Research Pkwy, College Station, TX 77845, United States*
stratos@tamu.edu

Abstract

Incorporating a high fidelity model that accurately describes a dynamical system in an optimization and control study may often lead to an intractable formulation, hence the use of model approximation is required. This computational study closely examines various approximation techniques in the context of multiparametric optimization and control with the use of key error metrics including: (i) open loop comparison of the high fidelity and approximate model, (ii) verification of step response profiles, and (iii) comparison of key features of the feasible space and objective function in the optimization formulation. Two systems are used as a basis for this study: a tank system utilized to highlight the main principles of this approach, and a Continuously Stirred Tank Reactor (CSTR) where the reaction mechanisms are manipulated to increase the model complexity.

Keywords: multiparametric programming, model reduction, model predictive control.

1. Introduction

Model based optimization for control, such as Model Predictive Control (MPC), has been gaining traction in the academic and industrial communities for more than 3 decades now (Camacho and Bordons, 2007). In recent years, more advanced models are being incorporated into these model predictive control frameworks (Santos et al, 2001). These direct formulations may typically result in an intractable, large scale, complex, non-convex optimization problem. Problems of this nature can be reduced to tractable forms via model approximation techniques (Diangelakis et al. 2017) that can then be solved offline explicitly using multiparametric programming (Bemporad et al., 2002).

Many techniques have been used for model approximation in Multiparametric Model Predictive Control (mpMPC), two common techniques include (piece-wise) linearization and system identification. Such model approximations are also at the heart of the PARameteric Optimization and Control (PAROC) framework for the derivation of these explicit/multiparametric controllers (Pistikopoulos et al., 2015). A key question that remains open within the PAROC framework is "*what constitutes a suitable approximate model for the derivation of explicit control strategies with multiparametric programming*?".

In this work, we present a computational study towards addressing this question. In particular, we study system identification (Ljung, 1998), and linearization. These model

approximation techniques are widely used to develop linear discrete time state space models that can be incorporated into an MPC formulation. An additional approximation technique based on forward Euler discretization is also studied to compare against the aforementioned techniques. Key error criteria will be used to ascertain the effectiveness of each of the approximation techniques. These error criteria include: (i) open loop error analysis comparing the approximate model against the high fidelity model, (ii) step response profile analysis for each of the approximate models, (iii) comparison of the closed loop trajectory against the optimal trajectory determined from dynamic optimization in the gPROMS® environment, (iv) comparison of the objective values from the different mpMPC formulations and the optimal value determined in gPROMS®, and (v) comparison of the feasible space and objective function for the mpMPC formulations.

2. Methodology

The three approximation techniques used in the work are linearization, system identification, and a technique based on forward Euler discretization. Linearization is performed using first order Taylor Series approximation, where the linearization point is chosen to be at the defined set point of interest. In addition, system identification was performed using the MATLAB routine N4SID which is a subspace based approach to determine the state space matrices that best represent a set of outputs based on a given set of inputs.

The last approximation technique is based on Forward Euler discretization, which approximates the derivative information by taking a finite difference, instead of letting the limit approach zero, as seen in Eq.(1).

$$\frac{\boldsymbol{dx}}{\boldsymbol{dt}} = \boldsymbol{f(x)} \Longrightarrow \frac{\boldsymbol{\Delta x}}{\boldsymbol{\Delta t}} \approx \boldsymbol{f(x_t)} \tag{1}$$

The benefit of this technique is that the derivate information is preserved through approximate discretization. As seen in Eq.(2), nonlinear state terms are treated as uncertain parameters, θ_t and ψ_t, to be used in an mpMPC formulation. In Eq.(2), the term $\psi_t u_t$ results in left hand-side uncertainty when formulated as an mpMPC problem. To avoid the challenges in solving a parametric programming problem with left hand-side uncertainty, this term can be grouped to form a new manipulated variable, as seen in Eq.(3). The new manipulated variables, $\tilde{u}$, will have varying bounds, which can be handled via multiparametric programming. However, grouping an uncertain parameter with a manipulated variable can only occur once for each manipulated variable.

$$\boldsymbol{x_{t+1} = f(x_t) + g(x_t)u_t = \theta_t + \psi_t u_t} \tag{2}$$

$$\boldsymbol{x_{t+1} = \theta_t + \tilde{u}_t} \tag{3}$$

The error criteria used in this work includes (i) root mean squared error deviation between the open loop profile for the high fidelity and approximate model, (ii) qualitative verification of the step response profiles of the approximate models, and (iii) comparison of the feasible space and objective function for the optimization formulations. Open loop analysis is performed using a random input profile on the high fidelity model and the approximate models developed. Step response profiles are verified based on how the real system is expected to perform. To compare feasible spaces of the optimization formulations, the volume of each feasible space is determined using Monte Carlo techniques, and to compare objective functions, the L^2 norm is used as seen in Eq.(4).

$$\|\boldsymbol{f}(\boldsymbol{x})\|_2 = \left(\int_{\boldsymbol{X}} |\boldsymbol{f}(\boldsymbol{x})|^2 \, \boldsymbol{dx}\right)^{\mathbf{0.5}} \tag{4}$$

3. Tank and Continuously Stirred Tank Reactor Examples

To demonstrate the criteria and approximation techniques used, a simplified tank system is examined. The tank has a fixed inlet flow and the flow out of the tank is manipulated via a valve on the exit of the tank. The control objective of the system is to maintain the level of the tank at a specified target. The system maintains a nonlinearity in the form of a square root, as seen in Eq.(5).

$$\frac{\boldsymbol{dh}}{\boldsymbol{dt}} = \frac{\boldsymbol{F}}{\boldsymbol{A}} - \boldsymbol{u}\sqrt{\boldsymbol{2gh}}\frac{\boldsymbol{A_{out}}}{\boldsymbol{A}} \tag{5}$$

The manipulated variable, u, maintains bounds between 0 and 1. The state of the system, h (m), must remain nonnegative. The parameters g (m/s^2), A (m^2), and A_{out} (m^2) are gravity, the area of the tank, and the outlet area of the tank respectively. The MPC formulation can be seen in Eq.(6).

$$\begin{aligned} &\min_{\boldsymbol{u}} \int_0^{\boldsymbol{\tau}} \left(\left(\boldsymbol{x} - \boldsymbol{x_{ref}}\right)^{\boldsymbol{T}} \boldsymbol{Q}\left(\boldsymbol{x} - \boldsymbol{x_{ref}}\right)\right) \boldsymbol{dt} \\ &\boldsymbol{s.t.}\ \ \dot{\boldsymbol{x}} = \boldsymbol{Ax} + \boldsymbol{Bu}, \quad \boldsymbol{x} \in \boldsymbol{X}, \boldsymbol{u} \in \boldsymbol{U} \end{aligned} \tag{6}$$

where Q is a cost matrix, x_{ref} is the set point of the system, x is the state of the system, and u is the input to the system. The MPC formulation can be discretized and converted to a multiparametric MPC where it is solved explicitly offline to determine the optimal control action as an affine function of the uncertain parameters, namely the initial state of the system and the set point.

As a second example, a single reaction, isothermal, and constant volume Continuously Stirred Tank Reactor (CSTR) is considered where the reaction rate complexity is increased. The reaction rate is either first order, second order, or third order. The CSTR has an adjustable flow rate of reactant into the reactor. Constant volume in the reactor is assumed and the reactor is considered to be isothermal. The control objective for this example problem is to maintain a specified reactant concentration level in the reactor while minimizing the use of the reactant flow.

$$\frac{\boldsymbol{dC_A}}{\boldsymbol{dt}} = \frac{\boldsymbol{F}}{\boldsymbol{V}}\left(\boldsymbol{C_{A_i}} - \boldsymbol{C_A}\right) - \boldsymbol{kC_A^{\alpha}} \tag{7}$$

$$\frac{\boldsymbol{dC_B}}{\boldsymbol{dt}} = \frac{\boldsymbol{F}}{\boldsymbol{V}}\left(-\boldsymbol{C_B}\right) + \boldsymbol{kC_A^{\alpha}} \tag{8}$$

The mass balance for the system can be described by Eq.(7) and Eq.(8), where the amount of reactant and product vary based on the inlet flow and the reaction mechanism. The manipulated flow into the reactor, F (m^3/s), is bounded between 0 and 1, and the states of the system, the reactant (C_A) and product (C_B) concentration (mol/m^3), are also bounded between 0 and 1. V (m^3) is the volume of the system, k is the rate constant of the system, C_{Ai} is the inlet reactant concentration, and the system order is α. The MPC formulation for this example problem, similar to Eq.(6), includes set point tracking on the reactant concentration, since there is a 1:1 correspondence between the reactant concentration and

the product concentration, and a penalty on the usage of the reactant material, or a penalty on the input to the system.

4. Results and Discussion

Both examples were simulated and dynamically optimized using the gPROMS® software. The dynamic optimization was performed using Control Vector Parameterization which discretizes the manipulated variable over a control interval, and over each interval the manipulated variable is assumed to follow a simple function. The optimal trajectory determined from gPROMS® is considered as the optimal trajectory and is compared against the mpMPC closed loop trajectories. The objective function used in the mpMPC formulation is the same in all formulations. For the mpMPC formulations, the output and control horizon is 3 and 2 respectively for the tank example and 5 and 2 respectively for the CSTR example.

Figure 1a shows the open loop response of the approximate models and the high fidelity model for the tank example, where the legend shows the root mean squared error. The open loop performance of system identification performs the worst out of the approximation techniques, and linearization performs the best. Figure 1b is a comparison of the step response profiles. All of the step response profiles show a decrease in the output for an increase in the input, which is consistent with how the system would react. The approximate model resulting from forward Euler discretization is a straight line because the state space representation is critically stable. Figure 2 shows trajectories from using different approximate models and control strategies, and the corresponding output of the system. The optimal trajectories are determined from the mpMPC formulation, a dynamically optimized PI controller, and open loop dynamic optimization on the full process model in gPROMS®. Table 1 shows a comparison of the objective function costs, volume of the feasible space, and distance between objective functions. From Table 1, based on the objective function costs, linearization has the most comparable performance to the open loop dynamic optimization results. This can be attributed to linearization performing well around the linearization point, which was chosen as the set point. Forward Euler has a similar cost and an objective function that more closely matches the 'real' objective function. System identification performs the worst, except it has a feasible space volume that is closer to the actual the feasible space volume.

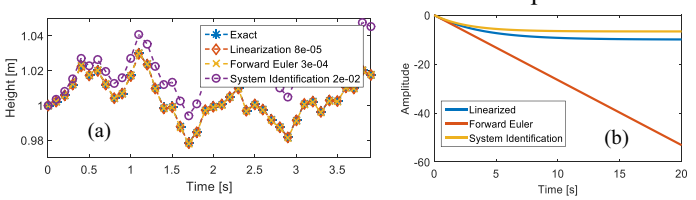


Figure 1. (a) Open loop comparison of output via root mean squared error (b) Step response profiles of the approximate models for the tank example

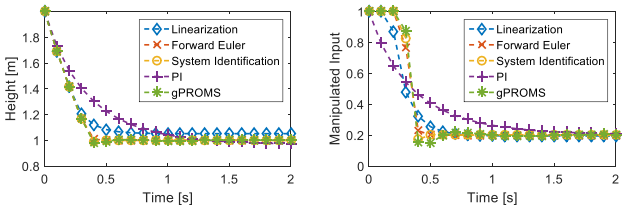


Figure 2. Closed loop performance comparison for the various controllers and open loop dynamic optimization for the tank example

Table 1. Quantitative results of different error criteria

	Linearization	System Ident	Forward Euler	Exact
Optimal Trajectory Cost (Tank)	1.1264	1.12738	1.12723	1.1261
Feasible Space Volume (Tank)	98	98.3	94.9	100
Objective Function L^2 (Tank)	5.19	13.42	0.1497	0
Optimal Trajectory Cost (CSTR),α=1	0.139	0.140	0.140	0.135
Feasible Space Volume (CSTR), α=1	0.47	0.44	0.26	1
Objective Function L^2 (CSTR), α=1	0.54	0.42	3.1604	0
Optimal Trajectory Cost (CSTR),α=2	0.112	0.113	0.112	0.110
Feasible Space Volume (CSTR), α=2	0.41	0.38	0.19	1
Objective Function L^2 (CSTR), α=2	0.67	1.54	4.22	0
Optimal Trajectory Cost (CSTR),α=3	0.174	0.181	0.174	0.173
Feasible Space Volume (CSTR), α=3	0.36	0.43	.22	1
Objective Function L^2 (CSTR), α=3	0.84	1.25	4.22	0

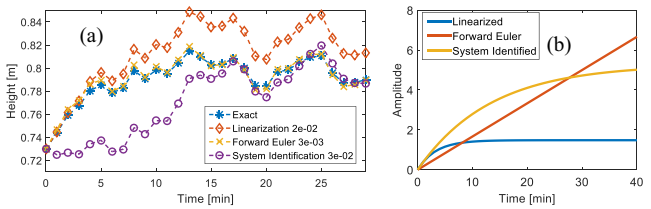


Figure 3. (a) Open loop comparison of output via root mean squared error (b) Step response profiles of the approximate models for the CSTR example

The open loop performance and step responses for the CSTR are only compared for an α value of 3. As seen in Figure 3a, the open loop performance of system identification performs the worst out of the approximation techniques. Figure 3b is a comparison of the step response profiles. All of the step responses show an increase in the output concentration for an increase in the input, as expected for this system. Forward Euler approximation has a straight line due because its state space model is critically stable. Figure 4 shows optimal trajectories from using different approximate models and techniques, and the corresponding output of the system. The optimal trajectories are determined from the mpMPC formulation, a dynamically optimized PI controller, and open loop dynamic optimization on the full process model. Comparison of the objective function costs, volume of the feasible space, and distance between objective functions can be seen in Table 1. For all of the different values of α, linearization performs the best, while maintaining a feasible space and objective function most similar to the real system. Because the system has a set point that matches the linearization point the linearized state space model can accurately represent the system in this region. Forward Euler performs well for all α even though it has a feasible space that is smaller than both the system identification and linearization feasible spaces. Its objective function distance is also farther away than the objective functions for both linearization and system identification. The feasible space being small is not enough on its own to categorize poor performance

and the distance in objective functions is not large enough to cause performance issues either.

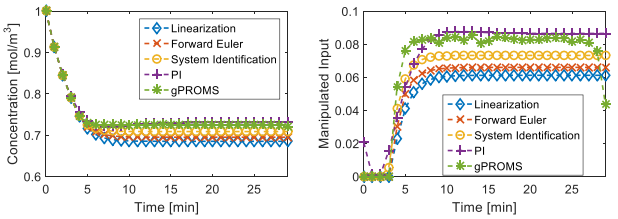


Figure 4. Closed loop performance comparison for the various controllers and open loop dynamic optimization for the CSTR example

5. Conclusions

Various model approximation techniques have been utilized to develop multiparametric model predictive controllers. These approximate models and resulting optimization formulations were assessed using various error criteria. For the example problems presented, and criteria used, the linearization technique provided an optimal trajectory that was closest to the 'desired' trajectory using dynamic optimization in gPROMS®. Future work is to apply these concepts on a system of significantly increased complexity and further development on the concepts of feasible space volume and distance between objective functions.

6. Acknowledgments

We acknowledge the financial support from the NSF (Grant no. 1705423), under the project "SusChEM: An integrated framework for process design, control and scheduling [PAROC]".

References

A. Bemporad, M. Morari, V. Dua, E.N. Pistikopoulos, 2002, *The explicit linear quadratic regulator for constrained systems*, Automatica 38 (1), 3-20

E.F. Camacho, C. Bordons, 2007, *Nonlinear Model Predictive Control: An Introductory Review*, Assessment and Future Directions of Nonlinear Model Predictive Control, Springer, 358, 1-16

N.A. Diangelakis, B. Burnak, J. Katz, E.N. Pistikopoulos, 2017, *Process Design and Control optimization: A simultaneous approach by multi-parametric programming*, AIChE Journal, 63 (11), 4827-4846

L. Ljung, 1998, *System Identification*, Signal Analysis and Prediction, 163-173

E.N. Pistikopoulos, N.A. Diangelakis, R. Oberdieck, M.M. Papathanasiou, I. Nascu, M. Sun, 2015, *PAROC-An integrated framework and software platform for the optimisation and advanced model-based control of process systems*, *Chemical Engineering Science*, 136, 115-138

L.O. Santos, P.A.F.N.A. Afonso, J.A.A.M. Castro, N.M.C. Oliveira, L.T. Biegler, 2001, *On-line implementation of nonlinear MPC: an experimental case study*, In Control Engineering Practice, Volume 9,847-857

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50105-1

Development of a gas composition soft sensor for distillation columns: A simplified model based and robust approach

Eliza H. C. Ito[a*], Argimiro R. Secchi[a] , Marcos V. C. Gomes[b], Carlos R. Paiva[a,b]

[a] *Chemical Engineering Graduate Program, COPPE, Universidade Federal do Rio de Janeiro, UFRJ, Cidade Universitária, Rio de Janeiro, RJ, 21941-972, Brazil.*
[b] *Petrobras Research Center, Rio de Janeiro, RJ, 21949-900, Brazil.*
eliza@peq.coppe.ufrj.br

Abstract

The product composition control in a multicomponent distillation column presents several challenges in the industry. The development of a soft sensor enables the monitoring of the product quality using an inferential model and on-line measured process variables. To be useful in an industrial application, the inferential model needs to provide fast sampling rates and with high accuracy. A new inferential model to estimate the gas composition product from a multicomponent distillation column is proposed. It is based on a model that combines a first-principles model at the equilibrium stage with heuristic rules based on the process knowledge. Moreover, strategies for key components selection and a model for the systematic correction of input variables are used to ensure the robustness on a real-time application. A rigorous simulation of a debutanizer distillation is implemented to compare with the proposed inferential model. The results of the soft sensor show a good capability of prediction and accuracy compared to the rigorous model, even using less input process variables.

Keywords: Multicomponent distillation columns, Key components selection, Soft sensor.

1. Introduction

Modern chemical processes have challenging profitability targets and require tight control on their product quality in real time. For that, accurate measurements of critical variables at fast sampling rates are of utmost importance. Usually, the disadvantage of using hardware sensors is either technical (e.g., with the time delay from gas chromatographs) or economic (a large number of online variables leads to high cost device sensors). Soft sensors enable the monitoring of important variables such as product compositions (dependent variables) from other easily measured variables (independent variables) such as temperature and pressure, provided by a mathematical model (Fortuna et al, 2005, Grbić et al. 2013, Jian et al. 2017). Evidently, the quality of this model is critical to the soft sensor performance.

To overcome online measurements difficulties, many soft sensor development techniques have been investigated. Model-driven soft sensors are based on first principle models (FPMs) and require a deep knowledge of the process. Its main advantage is its accurate extrapolation to different operating points (Luyben, 2012). However, building a rigorous model of a complex process can be time consuming. For this reason, industry has explored data-driven soft sensors more often in recent years.

Many authors have developed soft sensors using multivariate statistical methods, such as Principal Components Analysis (PCA), Partial Least Squares (PLS), Artificial Neural Networks (ANN) and Neuro-fuzzy systems (NFS) (Kadlec et al., 2011). However, these methods usually need a large set of past measurements, selection of appropriate inputs and quality assessment techniques to account for outliers, co-linearity and missing data (Rani et al., 2013; Grbić et al. 2013).

In this work, we present a new soft sensor for a gas multicomponent distillation column. This soft sensor is built by simplifying the rigorous model of a distillation column by two equilibrium stages and a strategy for the selection of the light key and the heavy key components. Based on the process knowledge, this method also combines an FPM with heuristic rules. To ensure robustness, we propose a methodology for correcting input temperatures. The model is tested with a debutanizer column and its results are compared with the analysers' outputs.

2. Description of the Problem

In natural gas processing plants (NGPPs), the recovery of sales gas and Liquefied Petroleum Gas (LPG) from natural gas uses several stages of separation units. The composition of natural gas from oil wells is usually majorly composed of methane and heavier hydrocarbons. The sales gas is obtained in the top product of the demethanizer column (primarily with methane and ethane) and the bottom product is a mixture of liquid hydrocarbons. This mixture is then processed in the debutanizer, where the LPG (primarily with propane and butanes) is separated from the natural gasoline (C_5^+).

The LPG composition requirements vary according to local regulations. In this work, we focus on monitoring the concentration of C_5^+ in the LPG product, whose maximum value is 2 % mol of total pentanes, and the Reid Vapor Pressure (RVP) of the gasoline product, with a maximum value of 15 psi (103.421 kPa). Therefore, the soft sensor needs to provide frequent estimations of these variables.

Figure 1 shows a schematic representation of the gas processing plant, highlighting the five regulatory control loops that mostly affect the variables of interest.

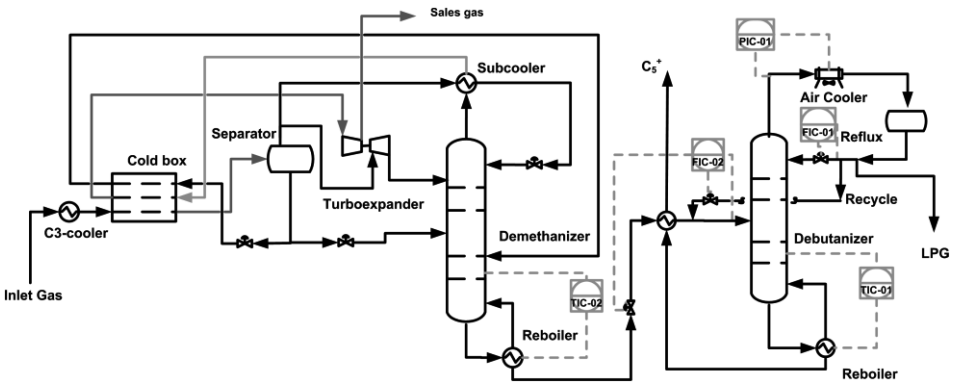


Figure 1. Simplified flow sheet representing the NGPP considered in this work.

In order to evaluate the estimation performance of the product composition, we developed and simulated a full FPM for the process represented in Figure 1 using Aspen Plus Dynamics®. The soft sensor then used data from such simulation as input.

3. Soft Sensor

3.1. General concept

Soft sensors have been used in distillation columns to obtain accurate online measurements of products compositions. As the product composition is a function of the temperature, methods based on the relationship between the temperature stages and the products compositions have been used to correlate these variables (Rani et al., 2013).

In this study, a Simplified Robust Soft Sensor (SRSS) is proposed to estimate the gas product composition of a multicomponent distillation column during its regular continuous operation. The soft sensor was implemented in EMSO, an Equation-Oriented process simulator. The data used as input to the SRSS is composed by:

- Composition of the debutanizer's inlet gas (C1, C2, C3, i-C4, n-C4, i-C5, n-C5, n-C6, n-C7, n-C8, n-C9), or the bottom product from the demethanizer;
- Temperature and pressure from the top outlet vapour stream.
- Temperature and pressure from the bottom outlet liquid stream.

The SRSS is based on flash models, which combine the vapor/liquid-equilibrium (VLE) condition, $\phi_i^V y_i = \phi_i^L x_i$, with the mass and energy balances, where x_i, y_i are the liquid and vapor compositions of component i, respectively, and the fugacity coefficients ϕ_i^V and ϕ_i^L are determined from the equation of state (for example, SRK or PR equations). These equations and the mass and energy balances have been suppressed due to lack of space.

The Two-Flash Process Model (TFPM) depicted by Figure 2 is used as a surrogate of the distillation column. The top flash unit, representing the column's top stage, is set as in dew point and the bottom flash unit, representing the column's bottom stage, is set as in bubble point. Therefore, inputting the temperatures and pressures for each flash is enough to determine the remaining degrees of freedom: the Heavy Key (HK) and Light Key (LK) flow split ratios. To determine which of the several components should be set as light and heavy keys, we propose heuristic rules based on the process knowledge.

3.2. Selection of light and heavy keys

First, it is necessary to know the key components governing the separation design of the distillation process. The objective of the debutanizer is to output all components lighter than the butanes as the top product, while a small amount of pentanes is admissible. In a regular operating point, it is expected that only butanes and pentanes will be switched between the top and the bottom products. Components lighter than the Light Key (LK) are set to go straight to the top product and the components heavier than the Heavy Key (HK) are set to go directly to the bottom product.

Finally, within both LK and HK, the proportion of each component needs to be in accordance with the product composition. These proportions should reflect the ones typically found in the products. If those are not available, a good starting point is using the proportions from the feed. In our work, we used the LK proportion from the feed but

the HK proportion (nC5/iC5) was tuned to 0.3 to better match the full FPM results. For general applications, these adjustments need to be made on each column. Once they are done, the soft sensor performance is guaranteed. Due to the small number of parameters that need to be adjusted, this is a simple task.

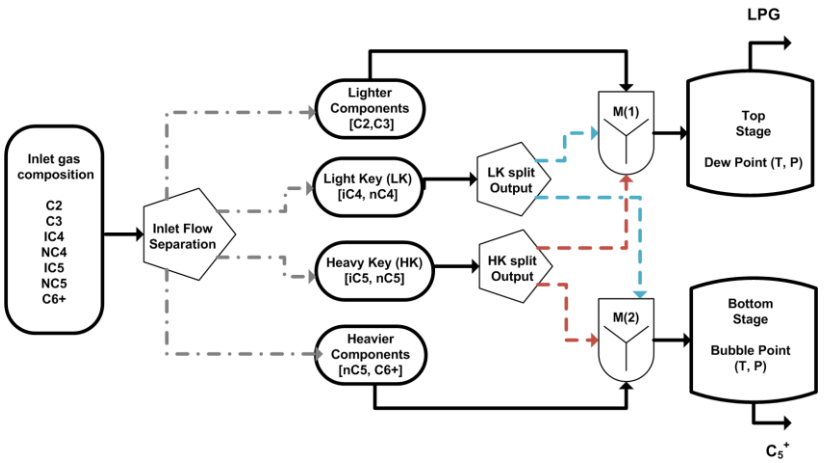


Figure 2. Scheme of the Two-Flash Process Model (TFPM) used for the SRSS.

4. Soft Sensor Robustness

4.1. Concept and implementation

The plant temperatures are subject to noises and transient behaviours that could fall out of the operating range of the TFPM. Hence, the soft sensor input variables need to be systematically corrected to avoid convergence issues in real time. Our concept involves using four copies of the TFPM to calculate temperature limits in parallel with the main TFPM. Each TPFM receives the plant pressure data as input, but assume different LK and HK splitting characteristics to calculate the temperature limit as follows:

1. Bottom Temperature Reference TFPM – Assumes perfect LK/HK split.
2. Top Stage High Temperature TFPM – Assumes that all HK goes to the top product. This is a maximum temperature under regular operation.
3. Top Stage Low Temperature TFPM – Assumes all LK flows to the top product. Only considered if the bottom temperature is above its reference.
4. Top Stage Very Low Temperature TFPM – Assumes all HK flows to the bottom product.

The TFPMs 2, 3 and 4 also receive the plant's bottom temperature as input.

The flowchart in Figure 3 explains how the variables output from each of the TFPMs above are used. Whenever necessary, the top stage temperature will be limited to one of the calculated temperature limits. In the flowchart, BS-Ref, TH-Lim, TL-Lim and TVL-Lim are the outputs of TFPMs 1, 2, 3 and 4, respectively.

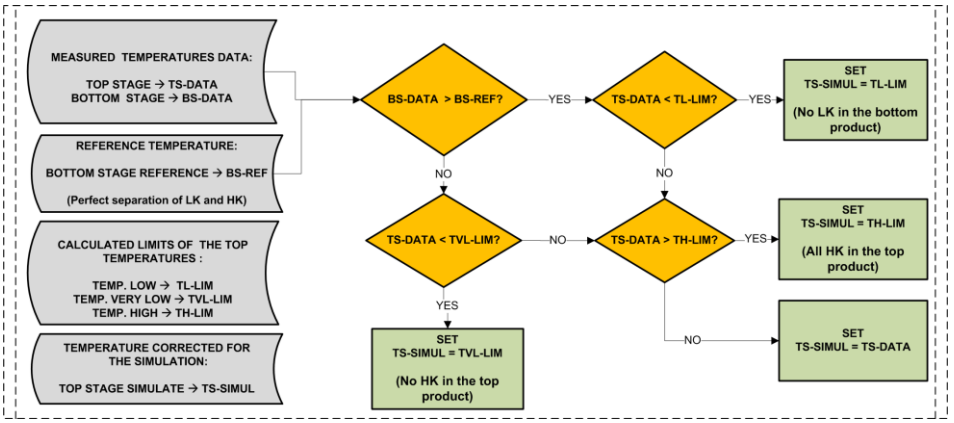


Figure 3. Temperature data correction scheme based on the top stage constraints.

4.2. Testing the soft sensor robustness

Figure 4 displays the results of a robustness test, with the top temperature (TS-Data) occasionally falling below the minimum limit (TVL-Lim). The bottom temperature (BS-Data) from the plant data can be higher or lower than its reference temperature (BS-Ref). In this test, the bottom temperature starts below the reference value. At minute 43, it raises above it, which switches the limit to TL-Lim. The online calculation of new limits ensures a robust estimation at different operations points.

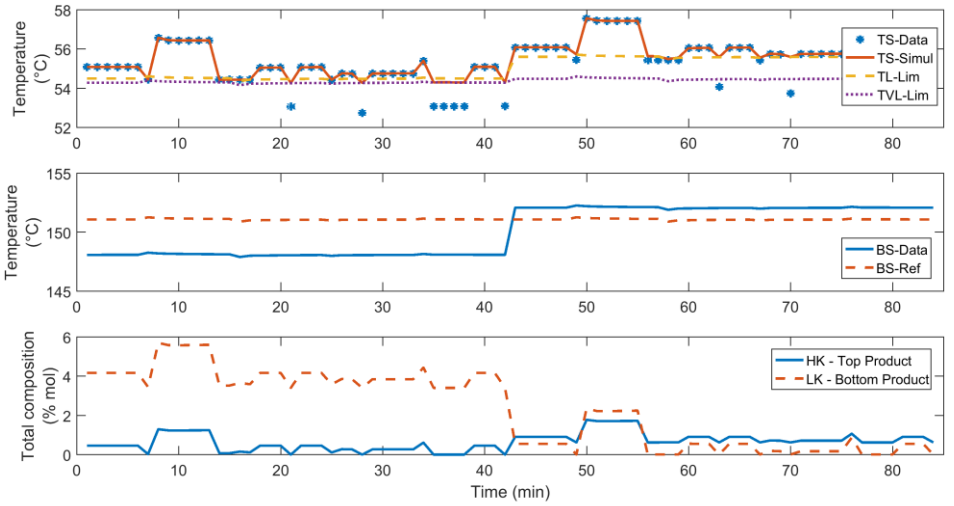


Figure 4. Tests to validate the top stage temperature correction

5. Soft sensor performance

Figure 5 compares the results from the SRSS with the full FPM throughout five step tests in the regulatory control loops. An offset correction was used in both variables inferred by the SRSS, using +0.24 % mol for the C5's composition in the top product and +0.64 psi for the RVP of the bottom product. The SRSS results have high correlation with the full FPM even subject to nonlinearities, as shown by the step test in

TIC-01 (temperature control of the debutanizer column's sensitive tray). The Root Mean Square Error (RMSE) was 0.072 % mol for C5's composition and 0.018 psi for the PVR.

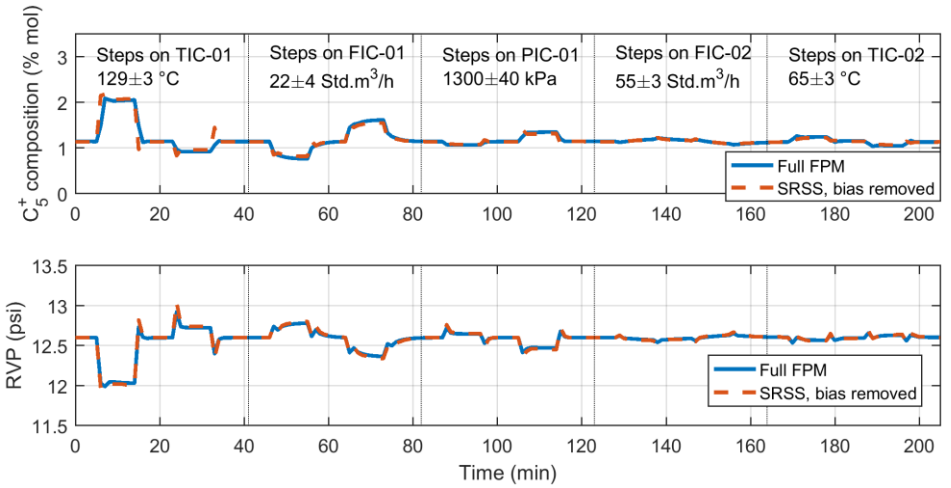


Figure 5. Prediction of total pentanes in the debutanizer overhead product and RVP from bottom product in different step tests in the regulatory controls of distillation units.

6. Conclusions

The proposed inference model accurately predicts critical variables of a multicomponent distillation column. The results indicate good accuracy during variations in the operating point when compared to a rigorous model, despite using fewer input variables. There were no convergence issues in any of the tested points, which is highly desirable for real-time applications. Tests with industrial data are being conducted. Future work includes generalizing the SRSS to other distillation columns.

References

L. Fortuna, S. Graziani & M. G. Xibilia, 2005. Soft sensors for product quality monitoring in debutanizer distillation columns. *Control Engineering Practice*, 13, pp.499-508.

R. Grbić, D. Slišković & P. Kadlec, 2013. Adaptive soft sensor for online prediction and process monitoring based on a mixture of Gaussian process models. *Computers and Chemical Engineering*, 58, pp.84-97.

W. Jian, L. Zhu, Z. Xu & X. Chen, 2017. A variable selection method for soft sensor development through mixed integer quadratic programming. *Chemometrics and Intelligent Laboratory Systems*, 167, pp.85-95.

P. Kadlec, R. Grbić & B. Gabrys, 2011. Review of adaptation mechanisms for data-driven soft sensors. *Computers and Chemical Engineering*, 35, pp.1-24.

W. L. Luyben, 2012. Rigorous dynamic models for distillation safety analysis. *Computers and Chemical Engineering*, 40, pp.110-16.

A. Rani, V. Singh & J. R. P. Gupta, 2013. Development of softsensor for neural network based control of distillation column. *ISA Transactions*, 52, pp.438-49.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50106-3

Diagnosis of Gas Leaks by Acoustic Method and Signal Processing

Luiza B. Fernandes, Flávio V. da Silva, Ana Maria F. Fileti*

School of Chemical Engineering, University of Campinas (UNICAMP), Av. Albert Einstein, 500, Campinas, 13083-862, Brazil
frattini@feq.unicamp.br

Abstract

The present work proposes a new application of algorithms of signal processing and machine learning known from the literature to identify, characterize and locate gas leaks through pipelines. Leakages are often dangerous and unpredicted events confronted daily in distributions of fluids. The purpose of this paper is to establish a methodology for the diagnosis of gas leaks. Among the techniques found in the literature for study of leakage patterns, the investigation of the sound produced was chosen in the present paper. The sound signals were studied in the frequency domain calculated by Fast Fourier Transform. For the identification and classification of patterns contained in the frequency data, two classes of machine learning were used independently: a supervised and an unsupervised learning algorithm. Both algorithms were able to distinguish the diameter of the leak and provide a notion of its location. However, the supervised method demonstrated difficulty in identifying and recognizing events with no leakage. In conclusion, it was possible to identify, characterize and locate leak events, specially using the unsupervised algorithm, with the proposed methodology.
Keywords: leaks, signal processing, acoustics

1. Introduction

The transportation of gas, as well as that of liquids and oils, is conducted primarily through pipelines and a majority of the leakage incidents occur during transportation. These incidents can cause property damage, ecological imbalance and financial debts. In the field of chemical engineering, a common situation of undesirable leakage involves the loss of reagents before entering the reactor. This hinders the control of the chemical reaction and can cause the entire production to be interrupted by safety measures. Thus, the study of a methodology capable of identifying leakage becomes important to avoid damages whenever leakage is not desired.

There are several works in the literature investigating methodologies to identify leaks. These investigations include principles of mass balance, volume balance, transient pressure behavior analysis, vibration monitoring techniques along the pipeline and, among others, acoustic response analysis. In this work, the purpose of choosing the acoustic method is to have an easy to install and inexpensive method to be investigated.

The acoustic method demands a simple experimental setup. Thus, considering that each event, leakage or absence of leakage, has its own acoustic characteristics, it becomes possible to classify and identify them through the system's sound behavior.

In order to make the identification, two machine learning algorithms were proposed: the Multi-Layer Perceptron (MLP) and the Kohonen neural network. They were

independently analyzed for the ability to identify leaks in occurrence, characterize the leak size and whether they can locate it in the pipeline.

2. Material and Methods

The data acquisition apparatus consisted of two acoustic sensors coupled to the pipeline (microphone 1 and microphone 2), a preamplifier, an analog-to-digital converter (National Instruments, NI cDAQ-9178) and a computer for storing the data. A recommended book on data acquisition and signal processing is (Drongelen, 2007).

The acoustic data were studied in the frequency domain calculated by Fast Fourier Transform (FFT). However, to reduce computational efforts in the system identification process, data compression was performed by Principal Components Analysis (PCA). As reported by Fernandes et al. (2016), PCA can be used for dimensional data compression preserving more than 97% of the original information.

Thus, the two proposed algorithms are evaluated in relation to their ability to identify the system. Neural networks have the ability to learn and improve performance through training following their learning algorithm, which can be classified as supervised or unsupervised. The MLP algorithm belongs to the supervised class, which means that input and output data must be provided for network training. The result of this network is numeric. The fundamental theory of neural networks is found in the book (Haykin, 1994).

Meanwhile, the Kohonen algorithm belongs to the unsupervised learning algorithm class. It is a self organizing map with the ability to dimensionally organize complex data into clusters (Kohonen, 2001). During training, only input data is required. The result of this training is visualized in the form of u-matrix (unified distance matrix) normally in 2-D. The u-matrix is a map of the patterns found in the input data (Vesanto et al., 2000). In the u-matrix, Euclidian distances are represented by a color scale and patterns are mapped according to their similarities.

Finally, to test the result of both networks, a set of data not present in the training is used. The quality of the neural network is considered good if the test result coincides with the training result and corresponds to the experimental values.

2.1. Data Acquisition and Signal Processing

The parameters for data acquisition were virtually configured in the LabView software, such as the sampling frequency at 30,000 Hz. Signal processing was also performed in this software, such as the realization of the Fast Fourier Transform and the use of a digital filter (to disregard frequencies above 10,000 Hz). However, the data compression and the analysis of the machine learning algorithms were performed using Matlab software.

2.2. Experimental setup

The experimental setup used compressed air as the fluid for the leak simulations. The pipeline was made of copper ½ inch in diameter. In all, it was 53 meters long and comprised thirty 90˚ elbows. For the simulations of the leaks, seven holes were drilled in different positions along the pipeline, they measured from 0.5 to 4 millimeter of diameter, as seen in Table 1. Also in Table 1, it is observed that ten different simulations were performed in total: three without leak, and one leakage simulation in each drilled hole. Each data acquisition lasted fifty seconds.

Table 1. Experimental Simulations, hole positions and diameters

Experiment number	Simulation	Hole name	Distance to gas inlet (m)	Hole diameter (mm)
1	No leakage; “nl”	-	-	-
2	No leakage, beats in the gas cylinder; “sc”	-	-	-
3	No leakage, beats in the pipeline; “sp”	-	-	-
4	Leakage in orifice 1	F1	0.2	1.0
5	Leakage in orifice 2	F2	1.0	0.5
6	Leakage in orifice 3	F3	1.6	2.0
7	Leakage in orifice 4	F4	2.4	1.0
8	Leakage in orifice 5	F5	3.4	4.0
9	Leakage in orifice 6	F6	4.8	3.0
10	Leakage in orifice 7	F7	5.6	0.5

Experiments 2 and 3 in Table 1 added sounds to the system that were not generated by leaks, but by beats in a gas cylinder and by beats in the pipeline, respectively. Microphone 1 (m1) is coupled to the pipeline at 0.1 m from the gas inlet, while microphone 2 (m2) is located at 19.2 m. Then, with the values in Table 1, m1 is closer to the holes than m2. Microphone 1 captures the sound signals of leakage within a distance of 5.5 m in all simulations. Meanwhile, microphone 2 from a distance of 13.4 m to 19 m.

2.3. Neural Networks Configurations

Both networks were tested for their ability to estimate the leakage dimension and to locate it. For the supervised network, the identification response (output) must be numeric. To identify the size of the leak, the leakage diameter in millimeters was chosen as the network response and, in situations without leakage, the output of the network would be zero, “0”. And to identify the location of the leak, the network response was chosen as “1” if the patterns belong to m1 and “2” if the patterns correspond to m2.

For the unsupervised net, a quadratic map was created and visualized as u-matrix. No output value is required for training. In both networks, the compressed frequency data (Fernandes et al., 2016) were used as input for training and testing.

3. Results

Both microphones were in range of distance capable of capturing the acoustic changes in the system in all experimental simulations. The sound response was evaluated in the frequency domain and compressed in terms of principal components. From the experimental observations points, 1,000 points were used to train the networks and 46, to test.

The results of the supervised network training is shown in Figure 1. The identification of the leakage diameter is seen to the left (A) and the identification of the microphone to the right in (B).

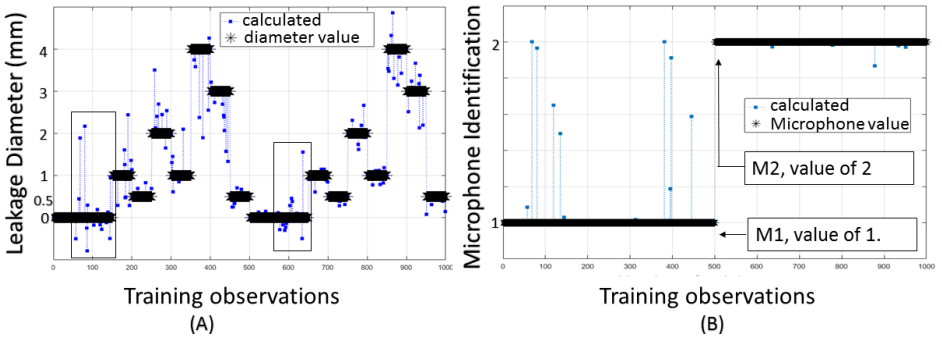


Figure 1. Values calculated by the supervised network and the observations used to train the network. (A) Leakage diameter value; (B) Microphone value

In Figure 1 (A), it is possible to notice that a large part of the calculated values are close to the diameter value. However, uncertainty specially around the value of zero does not facilitate the evaluation of events without leakage, situations in evidence by rectangles. Figure 1 (B) represents the identification of the microphone that is capturing the signal. If the distance of the captured signal to the leak hole is within a range of approximately 5.5 m, the network must recognize the first microphone, m1. If it is farther away- within 13.4 m to 19 m- it must recognize microphone 2. Figure 1 (B) showed good agreement for most of the calculated and expected values.

The results of the supervised network test are shown in Figure 2. In Figure 2 (A), two situations were used for test: observations with pipe strikes and leakage in F1, but different from those used in the network training. The expected results for these situations were zero "0" (no leakage) and one "1" (leakage diameter of 1 mm), respectively. However, in the situation without leakage, in evidence by rectangle, the calculated values are not consistent and vary from zero. Still in Figure 2 (A), for the test with the leak situation of 1 mm diameter, the network responses are approximately "1", as expected.

The represented test in Figure 2 (B) showed good results, the calculated values matched the expected values. For test observations captured by microphone 2, the value "2" was correctly calculated by the network. And, by microphone 1, the value "1", as expected.

Alternatively, Figure 3 represents the results of the unsupervised network. It represents the u-matrix resulting from the mapping of the patterns found in the frequency data.

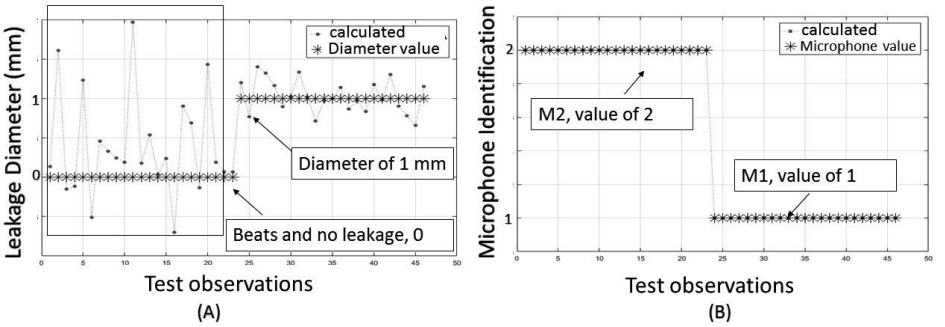


Figure 2. Values calculated by the supervised network and the observations used to test the network. (A) Leakage diameter value; (B) Microphone value

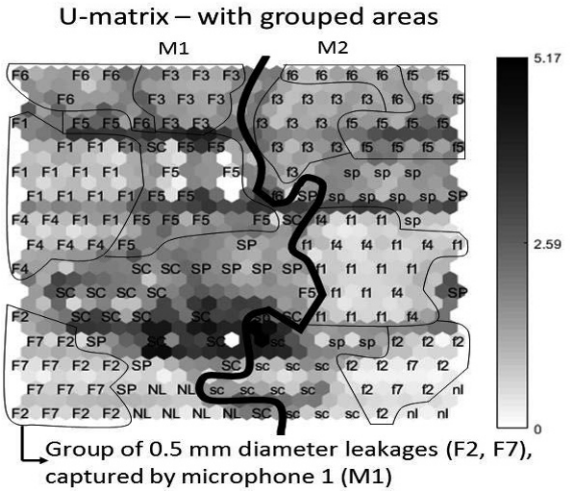


Figure 3. The self organizing map in u-matrix

In the resulting self-organized map in Figure 3, the areas corresponding to a given pattern are exposed and grouped according to their similarities. As in emphasis in Figure 3, in the same group there are "F2" and "F7" meaning areas activated by leakages of 0.5 mm. The same happens with "F1" and "F4", leaks of 1 mm. In addition, capital letters indicate that the signals come from microphone 1, they are seen mostly on the left side of the map. In contrast, lowercase letters indicate that they are from microphone 2 and are seen mostly on the right side of the map. The ticker line divides these two sides.

The simulations without leaks, with and without beating simulations, were not attributed to any explicit group, but are indicated in the map by "nl", "sc" and "sp", as in Table 1.

Then, examining Figure 3, the map created by the Kohonen algorithm can differentiate characteristics of different diameters and different distances between the source of leakage and the microphone.

The same test performed in the supervised network was repeated to the unsupervised one. In Figure 4, the self-organizing map is shown containing circles in the areas corresponding to the closest patterns found in test data. In Figure 4 (A), the test data used corresponds to observations with strikes in the pipeline. As a result, twelve areas were circled accused of having a pattern close to this tested data: six of strikes in pipeline ("sp"), three of absence of leakage ("nl"), one is not representing any feature, one with 1 mm ("f1") and one with 0.5 mm ("F2") leakage. The last two were mistakenly activated, but most of the activated areas are correctly informing that no leakage is occurring.

Figure 4 (B) depicts the second data test considering only 1 mm leakage acquired by microphone 2, m2. All the activated areas were circled. A total of 14 areas were activated: almost all are representing "f4" or "f1"- leakages of 1 mm. It is therefore correctly stating that the mapped pattern most similar to the test data is a 1 mm diameter leak. In addition, it is noted that these circled areas are on the right side of the map, also verifying that the test data have characteristics close to the data captured by m2.

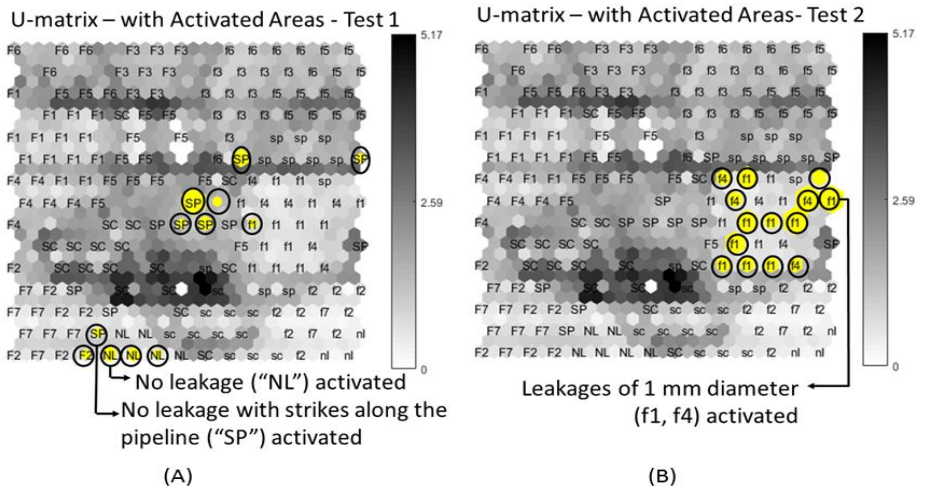


Figure 4. Self-organized map with test results in u-matrix, areas with patterns similar to test data are circled. Test data with (A) beats in the pipeline (B) 1 mm leakage

4. Conclusions

The proposed algorithms were able to identify leaks and characterize it. They also distinguished signals from one microphone to another, proving that the algorithms were able to differentiate characteristics in the signal caused by different distances from the leakage to the acoustic sensor. However, the unsupervised map was more accurate and has the advantage of not needing a numeric output to train. In the generated map, areas of similar events can be grouped and will be activated by test observations that contain characteristics similar to that group.

To identify situations without leakage but with beats in the simulation, the supervised net provided an uncertain response. On the contrary, the same test was performed on the unsupervised map, and the identification that it was an event without leakage was more accurate.

Then, it can be concluded that the signal processing performed in this work as well as the networks used to identify the system were successful in the identification. However, the unsupervised algorithm was found to be a more adequate choice.

References

W. van Drongelen, Signal Processing for Neuroscientists, First Edition, Elsevier, 2007.

L. B. Fernandes, R. B. Santos, A. M. F. Fileti, Principal Component Analysis in Multivariate Acoustic Response of Microphones Coupled to Metal Pipeline in the Presence of Simulated Leakage of Compressed Air, IASTED-MSI, DOI: 10.2316/P.2016.840-020, 2016.

S. Haykin, Neural Network A Copmprehensive Foundation, Upper Saddle River, NJ: Pretence Hall, 1994.

T. Kohonen, Self Organizing Maps, Third Edition, Berlin: Springer, 2001.

J. Vesanto, J. Himberg, E.Alhoniemi, J. Parhankangas, SOM Toolbox for Matlab 5 (Libella Oy, Espoo 2000).

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50107-5

Optimal Operation of a Combined Continuous–Batch Process

Cesar de Prada*, Rogelio Mazaeda, Smaranda Podar

Dpt. of Systems Engineering and Automatic Control, University of Valladolid, c/ Real de Burgos s/n, Sede Mergelina EII, Valladolid 47011, Spain
prada@autom.uva.es

Abstract

The paper deals with the optimal operation of processes that are composed by continuous and batch units tightly integrated by means of mass or energy flows. It uses as a reference process a benchmark developed in the EU project HYCON2: a simplified crystallization section of a sugar factory. The aim is to present a novel formulation that allows the coordination of the continuous and discrete processes generating feasible crystallizer schedules that can be easily integrated in a wider optimization scheme. The approach uses formal scheduling methods on continuous time to deal with the shared resources integrating the continuous dynamics.

Keywords: Real-time optimization, process scheduling, integration MPC-scheduling, distributed optimization, sugar crystallizers.

1. Introduction

Decisions about global production and plant–wide operation are increasingly being considered as highly relevant in the decision hierarchy of a factory. Nevertheless, the well-known control pyramid does not fit very well with the reality when the production process includes a combination of continuous and batch processes. In this case, the usual static models and non-linear programming tools of the RTO layer are useless due to the dynamic and discontinuous nature of the problems, which require coordination and scheduling of the operations over time in order to avoid creating bottlenecks, violating constraints or risking the safe operation of the plant.

The problems become much more complex due to their dynamic and discontinuous nature, which leads to large-scale mix-integer optimization problems that must be solved with sensible computing times in order to allow for real-time implementation. The subject is receiving increased attention, not only because of its importance, but also due to the opportunities opened by the integration of different decision layers and the use of improved optimization methods and tools, Harjunkoski, et al. (2014), Nie et al. (2015).

This paper intends to contribute to the advances in the field studying the special plant wide optimization problems that results when continuous plants must operate together with batch ones in series interchanging materials and energy and sharing resources, and presents a novel and efficient formulation of the scheduling problem combining continuous and discrete time domains. The approach is illustrated with an example taken from the sugar industry organized around the interface of the sections of evaporation and crystallization.

The paper is organized as follows: after the introduction, section 2 shows the case study that has been used to illustrate the approach. Then, section 3 describes the formulation of the scheduling problem and, finally, section 4 provides results corresponding to a test on the case study. The paper ends with some conclusions and references.

2. Sugar factory case study

In this section we consider a benchmark problem that was designed to experiment with plant-wide control and optimization strategies. It corresponds to a process of the sugar industry that combines units of continuous and semi-batch type, so that its operation and control requires taking into account continuous and discrete decisions. The model was proposed as a benchmark in the FP7, EU Network of Excellence "Highly-complex and Networked Control Systems" (HYCON2) and it is represented in the Figure 1.

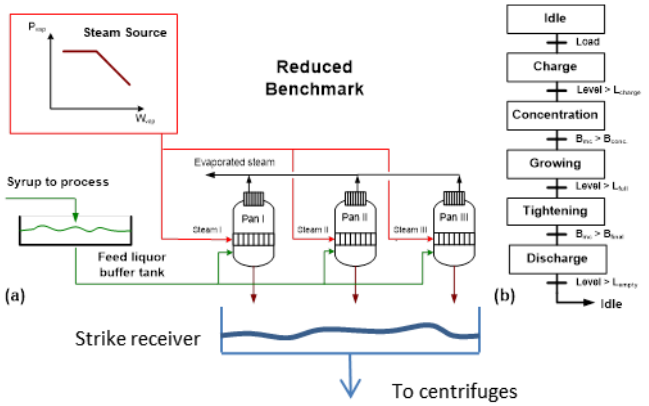


Figure 1. Schematic of the benchmark problem with two buffer tanks and three vacuum pans for sugar crystallization

It incorporates a buffer tank for supplying syrup to three semi-batch crystallizers that can operate in parallel following a certain recipe to produce sugar crystals. The operation in the crystallizers, also called "tachas" or vacuum pans, follows a cycle with several stages as in Figure 1(b), and uses steam from the evaporation section and syrup from the tank, which are shared resources. When the crystals reach the required size and the crystallizer is full, the strike ends and the resulting massecuite is discharged to the strike receiver tank. The target is to schedule the crystallizers optimally respecting the constraints in the shared resources: total amount of steam and level in the tanks, in spite of changes in the supply syrup.

3. Scheduling and coordination of the batch and continuous units

The joint operation of the crystallizers and the supply and strike receiver tanks can be formulated as a scheduling problem where one has to decide on the starting times of the crystallizers over time while maintaining the shared resources (syrup in the supply tank, massecuite in the strike receiver tank and total steam demand from the crystallizers) within admissible ranges at all times.

The approach we present in this paper to formulate the scheduling problem looks for computational efficiency in order to be applied in real-time, and enough flexibility as to

be adapted to the real process in a sensible way. This requires a linear mix-integer formulation in spite of the presence of clear non-linear profiles of many variables as well as large horizons that must cover at least a factory shift. With these aims in mind, the formulation of the problem combines three elements:

- An assignment problem, described in continuous time, to compute the scheduling of the crystallizers.
- A description of the non-linear profiles of the variables associated to the operation of the crystallizers in terms of a set of local linear approximations.
- A description of the dynamic constraints imposed by the shared resources using a discrete time grid synchronized with the continuous one.

The reason it uses two base times is linked to the fact that assignments in continuous time provide information only on the start and end times of each batch unit, but some shared continuous variables needs to check the fulfilment of its associated constraints at regular and more frequent time intervals.

Each crystallizer will process a certain number of batches of syrup from the supply tank. We will denote by I a set of indexes i, each one representing a batch or processed syrup lot. In the same way, we will denote by M the set of crystallizers m. The continuous inflow of syrup to the supply tank can be considered as composed of a series of syrup lots arriving to the tank to be processed later on. The size of each lot is equal to the amount of syrup consumed by a crystallizer for a whole batch. Depending of the characteristics of the syrup, we can associate to each lot a processing time or duration in the crystallizer that can be different in each crystallizer and that will be denoted as p(i,m). In parallel, we can assign a cost associated to the processing of a syrup lot i in a crystallizer m, denoted by c(i,m). In addition, the scheduling problem uses the following continuous time variables:
ts(i) denotes the start time of lot number i (continuous variable)
te(i) denotes the earliest time in which lot i can be processed
td(i) denotes the latest time in which lot number i should have been processed

Note that te and td are known parameters at time zero, when the scheduling problem has to be solved over the prediction horizon. For all lots present in the supply tank according to its capacity, te will be zero, as they can be processed in parallel, while the rest will have different values of te according to their expected arrival times. As well, two binary variables are defined following a general precedence approach:
y(i,m) will be equal to 1 if lot i is processed in crystallizer m and o otherwise
z(i, j) will be equal to 1 if lot i precedes lot j on the same crystallizer and zero otherwise

Then, the optimal allocation of lots and its sequencing within each crystallizer can be formulated easily if we select as optimality criterion to minimize the overall costs by means of equations (1). Here the first equation guarantees that only cost of each effective assignment is counted in the cost function. The first two sets of constraints assure that a lot I will not be assigned before the time te(i) and will be processed before the time td(i). The following equation guarantees that each lot will be processed in a crystallizer and only in one of them. Next equation serves to limit the number of lots that can be assigned to a single crystallizer. Finally, the first of the two final equations assures that if two lots, i and j, are assigned to the same crystallizer m, then either i precedes j or j precedes i. Then, the last equation stablish that if i precedes j on

crystallizer m, the start time ts of j cannot be earlier than the start time of I plus its processing time.

$$
\begin{aligned}
&\min_{y,tz} \sum_{i\in I}\sum_{m\in M} c(i,m)\,y(i.m) \\
&s.t. \quad ts(i) \geq te(i) \\
&\qquad ts(i) \leq td(i) - \sum_{m\in M} p(i,m)\,y(i,m) \quad \forall i \in I \\
&\qquad \sum_{m\in M} y(i,m) = 1 \quad \forall i \in I \\
&\qquad \sum_{i\in I} y(i,m)\,p(i,m) \leq max_i\{td(i)\} - min_i\{te(i)\} \quad \forall m \in M \\
&\qquad 1 \geq z(i,j) + z(j,i) \geq y(i,m) + y(j,m) - 1 \quad \forall i,j \in I,\ i > j,\ m \in M \\
&\qquad ts(j) \geq ts(i) + \sum_{m\in M} p(i,m)\,y(i,m) - M(1 - z(i.j)) \quad \forall i,j \in I,\ i \neq j
\end{aligned}
\tag{1}
$$

The above formulation can be used for assignment but it does not consider neither the constraints associated to the shared resources (syrup, steam and massecuite) nor the ones of the continuous elements (supply tank and strike receiver). Because of that, it needs to be expanded, first with a representation of the time evolution of flows associated to the shared resources and, then with the corresponding level or range constraints.

For the representation of the profile of the demand of syrup made from a crystallizer over time, as well as for the demand of steam and the discharge flow of a crystallizer, we define a set of time points tp(n) covering from $-t_N$ to t_N and approximate the continuous profile by a series of linear segments, each one with an associate binary variable w, as in Figure 2. Outside the range of the operation, the profile is fixed to zero.

This non-uniform discrete time base is linked to each batch and needs to be synchronized with the overall time scale. Also notice that the times ts(i) are only a relatively small number and likely not equally spaced over time, so that they are not adequate to check if a variable like the level in a tank is within the desired range. For these reasons, we need to define a third time base that consists of a larger set of time points t(k) equally distributed over the future horizon.

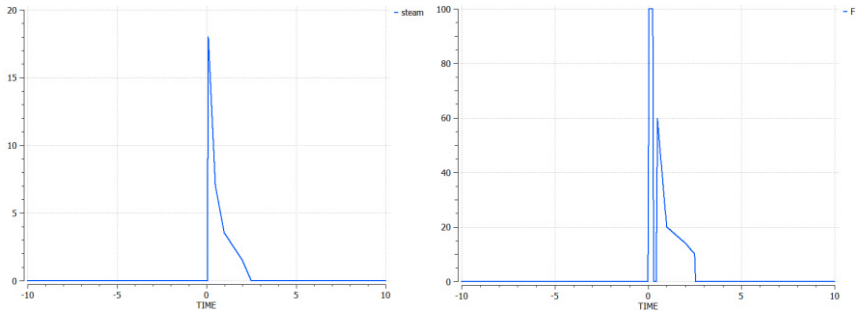


Figure 2. Profiles of the steam flow (left) and syrup demand (right) over the operation of a crystallizer approximated by piecewise linear segments at different time points tp(n)

For the synchronization of the different times, the following variables have been defined:
w(k,i,n) is a binary variable equal to 1 if the time instant t(k) – ts(i) belongs to the interval n of the corresponding profile and 0 otherwise.
α(k,i,n) is a coefficient between cero and one used for linear interpolation in the n piecewise profiles for each time instant t(k) and each crystallizer batch i

The following set of equations can be used then for computing the total flow of syrup Q(k), steam Qs(k) and discharge to the strike receiver Qd(k) at a given time instant t(k) by synchronizing the corresponding profiles and the start times and interpolating later on according to the particular value of the elapsed time.

$$\begin{aligned}
&\sum_n w(k,i,n) = 1 \qquad w(k,i,N) = 0 \\
&\sum_n \alpha(k,i,n) = 1 \qquad \alpha \geq 0 \\
&\alpha(k,i,n) \leq w(k,i,n-1) + w(k,i,n) \\
&t(k) - ts(i) = \sum_n \alpha(k,i,n) tp(n) \qquad \forall i \in I, \forall k \in K \\
&Q(k) = \sum_i \sum_n \alpha(k,i,n) F(i,n) \\
&Qd(k) = \sum_i \sum_n \alpha(k,i,n) Fd(n) \\
&Qs(k) = \sum_i \sum_n \alpha(k,i,n) Fs(i,n)
\end{aligned} \tag{2}$$

The first equation stablish that one and only one interval n of the profile can be active in a time instant k and a batch i. The following equations identify the interval n and the fraction of time α of the interval that corresponds to the time relative to the beginning of the batch i. The last set of equations computes the corresponding flows interpolating with that fraction α in the profiles. Notice that these can be particularized according to the characteristics of the syrup lot I being processed. Finally, once the global flows at the synchronization points k have been computed, mass balances in the tanks can be formulated and the constraints associated to levels or total steam demand are added to the model:

$$\begin{aligned}
&L(k) = L(k-1) + (q(k) - Q(k))\Delta t / A_t \\
&Lm(k) = Lm(k-1) + (Qd(k) - qc(k))\Delta t / A_m \\
&L_{max} \geq L(k) \geq L_{min} \qquad \forall k \in K \\
&Lm_{max} \geq Lm(k) \geq Lm_{min} \\
&Qs_{max} \geq Qs(k) \geq Qs_{min}
\end{aligned} \tag{3}$$

Here L(k) stands for the level of syrup in the supply tank, Lm for the corresponding level in the strike receiver, q(k) represents the incoming syrup flow to the tank and qc(k) the centrifuges demand to the strike receiver.

The resulting problem is a MILP one that can be solved with the appropriate algorithms, CPLEX in our case in 0.06 s. in the considered case study that involves 3252 equations 4439 variables, 2170 of them binaries.

4. Scheduling

The above approach has been tested in the case study using a variable syrup inflow to the supply tank as given in Figure 3 (left), which presents a strong increment after two hours and another decrement after five hours of operation. The capacity of the syrup tank is 77 m^3 and the one of the strike receiver 90 m^3. The level is constrained to operate in a range [10% , 90%] and the total maximum demand of steam is limited to 27 t/h. The problem considers six future lots and assumes that crystallizer 1 has started its operation one hour before the current time. Results of the corresponding scheduling can be seen also in Figure 3 (right) where the starting times of each batch of every crystallizer are also shown.

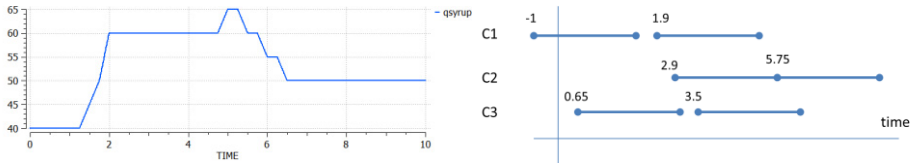


Figure 3. Time evolution of the incoming syrup to the supply tank in m^3/h

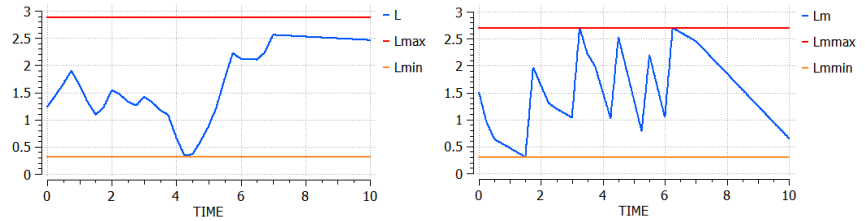


Figure 4. Time evolution of the levels in the supply tank (left) and strike receiver (right) with the corresponding upper and lower limits.

In the same way, Figure 4 shows the time evolution of the levels that are maintain within their allowed ranges in spite of the strong changes in the supply flow.

5. Conclusions

A formulation for a challenging problem involving the joint operation of continuous and batch processes have been presented. The approach uses MILP to obtain optimal schedules, providing solutions in 0.06 sec. using CPLEX and can be integrated in a predictive optimization environment.

Acknowledgements: The authors are grateful to the Spanish Government and EU FEDER for their support through project DPI2015-70795

References

Harjunkoski, I.; Maravelias, C. T.; Bongers, P.; Castro, P. M.; Engell, S.; Grossmann, I. et al. (2014) Scope for industrial applications of production scheduling models and solution methods. *Computers & Chemical Engineering*, 62, 161 – 193.

Mazaeda, R., Acebes, L.F., Rodríguez, A., Engell, S., Prada, C. (2014). Sugar Crystallization Benchmark. *Computer Aided Process Eng. series*, vol. 33, pp. 613-618.

Nie Y., Biegler L.T., Villa C.M., Wassick J.M. (2015) Discrete Time Formulation for the Integration of Scheduling and Dynamic Optimization, *Ind. Eng. Chem. Res*., 54 (16), pp 4303–4315

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50108-7

Variation and Risk Analysis in Tablet Press Control for Continuous Manufacturing of Solid Dosage via Direct Compaction

Qinglin Su[*, a], Yasasvi Bommireddy,[b] Marcial Gonzalez,[b]
Gintaras V. Reklaitis,[a] Zoltan K. Nagy[*, a]

[a] *Davidson School of Chemical Engineering, Purdue University, West Lafayette, IN 47907, US*
[b] *School of Mechanical Engineering, Purdue University, West Lafayette, IN 47907, US*
su130l@purdue.edu; znagy@purdue.edu

Abstract

A continuous rotary tablet press is a multi-stage process with many punch stations running in parallel, in which each punch undergoes the following steps: die filling and metering, pre-compaction, main-compaction, tablet ejection, and tablet take-off from lower punch. Process uncertainties or disturbances within a punch station or among stations in the tablet press are a major source of variation in final product quality attributes, e.g., hardness, weight, etc., which in turn imposes challenges for the real-time release in pharmaceutical continuous manufacturing of solid dosage. In this study, the direct compression line at Purdue University was investigated and a Natoli BLP-16 tablet press was used to characterize powder compressibility, system dynamics and variation, as well as the interaction effects on process control development. The compressibility of tablets made from a blend of Acetaminophen (API), Avicel Microcrystalline Cellulose PH-200 (excipient), and SiO_2 (lubricant) was found to be largely independent of tableting speed. By contrast, filling depth or dosing level, turret speed, feed-frame speed, and compression force were interacting and significantly affected the die-filling process and the final product quality attributes. Thus, the design of the process control structure plays an important role in reducing process and product quality variations. A hierarchical three-level control design was proposed and evaluated, consisting of Level 0 Natoli built-in control, Level 1 decoupled Proportional Integral Derivative (PID) cascaded control loops for tablet weight and production rate control, and Level 2 advanced model predictive control. Process variations, e.g., in powder bulk density changes, during continuous steady-state operation were also investigated. Finally, a risk analysis of the effects of the process dynamics on variation on the product quality control was briefly discussed and summarized.

Keywords: continuous manufacturing; tablet press; process control; risk analysis.

1. Introduction

A real-time release strategy is essential to the pharmaceutical manufacturing industry as it shifts from batch to continuous processing. Such a strategy requires effective on-line process monitoring and control of pharmaceutical critical process parameters (CPPs) and critical quality attributes (CQAs) so as to achieve steady-state and consistent production of pharmaceutical ingredients, intermediates, or final products. Systematic frameworks for the development and implementation of process monitoring and control

strategies in pharmaceutical continuous manufacturing have been extensively discussed in the past decade (Singh et al., 2014); however, reports of practical and thorough studies of the implementation and evaluation of monitoring and control systems in a physical pilot plant or manufacturing facility are still limited in the open literature. Specifically, the dynamic performance and risk analysis of the continuous manufacturing process in a control closed-loop operation are rarely reported (Bhaskar et al., 2017). In this study, our particular interest was focused on the control performance of a continuous rotary tablet press in the Purdue University pharmaceutical continuous manufacturing pilot plant. The continuous rotary tablet press is a multi-stage process, in which each station undergoes the following major steps: die filling and metering, pre-compaction, main-compaction, tablet ejection, and tablet take-off from lower punch. Process uncertainties or disturbances within a punch station or among stations in the tablet press are a major source of variation in final product quality attributes, e.g., hardness, weight, etc., which in turn challenges the real-time release decisions in continuous manufacturing of solid dosage. The Natoli BLP-16 tablet press was investigated within the direct compression line to characterize powder compressibility, system dynamics and variation, as well as their effects on process control development.

2. Process variation and characterization

The challenges of maintaining consistent product quality in the output of a tablet press are to identify the appropriate CQA and CPP variables, and to understand the effect of their variations on the target product profiles (TPPs). These critical-to-quality variables for the given system were identified as tablet weight, relative density, strength (hardness) and compaction force. For example, the weight of the tablet ultimately determines the API potency within a dose. Further, it also determines the main compression force at the set punch displacement and thus the relative density and tensile strength of the tablet, which in turn affect the final product attributes such as dissolution rate. Inline process characterization of the output of the tablet press was conducted using the Sotax AT4 automatic tablet testing system for a powder formulation of 10% Acetaminophen (API), 89.8% Avicel Microcrystalline Cellulose PH-200 (excipient) and 0.2% of SiO_2 (lubricant) by mass.

Tablet weight is the desired outcome of the die filling station, whose variation could be due to the changes in powder flowability and die filling time. It was found that powder bulk density is responsive to the variations in blend composition, moisture content, and temperature, which also result in variation in its flowability. Hence, the tablet weight variation was characterized by the following equation,

$$W_t = \tilde{\rho}_b \frac{\pi D^2}{4} DP(1 - cTS) \tag{1}$$

where W_t is the tablet weight, $\tilde{\rho}_b$ is the powder bulk density, D is the diameter of the die, DP is the dosing position, TS is the turret speed, and c is an efficiency parameter of powder flowing into the dies from the feed-frame.

Tablet relative density is the ratio of tablet bulk density to the powder true density, which can be calculated from tablet weight, as shown below and verified in Figure 1(a),

$$\rho_r = \frac{4W_t}{\pi D^2 t \rho_t} = \frac{\tilde{\rho}_b DP(1 - cTS)}{t\rho_t} \tag{2}$$

where ρ_r is the tablet relative density and t is the in-die tablet thickness and ρ_t is the known powder true density.

Tensile strength is a measure of tablet hardness which is independent of tablet dimensions. It can either be measured using a destructive test which fractures the tablets in between two platens (e.g., Brazilian test) or it can be estimated using the Leunberger equation (Kuentz and Leunberger, 2000), as described below,

$$\sigma_{t,m} = \frac{2F}{\pi D \bar{t}} \tag{3}$$

$$\sigma_{t,p} = \sigma_{max}\left[1 - \left(\frac{1-\rho_r}{1-\rho_{c,\sigma}}\right)\mathrm{e}^{(\rho_r - \rho_{c,\sigma})}\right] \tag{4}$$

where $\sigma_{t,m}$ is the tensile strength measured from the Brazilian test, $\bar{t}$ is the out-of-die tablet thickness, and F is the breaking force. $\sigma_{t,p}$ is the predicted tensile strength from the tablet relative density as in Eq. (2), $\rho_{c,\sigma}$ is the critical density of strength, and σ_{max} is the tensile strength at full density, i.e., at $\rho_r = 1$.

Powder compressibility is characterized by the relationship between main compression force CF during compaction and the resulting tablet relative density ρ_r. It is one of the critical material attributes (CMAs) that affect consistent tablet manufacturing. There are many models characterizing the compressibility of a powder at various degrees of compression. In this study, the Heckel relationship was employed to characterize the main compression force at high tablet relative density, as given below and verified in Figure 1(b),

$$\frac{CF}{1-\frac{\rho_c}{\rho_r}} = \frac{CF}{a} + \pi\frac{\frac{D^2}{4}}{ab} \tag{5}$$

where ρ_c is the critical density during compaction, parameters a and b are interpreted as the maximum degree of compression and the reciprocal of the pressure applied to attain the maximum degree of compression, respectively.

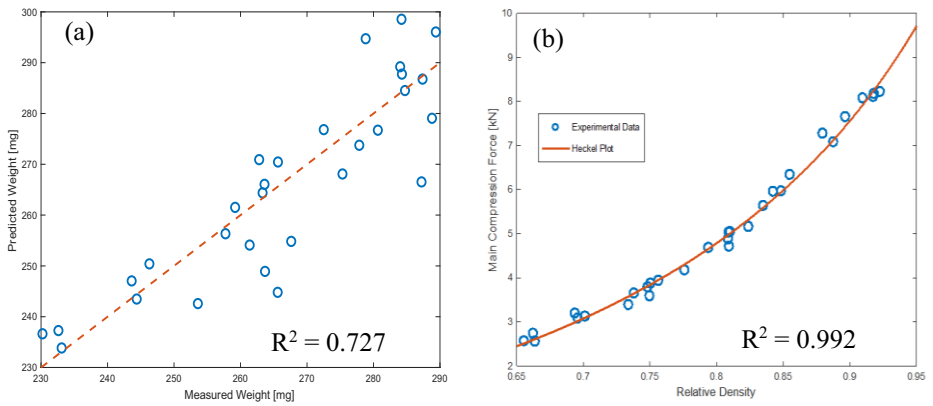


Figure 1. Comparisons between predicted and measured tablet weight (left) and main compression force (right).

3. A hierarchical control development

A hierarchical three-level controller design was implemented on the Natoli tablet press following to the previously proposed systematic framework for process control design and risk analysis in continuous pharmaceutical solid-dosage manufacturing (Su et al., 2017). The framework consists of a series of steps involving system identification, control design and analysis, hierarchical three-level control, risk mapping, accessing and planning, performance evaluation, etc. The Natoli BLP-16 tablet press has a built-in programmable logic control (PLC) system to manipulate the process parameters of fill depth, turret speed, and feeder speed, which are regarded as providing Level 0 control in this context. In light of the above process variation characterization, the system identification step using state-space model and the control design and analysis step using classical control metrics, e.g., Condition number, Relative Gain Array (RGA) analysis, etc., suggested a Level 1 control with decoupled PID control loops for a cascaded control of tablet weight, production rate, and main compression force by manipulating the set points of fill depth and turret speed at the Level 0 control. The identified state-space model was further employed to develop the Level 2 linear model predictive control (MPC) scheme for the tablet press, in which the main compression force was constrained and monitored as it is closely related to the tablet CQAs of hardness, tensile strength, and dissolution rate. The Emerson DeltaV Control Studio and DeltaV Predict toolbox were utilized for Level 1 and 2 control implementations. Details of control loops and operating panels can be found in Figure 2.

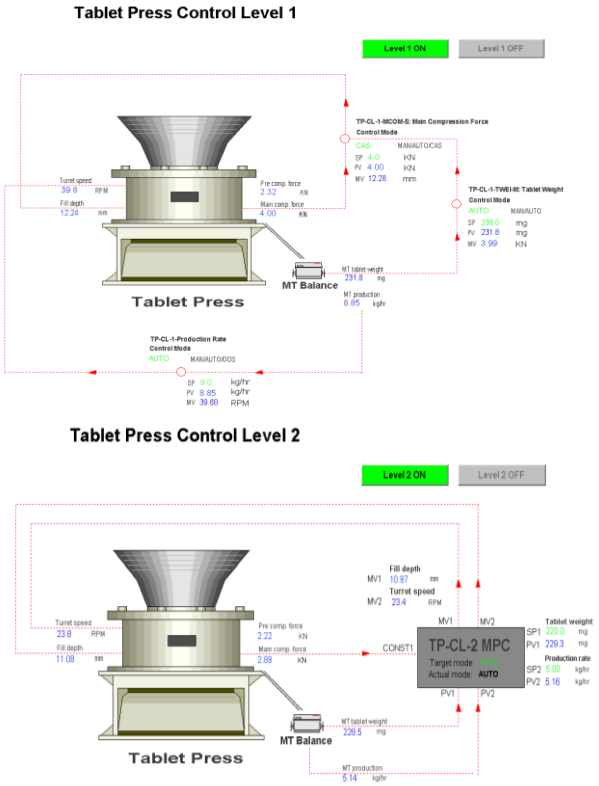


Figure 2. The hierarchical Level 1 (top) and Level 2 (bottom) control for tablet press.

4. Results and discussion

With the Level 0 Natoli built-in control of turret speed and dosing level (feeder speed was set constant and punch displacement was adjusted manually only), step changes in both turret speed and dosing level were conducted, as shown in Figure 3(a). It was observed that the measured tablet weight, as understood in process variation and characterization, strongly depended on dosing level after reaching steady-state. Step changes in turret speed, as highlighted with rectangles, showed only slight changes in tablet weight and main compression force due to changes in die filling time and flow dynamics, as described by Eq. (1). This is also consistent with the research findings of favourable compressibility of the powder formulation investigated (Tye et al., 2005), as described by Eq. (5). This variation was further studied when the tablet press was under closed-loop control operation, as shown in Figure 3(b). The Level 1 PID control for tablet production rate was deliberately set to be open, while a step change in turret speed was introduced. It was demonstrated that the cascaded control loop for tablet weight was capable of bringing the measured tablet weight to the set point of the master loop, as well as the main compression force of the slave loop. The variation of turret speed, viz., the die filling time, was compensated for by relatively minor adjustments of dosing level. Furthermore, this variation contributed insignificantly to the interaction of the decoupled control loops for tablet weight and production rate, as shown in Figure 4(a), when set point changes were made in tablet weight and production rate. It is worth noting that the turret speed was first increased then reduced in response to the first increase in production rate then the following increase in tablet weight, during which both the tablet weight and main compression force reached their set points rapidly and steadily, thus achieving the desired critical quality attributes in relative density, tensile strength, as discussed previously. Due to the moderate nonlinearity and relatively modest interaction of the studied tablet press and specific powder blend, comparable promising performances were achieved for both Level 1 PID control and Level 2 linear MPC control (with sampling time of 1 second) in terms of process set point changes and under the risk of powder bulk density disturbances (see Su et al., 2017, for disturbances due to changes in API mass fraction set point in a feeding and blending system).

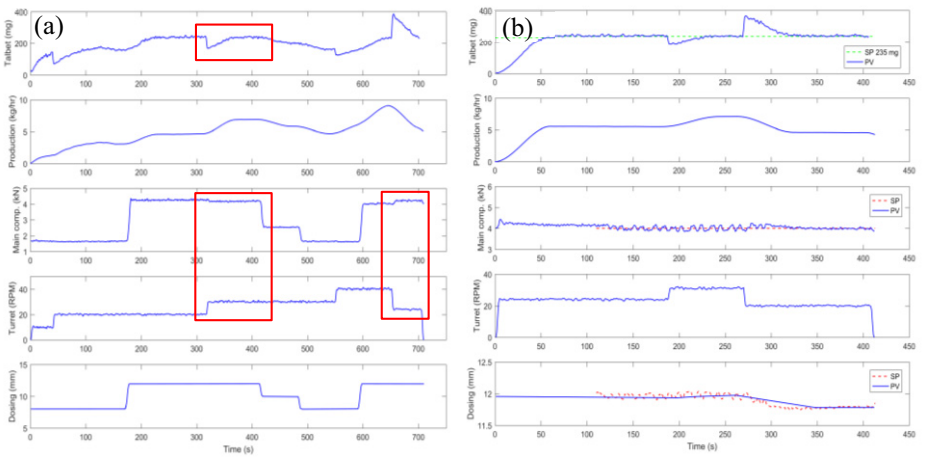


Figure 3. (a) Step changes in turret speed and dosing level under Level 0 control; (b) variation of the turret speed on tablet weight and main compression force control under Level 1 control when production rate control loop was deliberately set to be open.

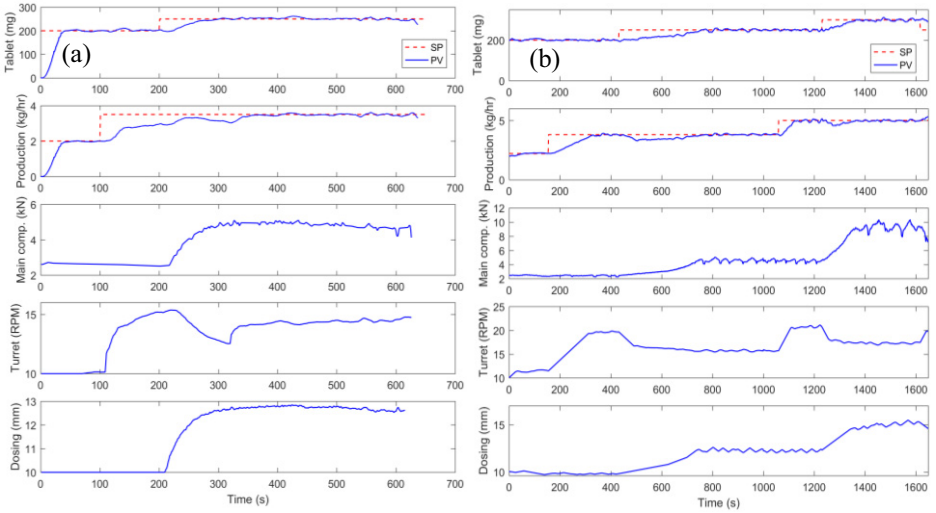


Figure 4. Control performance of Level 1 PID (a) & Level 2 MPC (b) control loop for set point changes.

5. Conclusions and future work

The system variation and its effect on the development of control strategies for a tablet press were investigated. It was found that the monitoring and control of the tablet weight and main compression force, two of the critical-to-quality variables, were important to achieving the consistent product properties under the process variations of turret speed and the risk of uncertainty in powder bulk density. In ongoing work, we are investigating additional powder formulations with different compressibility properties and implementing the control framework on an industrial scale tablet press.

Acknowledgement

The authors wish to acknowledge the financial support of the Food and Drug Administration under the grant number DHHS-FDA U01FD005535-01.

References

A. Bhaskar, F. N. Barros, R. Singh, 2017, Development and implementation of an advanced model predictive control system into continuous pharmaceutical tablet compaction process, International Journal of Pharmaceutics, 534, 159-178.

M. Kuentz, H. Leuenberger, 2000, A new model for the hardness of a compacted particle system, applied to tablets of pharmaceutical polymers, Powder Technology, 111, 145-153.

R. Singh, A. Sahay, F. Muzzio M. Ierapetritou, R. Ramachandran, 2014, A systematic framework for onsite design and implementation of a control system in a continuous tablet manufacturing process. Computers and Chemical Engineering, 66, 186-200.

Q. Su, M. Moreno, A. Giridhar, G. V. Reklaitis, Z. K. Nagy, A systematic framework for process control design and risk analysis in continuous pharmaceutical solid-dosage manufacturing, Journal of Pharmaceutical Innovation, 12, 327-346.

C.K. Tye, C. Sun, G. E. Amidon, 2005, Evaluation of the effects of tableting speed on the relationships between compaction pressure, tablet tensile strength, and tablet solid fraction. Journal of Pharmaceutical Sciences, 94, 465-472.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50109-9

Dynamic Simulation of a LNG Regasification Terminal and Management of Boil-off Gas

Vikas Singh Bisen[a], Iftekhar A. Karimi[a*], Shamsuzzaman Farooq[a*]

[a]*Department of Chemical & Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, 117585, Singapore, Singapore*
chevsb@nus.edu.sg

Abstract

The use of natural gas (NG) is increasing worldwide due to its cleaner nature and economic benefits. The distances between its supply and demand centers necessitate long distance transport, which is currently done as liquefied natural gas (LNG). LNG receiving terminals are critical elements of this supply network. They store and regasify the imported LNG and deliver NG at specified rates to the gas distribution network. A regasification terminal usually operates in three modes, unloading mode when LNG is transferred from carrier ship to tank, reloading when carrier ship is loaded from tank and holding or normal operational mode.

In this work, a detailed and rigorous dynamic model of a regasification terminal is developed in Honeywell UNISIM Design and validated against actual operational data of the real terminal. With this detailed and validated dynamic model, boil-off gas (BOG) minimization studies have been done to achieve a reduction of 25% in BOG generation from the base value. Furthermore, we propose an efficient compressor scheduling strategy taking all the modes of terminal operation into account. Equipped with a rigorous dynamic model, we have also included the effects of variable NG demand.

Keywords: LNG, BOG, dynamic model, regasification terminal, compressor scheduling.

1. Introduction

LNG receiving terminals are mainly composed of LNG storage tanks, LNG send out pumps, boil-off-gas (BOG) compressors, cryogenic pipelines, LNG booster pumps, LNG recondensers and LNG vaporizers. Heat leak to tanks and recirculation pipelines from the surroundings is steady influencing factor contributing to BOG generation in the terminal. In addition to this variable influencing factors like changing NG demand, ship unloading and ship reloading can amount for BOG generation several time larger than normal operating mode. During the unloading and reloading of an LNG carrier, a huge volume of BOG is generated due to volume displacement of onboard LNG, heat dissipation from unloading lines, unloading pump and incoming feed flash. BOG generated is handled by compressors which should be operated in a way that tank pressures remains within a safe range. However, excess compressor operation leads to excess BOG generation due to lowering of the tank pressures. Therefore, proper BOG management during all the operating modes is important to ensure effective and reliable terminal operation.

Previous work in BOG management and an effective operational strategy done by Li et al. [2016] and Lee et al. [2012] focuses mainly on unloading mode, assuming constant NG demand. Variable NG demand from terminal, recirculation strategy, unloading and reloading strategy and compressor scheduling together play a key role in framing strategy for the optimal operation of a terminal.

2. Dynamic Simulation

LNG regasification in a receiving terminal is a dynamic process and should be modelled and simulated accurately to study and suggest any improvement in existing operations. A general schematic of LNG receiving terminal as discussed by Park et al. [2010] is depicted in Figure 1. In the present study, UNISIM dynamics is employed to develop a detailed model of the terminal. Modules used to develop dynamic model are listed in Table 1.

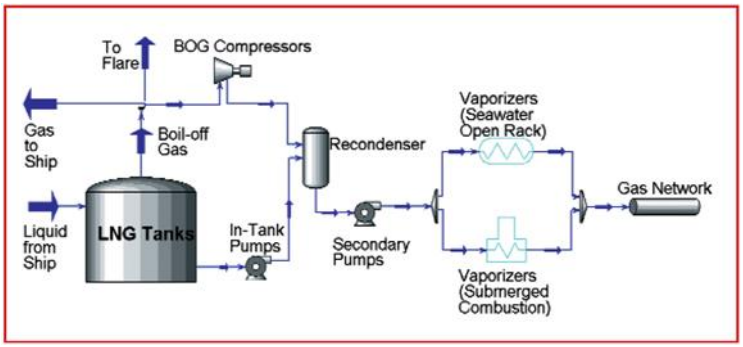


Figure 1. LNG regasification terminal

Table 1.Modules used in UNISIM model

Equipment	Module
Ships	Tank
LNG Tank	Tank
Compressors	Reciprocating Compressor
Recondenser	Separator
Pump	Centrifugal pump
Control Valves	Control Valves
Jetty arms	Valves and pipes
Vaporizers	Heat exchanger
Surge drums	Separator
Headers	Header
Safety Valve	Relief Valve

Developing an accurate dynamic model for the regasification terminal involves capturing all the physics involved in critical equipment like LNG tank, BOG recondenser, vaporizers etc. Accurate predictions of BOG generation during all the modes of operation is important, as BOG handling capability of regasification terminals

is one of the important measure of efficiency of the terminal. LNG tank modelling is critical and challenging, but accurately capturing the non-equilibrium behaviour between liquid and gas within the tank is important. BOG flow and temperature out of LNG tank are important parameters, and have a significant impact on compressor capacity. Similarly, modelling the BOG recondenser in UNISIM is a major challenge. BOG recondenser is a packed bed column for direct contact heat exchange between BOG and LNG. Modelling a packed bed column in UNISIM requires an in-depth understanding of the process as there is no built-in module for the packed column. Packing behaviour should be accurately captured to predict accurate level in the recondenser. Previous work in dynamic modelling of regasification done by Li et al. [2016] involved development of a generic dynamic model of the terminal with basic unit operations and control structure. In the present work, we have developed a much detailed and rigorous dynamic model involving all the plant specifics from ships, jetties, etc. up to the final natural gas metering station. All the equipment are modelled using the design datasheets, accurate pump curves, and all the piping and elevation details for accurate pressure-flow profiles. Piping and instrumentation diagram (P&ID) of the terminal is used to precisely model all the piping layout, mixing headers, valve positions, and controls. In addition to the design details, all the controllers are modelled accurately including the safety interlocks and control loops. This would allow us to perform control studies. This model can also be used to provide hands-on training to the engineers and operators as it includes all the minute details.

3. Validation

Validation of the dynamic model against the real plant data is important as it provides confidence to implement the findings of studies in a real plant operation. Validation of the tank model is done against the plant operational data for the shutdown. Tank pressure profile, tank level and tank bulk liquid temperature are matched against the plant data and shows an excellent agreement which validates the tank model. The data used for this validation cannot be disclosed due to a confidentiality agreement in our project. The validation plots are normalized and expressed as percentages of some base values and shown in Figure 2, 3 and 4.

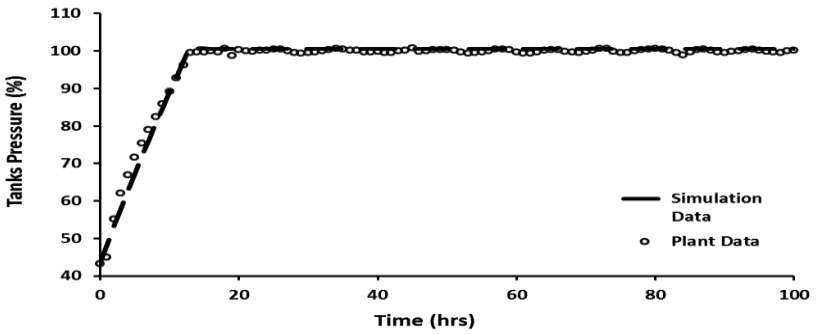


Figure 2. Tank Pressure profile

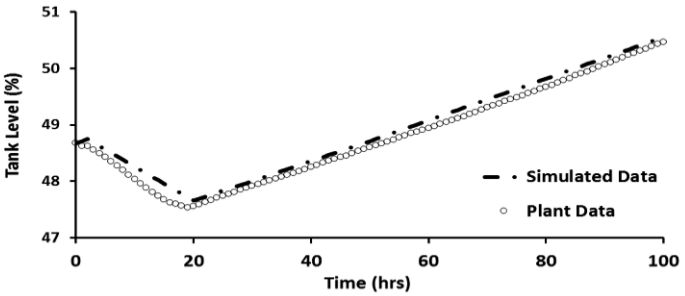


Figure 3. Tank Liquid Level profile

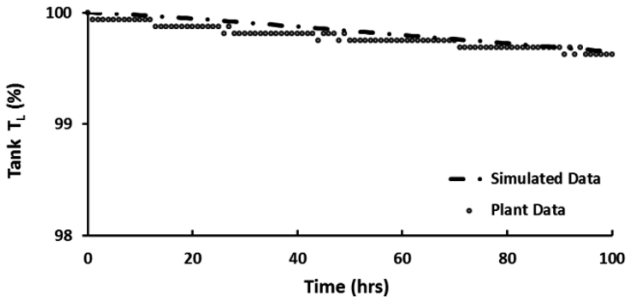


Figure 4. Bulk Liquid temperature profile inside tank

4. Case Study

The validated dynamic model can be used to test various operational and control strategies. We have used the model to perform studies related to BOG management, compressor scheduling, and recondenser operation for the three modes of terminal operation. In this work, we only discuss our study on the holding mode of operation. Our objective was to study various recirculation flow return configurations and their impact on BOG generation and compressor scheduling. These results promises a significant impact on BOG compressor power savings.

4.1. Recirculation flow configuration

Figure 5 below shows the flow scheme for send out and recirculation flow in terminal.

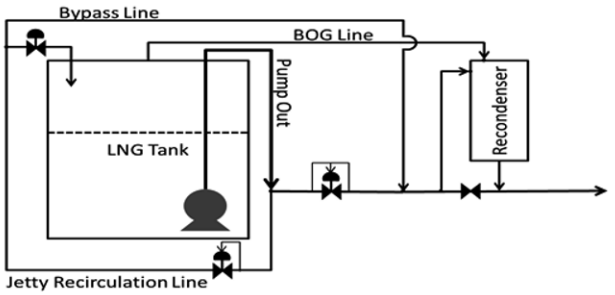


Figure 5. Flow scheme for recirculation and send out flow

A part of total flow from the tank is sent as recirculation flow to keep the unloading lines cool during normal holding mode of operation. This recirculation flow to unloading lines can be controlled using different valve settings. Recirculating LNG receives a lot of heat due to significant temperature gradient between LNG and atmosphere. Recirculation flow
can be completely sent back to LNG tank or a part of the flow can be mixed with send out flow using the bypass line shown in Figure 5 above. In present case study, we have tried different flow strategies by changing valve settings in the flow network and using the hydraulics to distribute the flow between tanks and bypass line. The results for the two configurations are discussed in section below.

4.2. Results and discussion

Figure 6 shows a comparison between complete pressure cycles for LNG tank.

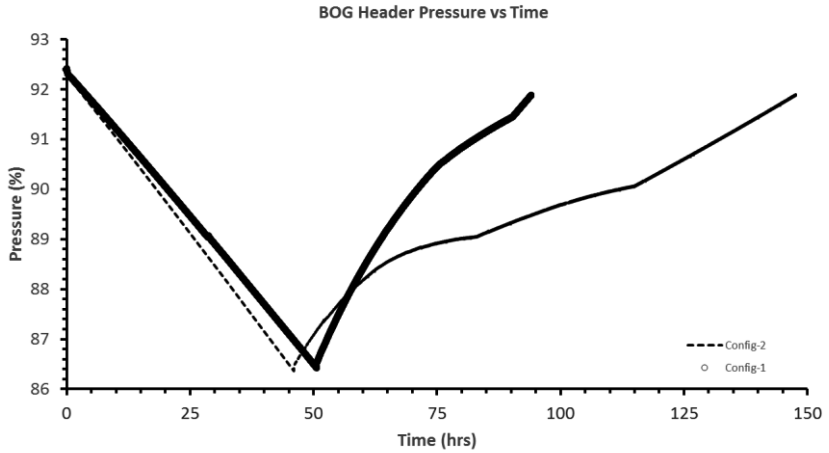


Figure 6. Pressure profile-LNG tank

One compressor is running at 100 % of its capacity for both the strategies. Configuration 1 takes around 52 hr. for compressor to drop pressure from 92.5% of its maximum value to 86.3 %. Configuration 2 takes 6 hours less than configuration 1 to bring down the pressure to the same extent. Once the pressure is lowered, the compressor is switched off and pressure is allowed to increase. Configuration 2 takes 40 hr. more for pressurization when compared to configuration 1, due to less BOG generation. Bypassing the recirculation flow results in lower addition of heat into the tank, and hence less BOG and less compressor running time. Configuration 2 will keep compressor off for 67% of the time and hence can save up to 50% of the compression power. Further study to propose an optimum compressor running capacity based on relative BOG generation rates for both the configurations is investigated and results are summarized in the Figure 7.

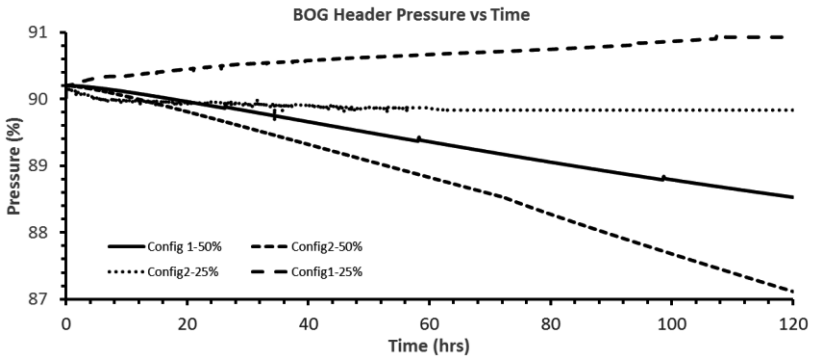


Figure 7. Pressure profile with variable compressor capacity

These results show different pressure profiles for different compressor capacities for both the configurations, and therefore, an optimum strategy for compressor operation can be devised to save maximum power.

5. Conclusion

Considering the dynamics involved in a regasification terminal, a robust and accurate model is necessary to study the process and suggest continuous operational improvements. Implementing any changes to ongoing plant operation is not plausible without foreseeing the risk involved. Having a detail validated dynamic model gives the confidence to apply and test operational changes and evaluate the risks involved. The case study discussed here aims at minimizing the BOG generation during holding mode of terminal operation. We are successful in reducing BOG generation by about 25 % and power consumption by about 36 %. Since there is substantial time gap between two ship unloading in terminal, an effective compressor operation strategy can minimize BOG generation as well as save power. The strategy proposed in this study shows a promising prospect in terms of an added benefit to existing operational practice. Testing controller configuration, response time, what if analysis and HAZOP studies can also be done using this model. Extending this model for expansion studies, testing start up shutdown procedures adds up to the future work.

6. Acknowledgements

This work was funded under Energy Innovation Research Program (EIRP) award NRF2014EWTEIRP003-008 administered by Energy Market Authority (EMA). The EIRP is a competitive grant call initiative driven by the Energy Innovation Program Office and funded by the National Research Foundation (NRF) of Singapore.

References

Li Yajun, Li Yue., 2016, Dynamic optimization of BOG fluctuations at an LNG receiving terminal, Journal of Natural Gas and Engineering, 30, 322-330.

Lee C.J., Lim Y, Han C., 2012, Operational strategy to minimize operating costs in LNG receiving terminals using dynamic simulation, Korean Journal of Chem. Engg.,29(4), 444-451.

Park et al., Optimization of recirculation operation in LNG receiving terminal, 2010, Journal of Taiwan Institute of Chemical Engineers, 41(4), 482-491.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50110-5

Modifier Adaptation with Model Adaptation in Iterative Real-Time Optimization

Afaq Ahmad*, Weihua Gao, Sebastian Engell

Process Dynamics and Operations Group, Department of Biochemical and Chemical Engineering, TU Dortmund, Emil-Figge Strasse 70, 44227 Dortmund, Germany. Afaq.Ahmad@tu-dortmund.de

Abstract

This paper deals with the iterative real-time optimization (RTO) of chemical processes under plant-model mismatch. Certain RTO methodologies, called modifier adaptation add bias and gradient correction terms to the underlying model based on measurements of the true values of the objective function and of the constrained function. These affine corrections lead to meeting the first-order necessary-conditions of optimality of the plant in spite of plant-model mismatch. However, accurately computing the gradient corrections is a limiting element here, and to improve the model parameters were possible will speed up convergence. Also, the additional terms do not guarantee satisfaction of the second-order optimality conditions. So modifier adaptation should be combined with model adaptation such that first- and second- order optimality conditions are met at the plant optimum. In this paper, we compare different iterative RTO schemes that perform modifier adaptation and model adaptation. The performance of the approaches is evaluated by means of simulation results for the Williams-Otto reactor benchmark problem.

Keywords: Real-Time Optimization; Modifier Adaptation; Effective Model Adaptation

1. Introduction

Real-time optimization (RTO) can make a significant contribution to profitable and resource efficient operation of processing plants. However, the feasibility and optimality of the proposed operating points relies on the accuracy of the plant model. Due to plant-model mismatch, the optimum of the real plant may differ significantly from the optimum that is computed from the available model and constraints may be violated by the theoretically optimal set-point. Therefore it is necessary to adapt the model or the optimization problem based on the available measurements. In recent years, several RTO methodologies have been developed which differ in how they utilize the information which is provided by the measurements.

In the two-step approach (Chen and Joseph, 1987), the model parameters are estimated at the current operating conditions in order to match the model prediction with the plant outputs. The optimization step is then performed based on the adapted model to update the operating point. This iterative two-step scheme may potentially converge to the process optimum if the plant-model mismatch is mainly of parametric nature, and there is sufficient excitation to estimate the model parameters accurately. However, if there is a structural mismatch, the two-step scheme may not converge to the true plant optimum. Therefore Roberts (1979) proposed a modified two-step scheme, known as *Integrated*

Acknowledgement: The research leading to these results has received funding from the Deutsche Forschungsgemeinschaft (DFG) under grant agreement number EN 152/41-1.

System Optimization and Parameter Estimation (ISOPE) in order to handle structural plant-model mismatch. The ISOPE approach converges iteratively to a Karush-Kuhn-Tucker (KKT) (also known as first-order necessary conditions) point of the plant by adding gradient correction terms to the cost function of the optimization problem.

Tatjewski (2002) realized that the step of the estimation of the model parameters in ISOPE is not required to satisfy the optimality condition of the plant. Based upon this work, Gao and Engell (2005) extended the approach to handle process dependent constraints and applied it to the optimization of a chromatographic separation process. Later, this approach was labelled as modifier-adaptation (MA) in Marchetti et al. (2009). The MA scheme adds bias and gradient modifiers to the cost and to the constraint functions of the model-based optimization so that the first-order necessary conditions of optimality of the adapted model match those of the plant. However, the approach does not ensure convergence if the model adequacy condition is not met as shown by Marchetti et al. (2009) and Gao et al. (2015).

A modification of the two-step approach is to estimate the model parameters such that the updated model predicts the gradient of the plant correctly at the current operating condition as proposed by Mandur and Budman (2015). However, the estimation of model parameters in order to fit the available plant measurements does not necessarily ensure that the model is adequate for optimization purposes (Forbes and Marlin, 1996). Model adequacy can be assured by using modifier adaptation with quadratic approximation (MAWQA) (Gao et al. 2016). However, the rate of the convergence depends on the collected data that is used to construct the quadratic approximation and is of course better when the model quality is high. Recently, Ahmad et al. (2017) proposed an effective model adaptation (EMA) scheme where an *adequate* process model is identified and used in the optimization problem. The key idea is to combine MA with EMA in the sense that the model that is used in the optimization problem is adapted only when the model with the updated parameter values is adequate, meaning that the first- *and* second-order necessary conditions of optimality are satisfied upon convergence.

In this paper, we investigate the reliability and efficiency of different iterative RTO schemes in the presence of plant-model mismatch. A simulation study is performed using the benchmark Williams-Otto reactor as a case study. The RTO schemes are compared based on their ability to converge to the plant optimum and the number of plant evaluations that are required to converge.

2. Iterative RTO schemes

The general model-based optimization problem at the (k)th iteration can be formulated as

$$\begin{aligned} \min_{\mathbf{u}} \quad & J_m(\mathbf{u}, \theta^{(k)}) := J(\mathbf{u}, \mathbf{y}_m(\mathbf{u}, \theta^{(k)})) \\ \text{s.t.} \quad & \mathbf{C}_m(\mathbf{u}, \theta^{(k)}) := \mathbf{C}(\mathbf{u}, \mathbf{y}_m(\mathbf{u}, \theta^{(k)})), \end{aligned} \tag{1}$$

where $\mathbf{y}_m$ is the vector of outputs which is a function of the input variables $\mathbf{u}$ and the set of model parameters $\theta^{(k)}$. $J_m(\mathbf{u}, \theta^{(k)})$ is the objective function that is minimized and $\mathbf{C}_m(\mathbf{u}, \theta^{(k)})$ represents the vector of the constraint functions. In the two-step scheme, the vector of model parameters $\theta^{(k)}$ is obtained as

$$\theta^{(k)} := \arg\min_{\theta} \|\mathbf{y}_p(\mathbf{u}^{(k)}) - \mathbf{y}_m(\mathbf{u}^{(k)}, \theta)\|^2, \tag{2}$$

where $\mathbf{y}_p(\mathbf{u}^{(k)})$ represents the measured variables at the current operating conditions $\mathbf{u}^{(k)}$. Since the conventional model adaptation via least-squares value fitting of the plant measurements does not guarantee the matching of the gradients, Mandur and Budman (2015) introduced an algorithm (called simultaneous model identification and optimization (SMIO)) in which the model is adapted in two steps in order to handle structural plant-model mismatch. In first step the values of the selected model parameters are computed via Eq.(2). The second step is based on gradient fitting, where the model gradients are matched to those of the plant by estimating the change of the model parameters $\Delta\theta^{(k)}$ as

$$\begin{aligned} \Delta\theta^{(k)} \quad &:= \arg\min_{\Delta\theta} \left(w_J \left| \nabla J_p^{(k)} - \nabla J_m^{(k)}(\theta^{(k)} + \Delta\theta) \right| + w_C \left| \nabla \mathbf{C}_p^{(k)} - \nabla \mathbf{C}_m^{(k)}(\theta^{(k)} + \Delta\theta) \right| \right) \\ \text{s.t.} \quad &\mathbf{y}_m(\mathbf{u}^{(k)}, \theta^{(k)} + \Delta\theta) := \mathbf{y}_m(\mathbf{u}^{(k)}, \theta^{(k)} + \Delta\theta) - c_k, \\ &\theta^{(k)} + \Delta\theta \in [\theta_{lb}, \theta_{ub}], \\ &\|\in^T\|_\infty = \in_{max}^T, \end{aligned} \tag{3}$$

where w_J and w_C are the weighting factors for the objective and constraint functions. $\nabla J_p^{(k)} := \nabla J(\mathbf{u}^{(k)}, \mathbf{y}_p(\mathbf{u}^{(k)}))$ and $\nabla \mathbf{C}_p^{(k)} := \nabla \mathbf{C}(\mathbf{u}^{(k)}, \mathbf{y}_p(\mathbf{u}^{(k)}))$ are the plant objective and constraint gradients at $\mathbf{u}^{(k)}$. A correction term $c_k := \nabla y(\theta^{(k)})\Delta\theta$ is introduced to compensate the deviation of the predicted steady-state values of the outputs due to the adjustment of the parameters in the second step, where $\nabla y(\theta^{(k)})$ is the Jacobian matrix of the model output with respect to the model parameters. The accuracy of the correction term can be quantified by using the relative truncation error

$$\in^T \quad = \frac{\mathbf{y}_m(\mathbf{u}^{(k)}, \theta^{(k)} + \Delta\theta^{(k)}) - \nabla y(\theta^{(k)})\Delta\theta^{(k)}}{\mathbf{y}_m(\mathbf{u}^{(k)}, \theta^{(k)})}. \tag{4}$$

In the MA scheme, the model-based optimization problem in Eq.(1) is adapted by the bias and first-order correction terms instead of model parameter adaptation. The updated optimization problem of the MA scheme can be written as follows:

$$\begin{aligned} \min_{\mathbf{u}} \quad & J_{ad}^{(k)}(\mathbf{u}, \theta^{(0)}) := J_m(\mathbf{u}, \theta^{(0)}) + J_p^{(k)} - J_m^{(k)} + \left(\nabla J_p^{(k)} - \nabla J_m^{(k)} \right)^T \left(\mathbf{u} - \mathbf{u}^{(k)} \right) \\ \text{s.t.} \quad & \mathbf{C}_{ad}^{(k)}(\mathbf{u}, \theta^{(0)}) := \mathbf{C}_m(\mathbf{u}, \theta^{(0)}) + \mathbf{C}_p^{(k)} - \mathbf{C}_m^{(k)} + \left(\nabla \mathbf{C}_p^{(k)} - \nabla \mathbf{C}_m^{(k)} \right)^T \left(\mathbf{u} - \mathbf{u}^{(k)} \right) \leq \mathbf{0}, \end{aligned} \tag{5}$$

where $J_{ad}^{(k)}(\mathbf{u}, \theta^{(0)})$ and $\mathbf{C}_{ad}^{(k)}(\mathbf{u}, \theta^{(0)})$ are the adapted objective and constraint functions, $\theta^{(0)}$ represents the fixed model parameter vector, $J_p^{(k)} := J(\mathbf{u}^{(k)}, \mathbf{y}_p(\mathbf{u}^{(k)}))$ is the process objective value and $\mathbf{C}_p^{(k)} := \mathbf{C}(\mathbf{u}^{(k)}, \mathbf{y}_p(\mathbf{u}^{(k)}))$ is the vector of constraint functions. The operating condition at the (k)th iteration is updated as:

$$\mathbf{u}^{(k+1)} = \hat{\mathbf{u}}^{(k+1)} + \mathbf{K}(\mathbf{u}^{(k)} - \hat{\mathbf{u}}^{(k+1)}), \tag{6}$$

where $\hat{\mathbf{u}}^{(k+1)}$ is the solution of Eq.(5) and $\mathbf{K}$ is a diagonal matrix of damping factors with

$0 \leq k_{ii} < 1$. The process gradients are estimated by finite difference approximation via perturbing the system in each iteration as proposed in Gao and Engell (2005).

3. Model adequacy

A criterion of model adequacy has been first discussed in Forbes and Marlin (1996). It states that a process model must produce a fixed point which is a local minimum for an RTO problem at the true process optimum u_p^*.

The model adequacy criterion for MA requires that u_p^* must meet the first- and second-order necessary conditions of optimality for the adapted optimization problem Eq.(5). As the MA scheme relies on the bias- and gradient corrections to ensure satisfaction of the first order KKT conditions of the plant, in addition the second-order necessary conditions of optimality need to be satisfied at u_p^*. This leads to the criterion that the matrix of the reduced Hessian $\nabla^2 \mathcal{L}_{ad,r}\left(\mathbf{u}_p^*, \theta, \mu^*\right)$ of the adapted optimization problem Eq.(5) should be positive definite. The Lagrangian function $\mathcal{L}_{ad}$ of the adapted optimization problem can be written as:

$$\mathcal{L}_{ad}(\mathbf{u}, \theta, \mu) := \quad J_{ad}(\mathbf{u}, \theta) + \mu^T \mathbf{C}_{ad}(\mathbf{u}, \theta), \tag{7}$$

where μ is the vector of Lagrange multipliers.

4. Modifier adaptation with model adaptation

A possible way to handle the issue of model adequacy is to combine the idea of MA with model adaptation. Ahmad et al. (2017) introduced the effective model adaptation (EMA) scheme to identify an adequate process model for the adapted optimization problem Eq.(5) and proposed the modifier-adaptation with EMA (in short MAWEMA) algorithm. EMA ensures and speeds up the rate of convergence to the optimum u_p^*. Let $\theta^{(k+1)}$ be the estimated model parameter vector at $\mathbf{u}^{(k+1)}$, then the updated parameters are accepted by EMA if and only if the following two conditions are satisfied:

- $\nabla^2 J_m(\mathbf{u}^{(k+1)}, \theta^{(k+1)})$ is positive definite
- $\left|J_{ad}^{(k)}(\mathbf{u}^{(k+1)}, \theta^{(k+1)}) - J_p(\mathbf{u}^{(k+1)})\right| < \left|J_{ad}^{(k)}(\mathbf{u}^{(k+1)}, \theta^{(k)}) - J_p(\mathbf{u}^{(k+1)})\right|$.

The first condition ensures the model adequacy criterion of the Lagrangian function $\mathcal{L}_{ad}\left(\mathbf{u}^{(k+1)}, \theta, \mu\right)$. It is sufficient that the Hessian matrix of the adapted objective function is positive definite which is the same as that reduced Hessian is positive definite. To motivate the second condition, we perform a Taylor series expansion of $J_{ad}^{(k)}(\mathbf{u}, \theta) - J_p(\mathbf{u})$ around $\mathbf{u}^{(k)}$, omitting terms of order higher than 2:

$$J_{ad}^{(k)}(\mathbf{u}, \theta) - J_p(\mathbf{u}) \approx \frac{1}{2}\left(\mathbf{u} - \mathbf{u}^{(k)}\right)^T \left(\nabla^2 J_m^{(k)} - \nabla^2 J_p^{(k)}\right)\left(\mathbf{u} - \mathbf{u}^{(k)}\right). \tag{8}$$

The prediction error of the adapted objective function at $\mathbf{u}^{(k+1)}$, that is, $J_{ad}^{(k)}(\mathbf{u}^{(k+1)}, \theta) - J_p(\mathbf{u}^{(k+1)})$, can be used to infer the difference of the Hessian matrices at $\mathbf{u}^{(k)}$. Hence, reducing the difference between $\nabla^2 J_m(\mathbf{u}, \theta)$ and $\nabla^2 J_p(\mathbf{u})$ ensures that the adapted model parameters accelerate the rate of convergence to the process optimum.

5. Simulation study

Different RTO schemes were tested on the Williams-Otto reactor benchmark problem which has been used to evaluate the performance of RTO schemes in many papers, e.g. Marchetti et al. (2010). The plant consists of an ideal CSTR in which three irreversible reactions take place whereas the presence of only two reactions is assumed in the model. The optimization objective is to maximize the profit which is defined as:

$$J(\mathbf{u}, \theta) = 1143.38X_P F + 25.92X_E F - 76.23F_A - 114.34F_B, \tag{9}$$

where $F := F_A + F_B$. The manipulated variables are the feed flow rate F_B and the reactor temperature T_R. The model equations and the description of the variables can be found in Marchetti et al. (2010). The meaning and the values of the model parameters are: $b_1 = -6 \times 10^3$ and $b_2 = -5 \times 10^3$ are the activation energies and $k_{1,0} = 1 \times 10^8$ and $k_{2,0} = 8 \times 10^8$ are the pre-exponential factors of the reactions. The initial model parameters b_1 and b_2 are calculated at the initial conditions of the manipulated variables, $F_B = 6.0$ and $T_R = 70$.

Figure 1 demonstrates the performance of the four RTO approaches presented above. Due to the presence of parametric and structural plant-model mismatch, the two-step approach does not converge to the true plant optimum. The standard MA approach where model parameters are not adapted moves near to the process optimum in the first iteration but then starts to oscillate since the available model is not adequate as it does not meet the second-order necessary condition of optimality. By updating the model at each iteration, the SMIO scheme performs better than the pure MA scheme. However, the SMIO scheme converges to a sub-optimal point after the eighteenth iteration, which is shown in the insert in Figure 1. This is due to the fact that parameter values do not exist which minimize the objective gradient error between the plant and model $\left|\nabla J_p^{(k)} - \nabla J_m^{(k)}(\theta^{(k)} + \Delta\theta^{(k)})\right|$. The best performance was obtained by the modifier adaptation with effective model adaptation (MAWEMA). After the initial iteration, EMA identifies an adequate process model for the modified optimization problem and ensures the convergence to the process optimum. Since the updated model at the first iteration also reduces the difference between the Hessians of the plant and the model, faster convergence to the process optimum is observed as compared to SMIO.

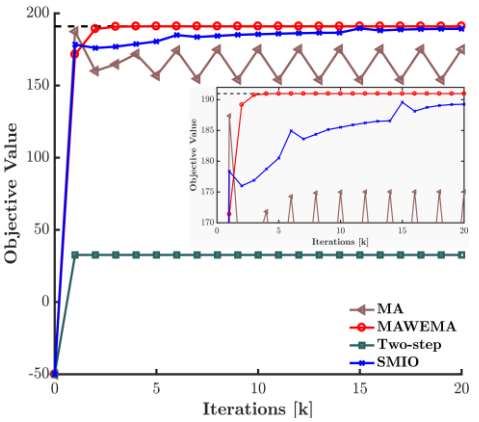


Figure 1: Comparison of the performance of the iterative RTO schemes

To illustrate the benefit of EMA, a simulation study was performed for different values of the initial feed flow rate and the reactor temperature. Figure 2 shows the comparison of the evolutions of the optimization objective over the iterations between SMIO and MAWEMA. Figure 2b shows that all trajectories obtained from the MAWEMA approach converge smoothly to the process optimum and satisfy the first- and second-

order optimality conditions whereas large oscillations can be observed in SMIO for some of the initial operating conditions. The SMIO scheme is not always able to reach the process optimum as the updated parameters cannot adapt the model gradients to the process gradient.

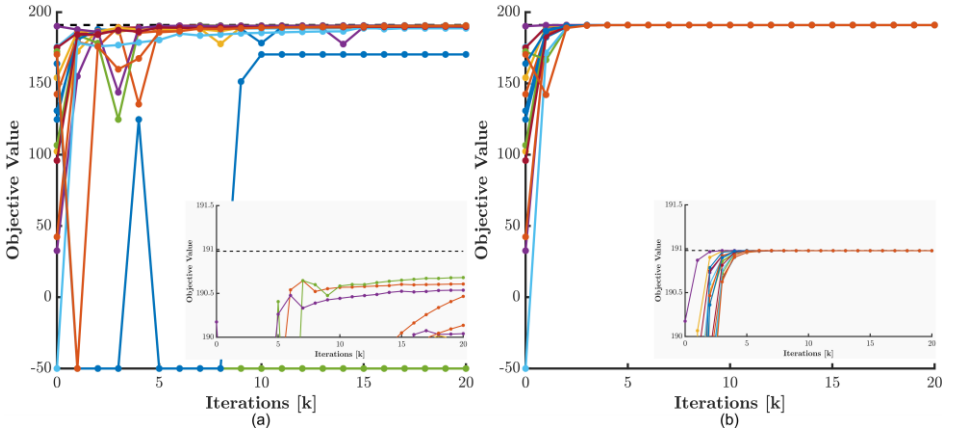


Figure 2: Optimization runs from arbitrary initial operating conditions. (a) SMIO, (b) MAWEMA

6. Conclusions

This paper highlights the importance of model adequacy in model adaptation which further enhances the performance of modifier adaptation. The key feature of MAWEMA is that a modified optimization problem is formulated to meet the first order KKT conditions and the second-order optimality conditions are satisfied via EMA.

References

Ahmad, A., Gao, W., Engell, S., 2017. Effective Model Adaptation in Iterative RTO. In: Computer Aided Chemical Engineering. Vol. 40. pp. 1717-1722.

Chen, C. Y., Joseph, B., 1987. On-line optimization using a two-phase approach: An application study. Ind. Eng. Chem. Res. 26(9), 1924-1930.

Forbes, J. F., Marlin, T. E., 1996. Design cost: A systematic approach to technology selection for model-based real-time optimization systems. Comput. Chem. Eng. 20 (6-7), 717–734.

Gao, W., Engell, S., 2005. Iterative set-point optimization of batch chromatography. Comput. Chem. Eng. 29 (6), 1401–1409.

Gao, W., Wenzel, S., Engell, S., 2015. Integration of gradient adaptation and quadratic approximation in real-time optimization. In: 34th Control Conference (CCC). pp. 2780–2785.

Gao, W., Wenzel, S., and Engell, S. 2016. A reliable modier-adaptation strategy for real-time optimization. Comput. Chem. Eng. 91, 318–328.

Mandur, J. S., Budman, H. M., 2015. Simultaneous model identification and optimization in presence of model-plant mismatch. Chem. Eng. Sci. 129, 106–115.

Marchetti, A. G., Chachuat, B., Bonvin, D., 2009. Modifier-adaptation methodology for real-time optimization. Ind. Eng. Chem. Res. 48 (13), 6022–6033.

Marchetti, A. G., Chachuat, B., Bonvin, D., 2010. A dual modifier-adaptation approach for real-time optimization. Journal of Process Control 20 (9), 1027–1037.

Roberts, P. D., 1979. An algorithm for steady-state system optimization and parameter estimation. Int. J. Syst. Sci. 10 (7), 719–734.

Tatjewski, P., 2002. Iterative optimizing set-point control - the basic principle redesigned. Proceedings of the 2002 IFAC World Congress 15 (1), 49–54.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50111-7

Closed-loop Economic Model Predictive Control for Scheduling and Control Problems

Michael J. Risbeck[a*], Christos T. Maravelias[a], James B. Rawlings[a]

[a]*Department of Chemical and Biological Engineering, University of Wisconsin–Madison*

risbeck@wisc.edu

Abstract

In this work, we show how scheduling and control problems can be treated within a unified optimization formulation. By transcribing discrete-time scheduling models into state-space form, a single dynamic model encompassing both layers can be optimized online as an instance of economic model predictive control (MPC), which inherits theoretical closed-loop properties from MPC theory. With the addition of suitable terminal constraints, a worst-case bound can be derived for the economic performance of the nominal closed-loop system, which helps to eliminate pathological behavior that can result from rescheduling too infrequently or without regard to recursive feasibility. Via an example system, we illustrate how constant feedback (i.e., reoptimization) and the inclusion of terminal constraints leads to better performance, even in the nominal case when no disturbances are present. With a unified treatment, the system is able to respond optimally and in real time to changing market conditions or process constraints, leading to improved economic performance and stability properties.

1. Introduction

In large-scale industrial environments, production-relevant decisions are typically split into multiple layers, with the most wide-reaching but lowest-resolution decisions at the top (e.g., design and campaign planning), slowly decomposing to more localized but detailed decisions at the bottom (e.g., regulatory control). Somewhere in the middle are layers commonly referred to as "scheduling" and "supervisory control" (Du et al., 2015). The former is primarily concerned with resource allocation, for example when to make a given product and in which unit to make it; the latter encompasses maintaining the process state at its desired operating point by choosing setpoints for localized regulatory controllers. Most decision hierarchies keep these layers completely separate, due to both the nature of the decisions (scheduling requires many discrete, "a or b" choices, while control considers continuous, "how much" process manipulations) as well as timescale separation (scheduling is typically performed over longer timescales than control). Furthermore, optimization of the scheduling layer is typically performed only infrequently, e.g., after some triggering event (Vin and Ierapetritou, 2000; Mendez and Cerdá, 2004), which means a schedule may become "stale" towards the end of its horizon by failing to consider future time points. Thus, reoptimizing more frequently often leads to improved performance but can also result in unexpected closed-loop behavior (Gupta and Maravelias, 2016). A successful technique for the supervisory control layer is model predictive control (MPC), which solves, repeatedly in real time, an optimization problem with an embedded process model to drive the system to its setpoint. Traditionally in industrial practice, MPC has been limited to continuous-valued decision variables. Recent research, however, has shown that discrete inputs can

be treated directly within the MPC formulation with little modification to the resulting stability and performance properties (Rawlings and Risbeck, 2017). An additional generalization is economic MPC, which replaces the artificial tracking objective with a tangible cost metric (Angeli et al., 2012). Thus, the online problem no longer minimizes distance to a static setpoint but rather dynamically optimizes the system's economic performance. These extensions allow the technique to be applied to systems with discrete decisions for which there is no natural steady-state operating point, e.g., cyclic production scheduling. Using ideas from Subramanian et al., (2012), standard scheduling models can be transcribed into the state-space form convenient for MPC, and underlying process dynamics can also be included in the model as necessary. In the remainder of this paper, we briefly discuss the economic MPC formulation and the theoretical cost bound afforded by including terminal constraints. Then, we apply this technique to a simple cyclic scheduling system to illustrate performance.

2. Economic MPC

We assume that the system of interest can be described in terms of internal states x which are affected by inputs u according to a state-space model $x(t+1) = f(x(t), u(t), t)$ with t denoting the discrete time. The goal of the controller is to choose u in a way that minimizes the economic cost function $\ell(x(t), u(t), t)$ while satisfying time-varying constraints $(x(t), u(t)) \in \mathbb{Z}(t)$. We assume further that there exists some feasible infinite reference trajectory $x_r(t)$, $u_r(t)$. Note that $x_r(t)$ could be constant (i.e., a steady state), periodic, or completely time-varying, but it must satisfy the model and constraints. For a given initial state x and initial time t, the following optimization problem is solved:

$$\min_{\hat{\mathbf{x}}, \hat{\mathbf{u}}} \quad \sum_{k=0}^{N-1} \ell\left(\hat{x}(k), \hat{u}(k), t+k\right) \tag{1a}$$

$$\text{s.t.} \quad \hat{x}(k+1) = f(\hat{x}(k), \hat{u}(k), t+k), \;\; k \in \{0, \dots, N-1\} \tag{1b}$$

$$(\hat{x}(k), \hat{u}(k)) \in \mathbb{Z}(t+k), \;\; k \in \{0, \dots, N-1\} \tag{1c}$$

$$\hat{x}(0) = x \tag{1d}$$

$$\hat{x}(N) = x_r(t+N) \tag{1e}$$

in which the decision variables are the (predicted) sequences $\hat{\mathbf{x}} \coloneqq (\hat{x}(0), \dots, \hat{x}(N))$ and $\hat{\mathbf{u}} \coloneqq (\hat{u}(0), \dots, \hat{u}(N-1))$. Letting $\kappa_N(x, t)$ denote an optimal value of $\hat{u}(0)$ in Eq. (1), the true system $x(t)$ evolves using the input $u(t) = \kappa_N(x(t), t)$; that is, the optimization problem is solved repeatedly and only the first input is used, rather than generating a single trajectory $\hat{\mathbf{x}}$, $\hat{\mathbf{u}}$ and following it until exhaustion. This closed-loop implementation imposes the requirement that each optimization problem can be solved within the discrete sample time, but suboptimal solutions are sufficient provided the cost is at least as good as the cost of the shifted solution from the previous timestep. Of particular note in is the terminal constraint Eq. (1e), which requires that the system terminate (in phase) on the reference trajectory $x_r(t)$. The purpose of this constraint is to ensure recursive feasibility of the predicted trajectory $\hat{\mathbf{x}}$. In particular, by joining $\hat{\mathbf{x}}$ and $\mathbf{x}_r$, it is possible to construct an infinite feasible sequence that the system could hypothetically follow. Thus, when the horizon is shifted forward, there is at least one

feasible solution, which then provides an upper bound for the cost of the re-optimized trajectory. More formally, one can show that if the cost function $\ell(\cdot)$ is bounded below, then the nominal closed-loop system satisfies

$$\limsup_{K\to\infty} \frac{1}{K} \sum_{t=0}^{K-1} \left(\ell(x(t), u(t), t) - \ell(x_r(t), u_r(t), t) \right) \leq 0 \tag{2}$$

That is, after an initial transient, the average cost of the real system's trajectory is no worse than the average cost of the reference trajectory. Thus, while the optimal control law $\kappa_N(\cdot)$ may move the system far from the nominal reference, in doing so, it improves the cost function for the system.

3. Example: Cyclic Production Scheduling

To illustrate the ideas from the previous section, we consider a single reactor that can produce multiple products (adapted from Flores-Tlacuahuac and Grossmann, (2006)). Orders of each product are due every 24 h, with excess sales permitted at any time. Product can be stored in inventory at a small inventory cost, while missed orders are heavily penalized and added to backlog that must be met at future times to avoid additional penalty.

3.1. Model Formulation

For each product $i \in \mathbf{I}$ and time $t \in \mathbf{T}$, there are nonnegative state variables S_{it} giving the current inventory and U_{it} tracking the accumulated backlog (both at the end of time period t). Although transition dynamics could be included in the process model alongside the scheduling considerations (possibly in approximate form as in Du et al., (2015)), for illustrative purposes we assume that changes between products can be described by fixed transition times $\tau_{ii'}$. To track product transitions, there are binary state variables X_{int} that are 1 if the reactor is $n \in \mathbf{N} \coloneqq \{0, \ldots, \max_{i,i'} \tau_{ii'}\}$ steps away from producing product i. Note that the X variables have been "lifted" (Subramanian et al., 2012) by the index n so that the values of X_{int} for a single value of t describes the entire state of the reactor. System inputs are how much product to sell V_{it} (nonnegative), how much to backlog H_{it} (continuous), and whether to transition to a new product $W_{ii't}$ (binary; 1 to begin transition and 0 otherwise). The states evolve at each time t according to

$$S_{i(t+1)} = S_{it} + \rho_i \left(X_{i0t} + X_{i1t} + \sum_{i' \in \mathbf{I}_{i0}} W_{i'it} - \sum_{i' \in \mathbf{I}} W_{ii't} \right) - V_{it}, \quad i \in \mathbf{I} \tag{3a}$$

$$U_{i(t+1)} = U_{it} + H_{it}, \quad i \in \mathbf{I} \tag{3b}$$

$$X_{in(t+1)} = X_{i(n+1)t} + \sum_{i' \in \mathbf{I}_{in}} W_{i'it}, \quad i \in \mathbf{I}, n \in \mathbf{N} \backslash \{0\} \tag{3c}$$

$$X_{i0(t+1)} = X_{i0t} + X_{i1t} + \sum_{i' \in \mathbf{I}_{i0}} W_{i'it} - \sum_{i' \in \mathbf{I}} W_{ii't}, \quad i \in \mathbf{I} \tag{3d}$$

with ρ_i giving the production rate of product i at the operating point and the auxiliary set $\mathbf{I}_{in} \coloneqq \{i' \in \mathbf{I} : \tau_{i'i} = n\}$. Constraints for each $t \in \mathbf{T}$ are then given by

$$\sum_{i \in \mathbf{I}} \sum_{n \in \mathbf{N}} X_{int} = 1 \tag{4a}$$

Table 1: Closed-loop profit for nominal system (normalized to $T = 48$ periodic solution).

Terminal Constraint	Rescheduling Period	Horizon		
		$N = 24$	$N = 48$	$N = 72$
None	12 h	−679.74%	−0.32%	−2.53%
None	1 h	−0.51%	−0.00%	+2.06%
$T = 48$	1 h	−0.02%	+0.29%	+0.44%
$T = 72$	1 h	+1.74%	+2.22%	+2.22%

$$\sum_{i' \in \mathbf{I}} W_{ii't} \le X_{i0t}, \quad i \in \mathbf{I} \tag{4b}$$

$$\delta_{it} - V_{it} \le H_{it} \le \delta_{it}, \quad i \in \mathbf{I} \tag{4c}$$

with δ_{it} giving the total orders that must be met or backlogged. Finally, the objective function is

$$\min \sum_{t \in \mathbf{T}} \sum_{i \in \mathbf{I}} (\beta_i U_{it} + \eta_i S_{it} - \pi_i V_{it}) \tag{5}$$

with β_i and η_i the backlog and inventory costs, and π_i the sales price. Note that backlog costs are 100 times inventory costs to discourage missed orders.

In state-space form, the states are $x(t) \coloneqq (S_{it}, U_{it}, X_{int})$ and inputs are $u(t) \coloneqq (V_{it}, H_{it}, W_{ii't})$. The system model $f(\cdot)$ is given by Eq. (3); the constraint sets $\mathbb{Z}(t)$ are Eq. (4) along with the sign and discreteness restrictions on the variables; and, the objective function $\ell(\cdot)$ is given by Eq. (5). The system timestep is 1 h.

3.2. Nominal Performance

To apply economic MPC, we start by determining a reference trajectory. Since orders are due every 24 h, it is natural to look for high-quality periodic solutions with cycle times T that are a multiple of 24 h. Such solutions can be found using Eq. (1) but replacing constraints (1d) and (1e) by the periodicity constraint $\hat{x}(0) = \hat{x}(N)$ and setting the prediction horizon N to the desired period T. Due to the transition times and required orders, there is no feasible $T = 24$ periodic solution, but optimal solutions can be found for $T = 48$, $T = 72$, etc. Note that longer periods generally have lower average cost, but the relationship is non-monotonic. To compare the performance of conventional scheduling with economic MPC, the nominal system is simulated for 30 days using different horizons, terminal constraints, and rescheduling frequencies. Table 1 shows closed-loop profit for each simulation case. In the first row, we see that

the traditional strategy of rescheduling infrequently without any terminal constraints can lead to poor performance. In case of the short $N = 24$ horizon, significant backlog is accumulated (on average, each order is 10.8 h late), which leads to extremely low profit. When the horizon is increased to $N = 48$, significantly less backlog is experienced, but the system still does not perform as well as it could. Interestingly, when the horizon is increased to $N = 72$, profit actually decreases due to the optimizer putting off production of less profitable products.

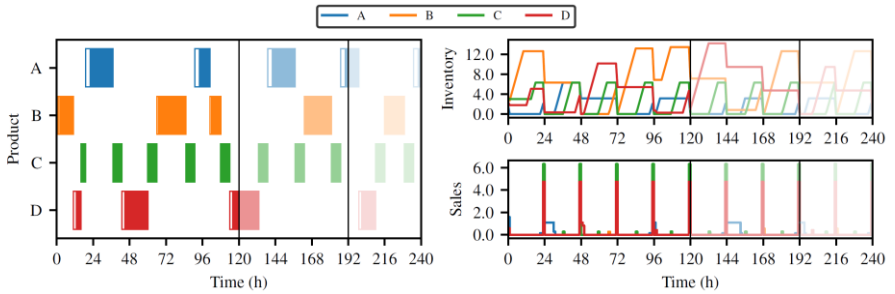


Figure 1: Beginning of closed-loop solution with horizon $N = 72$ and terminal constraint $T = 48$. Plots are split into three sections by vertical lines. First section shows closed-loop solution so far. Second section shows incumbent schedule. Third section shows reference trajectory on which the incumbent schedule terminates.

When the schedule is recomputed every hour, performance is improved. Even without terminal constraints, constant reoptimization is sufficient to prevent backlog in the $N = 24$ case, and profit is increased by almost 5% for the longer $N = 72$ horizon, as the system can react more quickly to new orders. However, without terminal constraints, there is no performance guarantee, and further, there is a lot more variation in schedules from one time period to the next, as new orders that enter the horizon every 24 h are essentially a disturbance not accounted for by the optimizer. By including the terminal constraints as suggested by economic MPC theory, system performance is improved. For the $N = 24$ horizon with the $T = 48$ periodic terminal constraint, Eq. (2) indicates that this value should asymptotically be no less than 0% (as costs were normalized to the $T = 48$ reference trajectory). The small negative value observed in Table 1 is due to aggressive control action in the initial transient required to reach the reference trajectory within 24 h. As the horizon is increased, average profit improves as the system is able to take a less aggressive (and more profitable) path to the reference trajectory. A plot of the system startup is shown in Figure 1. Finally, in all cases with the $T = 72$ terminal constraint, the system converges to the reference trajectory, and closed-loop profit is the highest among all of the simulated cases. Thus, in addition to providing a theoretical closed-loop cost bound, the inclusion of a quality terminal constraint leads to improved closed-loop performance even in the nominal case.

3.3. Effect of Disturbances

In the previous section, all simulations assume that the model is exact. However, real systems are subject to disturbances, either in the form of model mismatch or external forces affecting the system. While the cost bound from Eq. (2) no longer holds with disturbances, including the terminal constraint may still be helpful by keeping the system close to a nearly-feasible periodic solution. To examine this possibility, the reactor is subjected to random yield disturbances. The model prediction assumes that

the reactor production is at some nominal value, but at each timestep, the actual hourly yield can vary from the nominal value by $\pm 15\%$. Using a prediction horizon of $T = 48$ and rescheduling every timestep, the system is simulated 100 times with different yield disturbances both with and without terminal constraints. A kernel density estimate of profit and backlog costs is shown in Figure 2. Note that due to the high cost assigned to backlogs, even a small number of missed orders is sufficient to depress profit. The mean profit with terminal constraints is 4.25% higher than without terminal constraints, indicating that the terminal constraint is still helpful in the presence of disturbances.

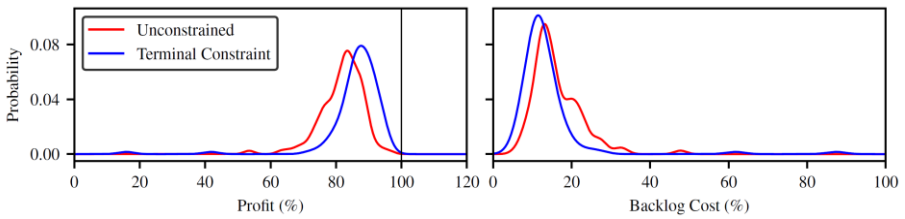


Figure 2: Kernel density estimate for closed-loop profit and backlog cost based on 100 simulations. Values normalized to profit of optimal $T = 48$ periodic solution.

4. Conclusions

In this work, we have demonstrated the formulation of cyclic production scheduling as an economic MPC problem. Even in the nominal case with no disturbances, rescheduling at every time point leads to improved cost over rescheduling only infrequently. In addition, using terminal constraints derived from periodic solutions ensures nominal recursive feasibility, which then gives a bound on the closed-loop cost for the system and leads to improved performance over scheduling without terminal constraints. Finally, when disturbances are present, the theoretical performance guarantees are lost, but as shown in simulation, the inclusion of terminal constraints still leads to higher closed-loop profit. Overall, the treatment of closed-loop scheduling via economic MPC helps to avoid pathological behavior that can result from naive rescheduling and facilitates analysis of closed-loop performance and stability properties.

References

D. Angeli, R. Amrit, and J.s B. Rawlings. On average performance and stability of economic model predictive control. *IEEE Trans. Auto. Cont.*, 57 (7): 1615–1626, 2012.

J. Du, J. Park, I. Harjunkoski, and M. Baldea. A time scale-bridging approach for integrating production scheduling and process control. *Comput. Chem. Eng.*, 79: 59–69, 2015.

A. Flores-Tlacuahuac and I. E. Grossmann. Simultaneous cyclic scheduling and control of a multiproduct CSTR. *Ind. Eng. Chem. Res.*, 45 (20): 6698–6712, 2006.

D. Gupta and C.s T. Maravelias. On deterministic online scheduling: Major considerations, paradoxes and remedies. *Comput. Chem. Eng.*, 94: 312–330, 2016.

C. A Mendez and J.e Cerdá. An MILP framework for batch reactive scheduling with limited discrete resources. *Comput. Chem. Eng.*, 28 (6-7): 1059–1068, 2004.

J. B. Rawlings and M. J. Risbeck. Model predictive control with discrete actuators: Theory and application. *Automatica*, 78: 258–265, 2017.

K. Subramanian, C. T. Maravelias, and James B. Rawlings. A state-space model for chemical production scheduling. *Comput. Chem. Eng.*, 47: 97–110, December 2012.

J. P. Vin and M. G. Ierapetritou. A new approach for efficient rescheduling of multiproduct batch plants. *Ind. Eng. Chem. Res.*, 39 (11): 4228–4238, 2000.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50112-9

Feedback control of proppant bank heights during hydraulic fracturing for enhanced productivity in shale formations

Prashanth Siddhamshetty[a,b], Abhinav Narasingam[a,b], Shuai Liu[c], Peter P. Valkó[c] and Joseph Sang-Il Kwon[a,b*]
[a]*Artie McFerrin Department of Chemical Engineering, Texas A&M University, USA*
[b]*Texas A&M Energy Institute, Texas A&M University, USA*
[c]*Department of Petroleum Engineering, Texas A&M University, USA*
kwonx075@tamu.edu

Abstract
In hydraulic fracturing of shale formations, compared to conventional reservoirs, the fracturing fluid is of low-viscosity, and hence, the proppant settles significantly during the pumping stage, forming a proppant bank. Motivated by this consideration, first, a section-based optimization method is employed to obtain key fracture design parameters for enhanced productivity in shale formations subject to given fracturing resources. Second, a reduced-order model is constructed using high-fidelity simulation data to design a Kalman filter. Lastly, a real-time model-based feedback control system is developed by explicitly taking into account actuator limitations, process safety and economic considerations. A series of closed-loop simulation results demonstrate that the proposed control scheme can regulate the uniformity of proppant bank heights across the fracture at the end of pumping.

Keywords: hydraulic fracturing, optimal pumping schedule, model predictive control.

1. Introduction
Unconventional natural-gas and oil resources are found in rock formations of ultra-low permeability (0.01 to 0.0001 mD or even less). Because of the low permeability of the formation, if a well were to be drilled into such a rock, it would take extremely long time to reach an economically viable amount of production. Horizontal drilling and hydraulic fracturing made the recovery of shale oil and gas economically viable, leading to shale boom in the U.S. (Economides and Nolte, 2000). In practice, while planning for hydraulic fracturing in unconventional reservoirs, very limited information is available. Therefore, the usual combination of sophisticated reservoir simulations and computationally extensive nonlinear optimization algorithms (Wilson and Durlofsky, 2013) might be difficult to justify at the planning stage. Motivated by these considerations, recently, Liu and Valkó (2017) proposed a section-based optimization method to maximize the dimensionless productivity index (PI) assuming constant pressure at the wellbore and infinite conductivity fracture (i.e., zero pressure drop in the fracture during the production stage). This approach converts a fixed amount of proppant into a total length of infinite conductivity fractures that in turn is used as a constraint to find the three decision variables: number of wells, number of fractures per well, and fracture half-length. These optimal design parameters can be used as set-point values to be satisfied during hydraulic fracturing processes. In conventional oil reservoirs, high-viscosity fracturing fluids ensure that most of the proppant remains in suspension during the hydraulic fracturing process. Thus, it is sufficient to regulate the

suspended proppant concentration along the fracture at the end of pumping. Some efforts have been made in this direction by developing model-based pumping schedule design techniques (Yang et al., 2017; Siddhamshetty and Kwon, 2017; Narasingam et al., 2017). In contrast, in unconventional oil reservoirs, predominantly low-viscosity ("slick-water") fluids are used and the proppant settles quickly forming a proppant bank. As a result, the proppant bank will continue to grow until it reaches the equilibrium height; a state when the rate of proppant washout on top of proppant banks due to the shear force is equal to the rate of bank formation via proppant settling. Thereafter, the injected proppant particles will flow over the top of the equilibrium-height proppant bank and travel farther to find a position at which the proppant bank has not reached the equilibrium height yet. This proppant transport behavior has been known for a long time (Kern et al., 1959) but has become dominant only recently, due to the widespread use of slick-water as fracturing fluids. We are not aware of incorporating it into a controller to compute optimal pumping schedules. Motivated by these considerations, we focus on the development of a model-based feedback control system to minimize the spatial variation of proppant bank heights along the fracture at the end of pumping by taking into account (a) actuator limitations that the pumping schedule consists of multiple pumping stages where the flow rate and the proppant concentration can be varied, (b) state constraints to prevent premature termination of the process, and (c) the desired propped fracture geometry obtained using the section-based optimization method.

2. Dynamic modeling of hydraulic fracturing processes

In this work, a hydraulic fracturing process is developed by considering the following three sub-processes: (1) fracture propagation, (2) proppant transport, and (3) proppant bank formation. More details regarding the modeling of hydraulic fracturing processes are given in Siddhamshetty et al. (2017).

3. Computation of optimal fracture configuration in unconventional reservoirs

3.1. Section-based optimization method

We employed a section-based optimization method, developed by Liu and Valkó (2017), to find the optimum number of wells n_c, number of fractures per well n_r, and fracture half-length x_f, that maximize the overall dimensionless PI subject to fixed fracturing resources. The section-based optimization is performed for planning the development of an unconventional reservoir with multi-stage fractured horizontal wells in a large square drainage area that we call a section (refer to Fig. 1). The section is evenly divided into multiple subsections such that there are n_c wells and n_r fractures per well. An individual subsection is considered as the drainage area of a single fracture (refer to Fig. 1), and all the subsections in one column represent a single horizontal well. In this approach, it is assumed that the created fractures will be of infinite conductivity. In addition, it is assumed that the selected total amount of fracturing materials is already enough to ensure that the stimulated volume will cover the whole section in which a uniform enhanced permeability will be established. The optimization problem to maximize the overall dimensionless PI, J_D, using the section-based optimization method is formulated as follows:

$$\max_{n_r, n_c} \quad J_D = n_f J_{D,f}(A_r, I_x) \tag{1a}$$

$$\text{s.t.} \quad J_{D,f} \text{ from the eigenvalue problem} \tag{1b}$$

$$A_r = \frac{n_r}{n_c}, \qquad I_x = \frac{l_{fD}}{n_r}, \qquad l_{fD} = \frac{l_f}{\sqrt{A_s}}, \qquad l_f = \frac{M_{prop}}{\rho_{sd} W_{opt} H_r} \tag{1c}$$

$$0 < x_f \leq x_{f,max} \tag{1d}$$

$$x_e - x_f \geq 0.5 d_{sep,min} \tag{1e}$$

$$x_f = \frac{l_f}{2 n_r n_c}, \qquad x_e = \frac{x_f}{I_x}, \qquad y_e = \frac{x_e}{A_r} \tag{1f}$$

where $n_f = n_c n_r$ is the total number of fractures in the section, l_f is the total fracture length (sum of $2x_f$ over the entire subsections) to be created in the section (injecting the available fracturing resources), and $J_{D,f}$ is the dimensionless PI for each fracture (i.e., a subsection). $J_{D,f}$ is a function of aspect ratio, A_r, and penetration ratio, I_x, of a subsection. x_e and y_e are the half-width and half-length of the subsection, l_{fD} is total fracture length divided by the section side length, A_s is the square drainage area of the section, M_{prop} is the total amount of proppant to be injected, W_{opt} is the average propped fracture width, and H_r is the reservoir thickness. More details regarding how to obtain $J_{D,f}$ can be found in Liu and Valkó (2017). The smallest distance between the tips of two fractures created from neighboring wells is called the separation distance d_{sep}. In this work, we consider $d_{sep,min} = 25$ m via Eq. (1f). The constraint Eq. (1e) imposes the limit $x_{f,max}$ on the fracture half-length. It is a crucial parameter, representing the current level of fracturing technology in the given unconventional formation, meaning that fractures can be routinely and reliably created and propped up to this half-length. In the current problem we use $x_{f,max} = 122$ m.

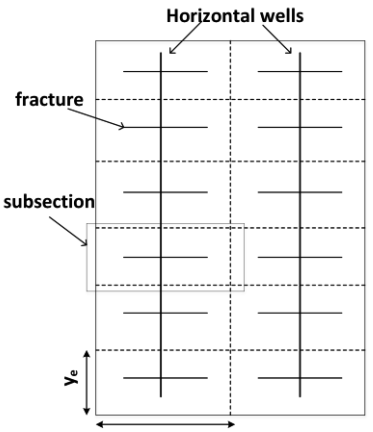


Figure 1. Illustration of a section for the case of two horizontal wells ($n_c = 2$) and six fractures per well ($n_r = 6$).

In the maximization problem, Eq. (1), there are only two decision variables, n_c and n_r. Once M_{prop}, W_{opt}, ρ_{sd} , H_r, and A_s are available, the total fracture length to be created in the section, l_f, is obtained using Eq. (1d). Since we use the concept of infinite conductivity fracture, the W_{opt} average propped fracture width is not the result of the optimization, rather an input, providing the minimum required propped width of a fracture to be considered as a propped fracture at all (often taken as three times the proppant grain diameter). In this work, we assumed the total amount of proppant for the section is $M_{prop} = 3.96 \times 10^7$ kg. Then, we computed the total fracture length in a section, $l_f = 79{,}248$ m, by substituting $W_{avg} = 2.9$ mm and $H_r = 60$ m for Eq. (1d). By solving the optimization problem, Eq. (1), we obtained that $n_c = 6$, $n_r = 55$, and $x_f = 120$ m. In summary, each fracture should be propped with $M_{prop,frac} = 120 \times 10^3$ kg proppant to achieve optimal fracture conductivity (i.e., productivity).

3.2. Treatment targets

As we already discussed, with the low-viscosity fracturing fluid, the proppant settles quickly and forms a proppant bank with an equilibrium height, h_{eq}. By assuming that

the proppant bank will cover the entire optimal fracture half-length, x_f, with the equilibrium height at the end of pumping, we can calculate the desired average fracture width, $W_{avg,target}$, as follows:

$$W_{avg,target} = \frac{M_{prop,frac}}{2\rho_p h_{eq} x_f (1-\phi)} \tag{2}$$

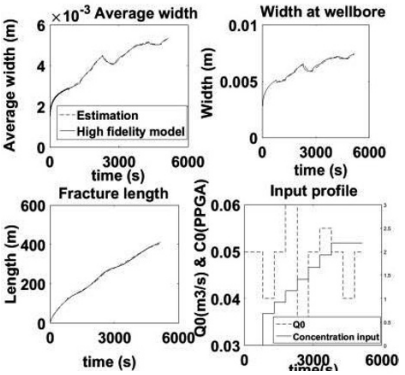


Figure 2. Comparison between the true values and the estimates.

where the equilibrium proppant bank height is $h_{eq} = 54$ m, which was calculated by using a correlation described in (Want el al., 2003). In our case, the calculated target fracture width at the end of pumping is $W_{avg,target} = 5.37$ mm. In the following section, we will focus on the development of a model-based feedback controller that aims to achieve the targeted length and width for developing the proppant bank.

4. Model predictive control for hydraulic fracturing processes

In this section, we first construct a linear time-invariant approximate model from the high-fidelity simulation data using the multi-variable output error state-space (MOESP), which is presented in the following form:

$$x(t_{k+1}) = Ax(t_k) + Bu(t_k) \tag{3a}$$

$$y(t_k) = H(t_k) \tag{3b}$$

where $u(t_k) = [Q_0(t_k), C_0(t_k)]^T$ is the flow rate and proppant concentration of the fracturing fluid at the wellbore (i.e., the manipulated input variables) and $y(t_k) = \left[W_{avg}(t_k), W_0(t_k), L(t_k)\right]^T$ denotes the vector of output variables where $W_{avg}(t_k)$ is the average fracture width, $W_0(t_k)$ is the fracture width at the wellbore, and $L(t_k)$ is the fracture length . In practice, the available real-time measurements are limited to the fracture width near the wellbore and the fracture length. This leaves the average fracture width to be estimated through a Kalman filter, which is presented in the following form:

$$\hat{x}(t_{k+1}) = A\hat{x}(t_k) + Bu(t_k) + M(t_k)(y_m(t_k) - \hat{y}(t_k)) \tag{4a}$$

$$M(t_k) = P(t_k)H^T(R(t_k) + HP(t_k)H^T)^{-1} \tag{4b}$$

$$P(t_{k+1}) = (I - M(t_k)H)P(t_k) \tag{4c}$$

where the notation $(\hat{\cdot})$ denotes the estimated variables, $M(t_k)$ is the Kalman filter gain, $P(t_k)$ denotes the covariance of the state estimation error, and $y_m(t_k) = [W_0(t_k), L(t_k)]^T$ are the available real-time measurements. Fig. 2 shows the comparison between the estimated and the actual average fracture width with time. It is observed that the estimated average fracture width quickly converges to the true value obtained from the high-fidelity process model. Then, a novel model-based feedback control system is designed to minimize the squared deviation of the average fracture width at the end of pumping from its set-point value. The proposed model-based feedback controller is formulated in the following form:

$$\max_{\substack{C_{0,k},\ldots,C_{0,9} \\ Q_{0,k},\ldots,Q_{0,9}}} \left(\widehat{W}_{avg}(t_f) - W_{avg,target}\right)^2 \tag{5a}$$

$$\text{s.t.} \quad \text{Kalman Filter, Eq. (4)} \tag{5b}$$

$$\widehat{W}_0(t_k) = W_0(t_k), \quad \hat{L}(t_k) = L(t_k) \tag{5c}$$

$$C_{0,k-1+m} \leq C_{0,k+m} \leq 2 \text{ ppga} \tag{5d}$$

$$Q_{min} \leq Q_{0,k+m} \leq Q_{max}, \quad m = 1, \ldots, 9-k \tag{5e}$$

$$\Delta\left(\sum_{k=1}^{9} 2Q_{0,k}C_{0,k}\right) = M_{prop,frac} \tag{5f}$$

where $W_{avg,target}$ is the set-point value for the average fracture width at the end of pumping, t_f is the total process operation time, $\widehat{W}_{avg}(t_k)$ is the predicted average fracture width obtained via Kalman filter which is then used in the approximate model (Eq. (3)) to find $\widehat{W}_{avg}(t_f)$, Δ is the sampling time, which is also the duration of each pumping stage, t_k is k^{th} sampling time, $W_0(t_k)$ and $L(t_k)$ are the available measurements, and $C_{0,k}$ and $Q_{0,k}$ are the inlet proppant concentration and the inlet flow rate of k^{th} pumping stage, respectively, which are to be determined by solving Eq. (5) with a shrinking prediction horizon $N_p = t_f - t_k$. The constraint of Eq. (5d) implies that the proppant concentration will be increased monotonically, and the maximum injected proppant concentration is less than 2 ppga, which stands for 1 pound of proppant added to one gallon of fluid and it is a standard concentration unit in petroleum engineering. The constraint of Eq. (5e) imposes the limits on the flow rate of fracturing fluids at the wellbore. The constraint of Eq. (5f) describes the amount of proppant to be injected to a single fracture (i.e., the material constraint).

5. Closed-loop simulation results under model-based feedback controller

In Section 3.2, we computed that to achieve the uniform equilibrium proppant bank height over the optimal fracture half-length, $x_f = 120$ m, the set-point value for the average fracture width that has to be satisfied at the end of pumping is $W_{avg,target} =$ 5.37 mm. For the simulations, the pad time, t_p, was fixed at 800 s in order to avoid

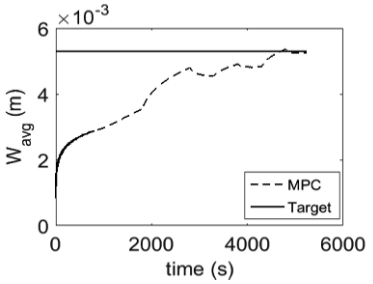


Figure 3. Average fracture width

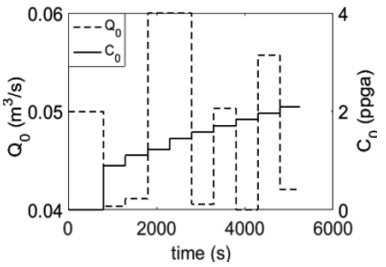


Figure 4. Optimal pumping schedule

premature termination of the hydraulic fracturing process before the fracture tip reaches the desired fracture half-length of $x_f = 120$ m. At $t = t_p$, the feedback control system and the Kalman filter were initialized. In the closed-loop simulations, Δ and t_f were chosen to be 500 s and 5300 s, respectively. The proppant pumping schedule was divided into 9 pumping stages and the duration of each pumping stage was 500 s, which was equal to the sampling time Δ. With this, the proposed model-based feedback controller was applied to the high-fidelity process model of hydraulic fracturing to achieve the desired average fracture width at the end of pumping. Fig. 3 presents that the average fracture width at the end of pumping is very close to the set-point value. Fig. 4 shows the optimal pumping schedule computed by the proposed model-based feedback controller. As we regulated the average fracture width at the end of pumping to the set-point value, for the given amount of proppant to be injected, the proppant

bank height along the fracture with the optimal half-length, x_f, was uniform with the equilibrium value as shown in Fig. 5.

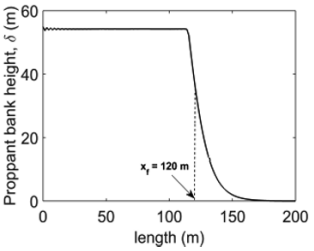


Figure 5. Spatial proppant bank height profile obtained at the end of pumping.

6. Conclusions

In this work, we developed a model-based feedback control system that computes an optimal pumping schedule to achieve the proppant bank with a uniform height over the optimal fracture length at the end of pumping. First, a section-based optimization method was employed to determine the optimal well-fracture configuration, maximizing the overall productivity achievable from a given amount of proppant to be injected. Second, using the high-fidelity simulation data, an approximate model was constructed for a Kalman filter to effectively estimate average fracture width. Third, MPC theory was applied for the design of a real-time model-based feedback control system, which was able to generate an online pumping schedule for a uniform proppant bank height along the targeted length while explicitly taking into account actuator limitations, state constraints for process safety, and economical considerations.

References

Z. Allen, Year, Article or Chapter Title, Journal or Book Title, Volume, Issue, Pages

S. Bhattacharya, M. Nikolaou, M. J. Economides, 2012, Unified fracture design for very low permeability reservoirs., J. of Natural Gas Sci. and Eng, 9, 184–195.

M. J. Economides, K. G. Nolte, 2000, Reservoir stimulation, John Wiley & Sons.

L. Kern, T. Perkins, R. Wyant, 1959, The mechanics of sand movement in fracturing., Journal of Petroleum Technology, 11, 55–57.

S. Liu, P. P. Valkó, 2017, Optimization of spacing and penetration ratio for infinite conductivity fractures in unconventional reservoirs- a sectional based approach., SPE Journal. doi:10.2118/186107-PA.

A. Narasingam, P. Siddhamshetty, J. Kwon, 2017, Temporal clustering for order reduction of nonlinear parabolic pde systems with time-dependent spatial domains: Application to a hydraulic fracturing process, AIChE Journal, 63, 9, 3818–3831.

P. Siddhamshetty, J. S. Kwon, 2017, Modeling of hydraulic fracturing and designing of online pumping schedules to achieve uniform proppant concentration., Computers & Chemical Engineering. https://doi.org/10.1016/j.compchemeng.2017.10.032.

J. Wang, D. Joseph, N. Patankar, M. Conway, R. Baree, 2003, Bi-power law correlations for sediment transport in pressure driven channel flows., Int. J. Multiphase Flow, 29, 475–494.

K. Wilson, L. Durlofsky, 2013, Optimization of shale gas field development using direct search techniques and reduced-physics models., Journal of Petroleum Science and Engineering, 108, 304–315.

S. Yang, P. Siddhamshetty, J. S. Kwon, 2017, Optimal pumping schedule design to achieve a uniform proppant concentration level in hydraulic fracturing., Computers & Chemical Engineering, 101, 138–147.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50113-0

Extremum-seeking Control of a Continuous Neutralization Reactor

Luís Felipe Safady[a], Flávio Vasconcelos da Silva[a], Martin Guay[b], Ana Maria Frattini Fileti[a*]

[a]*School of Chemical Engineering, University of Campinas, Av. Albert Einstein, 500, Campinas, SP, 13083-852 Brazil*
[b]*Faculty of Engineering and Applied Science, Queen's University, Dupuis Hall 406, Kingston, K7L 3N6, Canada*
frattini@feq.unicamp.br

Abstract

The potential of hydrogen, or pH, is an important factor to be controlled in various systems, especially on water treatment, bioprocesses, and sugar processing. However, its control can be a challenging task due to its strong non-linear behaviour and time-varying properties. Therefore, it is important to choose a control strategy that can detect or workaround these changes. Extremum-seeking control (ESC) is a model-free control strategy that has promising results in a range of applications such as increasing the power capture of wind turbines (Ciri et al., 2017) and controlling the position of high-altitude balloons (Vandermeulen et al., 2017), but it not yet very explored in the area of chemical engineering. A Single-Input Single-Output (SISO) and a Multiple-Input Single-Output (MISO) ESC strategies to control the pH of a continuous neutralization pilot-plant reactor are proposed on this paper and were able to successfully control its pH. Special care was taken with the values of the parameters as they greatly influence the rate of convergence of the algorithm.

Keywords: model-free; extremum-seeking control; process control, pH, continuous reactor.

1. Introduction

There are various types of automated control systems, from a simple on/off switch to set the level of a tank to a complex mechanistic model predictive controller to control fluid catalytic cracking. The choice of the right type of control system greatly depends on the system to be controlled.

The potential of hydrogen, or pH, is not only a difficult property to be controlled but also one with great industrial importance owing to the fact that it needs to be controlled in at least one operational unit on most industries. Some examples are wastewater treatment (Cheremisinoff, 2002), sugar processing (Chou, 2000), and for chemical and biochemical processes (Obut and Özgen, 2008).

The control strategies generally used in chemical plants are based on a mathematical model of the system, or mostly based on data that can be collected from them. This paper proposes the use of an extremum-seeking (ES) technique, a data-based strategy with no prior knowledge of the system, to directly control a neutralization reactor. Computational simulation of the system allowed prior performance assessment, while experimental tests showed the potential of ESC techniques in chemical process control.

This paper is organized as follows. Section 2 presents the experimental setup that was used to assess the control performance. Section 3 presents the simulation setup, including its model, software and solving algorithm. Section 4 presents the proposed control structure based on extremum-seeking. Section 5 presents both the results and discussion of the simulation and the experiments. Section 6 presents the conclusions of this work.

2. Experimental setup

A continuous neutralization pilot-plant available at the Laboratory of Automation and Process Control (LCAP) at University of Campinas (UNICAMP) was used to evaluate the strategy, and its instrumentation diagram can be seen on Figure 1, where the construction of the plant is shown along with the name of each unit and controller. Table 1 summarizes the information about all control loops.

The reactor in the plant is a continuous reactor in which a reaction between a solution of nitric acid, a solution of sodium hydroxide, and a solution of sodium bicarbonate take place.

Three liquid streams feeds the reactor, one carries a strong acid, another carries a strong base and a third carries a buffered medium that represents a theoretical effluent to have its pH corrected. The buffered medium is a solution of $NaHCO_3$ at 0.03 M, the acid, a, is a solution of HNO_3 at 0.0034 M, and the base, b, is a solution of NaOH at 0.0057 M. Three barrels are used to store each up to 100 L of each solution.

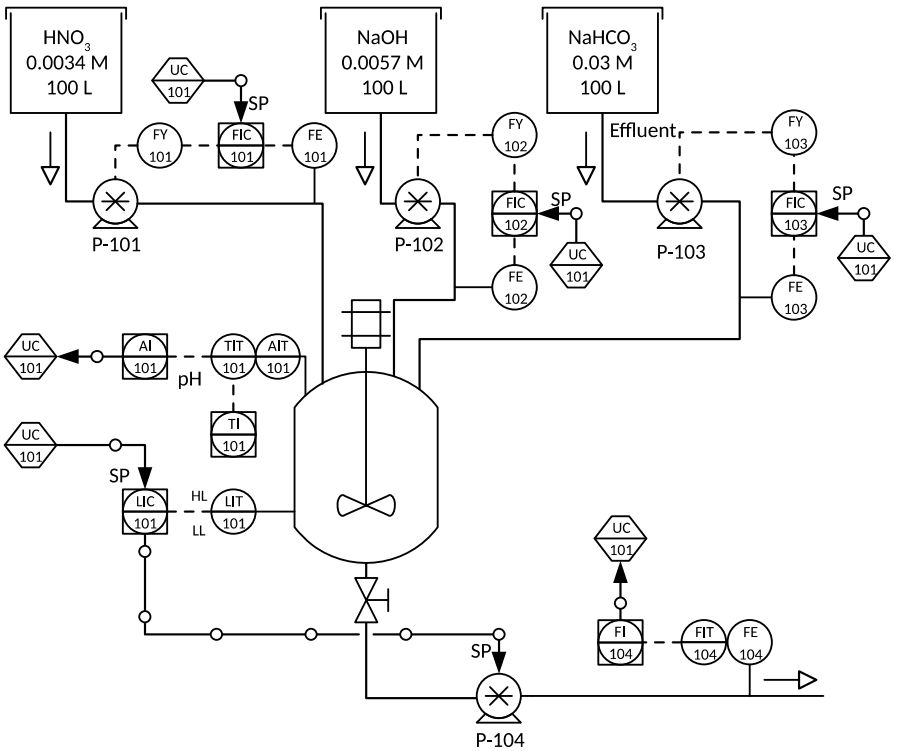


Figure 1 – Instrumentation diagram of the pH neutralization plant.

Table 1 – Summary of all control loops and strategies.

Tag	Controlled variable	Manipulated variable	Type
FIC-101	Acid flow rate	P-101 power	PI
FIC-102	Base flow rate	P-102 power	PI
FIC-103	Buffer flow rate	P-103 power	PI
LIC-101	Liquid level	P-104 power	PI
UC-101	pH	FIC-101 and/or FIC-102 set-points	SISO or MISO ESC

All data collection and management of the PI controllers are done by a PLC installed beside the plant, whose data can then be read and modified through the OPC protocol.

The management of the ESC controller and the monitoring of the plant is done through the computer UC-101, which is also responsible for the extremum-seeking strategy. It operates as an extremum-seeking controller whose slaves are FIC-101, FIC-102, or both at the same time depending if it is operating as a SISO or MISO controller. It can also set the set-point of LIC-101 to the desired level of the reactor. The reactor's height is 30 cm and its base area is 201.06 cm^2.

The free software SciLab and its XCos module along with a custom server software was used to communicate with the PLC and do all the calculations of the control strategy.

3. Process simulation and equations

A simulation of the system being controlled through Extremum-Seeking Control was used to assess the possible performance it can achieve and to have an approximated understanding of its transient response. The model used is based on the Reaction Invariant Model, and its equations are shown on Eq. (1), Eq. (2), Eq. (3), and Eq. (4). The simulations showed promising results, therefore the experimental procedure was designed.

On those equations A_b is the reactor base area, h the liquid level, q_a the flow rate of acid, q_b the flow rate of base, q_{buffer} the flow rate of the buffered medium, and q the flow rate at the reactor outlet.

The invariant $w_{1,a}$ is 0.0034 M, the invariant $w_{1,b}$ is -0.0057 M, the invariant $w_{1,buffer}$ is -0.03 M, the invariants $w_{2,a}$ and $w_{2,b}$ are both 0.00 M, the invariant $w_{2,buffer}$ is 0.03 M. They were calculated using conservation equations and equilibrium relations, similar to the work done by Henson and Seborg (1994). The anti-log of the equilibrium constants K_{a1} and K_{a2} from carbonic acid dissociation are given by pK_{a1}, which is 6.35, and pK_{a2}, which is 10.33, respectively, at 25 °C.

$$A_b h \frac{dw_1}{dt} = q_a(w_{1,a} - w_1) + q_b(w_{1,b} - w_1) + q_{buffer}(w_{1,buffer} - w_1) \quad (1)$$

$$A_b h \frac{dw_2}{dt} = q_a(w_{2,a} - w_2) + q_b(w_{2,b} - w_2) + q_{buffer}(w_{2,buffer} - w_2) \quad (2)$$

$$A_b h \frac{dh}{dt} = q_a + q_b + q_{buffer} - q \tag{3}$$

$$w_1 + 10^{pH-14} - 10^{-pH} + w_2 \frac{1 + 2 \times 10^{pH-pK_{a2}}}{1 + 10^{pK_{a1}-pH} + 10^{pH-pK_{a2}}} = 0 \tag{4}$$

4. Extremum-Seeking Control

Extremum-seeking has diverse literature, and the chosen strategy is based on the discrete variant as proposed by Killingsworth and Krstic (2006) for PI tuning. The cost function can be seen on Eq. (5), where $e(t, q_a, q_b)$ is the error between the set-point and the plant output, the value of the pH. The step size is given by h_D, which is the time in seconds taken between each adjustment of the input variables q_a and q_b.

$$J(q_a, q_b) \triangleq \frac{1}{h_D}[e(t, q_a, q_b)]^2 \tag{5}$$

The control loop can be configured to adjust only one of the input variables, or both at the same time. The parameters that need to be tuned are the integration factors γ_i, the dither amplitudes α_i and frequencies ω_i, and the high pass filter constant h_f.

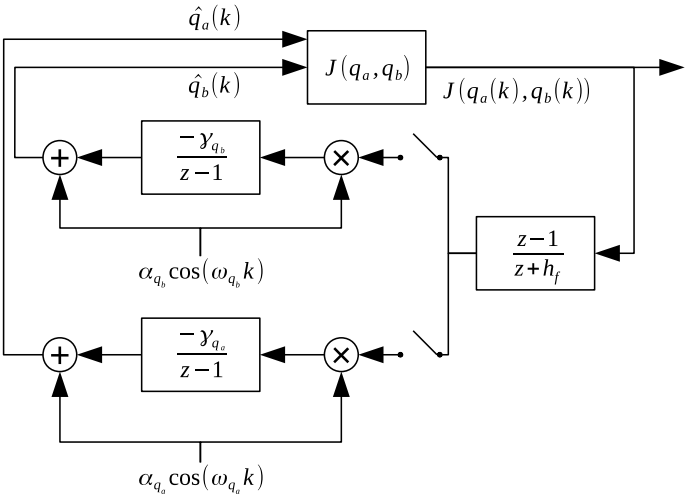


Figure 2 – Proposed Extremum-Seeking Control structure

5. Results and discussion

Prior simulations helped on finding the best parameters for the ESC algorithm on Figure 2, especially the sampling time, which is important for this type of algorithm.

On both experiments the buffered medium flow rate was set to 10 L/h, the reactor level to 60%, and the sampling time, h_D, was set to 15 s as the time-response of the system is slow.

5.1. SISO

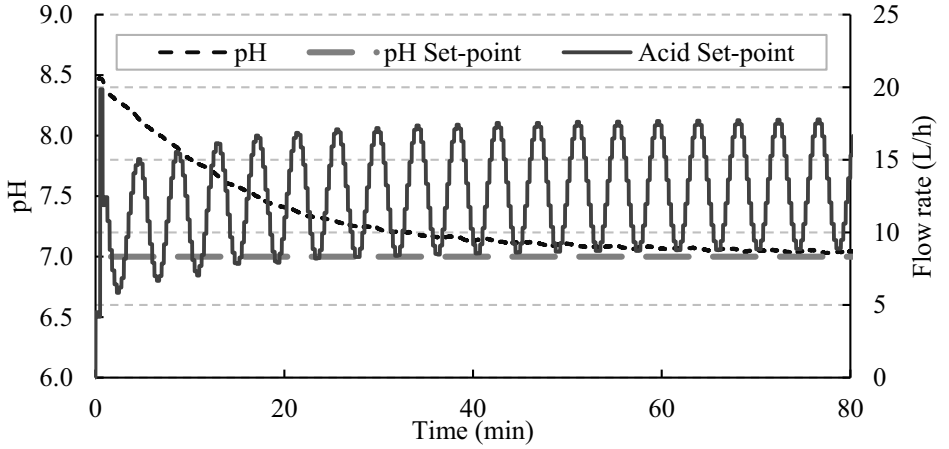


Figure 3 – SISO experimental results. Acid flow rate (FIC-101) as manipulated variable.

The parameters used for the ESC algorithm on this experiment can be seen on Table 2.

Table 2 – Parameters used for the SISO experiment

$h_f(-)$	$\gamma_{q_a}\,(L/h)$	$\alpha_{q_a}\,(-)$	$\omega_{q_a}\,(rad)$
0.5	55	4.5	$0.7^6\pi$

While greater dither amplitude α_{q_a} would increase the rate of convergence, it is also harder on the pump. Thus, the value of the integration factor γ_{q_a} was set high to compensate for the lack of dither. The frequency was set taking in mind the time-scale of both the plant and the pump, higher frequency could increase the convergence rate but the pump could not keep up. All parameters were set through trial and error considering all those factors.

The SISO algorithm managed to control the *pH* and find an optimum point where the cost function is almost zero. While small, the oscillations on the system response caused by the dither can be seen.

5.2. MISO

The parameters used for the ESC algorithm on this experiment can be seen on Table 3.

Table 3 – Parameters used for the MISO experiment

$h_f(-)$	$\gamma_{q_a}(L/h)$	$\gamma_{q_b}(L/h)$	$\alpha_{q_a}(-)$	$\alpha_{q_b}(-)$	$\omega_{q_a}(rad)$	$\omega_{q_b}(rad)$
0.5	55	35	4.0	3.5	$0.7^6\pi$	$0.7^5\pi$

The same observations from the SISO experiment are also valid here. Furthermore, the frequency of both dithers must be different. The integration factor of the base flow rate was set lower than for the acid flow rate as the effect it has on the *pH* is higher due to its higher concentration of base.

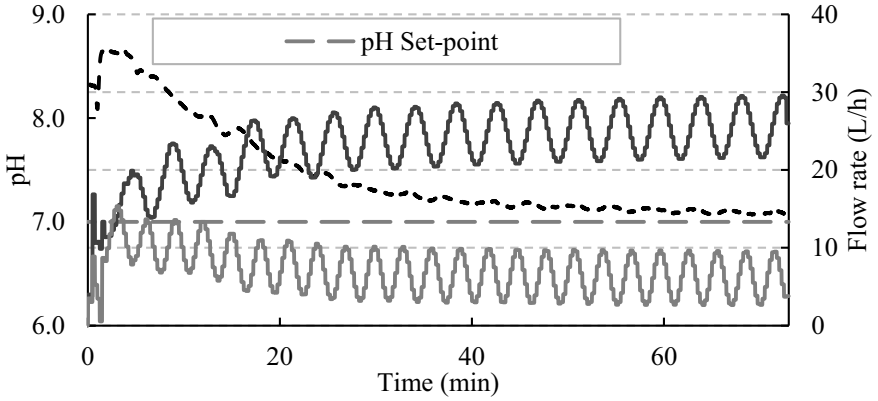


Figure 4 – MISO experimental results. Acid (FIC-101) and base (FIC-102) flow rates as manipulated variables.

The MISO algorithm was able to control the *pH* and find the optimum point where the cost function is almost zero. The oscillations on the system response caused by both dithers can be seen, and upon closer inspection it resembles the sum of both oscillations, as expected.

6. Conclusions

Two ESC algorithms were proposed in this work, a SISO and a MISO algorithms. The control strategy performance was first evaluated through simulations and then confirmed through experiments in a continuous neutralization plant. While the extremum-seeking control technique does not see much use on the chemical engineering area, it was able to properly control the pH of the continuous neutralization reactor, and show promising results for follow up works to increase the rate of convergence and reach the optimum, or set-point, faster. While a standard PI controller is able to control a system well when operating where it was tuned, ESC has the potential to control well on any range.

References

N.P. Cheremisinoff, 2002. Handbook of water and wastewater treatment technologies. Butterworth-Heinemann, Boston.

C. Chou -C. (Ed.), 2000. Handbook of sugar refining: a manual for design and operation of sugar refining facilities. Wiley, New York.

U. Ciri, M. Rotea, C. Santoni, S. Leonardi, 2017. Large-eddy simulations with extremum-seeking control for individual wind turbine power optimization: LES with ESC for power optimization. Wind Energy 20, 1617–1634. https://doi.org/10.1002/we.2112

M.A. Henson, D.E. Seborg, 1994. Adaptive nonlinear control of a pH neutralization process. IEEE Trans. Control Syst. Technol. 2, 169–182. https://doi.org/10.1109/87.317975

N.J. Killingsworth, M. Krstic, 2006. PID tuning using extremum seeking: online, model-free performance optimization. IEEE Control Syst. Mag. 26, 70–79. https://doi.org/10.1109/MCS.2006.1580155

S. Obut, C. Özgen, 2008. Online Identification and Control of pH in a Neutralization System. Ind. Eng. Chem. Res. 47, 4394–4404. https://doi.org/10.1021/ie070492p

I. Vandermeulen, M. Guay, P.J. McLellan, 2017. Formation control of high-altitude balloons experiencing real wind currents by discrete-time distributed extremum seeking control. IEEE, pp. 991–996. https://doi.org/10.23919/ACC.2017.7963082

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50114-2

Integrating Model Identification and Model-Based Control of Networked Process Systems

Amr Zedan, Da Xue, Nael H. El-Farra*

Department of Chemical Engineering, University of California, Davis, CA 95616, USA
nhelfarra@ucdavis.edu

Abstract

In this work, we present a methodology for the integration of model-based control and model identification for networked process systems subject to communication constraints. The methodology aims to enhance the stability and performance properties of the networked closed-loop system in the presence of process parameter variations while simultaneously reducing the overall rate of information transfer between the plant control subsystems. This is accomplished by incorporating within the time-triggered model-based control strategy an event-triggered model parameter update scheme that keeps the plant-model mismatch to a minimum and averts the need for increasing the communication rate in response to the destabilizing influence of process drift. A process monitoring scheme with an instability alarm is devised to trigger model identification whenever a certain instability threshold is breached. The results are demonstrated using a chemical process example.

Keywords: Networked process systems, model-based control, model identification.

1. Introduction

The problem of designing distributed and supervisory control systems for process networks has been the subject of significant research work in process control (e.g., Jillson and Ydstie, 2007; Rawlings and Stewart, 2008; Jogwar et al., 2009; Christofides et al. 2011; Cai et al., 2014). In addition to handling the dynamic interactions between the constituent subsystems, the management and optimization of information flow and communication between the plant subsystems are becoming increasingly important considerations in the control problem formulation (e.g., Sun and El-Farra, 2008; Zheng et al., 2015), especially in view of the growing reliance on networked sensor and control systems in plant operations and the increased emphasis placed on smart plant operations (Christofides et al., 2007). Beyond the economic savings achieved through the deployment and integration of wireless sensor networks in process monitoring and control systems, the increased operational flexibility, the ease of reconfiguration and real-time information coupling between the different plant layers enabled by this technology are appealing goals that promise to significantly enhance process operations.

An approach to address the integration of control and communication in the context of plant-wide control was presented in earlier work (Sun and El-Farra, 2008) where a quasi-decentralized model-based networked control structure that enforces closed-loop stability with minimal cross communication between the distributed control systems was developed. Within each local control system, a set of predictive models were used, in conjunction with local measurements, to generate the local control action at times when communication between the plant subsystems was suspended, and the models' states were updated when communication was permitted at discrete times. The

minimum communication rate required for closed-loop stability was shown to strongly depend on the size of the plant-model mismatch, with a larger mismatch generally requiring a higher rate of information transfer.

While a model-based control strategy generally helps reduce the control system's reliance on the communication medium, and therefore reduces its vulnerability to sensor failures and communication outages, the use of a single model with fixed parameters may limit the achievable savings in network resource utilization, especially when plant operating conditions experience variations and drift over time (e.g., due to catalyst deactivation in reactors and fouling in heat exchangers) causing increased plant-model mismatch. Such variations, if left unaddressed, could cause performance degradation and even instability. While increasing the communication frequency is one way of stabilizing the system (as it provides more frequent feedback that corrects for the increased uncertainty), this approach leads to higher communication costs and bandwidth over-utilization. An alternative approach is to update the model parameters instead, using model identification techniques (e.g., Overschee and Moor, 1996; Huang and Kadali, 2008). This allows keeping the plant-model mismatch to a minimum which ensures stability without the need to increase the communication frequency.

Motivated by this, we present in this work a framework for integrating model identification and model-based control in order to maintain closed-loop stability and reduce network resource utilization, while enhancing controller performance and accounting for plant drift and plant-model mismatch. Initially, a model-based controller with a well-characterized model state update rate is designed and implemented. An error detection scheme with a time-varying alarm threshold is devised to track the state evolution and trigger model parameter updates. When the instability threshold is breached, the communication rate is temporarily increased to avoid instability and, in the meantime, the data collected during this time period are used to identify a new model online based on subspace identification techniques. The networked closed-loop stability region associated with the new model is characterized and used to identify a suitable model state update rate that can restore the communication frequency to its original level. Finally, the results are illustrated using a chemical process example.

2. Preliminaries

We consider a plant comprised of a network of interconnected subsystems with the following state-space representation:

$$\dot{x}_i = A_i x_i + B_i u_i + \sum_{j=1, j\neq i}^{n} A_{ij} x_j \tag{1}$$

where x_i is the state vector, u_i is the manipulated input vector, A_i and B_i are the state and input matrices, respectively, all of the *i*-th subsystem, A_{ij} is the interconnection matrix which captures the coupling between the *i*-th and *j*-th subsystems, and *n* is the total number of subsystems. The various subsystems have local control systems that exchange state measurements over a resource-constrained communication medium. The control objective is to stabilize the overall system at the desired steady state, while simultaneously reducing unnecessary utilization of the communication medium. To simplify the presentation of the main results, we will focus on the state feedback control problem where the states of all the units are assumed to available as measurements.

3. Model-Based Control with Time-Triggered Communication

3.1. Local controller synthesis

To address the control objective, we consider a model-based control strategy similar to the one considered in Sun and El-Farra (2008). The main idea is to embed within the local control system of each unit a set of dynamic models which are used to provide estimates of the states of the neighbouring units in the event when communication between the plant subsystems is suspended. When communication is re-established, the model states are updated using the plant states, at discrete time instances, based on the sensor measured values. The resulting control and update laws are given by:

$$u_i(t) = K_i x_i(t) + \sum_{j=1.j\neq i}^{n} K_{ji}\,\hat{x}_j^i(t),\ t \in (t_k, t_{k+1}),\ \ i \in \{1,2,\cdots,n\} \tag{2}$$

$$\begin{aligned}\dot{\hat{x}}_j^i(t) &= \hat{A}_j\hat{x}_j^i(t) + \hat{B}_j\hat{u}_j^i(t) + \hat{A}_{ji}x_i(t) + \sum_{l=1,l\neq i,l\neq j}^{n} \hat{A}_{jl}\hat{x}_l^i(t)\,,\ t \in (t_k, t_{k+1})\\ \hat{u}_j^i(t) &= K_j\hat{x}_j^i(t) + K_{ji}x_i(t) + \sum_{l=1,l\neq i,l\neq j}^{n} K_{jl}\hat{x}_l^i(t), t \in (t_k, t_{k+1})\end{aligned} \tag{3}$$

$$\hat{x}_j^i(t_k) = x_j(t_k), \qquad k\ \in \{0,1,2,\cdots\} \tag{4}$$

where K_i is the local feedback gain, K_{ij} is the controller gain that compensates for the interactions, $\hat{x}_j^i$ is the state of the model capturing the dynamics of the *j*-th unit embedded in the *i*-th unit, $\hat{A}_j$, $\hat{A}_{ji}$, $\hat{A}_{jl}$ and $\hat{B}_j$ are constant matrices associated the model of the *j*-th unit, and t_k is the update time. Note that, for simplicity, the states of all models are assumed to be updated at the same time.

3.2. Closed-loop stability analysis

An important measure of the extent of network utilization is the update period, *h*, which is defined as the difference between two successive update times, i.e., $h = t_{k+1}$-t_k. To achieve closed-loop stability with minimal communication, it is important to characterize the maximum allowable update period for the control system which dictates the minimum communication frequency required for stability. This characterization can be obtained through a closed-loop stability analysis. Specifically, it can be shown (see Sun and El-Farra (2008) for the details) that the response of the closed-loop system of (1)-(4) in terms of the update period is given by:

$$\xi(t) = e^{\Lambda(t-t_k)}(I_s e^{\Lambda h} I_s)^k \xi_0 = e^{\Lambda(t-t_k)} M^k \xi_0,\ \ t \in (t_k, t_{k+1}) \tag{5}$$

for $k\ \in \{0,1,2,\cdots\}$, where $\xi(t) \coloneqq [\boldsymbol{x^T}(t) \quad \boldsymbol{e^T}(t)]^T$ is an augmented vector of the plant states and model estimation errors between the model and plant states, Λ is the augmented closed-loop matrix which depends on the plant and model matrices as well as the controller gains, I_s is an augmented identity matrix, and $\xi_0 = [\boldsymbol{x^T}(t_0) \quad \boldsymbol{0}]^T$ is the initial condition. By analysing this response, it can be verified that a necessary and sufficient condition for closed-loop stability is to have all eigenvalues of the stability test matrix *M* lie strictly inside the unit circle. This is represented by $\lambda_{max}(M)[A_i, A_{ij}, B_i, \hat{A}_i, \hat{A}_{ij}, \hat{B}_i, K_i, K_{ij}, h] < 1$, where λ_{max} is the maximum eigenvalue magnitude of *M*. This condition can be used to explicitly characterize the maximum allowable update period in terms of the plant-model mismatch and controller gains.

4. Integrating Model Identification and Model-Based Control

In this section, an overview of the proposed methodology is presented first, and is then followed by a discussion of the key implementation issues.

4.1. Overview of the proposed methodology

Fig. 1 summarizes the proposed approach for integrating model identification and model-based control. Initially, the control system is operated at a suitably chosen update period that minimizes communication while ensuring closed-loop stability. The evolution of the closed-loop state is continuously monitored and checked against an instability alarm threshold to determine when the model parameters need to be updated. In the event of a threshold breach, a "safe-parking" protocol is activated to momentarily stabilize the closed-loop system while additional data is collected and a new model is identified. The new model is then checked for stability, and the corresponding communication rate is determined. Finally, the model parameters are updated and the new communication rate is implemented. The algorithm is repeated every time the instability alarm threshold is breached.

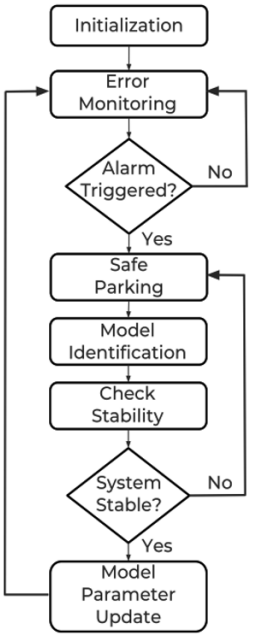


Fig. 1. Flowchart of the algorithm integrating model identification and model-based control.

4.2. Triggering criterion

To determine when a new model needs to be identified, a process monitoring scheme is implemented whereby the closed-loop states are continuously checked to detect potential instabilities caused by the drift in plant parameters. An instability alarm can be designed on the basis of the nominal closed-loop response in (5). Specifically, it can be verified that, in the absence of process parameter variations, the following time-varying bound is satisfied:

$$\|x(t)\| \leq \alpha \cdot e^{-\beta(t-t_0)} \cdot \|x(t_0)\| \tag{6}$$

where $\alpha > 1$ and $\beta > 0$ are tunable constants that depend on the choice of the controller design parameters. Using this bound as a time-varying alarm threshold, instability can be detected once the norm of the state violates the threshold. This acts as a trigger for subsequent steps in the process, culminating in the update of the model parameters. Note that, unlike the model state updates which are time-triggered periodically, the model parameter updates are event-triggered and thus occur less frequently.

4.3. Safe parking and model identification

Once the instability threshold is breached, the plant is switched temporarily to a safe-parking mode of operation where the model state update period is reduced for a certain period of time. The temporary increase in the communication rate serves to maintain closed-loop stability and to collect sufficient data for the model identification step. Model identification is then triggered and new model parameters are obtained. While the choice of the identification method depends on the particular application, subspace techniques are used in this study to identify the new model matrices based on the closed-loop input and state data to keep plant-model mismatch to a minimum. Note that

in the case of time-varying uncertainty where the mismatch continues to deteriorate over time, the safe-parking step may need to be repeated to ensure closed-loop stability.

4.4. Stability check

Once a new model is identified, the closed-loop stability properties are analysed based on the eigenvalues of the new stability test matrix M to determine the range of stabilizing update periods associated with the new model. If either no stabilizing update period exists, or the maximum allowable update period is less than the safe-parking update period, the plant is kept in the safe-parking mode until a stabilizing model with a lower update frequency requirement can be found (possibly using a different identification method); otherwise, the model parameters are updated and the corresponding update period is implemented.

5. Application to a chemical process example

We consider a networked plant comprised of two interconnected non-isothermal continuous stirred tank reactors with a material recycle stream. The control objectives are to stabilize the system at the open-loop unstable steady-state using the heat input rates as the manipulated variables, to minimize the required communication between the two control systems and to account for process parameter variations. The process modelling and controller synthesis details are omitted due to space limitations.

To simulate plant parametric drift, we introduce uncertainty in the plant parameters. We define $\delta = (\Delta H^m - \Delta H)/\Delta H$, where ΔH^m is the nominal values of the heat of reaction used in the plant models. The operating conditions were chosen such that the model states are updated every 4.8 minutes and that there is initially no mismatch between the model and plant states; that is, $h = 0.080$ h and $\delta_1 = 0$. After 1 hour of closed-loop stable operation (see Fig. 2), an uncertainty of $\delta_1 = -0.178$ is introduced in the plant causing a mismatch between the plant and model parameters. It can be seen in Fig. 2 that, starting at approximately the 4th hour of operation, the reactor temperature starts to diverge from the operating steady-state in an oscillatory fashion until the instability alarm threshold is breached and the safe-parking mode is triggered to regain stability.

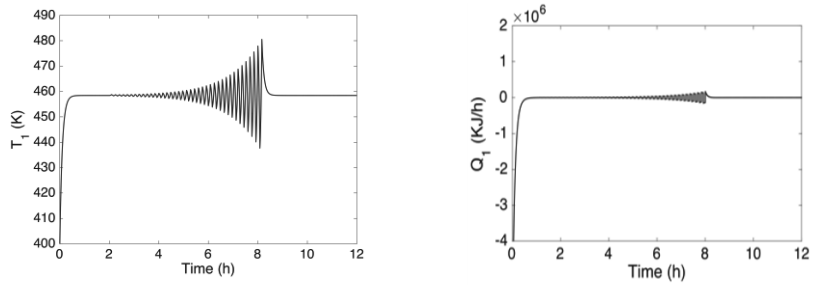


Fig. 2. Closed-loop temperature (left) and heat input rate (right) profiles for the first reactor.

The safe-parking update period is determined from the stability plot in Fig. 3 (see dashed line) which shows that the update period can be reduced to $h = 0.073$ h to stabilize the system. It is noted that it is sufficient to increase the communication frequency to achieve stability; however, the objective here is to maintain stability with low communication frequencies, and model re-identification allows us to continue achieving this balance even after the process drifts. After data has been collected during the safe-parking period, the plant model is re-identified at time = 10 h. At that point, the

stability region of the new model is characterized based on the new stability matrix M and the update period is increased to $h = 0.084$ h to switch the plant out of the safe-parking mode; and, as seen in Fig. 2, maintain stability for the rest of the simulation.

The effect of uncertainty and model re-identification on closed-loop stability can be seen in Fig. 3 which shows the stabilizing ranges for the update period before the uncertainty was introduced, after the error was introduced and the safe-parking mode was invoked, and after the model re-identification took place. Eigenvalues below the critical unit line correspond to stable update periods. It can be seen that when the error is introduced, the stability region shrinks; however, after model re-identification takes place, the stability region increases, thus yielding a larger feasible range of update periods. This translates into stabilization with a lower communication frequency after model uncertainty is introduced.

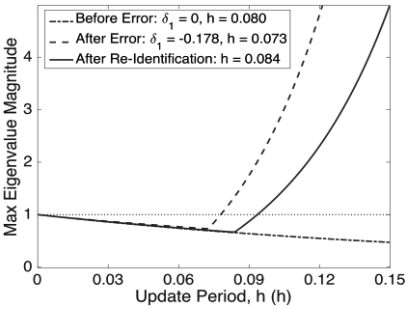


Fig. 3. λ_{max} (M) as a function of the update period under normal operation (dash-dotted), under plant-model mismatch (dashed) and for the identified model (solid).

6. Conclusion

A framework for the integration of time-triggered model-based control and event-triggered model identification for networked process systems subject to process parameter variations and communication constraints was presented in this work. It was shown that by minimizing the plant-model mismatch, model identification techniques can help enhance closed-loop stability with reduced communication requirements. Future work will focus on generalizing this approach to address the output feedback control problem and also explore the effect of the particular choice of the model-identification method on the achievable savings in communication resource utilization.

References

X. Cai, M. Tippett, L. Xie, and J. Bao, 2014, Fast Distributed MPC Based on Active Set Method, Comp. Chem. Eng., 71, 158–170.

P. Christofides, J. Davis, N. El-Farra, D. Clark, K. Harris, and J. Gipson, 2007, Smart Plant Operations: Vision, Progress and Challenges, AIChE J., 53, 11, 2734-2741.

P. Christofides, J. Liu, and D. de la Pena, 2011, Networked and Distributed Predictive Control: Methods and Nonlinear Process Netwrok Applications, Springer-Verlag, London.

B. Huang and R. Kadali, 2008, Dynamic Modeling, Predictive Control and Performance Monitoring: A Data-driven Subspace Approach, Springer-Verlag, London.

K. R. Jillson and B. E. Ydsite, 2007, Process Networks with Decentralized Inventory and Flow Control, J. Proc. Contr., 17, 399–413.

S. Jogwar, M. Baldea, and P. Daoutidis, 2009, Dynamics and Control of Process Networks with Large Energy Recycle, *Ind. Eng. Chem. Res.*, 48, 6087–6097.

P. Overschee and B. Moor, 1996, Subspace Identification for Linear Systems: Theory, Implementation, Application, Kluwer Academic Publishers, Boston, MA.

J. Rawlings and B. Stewart, 2008, Coordinating Multiple Optimization-Based Controllers: New Opportunities and Challenges, J. Proc. Contr., 18, 839–845.

Y. Sun and N. H. El-Farra, 2008, Quasi-decentralized Model-Based Networked Control of Process Systems, Comp. Chem. Eng., 32, 9, 2016-2029.

C. Zheng, M. Tippett, J. Bao, and J. Liu, 2015, Dissipativity-Based Distributed Model Predictive Control with Low Rate Communication, AIChE J., 61, 3288–3303.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50115-4

Reduced Order Nonlinear Multi-parametric Model Predictive Control of Large Scale Systems

Panagiotis Petsagkourakis[a], Constantinos Theodoropoulos[a*]

[a]*School of Chemical Engineering and Analytical Science,The University of Manchester, M13 9PL, UK*
[*]*k.theodoropoulos@manchester.ac.uk*

Abstract

Multiparametric MPC (mp-MPC) is an efficient methodology for computing fast control actions. Control of distributed parameter systems (DPS) remains a challenging task, as the system dynamics are infinite-dimensional. Model reduction of such systems may produce instabilities and thus it is essential that the model reduction methodology used is robust. In this work, a Galerkin-based, efficient model reduction is employed for DPS. The reduced order model is then utilized by the mp-MPC, introducing a novel strategy for computing the critical regions of the approximated multiparametric nonlinear problem. This strategy initializes the solution effectively, reducing the number of the required critical regions and the corresponding computational time. The effectiveness of the proposed algorithm is demonstrated for a tubular reactor where an exothermic reaction takes place.

Keywords: Model reduction, Distributed parameter systems, Nonlinear MPC

1. Introduction

Constrained model predictive control (MPC) is a powerful control strategy, where on-line optimization and a receding horizon are employed to compute the optimum trajectory for the manipulated variables. The use of MPC for DPS, consisting of partial differential equations (PDEs) is computationally expensive (Theodoropoulos 2011, Edgar et al. 1999). The dissipative nature of such systems can be exploited for model order reduction (MOR). Dissipativity can be expressed as the separation of scales in the eigenvalues of the linearized models (Xie et al. 2015). Nevertheless, the computation of the online control action can be problematic even with the use of a reduced model. Mp-MPC has proven to be a powerful control methodology that computes off-line the control actions as functions of the states (Pistikopoulos 2012), overcoming the online computational burden of the classic MPC. Despite the fact that the majority of computations are performed off-line, the large size of the physical system may produce an intractable problem. To tackle this, Rivotti et al. (2012) combine nonlinear mp-MPC (Dominguez and Pistikopoulos 2011) with nonlinear MOR, which is based on balancing of empirical gramians to reduce the size of the system. However, this approach can still be prohibitive for complex large-scale systems, modelled by large sets of PDEs.

In this work, an efficient Galerkin projection-based MOR is applied in order to i) reduce infinite dimensional systems into considerably smaller models and ii) ensure their robustness. The latter is crucial since small open-loop errors may be amplified under the closed-loop conditions, creating spill-over effects (Christofides 2012). Nevertheless, even a reduced system may be computationally expensive, as the number of the critical regions increases steeply with problem size. A novel method is proposed in this work

that can initialize the solution effectively, whilst reducing the number of critical regions and the computational time.

2. Galerkin-based Proper Orthogonal Decomposition

The focus of this work is on feedback control of spatially distributed processes described by dissipative PDEs with the following state-space description, with mixed boundary conditions:

$$\frac{\partial x}{\partial t} = \mathcal{A}\left(\frac{\partial x}{\partial z}, \frac{\partial^2 x}{\partial z^2}, \dots, \frac{\partial^n x}{\partial z^n}\right) + B(z)u + F(x), y = \int_V c(z)xdz \tag{1}$$

where $\mathcal{A}$ is a dissipative, possibly nonlinear, spatial differential operator, $F(x)$ is a nonlinear function, V the spatial domain of the physical system, x the system states, u the manipulated, and y the output variables. The physical system that is represented in Eq. 1 is infinite dimensional and a robust finite dimensional approximation is needed so that spill-over effects are avoided. This is achieved by employing proper orthogonal decomposition (POD) with Galerkin projection. POD is one of the most efficient model reduction techniques, where data compression retains the most 'energy' of the system. The 'energy' of a given mode is related to the magnitude of the eigenvalue corresponding to that mode. The global eigenfunctions (ϕ_i) that efficiently span the system are orthonormal and can be computed with the method of snapshots (Sirovich 1987) after collecting data points using either an experimental set-up or a dynamic simulator. Every basis function is computed as the solution of an integral eigenvalue problem (Xie et al. 2015) and each eigenvalue is associated with a global eigenfunction. Then, the finite dimensional approximation of the infinite dimensional PDE is derived using the Galerkin projection. The infinite dimensional states can be approximated as follows:

$$x(t,z) = \sum_{i=1}^{\infty} a_i(t)\phi_i(z) \simeq \sum_{i=1}^{l} a_i(t)\phi_i(z) + \bar{x} \tag{2}$$

where l is the number of modes that captures the most of the energy, a_i the time coefficients, ϕ_i the orthonormal eigenfunction obtain by POD. Substituting Eq. 2 to Eq.1 and computing the Galerkin projection (taking the $\mathcal{L}_2$ inner product with the eigenfunctions) we get the following for $i = 1 \dots l$:

$$\frac{da_i}{dt} = \int_V \phi_i(z)[\mathcal{A}\left(\sum_j a_j(t)\,\phi_j(z) + \bar{x}\right) + B(z)u + F\left(\sum_j a_j(t)\,\phi_j(z) + \bar{x}\right)]dz \tag{3}$$

The last equation is exploiting the orthonormality of the eigenfunctions in the sense of $\mathcal{L}_2$ inner product. The system now is approximated using a small finite number of modes a_i. The continuous time system in Eq.3 is converted in discrete form using an appropriate integrator (implicit or explicit).

3. Multiparametric Nonlinear MPC

The general mp-MPC case aims to solve the optimization problem for a varying vector of parameters, θ, that are contained in a fixed set, ϑ. The nonlinear continuous multiparametric problem can be stated as:

$$J(\theta) = \min_{\chi} f(\chi, \theta) \tag{4}$$
$$G(\chi, \theta) = 0$$
$$h(\chi, \theta) \leq 0$$

$\theta \in \vartheta$

where χ are the degrees of freedom of the problem, J is the optimum solution, f the nonlinear objective function, G are the equality constraints that usually include the equations describing the physical system, and h the inequality constraints, which include any bounds for the degrees of freedom. MPC for the reduced order system can be written as a nonlinear mp-optimization problem with $\theta = \begin{bmatrix} \alpha^T & u_{-1}^T & y_{ref}^T \end{bmatrix}^T$:

$$J(\theta) = \min_{\alpha,u} h(\alpha, u, \theta) = \min_{\alpha,u} \sum_{i=1}^{N} \left\| y_i - y_{ref} \right\|_Q^2 + \sum_{i=1}^{M} \|\Delta u_i\|_R^2 \quad (5)$$
$$u_i = u_{i-1} + \Delta u_i,\ y_i = C\alpha_i$$
$$G(\alpha, u, \theta) = \alpha_{i+1} - g(\alpha_i, u_i) = 0$$
$$h(\alpha, u, \theta) \leq 0$$
$$A\theta \leq b$$

where $\alpha_i = \alpha(t_i) = [a_1(t_i) \dots a_l(t_i)]^T$, N, M are the prediction and control horizon, respectively, u_{i-1} the manipulated variable at the previous time point, y_{ref} the set-point and g the temporal discretization of Eq. 3.

Mp-nonlinear methodologies, such as the one in (Jahansen 2002), which can be used to solve Eq.5, aim to divide the parameter space in many partitions and approximate Eq. 5 with multiple mp-QP. The state space is partitioned into hyper-rectangles and the main advantage for convex problems is that the approximation error can be a priori computed. In the case of non-convex problems, it is proposed (Johansen 2002) to compare the solution of the implicit NLP and the mp-QP problem at the vertices of each region. This approach appears to approximate the nonlinear problem adequately well, although the computational cost increases steeply with the size of the system. In this work, we develop a number of novel solutions in order to improve the computational efficiency and robustness of nonlinear mp-MPC.

3.1 *The proposed algorithm*

Our proposed algorithm begins with the multiple linearization of the system's dynamics considered to be the main contribution to the nonlinearity of the problem. This means that the approximation of Eq. 5 with multiple QPs can be seen as optimization with a multi-linear model (Bonis et al. 2014) for the corresponding system. To take advantage of this approach, a piecewise affine (PWA) identification problem is considered, where the reduced order system is approximated by a PWA function by employing clustering and piecewise linearization. The collected trajectories (snapshots) for the POD-basis construction can be also used to identify a number of clusters using k-means (Hastie et al. 2009). Cluster centroids can be utilized as linearization points and cluster polyhedral partitions (given by the Voronoi diagram of the clusters' centroids) are used to compute the boundaries of each linear model (multi-category discrimination can be also applied (Breschi et al. 2016)). This step will give a good approximation of the system for different convex regions without separating the phase space in hundreds of partitions as in (Johansen 2002). The solution of mp-multi-linear MPC for the PWA approximation can be then computed. The exact solution of hybrid MPC (Oberdieck and Pistikopoulos 2015) could also be used as an initial approximation but the computational cost would increase. After constructing the initial solution for the critical regions, the approximated solution is compared at the vertices of each critical region with the implicit NLP (Eq. 5). If the result is not satisfactory with respect to ta chosen tolerance, then a refinement step follows.

We, subsequently apply a multiple shooting method to improve numerical stability of the algorithm (side-stepping sequential approaches that are not appropriate for stiff problems) at a point inside each partition (computed through the Chebyshev center). Then, an efficient procedure is built to construct and solve the local mp-QP problem (one for each partition) in the form of Eq. 6.

$$J(\theta) = \min_{U} \frac{1}{2}\Delta U^T H\, \Delta U + (B + F(\theta - \theta_0)^T \Delta U + J^* \tag{6}$$
$$C\Delta U \leq d\Delta\theta + E$$
$$A\theta \leq b$$

$H = \nabla^2_{UU} J|_{\theta_0,U^*}$, $F = \nabla^2_{U\theta} J|_{\theta_0,U^*}$, $B = \nabla_U J|_{\theta_0,U^*}$, $C = -\nabla_U h|_{\theta_0,U^*}$, $d = -\nabla_\theta h|_{\theta_0,U^*}$, $E = -h$ all computed for the optimum set of parameters. The number of the new partitions depends on the size of the critical region but not the size of the parameter space. First and second order derivatives with respect to degrees of freedom and the varying parameters are computed, through numerical perturbations of the equality constraints, and are then substituted in the objective function and in the inequality constraints to produce Eq. 6. If J^* is the optimal solution obtained with the multiple shooting method, for parameters θ_0 and optimal degrees of freedom U^* then, for example, the derivative of the objective function with respect to parameter θ_0 is computed as follows:

$$\frac{\partial h}{\partial \theta_0^i} = \frac{h\left(\alpha_{U^*,\theta_0+\epsilon}, U^*, \theta_0 + \epsilon\right) - h(\alpha_{U^*,\theta_0-\epsilon}, U^*, \theta_0 - \epsilon)}{2\epsilon} \tag{7}$$

where α_{U^*,θ_0} is the solution of $G(\alpha, U^*, \theta_0)$=0 and $\alpha_{U^*,\theta_0+\epsilon}$ the solution of $G(\alpha, U^*, \theta_0 + \epsilon)$=0. This refinement procedure continues until an a priori tolerance is satisfied. If the new mp-approximation creates only one critical region that still does not provide an appropriate solution then the region should be divided according to a geometrical criterion (in this work a triangulation is employed). This approach may produce complex convex geometrical regions; however, the number of regions does not increase as steeply as in (Johansen 2002), and because of the good initial approximation the result doesn't produce many additional critical regions. In the next section, this algorithm is employed to an infinite dimension problem in order to illustrate its computational capabilities.

4. Application

To illustrate the effectiveness of the proposed algorithm, a chemical engineering application is considered, particularly an exothermic tubular reactor. The physical system is described by two partial differential equations with Neumann boundary conditions:

$$\frac{\partial c}{\partial t} = \frac{1}{Pe_1}\frac{\partial^2 c}{\partial z^2} - \frac{\partial c}{\partial z} - Da\, c\, e^{\gamma \frac{T}{1+T}} \tag{8a}$$
$$\frac{\partial T}{\partial t} = \frac{1}{Pe_2}\frac{\partial^2 T}{\partial z^2} - \frac{\partial T}{\partial z} - BDa\, c\, e^{\gamma \frac{T}{1+T}} + b(T_w - T) \tag{8b}$$

Here c and T are the dimensionless concentration and temperature respectively, while T_w is the temperature of the cooling zone. The system's parameters are $Pe_1 = Pe_2 = 7$,$Da = 0.1$, $B = 2.5$, $b = 2$,and $\gamma = 6$. The eigenfunctions are computed employing the method described in section 2, where after the eigenanalysis 2 dominant modes capturing 99.8% of the system's energy are computed. It should be noted that the

number of dominant modes depends on the accuracy required, as well as on the phenomena that dominate the physical system. When diffusion is dominant, a small number of modes is required. When convection dominates, a (slightly) larger number of modes is typically required to capture the system's energy. Then, Eq. 8a, b are projected onto the slow modes using Galerkin projection in order to produce an accurate reduced-order discretized model from the infinite-dimensional one. The reduced model is then used to construct the NLP problem (with Q= Identity matrix (I) and R=0.01I) at each Chebyshev center with $-1 \leq u_i \leq 1$ and $-1 \leq \alpha_i \leq 1$ with control and prediction horizon being 2 time steps. The collected trajectories can then be used to identify the clusters as in section 3.1 and the, mp problem is subsequently solved. The facet in Fig. 1a depicts two reduced variables after the mp-multi-linear MPC solution with 23 critical regions and the final solution is depicted in Fig.1b with tolerance=0.002. Overall 66 critical regions (~1 CPU-hour) are constructed through the refinement process. For demonstration purposes the method of Johansen is also applied here, producing 841 critical regions (over 4 CPU-hour). Then, the controllers are validated for different set-points and the results are depicted in Fig.2. The solution given by the multi-linear MPC has a large offset, but the final solution is adequately close to the actual implicit nonlinear MPC, and as a matter of fact closer to the one provided by Johansen (2002) for the same a priori error bound. This shows the effectiveness of the proposed method. For demonstration purposes, 4 modes were selected for the model reduction and the same algorithm was applied. The algorithm produced 126 critical regions in ~3 CPU-hours.

All the simulations are carried out using a i7 and 16 RAM single threaded machine. All the calculations are implemented in MATLAB, and each mp QP as well as each NLP problem, is solved using the toolbox POP (Pistikopoulos 2012) with NAG Mark 25. It should be noted that global optimization technique will be considered in a future publication.

5. Conclusions

We have developed a reduced order mp MPC suitable for infinite dimensional systems such as ones described by PDEs. POD-Galerkin is employed to produce a reduced order system which is then employed within a new nonlinear mp-MPC algorithm. The effectiveness of the proposed methodology is illustrated through the MPC of an exothermic tubular reactor described by complex nonlinear PDEs. The offline complexity is reduced and the approximated solution is obtained with computational efficiency using a multi-linear solution as an initial partition.

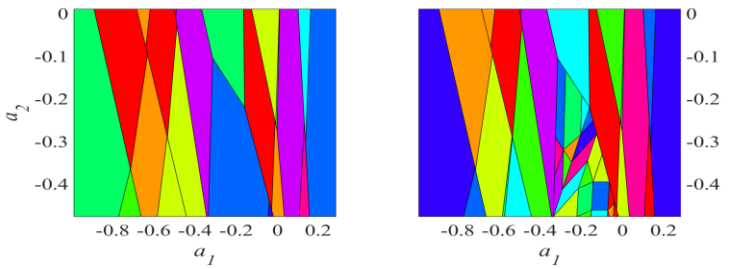


Figure 1: Critical regions for (a) Multi-Linear model (b) Proposed method

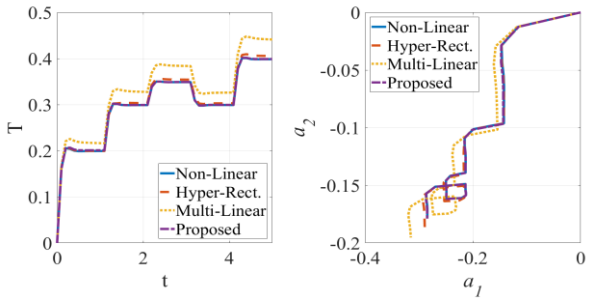


Figure 2: (a) solution for different setpoints and (b) phase space of 2 reduced variables and

References

I. Bonis, W. Xie, C. Theodoropoulos, 2014, Multiple model predictive control of dissipative PDE systems, IEEE Transactions on Control Systems Technology, 22(3), 1206-1214

V. Breschi, D. Piga, A. Bemporad, 2016, Piecewise affine regression via recursive multiple least squares and multicategory discrimination, Automatica, 73, 155-162

P. D. Christofides, 2012, Nonlinear and robust control of PDE systems: Methods and applications to transport-reaction processes, Springer Science & Business Media.

L. F. Domínguez and E. N. Pistikopoulos, 2011, Recent Advances in Explicit Multiparametric Nonlinear Model Predictive Control, Ind. Eng. Chem. Res., 50 (2), pp 609–619T. F. Edgar, W. J. Campbell, C. Bode, 1999, Model-based control in microelectronics manufacturing, Proceedings 38th IEEE Conf Decis Control 4:85–91.

T. Hastie, R. Tibshirani, J. Friedman, 2009, The Elements of Statistical learning: Data Miningm Inference and Prediction, Springer, 2

T. A. Johansen, 2002, On multi-parametric nonlinear programming and explicit nonlinear model predictive control, Proceedings of the 41st IEEE Conference on Decision and Control, 3, 2768-2773.

R. Oberdieck, E. N. Pistikopoulos, 2015, Explicit hybrid model-predictive control: The exact solution, Automatica, 58, 152-159

E. N. Pistikopoulos, 2012. From multi-parametric programming theory to MPC-on-a-chip multi-scale systems applications. Computers and Chemical Engineering

P. Rivotti, R. S. C. Lambert, E. N. Pistikopoulos, 2012, Combined model approximation techniques and multiparametric programming for explicit nonlinear model predictive control. Computers and Chemical Engineering, 42, 277–287.

L.Sirovich, 1987, Turbulence and the dynamics of coherent structures, parts i–iii. Quarterly of Applied Mathematics, 45, 561–590

C. Theodoropoulos, 2011, Optimisation and Linear Control of Large Scale Nonlinear Systems: A Review and a Suite of Model Reduction-Based Techniques, Lecture Notes in Computational Science and Engineering, vol 75. Springer, Berlin

W. Xie,I. Bonis, C. Theodoropoulos, 2015, Data-driven model reduction-based nonlinear MPC for large-scale distributed parameter systems, Journal Process Control 35:50–58.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50116-6

Partitioned Based Cooperative Distributed Model Predictive Control for Large-Scale Nonlinear Systems

Rosiane R. Rocha[a], Luís C. Oliveira-Lopes[b*]

[a] *Federal Institute of Espírito Santo, ES-010 km 6.5, Manguinhos, CEP 29173-087, Serra, Espírito Santo, Brazil*
[b] *School of Chemical Engineering, Federal University of Uberlândia, Av. João Naves de Ávila, 2121, Santa Mônica, CEP 38400-902, Uberlândia, Minas Gerais, Brazil*
[*] *lcol@ufu.br*

Abstract

A distributed model predictive control (DMPC) strategy brings interesting features of topology, flexibility and maintenance to large-scale nonlinear systems. This work introduces two new cooperative distributed nonlinear model predictive control strategies with automatic partitioning. The first approach is based on oriented graphs of the linearized model of the plant and the second one based on incidence matrices of the nonlinear model of the plant. The proposed strategies were applied to a benchmark plant and compared to other control structures, obtaining promising results.

Keywords: Cooperative Distributed Model Predictive Control, Large-Scale Systems and Partitioning.

1. Introduction

Controlling of large-scale nonlinear systems is still a challenging problem. A distributed model predictive control (DMPC) strategy brings interesting features of topology, flexibility and maintenance to this scenario. The distributed control strategy can be divided into two types: non-cooperative, where each local controller optimizes a local objective function and the cooperative, where each local controller optimizes a global objective function. This paper focuses on the study of the cooperative approach, which may improve performance when compared to the non-cooperative structure by requiring each subsystem to consider the effects of local control actions on all subsystems in the network as presented in the study of Venkat et al., (2008).

The selection of pairs of manipulated input and controlled outputs belonging to each subsystem is still made in most cases based on the physical structure of the plant (Christofides at al., 2013; Jogwar et al., 2009; Kumar & Dautidis, 2002). This is simple for a small-scale system but can become an arduous task in large-scale systems applications. So, the study of a systematic methodology of partitioning presents great applicability in applications of DMPC, mainly in large scale plants.

The objective of this study is to introduce and evaluate two new cooperative distributed nonlinear model predictive control strategies with automatic partitioning. The paper is organized through the following sections: section 2 has the problem formulation, presenting the decomposition methodologies and the cooperative DMPC controller

design. In section 3 a case study is presented along with the results and discussions. Finally, we draw the main conclusions in section 4.

2. Problem Formulation

Let a process be assumed to be described by the following nonlinear ordinary differential equation system: $\dot{\mathbf{x}} = \mathbb{f}(\mathbf{x}, \mathbf{u})$ and $\mathbf{y} = \mathbf{g}(\mathbf{x})$, where $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{u} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^l$ are state variables vector, manipulated variables vector and controlled variables vector, respectively.

This paper proposes two new cooperative distributed nonlinear model predictive control strategies with automatic partitioning. The approached are presented in Figure (1), where the controllers perform the steps presented to calculate their respective control actions. The control methodologies, using the respective partitioning method will be presented in the next section.

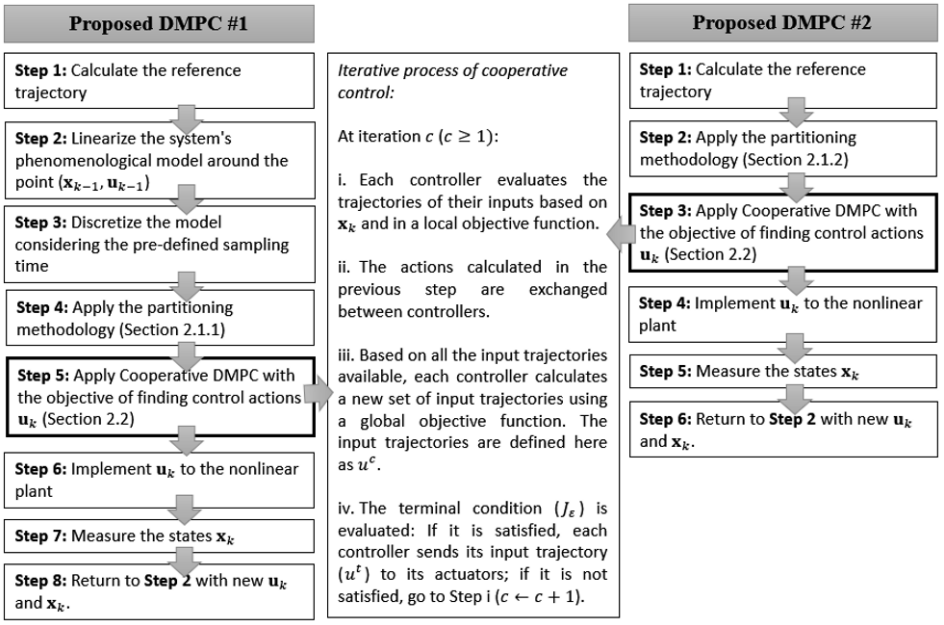


Figure 1: Cooperative DMPC schemes proposed in this work, with emphasis on the iterative process of cooperative control methodology.

2.1. The partitioning methodology

This section presents the partitioning methodologies proposed in this work. The first approach refers to the partitioning from a linear local model of the plant and the second approach from the nonlinear model.

2.1.1 Proposed Cooperative DMPC #1 - Partitioning for linear local model

Partitioning from the linear model of the plant can be applied when the model is linearized around a point of interest. For application in non-linear processes, the non-linear model is linearized around the condition ($\mathbf{x}_{k-1}, \mathbf{u}_{k-1}$) and discretized as a function of the sampling time T_S, obtaining the locally discrete time representation of the process model. The details of the obtained model can be found in Rocha and Oliveira-Lopes

(2016b) and the details of the applied partitioning process are presented in Rocha and Oliveira-Lopes (2016a).

2.1.2 Proposed Cooperative DMPC #2 - Partitioning for non-linear model

In this approach, the partitioning will be applied directly to the non-linear model of the process. In order to perform that consider the following definitions:

Definition 1: Define the incidence matrix of manipulated inputs (**E**), identifying at each $\dot{x}_i = \mathbb{f}(\mathbf{x}, \mathbf{u})$, for $i = 1, \cdots, n$ of the phenomenological model, which input affects $\dot{x}_i$. If there is an effect, the matrix element $e_{ij} = 1$, otherwise $e_{ij} = 0$, where e_{ij} is the *i*-th row and *j*-th column of the incidence matrix of the manipulated inputs, **E**.

Definition 2: Define the incidence matrix of the states (**S**), identifying at each $\dot{x}_i = \mathbb{f}(\mathbf{x}, \mathbf{u})$, for $i = 1, \cdots, n$ of the phenomenological model, which state affects $\dot{x}_i$. If there is an effect, the matrix element $s_{ij} = 1$, otherwise $s_{ij} = 0$, where s_{ij} is the *i*-th row and *j*-th column of the incidence matrix of states, **S**.

Definition 3: Define the incidence matrix of the controlled outputs (**C**), identifying at each $y_k = \mathbf{g}(\mathbf{x})$, for $k = 1, \cdots, l$ of the phenomenological model, which state affects y_k. If there is an effect, the matrix element $c_{kj} = 1$, otherwise $c_{kj} = 0$, where c_{kj} is the *k*-th row and *j*-th column of the incidence matrix of controlled outputs, **C**.

Definition 4: Define a set $\boldsymbol{\varphi}_j = \{\alpha | e_{\alpha j} = 1, \alpha = 1, \cdots, n\}$, in which $e_{\alpha j}$ is the α-th row and *j*-th column element in the **E** matrix. The set $\boldsymbol{\vartheta}_j = \{x_\alpha, \alpha \in \boldsymbol{\varphi}_j\}$ represents the states directly controlled by the input u_j.

Definition 5: Define a set $\boldsymbol{\xi}_j = \{\alpha | s_{\alpha j} = 1, \alpha = 1, \cdots, n\}$, in which $s_{\alpha j}$ is the α-th row and *j*-th column element in the **S** matrix. The set $\boldsymbol{\varrho}_j = \{x_\alpha, \alpha \in \boldsymbol{\varphi}_j\}$ represents the states directly affected by the state x_j.

Definition 6: Define a set $\boldsymbol{\pi}_j$ with the dynamic behaviors of the states ($\dot{x}_i = \mathbb{f}(\mathbf{x}, \mathbf{u})$) that must be included in the submodel $\mathbf{M}\{\boldsymbol{\pi}_j\}$. Each submodel $\mathbf{M}\{\boldsymbol{\pi}_j\}$ will be defined based on the mapping of each state belonging to the set $\boldsymbol{\varphi}_j$ and affecting the output space (from matrix **C** previously defined). If any controlled output is not directly affected by any input, the set $\boldsymbol{\xi}_j$ must be added to the submodel thus including the indirect effects. This setting ensures that the partitioning method is based on the effects of each input on the controlled outputs.

The system partition proposed is based on the Definitions (1-5). A step-by-step algorithm for the introduced decomposition strategy is given as follows:

Step 1: create the set $\boldsymbol{\varphi}_j$ and $\boldsymbol{\xi}_j$, as well as their associated sets $\boldsymbol{\vartheta}_j$ and $\boldsymbol{\varrho}_j$. $\boldsymbol{\vartheta}_j$ is based on the incidence matrix of the manipulated inputs **E**. $\boldsymbol{\varrho}_j$ is based on the incidence matrix of states **S**;

Step 2: create the set $\boldsymbol{\pi}_j$ with the states belonging to the sets $\boldsymbol{\vartheta}_j$ and $\boldsymbol{\varrho}_j$ affecting directing the states that affect the controlled outputs space;

Step 3: compose submodel $\mathbf{M}\{\boldsymbol{\pi}_j\}$ with elements of the set $\boldsymbol{\pi}_j$;

Step 4: merge $\mathbf{M}\{\boldsymbol{\pi}_j\}$ and $\mathbf{M}\{\boldsymbol{\pi}_p\}$ $(j \neq p)$ submodels if $\boldsymbol{\pi}_j$ and $\boldsymbol{\pi}_p$ are identical sets. Create a new set of inputs by merging u_j and u_p inputs.

Step 5: check if the submodel $\mathbf{M}\{\boldsymbol{\pi}_j\}$ is controllable, otherwise additional elements (states) are augmented to $\boldsymbol{\pi}_j$ until that requirement is satisfied.

The described partitioning process will generate submodels $\mathbf{M}\{\boldsymbol{\pi}_j\}$ of the type:

$$\dot{\mathbf{x}}_i = \mathbb{f}_i(\mathbf{x}_i, \mathbf{u}_i) \tag{1}$$

$$\mathbf{y}_i = \mathbf{g}_i(\mathbf{x}_i) \tag{2}$$

where $i = 1, \cdots, M$; $\mathbf{y}_i(k) \in \mathbb{R}^{l_i}$; $\mathbf{x}_i(k) \in \mathbb{R}^{n_i}$; $\mathbf{u}_i(k) \in \mathbb{R}^{m_i}$ are output, state and input vectors of subsystem $\mathbf{M}\{\boldsymbol{\pi}_j\}$.

2.2. Controller Design

The cooperative methodology that occurs in Step 5 (for Proposed DMPC #1) and Step 3 (for Proposed DMPC #2) of the algorithm is shown in detail in Figure (1). At the iteration described in the item *iii*, each controller solves the optimization problem for proposed DMPC #2 shown in Figure (2). The optimization problem performed for the DMPC # 1 approach, as well as the details of the controller design and the iteration structure of cooperative control are presented in Rocha and Oliveira-Lopes (2016b).

$\min_{u_i(k+j\|k),j=0,\cdots,H_{ui}-1} \quad V(k) = \sum_i V_i(k)$ s.t. Eqs. (1) - (2) $\mathbf{y}_i(k+j\|k) \in Y, j > 0, \mathbf{y}_i(k+H_{pi}\|k) \in Y_f$ $\mathbf{u}_i(k+j\|k) \in \Lambda_i, j = 0, \cdots, H_{ui} - 1$ $\mathbf{u}_l(\mathrm{k}+j\|k) = \mathbf{u}_l(k+j\|k)^{c-1}, \forall l \neq i$	with $V_i(k) = \sum_{j=H_{wi}}^{H_{pi}} \left\|\bar{\mathbf{y}}_i(k+j\|k) - \mathbf{r}_{yi}(k+j)\right\|^2_{\mathbf{Q}_i(j)} +$ $+ \sum_{j=0}^{H_{ui}-1} \|\bar{\mathbf{u}}_i(k+j\|k) - \mathbf{r}_{ui}(k+j)\|^2_{\bar{\mathbf{R}}_i(j)} +$ $+ \sum_{j=0}^{H_{ui}-1} \|\Delta\bar{\mathbf{u}}_i(k+j\|k)\|^2_{\mathbf{W}_i(j)}$

Figure 2: Optimization problem solved by each controller in the cooperative structure for proposed DMPC #2 (adapted from Christofides at al., 2013).

where $\mathbf{Q}_i > 0$, $\mathbf{R}_i > 0$ and $\mathbf{W}_i \geq 0$ are the state, input and variation of inputs weighting matrices, respectively. H_{wi} represents the beginning of the considered prediction horizon (useful for delayed systems); H_{pi} is the prediction horizon and H_{ui} is the control horizon. $\mathbf{r}_{yi}$ is the reference trajectory defined for the controlled outputs and $\mathbf{r}_{ui}$ is the reference trajectory defined for the manipulated inputs.

3. Application

Consider a plant present in Stewart et al. (2010). The model was adapted from Stewart et al. (2010) resulting in the equations and parameters shown in Rocha and Oliveira-Lopes (2016b). The controlled outputs and manipulates inputs are denoted, respectively $\mathbf{y} = [H_1 \;\; T_1 \;\; H_2 \;\; T_2 \;\; H_3 \;\; T_3]^{\boldsymbol{T}}$ and $\mathbf{u} = [Ff_1 \;\; Q_1 \;\; Ff_2 \;\; Q_2 \;\; F_R \;\; Q_3]^T$.

The sampling time is $T_s = 0.5$ min. The manipulated input constraints for Q_1, Q_2 and Q_3 are defined in the region (0;500) kJ/min and $|\Delta u| \leq 10$ kJ/min and (0;500) kg/min and $|\Delta u| \leq 10$ kg/min for the inputs Ff_1, Ff_2 and F_R. The simulations presented are performed for a prediction horizon $H_p = H_u = 3$; the weighting matrices for centralized, cooperative DMPC and non-cooperative DMPC are set as $\mathbf{Q} = \mathbf{I}_{l\,x\,l}$, $\mathbf{R} = \mathbf{I}_{m\,x\,m}$ and $\mathbf{W} = \mathbf{I}_{m\,x\,m}$, with adequate dimensions. The terminal condition that stops the process of cooperation between controllers is defined as the successive relative variation

of the objective function at each iteration, which should be less than J_ε, in this work a value of $1 \cdot 10^{-2}$ was used.

The purpose of control is to keep the controlled variables at steady state *ss1* until the instant $k = 10$, from this point take it and keep it at steady state *ss2* until the instant $k = 90$, and then return it to the initial condition up to $k = 160$, when the simulation ends.

Table (1) represents the evolution of the partitions for cooperative DMPC proposed #1 and the intervals of sampling time that each partitioning occurs (I_{St}). The model and partitioning update were done at every sampling time, and the number of distinct partitions found by the proposed methodology was updated as necessary throughout the operating space composed by ten different types of graphs representing the process during this transition.

Table 1 – Evolution of partitions for the case study for cooperative DMPC proposed #1

Graphs	***Partition***	I_{St}
G_1:$\{u_1\} \Rightarrow \{x_1,\ x_4,\ x_5,\ x_8\ and\ x_{12}\}$ G_2:$\{u_2\} \Rightarrow \{x_4\}$ G_3:$\{u_3\} \Rightarrow \{x_4,\ x_5,\ x_8,\ x_9\ and\ x_{12}\}$ G_4:$\{u_4\} \Rightarrow \{x_8\ and\ x_{12}\}$ G_5:$\{u_5\} \Rightarrow \{x_1,\ x_4,\ x_5,\ x_8\ and\ x_9\}$ G_6:$\{u_6\} \Rightarrow \{x_4\ and\ x_{12}\}$ G_7:$\{u_5\} \Rightarrow \{x_1,\ x_4,\ x_5,\ x_8,\ x_9\ and\ x_{12}\}$ G_8:$\{u_5\} \Rightarrow \{x_1,\ x_4,\ x_5,\ x_9\ and\ x_{12}\}$ G_9:$\{u_6\} \Rightarrow \{x_{12}\}$ G_{10}:$\{u_2\} \Rightarrow \{x_4\ and\ x_8\}$	**Partition #1:** $\{G_1,\ G_2,\ G_3,\ G_4,\ G_5$ and $G_6\}$	1-11; 129-130; 142-160
	Partition #2: $\{G_1,\ G_2,\ G_3,\ G_4,\ G_7$ and $G_6\}$	12-13; 131-141
	Partition #3: $\{G_1,\ G_2,\ G_3,\ G_4,\ G_7$ and $G_9\}$	14-18
	Partition #4: $\{G_1,\ G_{10},\ G_3,\ G_4,\ G_7$ and $G_6\}$	19-53; 56-59; 91-128;
	Partition #5: $\{G_1,\ G_{10},\ G_3,\ G_4,\ G_5$ and $G_6\}$	54-55; 60-90;

Figure (3) presents the dynamics of the controlled outputs. They show the responses obtained by centralized control, non-cooperative DMPC, cooperative DMPC proposed #1 and cooperative DMPC proposed #2.

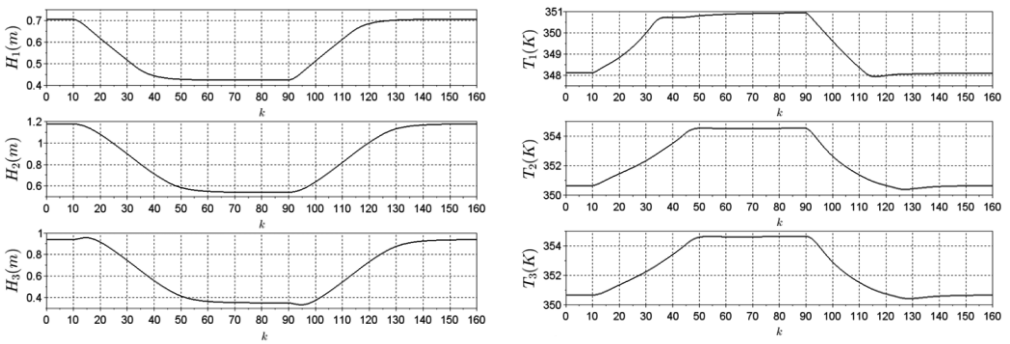


Figure 3: Dynamic evolution of controlled outputs obtained by centralized control and proposed DMPC approaches (overlapping curves).

From the graphical results for the controlled outputs, one can notice that the responses obtained for controlled outputs by the DMPC control structures (non-cooperative, cooperative # 1 and cooperative # 2) are close to those obtained by centralized control (curves overlapping). The performance of each control strategy was evaluated by comparing the absolute error sum of the groups of variables with the same dimension, including the manipulated inputs agreeing with the results Figure (3). The results obtained show the similarity of the performances between the approaches for distributed control and the centralized control. The parameter that will in fact discriminate such

strategies will be the processing time used in the calculation of control actions. The parameter in question is displayed in the Table (2).

No effort for optimizing the computer codes or controller tuning was made. Analyzing Table (2), one can notice that the non-cooperative structures and cooperative DMPC# 2 are the ones with the shortest computer processing time, when compared to centralized control. On the other hand, there was an increase in the processing time for the cooperative control DMPC # 1. This is due to the increase in the number of iterations in the structure in the cooperative process to achieve the same performance.

Table 2 - Comparison of computer processing time

Control Strategy	***Processing Time***
Centralized	$T_C = 57.77$ s
Non-cooperative DMPC	$0.11 \cdot T_C$
Cooperative DMPC #1	$1.52 \cdot T_C$
Cooperative DMPC #2	$0.77 \cdot T_C$

4. Conclusion

This work introduces two new cooperative distributed nonlinear model predictive control strategies with automatic partitioning. The first approach is based on oriented graphs of the linearized local model of the plant and the second one is based on incidence matrices of the nonlinear model of the plant. While the first approach solves a set of linear optimization problem, it requires a larger number of iterations for same performance. The proposed strategies were applied to a benchmark plant and compared to other control structures. The results indicate good performance of the proposed DMPC structures when compared to centralized control. In addition, they presented acceptable computer processing time for the case study evaluated.

References

P. D. Christofides, R. Scattolini, D. M. noz de la Peña, J. Liu. 2013. Distributed model predictive control: A tutorial review and future research directions. Computers & Chemical Engineering 51, 21- 41.

S. S. Jogwar, M. Baldea, P. Daoutidis, 2009. Dynamics and control of process networks with large energy recycle. Industrial & Enginnering Chemistry Research, 48, 6087-6097.

A. Kumar, P. Doutidis, 2002. Nonlinear dynamics and control of process systems with recycle. Journal of Process Control, 12, 475-484.

J. Liu, D. M. noz de la Peña, P. D. Christofides, 2010. Distributed model predictive control of nonlinear systems subject to asynchronous and delayed measurements. Automatica 46 (1), 52-61.

R. R. Rocha, L. C. Oliveira-Lopes, 2016a. A Distributed Model Predictive Controller for Linear Systems with Automatic Partitioning. In: The 15th IASTED International Conference on Intelligent Systems and Control, Calgary: ACTAPRESS, Campinas, Brazil, August 16-18.

R. R. Rocha, Luís C. Oliveira-Lopes, 2016b. A Cooperative Distributed Model Predictive Control for Nonlinear Systems with Automatic Partitioning. In Computer Aided Chemical Engineering, Elsevier, Volume 38, Pages 2205-2210.

B. T. Stewart, A. N. Venkat, J. B. Rawlings, S. J. Wright, G. Pannocchia, 2010. Cooperative distributed model predictive control. Systems and Control Letters 59 (8), 460-469.

A. N. Venkat, I. A. Hiskens, J. B. Rawlings, S. J., 2008. Wright. Distributed MPC strategies with application to power system automatic generation control, IEEE Control Syst. Tech. 16 (6) 1192-1206.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50117-8

POD-based EnKF estimation of heterogeneous reservoir parameters for feedback control of hydraulic fracturing

Abhinav Narasingam[a,b] , Prashanth Siddhamshetty[a,b] , Joseph Sang-Il Kwon[a,b*]

[a]*Artie mcFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77845, USA*
[b]*Texas A&M Energy Institute, Texas A&M University, College Station, TX 77845, USA*
kwonx075@tamu.edu

Abstract

Accurate characterization of reservoir properties is of central importance to achieve a desired fracture geometry during a hydraulic fracturing process. However, the estimation of spatially varying geological properties in hydraulic fracturing is inherently ill-posed due to a limited number of measurements. In this work, parameterization is performed to reduce the dimensionality of spatially varying Young's modulus profiles via proper orthogonal decomposition (POD), and a data assimilation technique called Ensemble Kalman filter (EnKF) is used to estimate the parameter values in the reduced low-dimensional subspace. Through a series of simulation results, it is demonstrated that the POD-based EnKF technique provides accurate process models with updated spatially varying geological parameters. Next, we use the updated high-fidelity process model in a model predictive control framework to construct a closed-loop system that achieves uniform final proppant concentration in a hydraulic fracturing process.

Keywords: parameter estimation, ensemble Kalman filter, proper orthogonal decomposition, hydraulic fracturing.

1. Introduction

In recent years, the extraction of underground resources such as shale gas and oil, which are trapped in ultra-low permeability formations, has become economically viable due to the advances in well-stimulation techniques such as hydraulic fracturing. In a typical hydraulic fracturing process, water and viscosifying agents such as proppant (e.g., sand or ceramic materials) are introduced to propagate fractures in a rock formation. This process will create a highly-conductive channel through which the oil can be produced from the reservoir (Economides and Nolte 2000). The final fracture conductivity depends on the propped fracture geometry and the proppant distribution throughout the fracture.

Several homogeneous reservoir assumptions have been proposed for the design of optimal control strategies to achieve the desired fracture geometry and uniform proppant concentration at the end of pumping (Gu and Hoo, 2015; Siddhamshetty et al., 2017a). However, field data indicates that even within the same rock formation, the performance of hydraulic fracturing can be notably different. This variability can, in part, be attributed to the spatially varying rock mechanical properties such as the Young's modulus. As a consequence, one of the key tasks, which has to be performed

prior to the controller design, is the characterization of spatially varying rock mechanical properties.

A well-known technique for estimation of states and parameters of a nonlinear system is the ensemble Kalman filter (EnKF) (Evensen, 1994). The EnKF has proven to be very successful in solving large and complex history matching problems within the petroleum industry (Aanonsen et al., 2009). The EnKF assimilates real time data sequentially to provide an updated process model. In a typical hydraulic fracturing process, real time measurements may include downhole pressure and micro-seismic data which will be processed to provide the fracture width at the wellbore and the fracture length, respectively. However, the number of parameters to be estimated far exceeds the measurements owing to the high-resolution required to describe the nonlinear hydraulic fracturing process. To deal with this unidentifiability issue, in this work, we perform parameterization (Afra and Gildin, 2016) using Proper Orthogonal Decomposition (POD) (Holmes et al., 1996) to reduce the number of unknown model parameters, prior to estimation via the EnKF. In particular, POD has been employed to compute the basis functions that retain the dominant spatial trends in the underlying geological properties.

The high-fidelity model of hydraulic fracturing with the updated spatially varying Young's modulus profile is simulated to generate the data required for model order-reduction. The local dynamic mode decomposition with control (LDMDc) methodology, which is developed in our previous work (Narasingam and Kwon, 2017), is then applied to the high-fidelity simulation data to derive reduced-order models. It has been demonstrated that the proposed LDMDc methodology provides accurate approximations to the high-fidelity solutions of nonlinear distributed parameter systems, particularly in the presence of time-dependent spatial domains. These reduced-order models are used in the model predictive controller (MPC) to drive the fracture geometry and the proppant concentration at the end of pumping to their respective set-point values by manipulating the flow rate and inlet proppant concentration at the wellbore.

The rest of the paper is organized as follows. First, we provide a brief description of the the POD-based EnKF methodology for the parameterization and estimation of spatially varying Young's modulus profiles. This is followed by the development of a model-based feedback control system to compute the optimal pumping schedule. Finally, the closed-loop simulation results of hydraulic fracturing process are presented.

2. POD-based EnKF

This section provides a brief description of the POD-based EnKF methodology employed for characterizing the spatially varying Young's modulus profile of the reservoir. Hydraulic fracturing can be modeled as a dynamic process describing the fracture propagation, fluid flow, and proppant transport inside the fracture. For a detailed description of the high-fidelity model, we refer the readers to Siddhamshetty et al. (2017b). While the true values of the spatially varying parameters are unknown, we can assume that we have access to the first two statistical moments (i.e., mean and covariance) of the underlying geological features of the formation, which are often available by analyzing historical data. Using this a priori knowledge, an initial ensemble of R realizations is constructed by adding R different perturbations to the mean of the spatially varying parameter distribution. Now, we apply POD to the ensemble R realizations to approximate its members by a set of basis functions as follows:

$$E(z) \approx \sum_{i=1}^{d} \mu_i \varphi_i(z) \tag{1}$$

Here φ_i denote the orthogonal basis functions that contain the key spatial patterns, μ_i are real-valued coefficients that define the appropriate linear combination of basis functions, and d is the dimension of the reduced parameter subspace. These basis functions induce a parametrization of the spatially varying Young's modulus profile if we allow the coefficients μ_i to be free. Below is the sequence of steps involved in the estimation of spatially varying parameters using the POD-based EnKF scheme.

1. Generate R realizations of the unknown parameter fields, $E = \{E^1, \dots, E^R\}$, using the mean and covariance functions available from the prior knowledge.
2. Parameterize the ensemble of realizations via POD, and compute the spatial basis functions $\mathbf{\Phi} = \{\varphi_i(z), \dots, \varphi_d(z)\}$, leaving the coefficients, $\mathbf{M} = \{\boldsymbol{\mu}_k^1, \dots, \boldsymbol{\mu}_k^R\}$, to be free.
3. Given a vector of true measurements, $\mathbf{d}_k$, available at every sampling time t_k, generate R artificial observations for data assimilation.

$$\mathbf{d}_k^j = \mathbf{d}_k + \epsilon_k^j, j = 1, \dots, R;\ \epsilon_k^j \in N(0, \mathbf{C}_D) \tag{2}$$

which can be stored in the columns of $\mathbf{D}_k = \{\mathbf{d}_k^1, \dots, \mathbf{d}_k^R\}$. Here ϵ is a random perturbation drawn from a normal distribution such that $\epsilon \in N(0, \mathbf{C}_D)$, where $\mathbf{C}_D$ is the covariance matrix of the measurement noise.

4. *Forecast* : At time t_k, for every realization in the set $\mathbf{M}$, the forecast of the ensemble, $\mathbf{X}_k^f = \{\mathbf{X}_k^{1,f}, \dots, \mathbf{X}_k^{R,f}\}$, is generated according to:

$$\mathbf{X}_k^{j,f} = \begin{bmatrix} \boldsymbol{\mu}_k^{j,f} \\ \mathbf{y}_k^{j,f} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_{k-1}^{j,a} \\ \mathbf{h}(\boldsymbol{\mu}_{k-1}^{j,a}, \mathbf{u}_{k-1}) \end{bmatrix} \tag{3}$$

where $\mathbf{X}_k$ is an augmented state vector that contains unknown model parameters $\boldsymbol{\mu}_k$ and the corresponding outputs $\mathbf{y}_k$ given by a nonlinear operator $\mathbf{y}_k = \mathbf{h}(\boldsymbol{\mu}_k, \mathbf{u}_k)$ described by the high-fidelity model and the superscripts f and a denote the forecast and assimilation steps, respectively. Since the predicted outputs and the model parameters are nonlinearly related, by including them in the augmented state vector as above, the relationship between $\mathbf{y}_k$ and $\mathbf{X}_k$ is given by the following linear expression.

$$\mathbf{y}_k = \mathbf{H}\mathbf{X}_k = [\mathbf{0} \quad \mathbf{I}] \begin{bmatrix} \boldsymbol{\mu}_k \\ \mathbf{h}(\boldsymbol{\mu}_k, \mathbf{u}_k) \end{bmatrix} \tag{4}$$

5. *Assimilation* : The covariance matrix, $\mathbf{C}_k$, associated with the ensemble of augmented states at time t_k is computed as:

$$\mathbf{C}_k = \frac{1}{R-1}\left(\mathbf{X}_k^f - \overline{\mathbf{X}}_k^f \cdot \mathbf{1}_R\right)\left(\mathbf{X}_k^f - \overline{\mathbf{X}}_k^f \cdot \mathbf{1}_R\right)^T, \quad \overline{\mathbf{X}}_k^f = \frac{1}{R}\sum_{j=1}^{R} \mathbf{X}_k^{j,f} \tag{5}$$

where $\overline{\mathbf{X}}_k^f$ denotes the ensemble average of the augmented states and $\mathbf{1}_R$ is a row vector whose elements are all equal to 1. The update to the model parameters and predicted outputs is then computed as:

$$\mathbf{K}_k = \mathbf{C}_k\mathbf{H}^T\left(\mathbf{H}\mathbf{C}_k\mathbf{H}^T + \mathbf{C}_\mathrm{g}\right)^{-1}, \quad \mathbf{X}_k^a = \mathbf{X}_k^f + \mathbf{K}_k\left(\mathbf{D}_k - \mathbf{H}\mathbf{X}_k^f\right) \tag{6}$$

where $\mathbf{K}_k$ is the Kalman gain of the ensemble at time t_k.

6. The sequence of forecast and assimilation steps is repeated until all the available measurements are assimilated. The mean of the final ensemble along with the previously obtained basis functions are used to compute the unknown spatially varying Young's modulus profile, according to Eq. (1).

3. Optimal pumping schedule design

Following the parameter estimation, we focus on the design of a model-based feedback controller for achieving uniform final proppant concentration in the hydraulic fracturing process. However, due to the infinite-dimensional nature of the high-fidelity model, it cannot be directly used for the controller design. Alternatively, we seek computationally efficient reduced-order models generated by employing the LDMDc methodology developed in our previous work. Once the reduced-order models are generated, we formulate the following MPC problem to design an optimal pumping schedule that will allow us to achieve uniform concentration of the proppant across the fracture at the end of pumping.

$$\begin{aligned}
\underset{C_{0,k}}{Min} \quad & \left(\hat{\mathbf{C}}(t_f) - \mathrm{C}_{target}\right)^T \mathbf{Q}_\mathbf{c}\left(\hat{\mathbf{C}}(t_f) - \mathrm{C}_{target}\right) \\
s.t \quad & \text{Reduced} - \text{order models} \\
& \hat{\mathbf{x}}(t_k) = \mathbf{x}(t_k) \\
& \mathbf{C}_{\min} \le \hat{\mathbf{C}}(t_k + j\Delta) \le \mathbf{C}_{max}, \qquad \forall j = 0, \dots, 10-k \\
& C_{0,k-1+m} \le C_{0,k+m} \le C_{0,k-1+m} + 4, \qquad m = 1, \dots, 10-k \\
& 2Q_0\Delta\left(\sum_k C_{0,k}\right) = M_{prop} \\
& L(t_f) = L_{opt}, \qquad W_0(t_f) \ge W_{opt}
\end{aligned} \tag{7}$$

where $\hat{(\cdot)}$ indicates the predicted state trajectory, C_{target} is the desired set-point value for the final proppant concentration, $\mathbf{Q}_\mathbf{c}$ is a positive definite matrix used to compute the weighted norm, t_f denotes the total treatment time, Δ is the sampling time, t_k is the current time, $\hat{\mathbf{C}}$ is the vector of proppant concentrations, and $C_{0,k}$ is the inlet proppant concentration (i.e., manipulated input) corresponding to k^{th} time interval i.e., $t \in [t_k, t_{k+1})$. The constraints in the above optimization problem correspond to optimality, safety, and economical considerations that need to be accounted for in the controller design. Please refer to Narasingam and Kwon (2017) for complete details on the MPC formulation adopted above.

4. Simulation results

For the high-fidelity simulation, we generated a synthetic spatially varying Young's modulus profile, which is shown in Fig. 1 (dashed line), to be the true parameter profile. We first employed the POD-based EnKF methodology to estimate this true parameter profile. The predicted outputs were compared against the true measurements at every 25 s. The total duration of the process was 1000 s, which gives $k = 40$ time steps. In the numerical experiments, measurement errors were modeled as white noise with the covariance $\mathbf{C}_{\mathrm{g}}$, the value of which was taken as $diag(0.005; 0.008)$. The values of the key process parameters used in the numerical experiments in this work can be obtained from Siddhamshetty et al. (2017b). The results of parameter estimation are shown in Figs. 1 and 2. Specifically, Fig. 1 shows the initial ensemble mean and Fig. 2 shows the final ensemble mean. It is evident from the figure that the POD-based EnKF method provided a very good estimate for the true spatially varying Young's modulus profile after assimilating all the available measurement data. The estimated parameters in Fig. 2 retain the major spatial trend of the true Young's modulus profile.

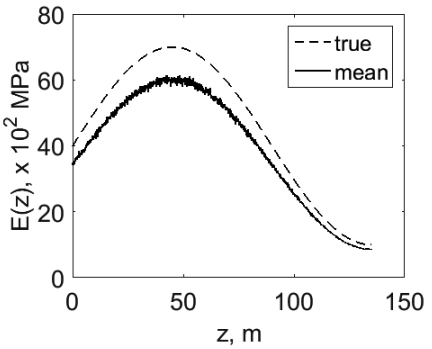


Figure 1. Initial ensemble mean

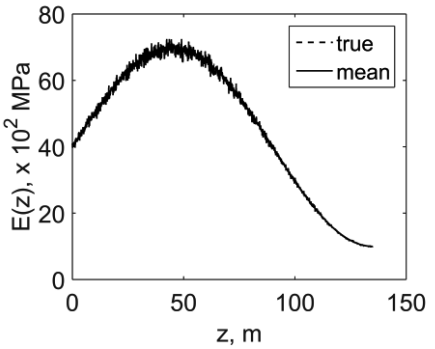


Figure 2. Final ensemble mean

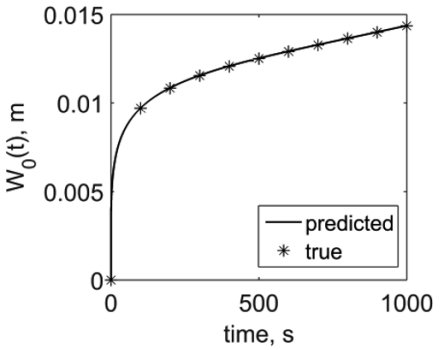


Figure 3. Predicted W_0 and its true measurements

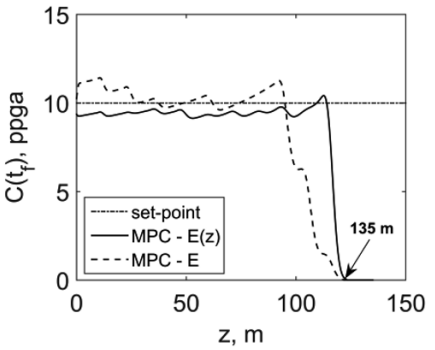


Figure 4. Closed-loop concentration profiles

In addition to finding a good approximation between the true and estimated parameter values, one of the primary objectives of parameter estimation of dynamic models is to accurately predict the process dynamics. This is an important feature in order to design model-based control systems or perform dynamic optimization of the process. The predictive capability of the updated model is evident from Fig. 3, which shows a good match between the true and predicted values of the fracture width at the wellbore, W_0.

With the updated spatially varying Young's modulus profile, reduced-order models were developed via LMDMC, which was then used to design a model-based feedback control system. For the sake of illustrating the effect of not characterizing the spatially varying Young's modulus profile prior to the design of feedback control systems, we have also designed a closed-loop system under MPC based on a constant Young's modulus profile, which was fixed at $E = 50 \times 10^2$ MPa across the fracture. The results are presented in Fig. 4 which shows a comparison of the final concentration profiles achieved by the respective closed-loop systems under MPC. Clearly, the proposed controller with the updated parameter profile drives the final proppant concentration (solid line) towards the target value (dotted line) while propagating the fracture to the desired length (arrow). On the other hand, when the rock properties are not characterized accurately, it can result in a poor controller performance, i.e., non-uniform final concentration profiles (dashed line), thereby diminishing the productivity.

5. Conclusions

In the present work, we have combined POD and the EnKF techniques to successfully estimate the unknown Young's modulus distribution in a hydraulic fracturing process. Parametrizing via POD resulted in a significant increase in the identifiability of the original parameter estimation problem, i.e., the number of variables to be estimated in the data assimilation framework is reduced considerably. We also observed significant predictive capabilities from the updated parameter profile while still retaining the underlying spatial trends in the Young's modulus profile. The update parameter profile is then used in the design of a closed-loop controller to drive the proppant concentration profile at each spatial location to a uniform target value at the end of pumping. Comparison of closed-loop simulation results shows the significance of accurately characterizing the spatial variability in the reservoir parameters on the productivity of hydraulic fracturing.

References

S. I. Aanonsen, G. Nævdal, D. S. Oliver, A. C. Reynolds, B. Vallès, 2009, The ensemble Kalman filter in reservoir engineering - a review, SPE Journal, 14(3), 393-412.

S. Arfa, E. Gildin, 2016, Tensor based geology preserving reservoir parameterization with Higher Order Singular Value Decomposition (HOSVD), Computers & Geosciences, 94, 110–120.

M. J. Economides, K. G. Nolte, 2000, Reservoir stimulation, John Wiley & Sons.

G. Evensen, 1994, Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research, 99(C5), 10143–10162.

Q. Gu, K. A. Hoo, 2015, A Model-based closed-loop control of the hydraulic fracturing process. Industrial & Engineering Chemistry Research, 54(5), 1585–1594.

P. Holmes, J. L. Lumley, G. Berkooz, 1996, Turbulence, Coherent Structures, Dynamical Systems and Symmetry, Cambridge University Press.

A. Narasingam, J. S. Kwon, 2017, Development of local dynamic mode decomposition with control: Application to model predictive control of hydraulic fracturing. Computers & Chemical Engineering. 106, 501–511.

P. Siddhamshetty, J. S. Kwon, 2017a, Modeling of hydraulic fracturing and designing of online pumping schedules to achieve uniform proppant concentration in conventional oil reservoirs, Computers and Chemical Engineering. https://doi.org/10.1016/j.compchemeng.2017.10.032.

P. Siddhamshetty, J. S. Kwon, S. Liu, P. P. Valkó, 2017b, Feedback control of proppant bank heights during hydraulic fracturing for enhanced productivity in shale formations, AIChE Journal, doi: 10.1002/aic.16031.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50118-X

Optimal Control of Surfactant containing Multiphase Systems – Challenges and Solution Strategies for a stable Mini-Plant Operation

Markus Illner[a*], Erik Esche[a], Jens-Uwe Repke[a]

[a]*Technische Universität Berlin, Process Dynamics and Operations Group, Sekr. KWT-9, Straße des 17. Juni 135, D-10623, Germany*
markus.illner@tu-berlin.de

Abstract

In this contribution, a model-based approach for the control and stable operation of a mini-plant for homogeneously catalysed reactions in microemulsions is presented. Within these systems, the control of the crucial reaction and phase separation steps is hindered by sensitive shifting operation windows and immeasurable concentrations. Combining the optical observation of the separation state and an underlying phase separation model, a soft-sensor is created to identify these relevant concentrations. Together with plant measurements, a moving horizon state estimation is used to calculate a consistent and validated state of the full plant model, which is then used in a dynamic optimization to calculate optimal trajectories for the mini-plant operation. Hereby, a stable separation is achieved for up to 200 h mini-plant campaigns, alongside an efficient reaction performance with up to 38 % yield and a chemoselectivity of 95 %.

Keywords: Soft-sensor, State Estimation, Microemulsion, Mini-plant.

1. Introduction and Motivation

In recent years, tunable solvent systems, which especially enable new synthesis paths for the conversion of usually long-chained bio-based starting materials, have emerged. Substitutionary, a surfactant containing system for the hydroformylation of long-chained alkenes towards the corresponding aldehydes, using aqueous catalyst solutions is regarded within this contribution (see Figure 1). The induced microemulsion provides a large interfacial area for increased reaction rates in a reactor. Subsequently, a gravity settler allows for an efficient separation of the products (oil phase), as well as the catalyst recycling (aqueous phase) using a gravity settler.

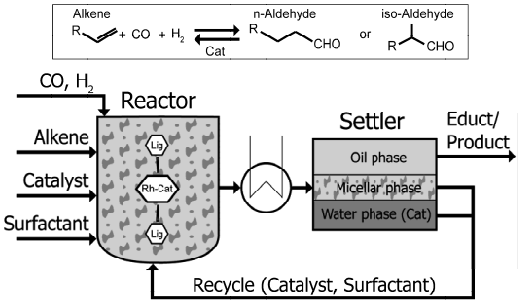


Figure 1: General process concept for the hydroformylation reaction in microemulsions

To test this concept, a mini-plant has been built and operated (Illner et al., 2016). However, the operability of such a system remains critical due to the complexity of the component system and immeasurable states. Since standard automation approaches fail in this case, a stable operation can be achieved with model-based strategies. This comprises a detailed investigation of the phase separation system, from which a soft-sensor is derived to gain relevant information on concentrations in the plant. Subsequently a state estimation can be carried out to calculate initials for the complete plant model. A dynamic real-time optimization is then used to generate optimal trajectories for the set points of the mini-plant's base automation controllers. To prove this concept, long-term min-plant campaigns of up to 200 h were carried out, aiming for a stable operation of especially the crucial phase separation and the reaction step.

2. System Information and Technical Application

Investigations on the afore mentioned process concept are carried out for the exemplarily alkene 1-dodecene, the surfactant Marlipal® 24/70, and a catalyst solution consisting of water, a rhodium precursor (CAS: 14874-82-9) and SulfoXantPhos.

2.1. Mini-plant Set-up

The constructed mini-plant consists of a feed section, holding feed tanks for the used components and corresponding feed pumps, which supply a high-pressure reactor, equipped with a gassing stirrer. The separation is carried out in a self-constructed settler unit augmented with gauge glasses and three drains. Thus, three developed phases can be recycled individually within the process, whereas the oil phase is also conveyed towards a product tank. The plant is fully automatized with SIEMENS PCS 7 and offline gas chromatography is used for the sampling of liquids. Further information on the system, safety layers, and analytics can be found elsewhere (Illner et al., 2016).

2.2. Separation Behaviour of Microemulsion Systems and Operation Window

A ternary mixture of water, oil, and a surfactant, is characterized by the formation of microscopic scale structures, so called micelles. The formation of micelles, their type, and the development of excess phases is influenced by temperature and concentration (Figure 2, left). For the depiction hereof, the Kahlweit's Fish plot in Figure 2 is used.

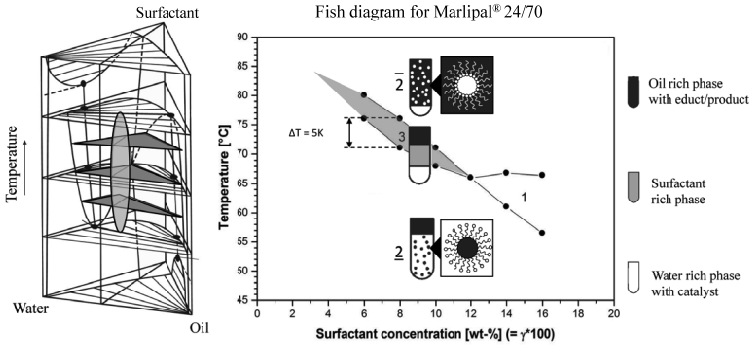


Figure 2: Left: Gibb's phase prism for oil, water, and a nonionic surfactant. Grey areas indicate three phasic miscibility gaps. Taken from (Stubenrauch et al., 2009). Right: Kahlweit's Fish plot with temperature dependent separation regimes for the applied component system. Adapted from (Pogrzeba et al., 2017)

Regarding the process concept, only the three-phase region (3) is feasible, due to the high catalyst loss towards the oil phase in the $\overline{2}$-state and the nonviable slow separation for the $\underline{2}$-state. Thus, an operation region for a successful phase separation is identified.

2.3. Operation Challenges

Controlling the phase separation state for a plant operation requires the identification of relevant influence parameters, their measurability and controllability. Table 1 hereof lists relevant process variables, which directly influence the position of the three-phase region. Besides the separation temperature, which is easily accessible and measured in sufficient accuracy, relevant influence parameters are only available with a large delay (concentration measurements) or are not accessible at all (surfactant concentration). Especially the latter hinders the controllability of the system dramatically, since also the influence on the position and expanse of the phase separation operation window is the largest (see Figure 2). Thus, frequent failures of the separation are to be expected, even with only small perturbations of a steady state operation point. Moreover, Müller (2017) and Illner (2016) reported an affection of the hydroformylation reaction through the separation state, resulting in a reduction of the selectivity towards the target product.

Table 1: Process variables influencing the phase separation, their sensitivity on the separation operation region location, measurability, and controllability. STD: Relative standard deviation of nominal measurement value at standard operation point.

Process variable	Sensitivity on separation	Measurability	STD in %	Sampling frequency	Control element
Temperature T	high	Temperature sensors	0.003	1 s	Heater settler
Oil to water ratio α	moderate	Offline gas chromatography	0.7	1 h	Feed pump
Product concentration X	moderate	Offline gas chromatography	0.7	1 h	Feed pump, residence time
Surfactant concentration γ	very high	Not available	-	-	Feed pump

3. Model Development and Control Strategy

To enable controllability and stable operation a model-based description of the separation behaviour is developed and a soft-sensor for relevant concentrations is derived thereof. To account for the dynamic and non-steady state process behaviour, the development of operation trajectories from a dynamic optimization is employed.

3.1. Phase Separation Model and Surfactant Soft-Sensor

Since no generic thermodynamically driven description of microemulsion systems is available, an empiric model is created based on experimental data. The main idea is to exploit the separation characteristics for the desired three-phase region (3). For constant temperature T and pressure p, the mixture of oil, water, and surfactant, separates into three phases with certain compound concentrations c_i^{Phase}. This corresponds to the corners of the triangle miscibility gap in the ternary diagram (see dark grey area in the Gibb's phase prims in Figure 2). If the mixture concentration is varied, the same individual phases with an equal composition will occur, but the volume fractions will be

shifted. Thus, a correlation between the compound concentrations in the initial mixture $c_i^{Mixture}$, the volume fractions φ^{Phase} and compound concentrations c_i^{Phase} is derived:

$$c_i^{Mixture}(T,p) = f\left(c_i^{Phase}(T,p), \varphi^{Phase}(T,p)\right) \tag{1}$$

A full factorial design of experiments over the relevant influence factors (Table 1) at varying levels (e.g. 4 for α and 10 for T) according to the factor's sensitivity is used to gather results on φ^{Phase}. Using this, Equation 1 is transferred into a smooth polynomial model with the inputs φ^{Phase}, T, and p and a respective maximum order (2, 2, 1), as well as fixed parameters c_i^{Phase} and P_n. The component concentrations c_i^{Phase} are gathered using offline gas chromatography measurements as well as the determination of density and the temperature dependent critical micelle formation concentration, which represents the surfactant concentration in the oily and aqueous excess phase. Subsequently, the surfactant concentration in the settler becomes accessible with a soft-sensor. The combination of the optical observation of the separation performance through gauge glasses via a webcam and an automated image processing results in frequent information on developed phase heights in the settler. Together with the geometric information of this unit, φ^{Phase} is calculated and used together with the settler temperature measurement to identify the settler mixture concentrations $c_i^{Mixture}$.

3.2. Calculation of optimal trajectories

With this approach, the first-time online identification of relevant surfactant concentrations for the microemulsion system at hand is possible and enables further control strategies. Therefore, a discretized dynamic mini-plant model described by Müller et al. (2017) is used. This also includes the transition to a failing separation, resulting in nonlinear, non-smooth (optimization) problems, which is dampened, using sigmoid functions. As a first step, measurement data from the mini-plant, offline gas chromatography sampling and the surfactant soft-sensor is collected and processed in a data conciliation and moving horizon state estimation framework. Extending the work of Hoffman et al. (2016) for dealing with gross errors, the handling of different sampling rates (see Table 1) is incorporated here. This is done by first solving the optimization problem for a reduced plant model containing only component mass balances to estimate stream and level information. Subsequently, when additional deferred concentration measurements are available, the full problem is solved and updated. Gaining validated initials of the full mini-plant model, a dynamic optimization is then carried out. The objective is to maximize the reaction product content in the product stream within the next 4 h. Constraints apply on the reactor and settler surfactant concentration and oil to water ratio in order to maintain an efficient reaction and a stable separation. This results in optimal trajectories for the controller set points of feed streams, recycle streams, and settler temperature (decision variables).

4. Application on Mini-plant Operation

A mini-plant campaign was conducted for a total of 180 h, with several transitions. The startup was realized by feeding the required compounds and initializing a full recycle operation. The reaction was started by the insertion of syngas. In parallel, the optimizer was started and calculated trajectories were applied within the process control system.

4.1. Separation performance

The overall separation performance is evaluated by the amount of oily substances (reaction educts and products) in the settler oil phase, as well as the catalyst leaching. Here, an excellent separation was achieved with a mean oil phase purity of > 99.5 % and a catalyst leaching below 0.1 ppm. Additionally, the effect of the applied optimization and soft-sensor approach is depicted in Figure 3. For a reference time frame of 2 h, the separation performance was observed via a gauge glass. The corresponding evaluation of the soft-sensor implies a reduction of the surfactant concentration, which consequently established with a reduction of the middle emulsion phase fraction. As a control action, the optimizer calculated the addition of surfactant to retain the desired operation point.

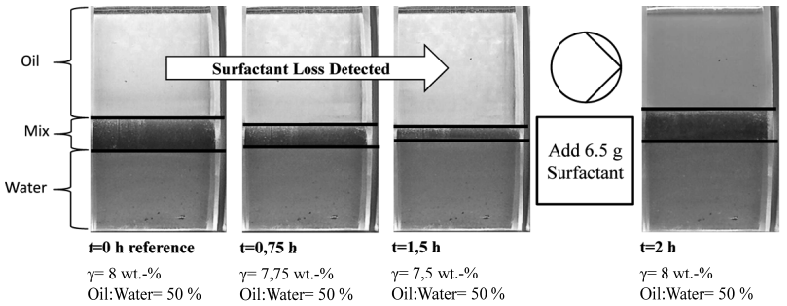


Figure 3: Gauge glass images of the settler over an operation horizon of 2 h and results of the soft-sensor calculations. The right picture shows the separation state after a performed control action derived from the dynamic optimization.

4.2. Reaction

Figure 4 shows the mini-plant reaction performance for the applied reaction residence time of 7 h, whereas a steady state conversion of the applied alkene 1-dodecene of 40 % was achieved. Moreover, a chemoselectivity towards the main product tridecanal of 95 % was observed. This meets results from reference lab scale investigations and shows a successful transfer of the process concept into the mini-plant.

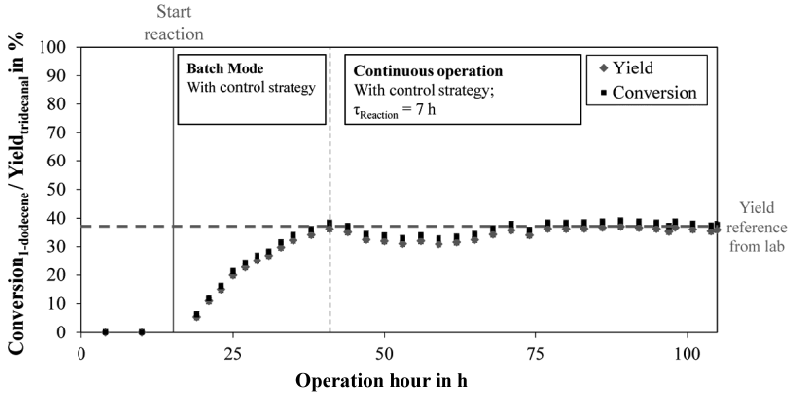


Figure 4: Mini-plant operation results for the conversion of 1-dodecene and the yield of the main product tridecanal for different operation modes.

5. Conclusions and Outlook

Within this contribution, a model-based approach has been presented to cope with the highly dynamic and non-steady state process behaviour of a hydroformylation process in microemulsions. Herein, an empiric model for the separation state and a soft-sensor for the identification of immeasurable surfactant concentrations are used to calculate valid initials for a full mini-plant model. Using calculated trajectories from a dynamic optimization, controller set-point of the process control system a stable mini-plant operation with successful reaction and separation was achieved.

Future works will focus on the online application of the presented framework, using OPC UA to directly interface the process control system. Additionally, the state estimator is to be extended towards a simultaneous handling of multiple sampling rates. Finally, further mini-plant operations will focus on testing the robustness of the used optimization framework regarding multiple or large perturbations.

6. Acknowledgements

This work is part of the Collaborative Research Centre 'Integrated Chemical Processes in Liquid Multiphase Systems' (subproject D2/D4) coordinated by the Technical University of Berlin. Financial support from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) is gratefully acknowledged (TRR 63). Furthermore, the authors gratefully acknowledge the support of Umicore N.V. for sponsoring the rhodium catalyst precursor 'Acetylacetonatodicarbonylrhodium(I)' (CAS: 14874-82-9), Sasol Ltd. for the surfactant used in the described experiments, the support of SIEMENS AG for sponsoring the entire process control system, SIMATIC PCS7 for the automation of the mini-plant, and Rhodius GmbH for sponsoring the knitted fabrics.

References

C. Hoffmann, M. Illner, D. Müller, E. Esche, L. T. Biegler, G. Wozny, 2016, Moving-horizon State Estimation with Gross Error Detection for a Hydroformylation Miniplant, Computer Aided Chemical Engineering, 38, 1485-1490.

M. Illner, D. Müller, E. Esche, T. Pogrzeba, M. Schmidt, R. Schomäcker, G. Wozny, J.-U. Repke, 2016, Hydroformylation in microemulsions: proof of concept in a mini-plant, Industrial & Engineering Chemistry Research, 55, (31), 8616 – 8626.

M. Müller, Y. Kasaka, D. Müller, R. Schomäcker, G.Wozny, 2013, Process Design for the Separation of Three Liquid Phases for a Continuous Hydroformylation Process in a Miniplant Scale, Industrial & Engineering Chemistry Research, 53, 7259 – 7264.

D. Müller, M. Illner, E. Esche, T. Pogrzeba, M. Schmidt, R. Schomäcker, L. Biegler, G. Wozny, J.-U. Repke, 2017, Dynamic real-time optimization under uncertainty of a hydroformylation mini-plant, Computers and Chemical Engineering, (2017), 108, 836-848.

T. Pogrzeba, M. Schmidt, N. Milojevic, C. Urban, M. Illner, J.-U. Repke, R. Schomäcker, 2017, Understanding the Role of Nonionic Surfactants during Catalysis in Microemulsion Systems on the Example of Rhodium-Catalyzed Hydroformylation, Industrial & Engineering Chemistry Research, 56 (36), 9934 – 9941.

C. Stubenrauch, 2009, Microemulsions: Background, New Concepts, Applications, Perspectives, John Wiley & Sons, Ltd: Chichester, UK

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50119-1

Validating control of extreme disturbance of an organic Rankine cycle using VMGsim

Matthew Lie[a], Boaz Habib[b], Wei Yu[a]*, David Wilson[c], Brent R. Young[a]

[a]*Department of Chemical and Materials Engineering, University of Auckland; Auckland, New Zealand*
[b]*Heavy Engineering Research Association, Auckland,New Zealand*
[c]*School of Engineering, Auckland University Technology, Auckland, New Zealand*
w.yu@auckland.ac.nz

Abstract

The Organic Rankine Cycle (ORC) is a thermodynamic process capable of utilising low temperature heat sources such as geothermal reservoirs and industrial waste heat. For an ORC system to be effectively controlled and operated, it is important to develop an understanding of the dynamic behaviour of the system and its components. One such aspect is the response of the plant in off-design scenarios that may not occur during normal day-to-day operation. Developing this understanding by performing tests on a real plant can be impractical as plant safety may be compromised. In this paper, dynamic simulation models are used as tools for plant and control design to partially circumvent the need for these potentially unsafe tests. Using a 75 kW waste heat ORC plant as a basis, a dynamic simulation model of an ORC plant was developed using VMGSim, a commercial flowsheet simulator.

A control strategy using conventional PID controllers was designed to maintain safe operation of the plant by regulating the expander inlet pressure and degree of superheat in the working fluid vapour. The response of the plant to a major process disturbance from an up-stream process in the form of a decreased exhaust gas flow rate was studied by performing step tests of varying magnitudes. The dynamic testing revealed that the superheat response exhibited non-linear behavior linked to the vaporizer. Despite this, the proposed PID-based control system was shown to be sufficient in keeping the plant in safe operating conditions under process disturbances that occur during day-to-day operation. Alterations to plant operation are suggested to account for different cooling air temperatures across the year to achieve consistent plant performance. Finally, simulations of a major process disturbance in the form of a decreased exhaust gas flow rate established the limit of operation of the proposed control system.

Keywords: Organic Rankine Cycle, Dynamic modelling, Process Control.

1. Introduction

Between the rapid growth of the world's population, and the advancement of developing countries, the stress placed on conventional energy sources is rising. In response, the amount of research and development of various alternative energy sources has increased. Biomass, solar, geothermal and industrial waste heat are just some of many alternative sources that have shown promise in their ability to alleviate the world's energy demand.

However, conventional power generation methods are unable to efficiently convert the above heat sources into power due to their lower temperatures. For instance, a standard steam Rankine cycle does not allow efficient utilisation of heat below 370 ◦C. One

solution to this problem that has been investigated is the use of the Organic Rankine Cycle (ORC). The ORC applies the same operating principles as the steam Rankine cycle to generate energy, but instead uses an organic working fluid, as opposed to water, that has a lower boiling point than water, allowing for the recovery of usable energy from these lower temperature heat sources.

The control system which regulates the plant is a key component of any ORC system. For an ORC system to be effectively controlled and operated, it is important to develop an understanding of the dynamic behaviour of the system and its components. One such aspect is the response of the plant in off-design scenarios that may not occur during normal day-to-day operation. Developing this understanding by performing tests on a real plant can be impractical as plant safety may be compromised.

In this paper, dynamic simulation models are used as tools for investigating these major process disturbances, circumventing the need to perform potentially unsafe tests. Using a 75 kW waste heat recovery ORC pilot plant as a basis, a dynamic simulation model of an ORC plant is built using VMGSim, a commercial flowsheet simulator. From this model, a control strategy is designed to maintain safe operation of the plant. Finally, the impact of a major process disturbance in the form of a decreased exhaust gas flow rate on the plant's performance is examined using the ORC plant model.

1.1. Simulation of off-design performance

Most studies use dynamic models to explore behaviour of ORC plants under scenarios that might be experienced under normal day-to-day operation. Zhang et al. (2013) utilised a dynamic model to assess the set point tracking ability of their proposed control system while Peralez et al. (2013) used mappings of exhaust gas temperature and flow rates to test the disturbance rejection ability of the proposed control system. However, there is little research on the simulation of an ORC plant outside of normal day-to-day operation, such as start-ups and shut downs, or major process disturbances outside of what can reasonably be anticipated. One example that does is a paper by Antonelli et al. (2015), which models the start-up of an ORC through the ramping of available thermal energy from the heat source, and assessing the time required to reach steady state to compare control systems. More generally, Ramzan and Witt (2007) use dynamic simulation in combination with Hazop to develop an Extended Hazop methodology that provides quantification of the risk posed by operation disturbances in a plant. In a similar vein, a dynamic model of an ORC plant can be used to study its response to major process disturbances, with the results of the simulation then used to provide quantification of risk and consequence in safety analysis.

2. System description

The Above Ground Geothermal and Allied Technologies (AGGAT) programme in New Zealand brings together industry, academic and international partners to deliver new AGGAT products using an established platform of research and development and a functioning innovation pipeline. One aim of the AGGAT programme is to develop an ORC manufacturing industry in New Zealand. A pilot waste heat ORC developed as part of the AGGAT programme is used as the basis of this study. A schematic of this ORC plant is shown in Figure 1.

In this ORC, R245fa is used as the working fluid, being heated by two sources: hot jacket water followed by hot exhaust gas from a set of landfill gas engines. The working fluid is

vaporised at constant pressure to a superheated vapour state before it is passed through the turbine to generate power. The fluid at the turbine outlet is passed through an air cooled condenser to condense it to a liquid. The liquid is then pressurised in a pump before being passed back to the vapouriser. This pilot ORC is designed to produce approximately 75 kW of power, similar to other small-scale waste heat recovery ORCs (Hou et al., 2014).

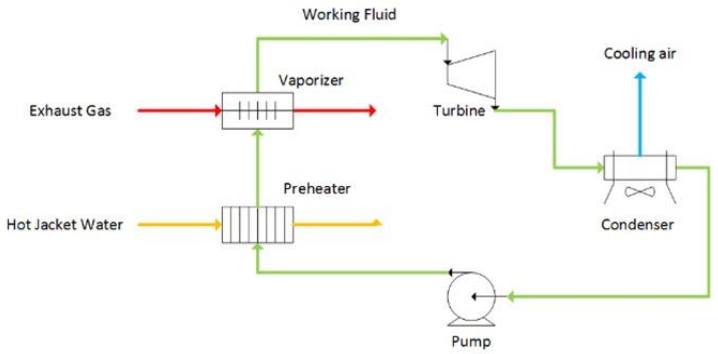


Figure 1: Schematic diagram of the ORC pilot plant

3. Dynamic modelling

The ORC system studied in this paper is modelled using VMGSim, a commercial flowsheet simulator. VMGSim contains a variety of pre-built unit operations and thermodynamic data that has been validated by the software owners and extensive use in different industrial applications. The models within VMGSim are much simpler than their numerically modelled counterparts from literature, allowing for a simpler and more accessible modelling process that requires less computational effort. Although the VMGSim unit operations lack some of the complexity present in numerical models, they have been utilised previously to model a large scale geothermal ORC plant in Proctor *et al*. (2016) with a high level of accuracy when compared with real plant data.

In order to model the ORC plant in VMGSim, the relevant unit operations are chosen and connected together by various material streams, using the plant P&ID as a guide. Equipment and stream specifications are then provided to the model. It should be noted that not all stream specifications are required for the system to be modelled as they will be solved by the relevant thermodynamic calculations performed by VMGSim. The model also includes all bypass lines within the plant.

3.1 Heat exchanger models

The heat exchangers are the most simplified unit operation used in the VMGSim model when compared with models in literature. Whilst the pre-heater, vapouriser and condenser are all different types of heat exchanger, they are all modelled using the same heat exchanger unit operation. This unit operation contains some simplifying assumptions:

- That the heat exchanger is a lumped parameter model;
- That the *UA* value (overall heat transfer coefficient, *U*, multiplied by the heat transfer area, *A*) of the heat exchanger is a constant value; and

- That the pressure drop across any side of the heat exchanger can be computed a '*k*' value, representing flow resistance (defined in Eq.(1))

$$m = \mathrm{k} \times \sqrt{\Delta P \cdot \rho} \tag{1}$$

where m represents the mass flow rate, ΔP is the pressure drop across the heat exchanger, and ρ is the density of the inlet fluid.

Based on these assumptions, each heat exchanger is specified by supplying a constant overall UA value and a 'k' value for each side of the heat exchanger. The UA value is calculated using Eq.(2) and Eq.(3) using values from the plant design point.

$$Q = \mathrm{UA} \cdot \mathrm{LMTD} \tag{2}$$

where, for counter-current heat exchangers,

$$LMTD = \frac{(T_{h,in} - T_{c,out}) - (T_{h,out} - T_{c,in})}{\ln \frac{(T_{h,in} - T_{c,out})}{(T_{h,out} - T_{c,in})}} \tag{3}$$

where the subscripts 'h' and 'c' represent the hot and cold streams of the heat exchanger respectively, and the subscripts 'in' and 'out' represent the inlet and outlet streams respectively. Q represents the amount of heat transferred per unit time. It should be noted that the method used to obtain the constant UA value of each heat exchanger does not rely on knowledge of the heat exchanger geometry, significantly reducing the complexity of heat exchanger modelling.

3.2 Turbine and pump models

The best accuracy in modelling the performance of the turbine and the pump used in the plant can be achieved by using manufacturer data. However, this information was not available at the time of writing and therefore other methods were needed. The performance of the turbine and pump was modelled using a built-in curve in the VMGSim unit operation that can predict off-design behaviour. A design point (including speed, power, flow and efficiency) was then specified for the equipment. This was used alongside a 'Simple Curve' derived from Stepanoff (1957) obeying fan laws. Whilst the curves may be applicable to a wide variety of rotating equipment, it is not specific to any particular turbine or pump and thus may be increasingly inaccurate at conditions far away from the design point.

4. Control system

Typically, the control system in an ORC plant has two main objectives: to regulate power generation, as well as ensure process safety. The power generation of the ORC plant is dictated by the pressure ratio through the turbine, which in turn is largely dictated by the high end pressure.

The papers in literature state the need to protect the turbine from the formation of liquid droplets in the turbine, and furthermore identify the effective control of the degree of superheat of the working fluid at the turbine inlet as a key aspect to ensuring plant safety (Zhang et al., 2012; Peralez et al., 2013).

It should be noted that this is normally achieved by controlling the vapouriser level. However, the vapouriser in this system is a finned-tube heat exchanger, with the working fluid flowing in the tube side to counteract the poor convective heat transfer coefficient of the

exhaust gas. The arrangement of the tubes in the finned-tube heat exchanger makes it difficult to implement level control. Therefore, superheat control is used instead. The working fluid flow rate (manipulated variable (MV)) is used to control the degree of superheat (controlled variable (CV)), whilst the exhaust gas flow rate (MV) is used to control the turbine inlet pressure (CV). Simple PID controllers are used to carry out this proposed control strategy. In addition, an exhaust gas bypass controller was included to keep the exhaust gas line pressure constant.

5. Simulation of a major process disturbance

A major process disturbance here is considered as one which causes the plant to be taken out of the range of operation of the control system, but does not trigger an independent layer of protection. The major process disturbance chosen for investigation in this paper is a decrease in the exhaust gas flow rate. This has the greatest impact on the heat input to the system and can be easily affected by changes in the upstream processes.

In order to examine the effect of a decrease in the exhaust gas flow rate, the flow rate was stepped down from the normal operating point of 6,000 kg/h to 4,500 kg/h and then again to 3,500 kg/h. Figure 2 shows the pressure and superheat response of the plant model to step decreases in exhaust gas flow rate to 4,500 kg/h and 3,500 kg/h.

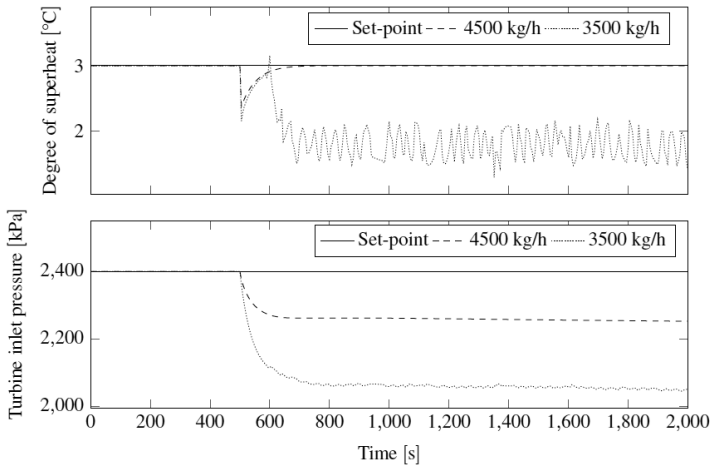


Figure 2: Response of process variables to significant decreases in exhaust gas flow rate

In both cases, there is insufficient heat input to the system for the pressure to return to the set point value, causing the pressure to settle to a value below the set point. Despite this, the working fluid flow rate still adjusts to maintain a constant degree of superheat at the turbine inlet by reducing the valve opening. The stability observed in the change to 4,500 kg/h compared with the instability of the superheat response in the change to 3,500 kg/h indicates that, at some flow rate between 4,500 kg/h and 3,500 kg/h, there is no further decrease in the working fluid valve opening possible, causing the system to become unstable. Interestingly, even though the system is clearly unstable, the degree of superheat remains above 0, indicating that the working fluid is maintained in a superheated vapour state and therefore may not trigger the plant's safety system, such as a turbine bypass.

Further simulations can be performed to more accurately quantify this transition point. Understanding where this point occurs can be useful for quantifying the probability of this situation occurring during real plant operation. By characterising the exhaust gas availability over a long time period, it would be possible to predict the probability that the system would become unstable due to a drop in exhaust gas flow rate, as well as quantify the impact of said disturbance on key plant performance parameters such as power generation. This information can then be used in safety assessments such as layer of protection analysis (LOPA), which in turn would inform decisions on the required layers of protection to ensure an acceptable level of plant safety is achieved. (Center for Chemical Process Safety, 2001). The analysis demonstrated in this paper can be extended to encapsulate other forms of major process disturbances, circumventing the need to perform these tests on a real plant and providing an increased amount of information available in the plant design process.

6. Conclusions

In this paper, a dynamic model for an ORC plant was successfully built using VMGSim and used to analyse the impact of a decrease in exhaust gas flow rate on the system. It was shown the plant was taken out of the range of operation of the control system when the exhaust gas flow rate drops past some value between 4,500 kg/h and 3,500 kg/h. This analysis can be used to quantify the probability of the plant becoming unstable due to this major process disturbance and the impact on the plant's performance, which in turn can be used to inform and improve plant safety and design.

References

Antonelli, M., Baccioli, A., Francesconi, M., Psaroudakis, P., Martorano, L., 2015. Small Scale ORC Plant Modeling with the AMESim Simulation Tool: Analysis of Working Fluid and Thermodynamic Cycle Parameters Influence. Energy Procedia 81, 440–449.

Center for Chemical Process Safety (CCPS), 2001. *Layer of protection analysis : simplified process risk assessment*, American Institute of Chemical Engineers, New York, NY.

Hou, G., Bi, S., Lin, M., Zhang, J., Xu, J., 2014. Minimum variance control of organic Rankine cycle based waste heat recovery. Energy Conversion and Management 86, 576–586.

Peralez, J., Tona, P., Lepreux, O., Sciarretta, A., Voise, L., Dufour, P., Nadri, M., 2013. Improving the Control Performance of an Organic Rankine Cycle System for Waste Heat Recovery from a Heavy-Duty Diesel Engine using a Model-Based Approach. In: 2013 IEEE Conference on Decision and Control (CDC).

Proctor, M. J., Yu, W., Kirkpatrick, R. D., Young, B. R., 2016. Dynamic modelling and validation of a commercial scale geothermal organic rankine cycle power plant. Geothermics 61, 63–74.

Ramzan, N., Witt, W., 2007. Combining disturbance simulation and safety analysis techniques for improve- ment of process safety and reliability. Computer Aided Chemical Engineering 24, 1259–1264.

Stepanoff, A. J., 1957. Centrifugal and axial flow pumps : theory, design and application. John Wiley & Sons, New York.

Zhang, J., Zhang, W., Hou, G., Fang, F., 2012. Dynamic modeling and multivariable control of organic Rankine cycles in waste heat utilizing processes. Computers and Mathematics with Applications 64 (5), 908–921.

Zhang, J., Zhou, Y., Li, Y., Hou, G., Fang, F., 2013. Generalized predictive control applied in waste heat recovery power plants. Applied Energy 102, 320–326.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50120-8

Hybrid Model Based Control of Propylene Copolymerization in Fluidized Bed Reactor

Nazratul Fareha Salahuddin[a], Ahmad Shamiri[b], Mohd Azlan Hussain[a,*] and Navid Mostoufi[c]

[a]*Chemical Engineering Department, Faculty of Engineering, University Malaya, 50603 Kuala Lumpur, Malaysia.*
[b]*Chemical & Petroleum Engineering Department, Faculty of Engineering, Technology and Built Environment, UCSI University, 56000 Kuala Lumpur, Malaysia.*
[c]*Multiphase Systems Research Lab. School of Chemical Engineering, College of Engineering, University of Tehran, P.O. Box 11155/4563, Tehran, Iran.*
mohd_azlan@um.edu.my

Abstract

A modified two-phase model for gas phase ethylene and propylene copolymerization was used to describe the process in a fluidized bed reactor. The process model considered solid entrainment in the reactor as an improvement to the existing two-phase model assumptions. Conventional controller such as PID is not capable enough to effectively control the process because of its non-linearity. An advanced control strategy (an integration of simple designed fuzzy logic controller (FLC) and generic model control (GMC)) is proposed to control the reactor temperature. This hybrid control system performance was compared to the GMC and PID controller. The simulation results proved that the hybrid Fuzzy-GMC controller performed better than GMC and PID in terms of both set point tracking and disturbance rejection.

Keywords: copolymerization, solid entrainment, modified two-phase model, generic model control, fuzzy logic controller.

1. Introduction

This present work will focus on the copolymerization of propylene and ethylene in fluidized bed reactor (FBR) with two sites Ziegler-Natta catalyst and triethyl aluminium as co-catalyst. The strong interaction between the reactor variables gave additional complexity to the fundamental control problem. Advanced control strategies are needed to cope these problems of which conventional controller are incapable to handle (Ho et al., 2012). Although the studies on the modeling and control of polymerization process in fluidized bed reactor are available in the literature (Akbari et al., 2015, Akbari et al., 2015, Alamolhoda et al., 2015; Ali et al., 2007; Ibrehem et al., 2008; Vahidi et al., 2008) but only a few work has been done especially on the copolymerization of propylene and ethylene, for both modeling and control.

Figure 1 shows the simplified schematic of fluidized bed propylene and ethylene copolymerization reactor. The recirculation of gas will help to remove the heat of polymerization process. The heat removal is crucial since the process is highly exothermic. Hence, it is important to stabilize the reactor temperature above the dew

point in order to achieve not only optimal production rate but also to prevent agglomeration of polymer and reactor thermal runaway.

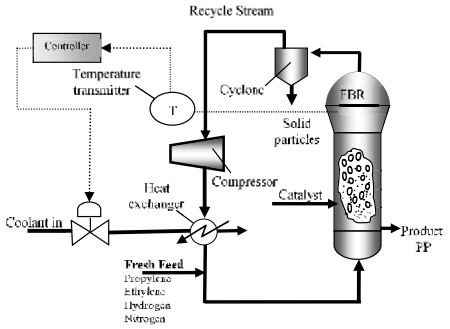


FIGURE 1. Simplified schematic diagram of fluidized bed propylene and ethylene. (from Shamiri et al., 2015)

At the beginning, GMC was chosen to control the temperature of the polymerization reactor since it is easy to implement, as the process model can be applied directly. Fuzzy logic controller was then added to the control strategy since the GMC alone was not sufficient to control the process efficiently. To the best of authors' knowledge, this is the first time that this hybrid Fuzzy-GMC controller is applied to the propylene copolymerization system. A comparative study has been done with the conventional PID controller as well as GMC controller to compare their performance in both servo and regulatory control.

2. Mathematical model

In this study, a modified model developed by Shamiri et al., (2015) is chosen to represent the copolymerization process. It is a modified form of the two-phase model by Shamiri et al., (2011) where in this model, solid entrainment at the top of the reactor has been taken into account.

The energy balance for emulsion phase is as follows:

$$\begin{aligned} &U_e A_e (T_{e,in} - T_{ref}) \sum_{i=1}^{m} [M_i]_e C_{p,i} - U_e A_e (T_e - T_{ref}) \sum_{i=1}^{m} [M_i]_e C_{p,i} - R_v (T_e \\ &\quad - T_{ref}) \left(\sum_{i=1}^{m} \varepsilon_e [M_i]_e C_{p,i} + (1-\varepsilon_b) \rho_{pol} C_{p,pol} \right) + (1-\varepsilon_b) R_{p,e} \Delta H_R \\ &\quad - H_{be} V_e \left(\frac{\delta}{1-\delta} \right) (T_e - T_b) - V_e \varepsilon_e (T_e - T_{ref}) \sum_{i=1}^{m} C_{p,i} \frac{d}{dt} ([M_i]_e) \\ &\quad - \frac{K_e A_e}{W_e (T_e - T_{ref})} \left(\sum_{i=1}^{m} \varepsilon_e [M_i]_e C_{p,i} + (1-\varepsilon_b) \rho_{pol} C_{p,pol} \right) \\ &\quad = V_e \left(\sum_{i=1}^{m} \varepsilon_e [M_i]_e C_{p,i} + (1-\varepsilon_b) \rho_{pol} C_{p,pol} \right) \frac{d}{dt} (T_e - T_{ref}) \end{aligned} \tag{1}$$

3. Control strategy

In this work, generic model control (GMC) controller was used to control the fluidized bed reactor temperature. It is chosen since the process model can be applied directly to formulate the GMC equation, which makes its implementation easier. GMC alone was

observed to be ineffective for the temperature control of the polymerization reactor. Fuzzy logic controller (FLC) is used to solve the oscillation problem in the response by GMC. FLC compensated the outcome from GMC, hence improved the control scheme by manipulating the rate of the cooling water (F_c) to control the reactor temperature.

3.1. Generic Model Control

GMC is highly dependent on the model and is only limited for the systems with linear in control. Implementing the GMC strategy for the propylene copolymerization process, $F(X)$ and $G(X)$ are mass and energy balances of the emulsion phase, U is the coolant flow (F_c), which is the manipulated variable to control the reactor temperature, and $\varepsilon(t)$ is the difference between the emulsion temperature and the desired set point. Based on the energy balance for the emulsion phase from model developed by Shamiri et al., (2015), hence, for the current process, GMC formulation is as follows:

$$F_c = \frac{\propto [K_1\varepsilon(t) + \int_0^{t_f} K_2\varepsilon(t)dt] - F(X)}{G(X)} \tag{2}$$

Where

$$\propto = V_e\left(\sum_{i=1}^{m} \varepsilon_e[M_i]_e C_{p,i} + (1-\varepsilon_b)\rho_{pol}C_{p,pol}\right) \tag{3}$$

$$\begin{aligned} F(X) = {} & (1-\varepsilon_b)R_{p,e}\Delta H_R + H_{be}V_e\left(\frac{\delta}{1-\delta}\right)(T_e - T_b) + U_eA_e(T_e - T_{ref})\sum_{i=1}^{m}[M_i]_e C_{p,i} \\ & + R_v(T_e - T_{ref})\left(\sum_{i=1}^{m} \varepsilon_e[M_i]_e C_{p,i} + (1-\varepsilon_b)\rho_{pol}C_{p,pol}\right) \\ & + V_e\varepsilon_e(T_e - T_{ref})\sum_{i=1}^{m} C_{p,i}\frac{d}{dt}([M_i]_e) \\ & + \frac{K_eA_e}{W_e(T_e - T_{ref})}\left(\sum_{i=1}^{m} \varepsilon_e[M_i]_e C_{p,i} + (1-\varepsilon_b)\rho_{pol}C_{p,pol}\right) + (333.15 \\ & - T_{ref})(U_eA_e\sum_{i=1}^{m}[M_i]_e C_{p,i}) \end{aligned} \tag{4}$$

$$G(X) = \frac{Ke^{-\tau_D s}}{\tau s + 1}U_eA_e\sum_{i=1}^{m}[M_i]_e C_{p,i} \tag{5}$$

GMC performance specifications, K_1 and K_2 is obtained from the generalized GMC profile specification graph by Lee and Sullivan (1988). These values are selected based on the desired response suitable for the process. For the copolymerization process, K_1 = 9 x 10^{-3} and K_2 = 2.25 x 10^{-6}.

3.2. Fuzzy-GMC Hybrid Control

The purpose for the hybrid control scheme is to modify the outcome from GMC, in order to reduce the overshoot problem. A fuzzy-logic controller is used to do the modification then coupled with GMC to control the reactor.

FLC is strongly based on its fuzzy sets, linguistic variables and the approximate reasoning. A set of fuzzy rules are set based on the input and output variables. Its input variables are error, *e* (the difference between set-point and reactor temperature) and $\Delta T_e/F_c$ (quotient of the changes of the reactor temperature and the associated coolant valve rate). The output variable is a factor, f where this value is used to compensate the outcome from GMC ($\Delta F_c = f \times F_c$).

This is a simple designed fuzzy logic controller by using Mamdani method (Mamdani 1976). The membership functions are in triangular form.

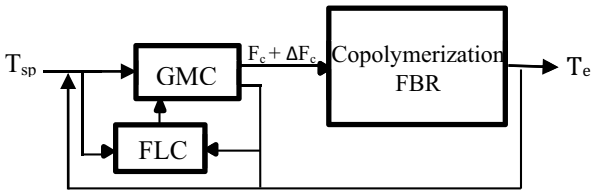


FIGURE 2. Fuzzy-GMC hybrid control scheme

4. Control system analysis

Set point tracking

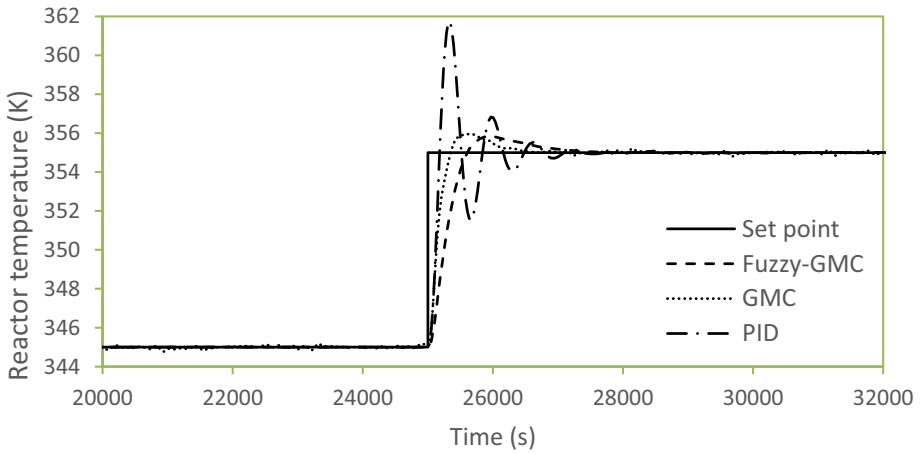


FIGURE 3. Performance comparison of GMC, Fuzzy-GMC Hybrid and PID controller in set point tracking. (PID setting are as follows: $K_c = 24.24$, $\tau_I = 129.5$, $\tau_D = 32.37$)

Figure 3 shows the performances of the controllers tracking the set-point changes. PID controller shows a very high overshoot and low settling time. GMC on the other hand shows lower overshoot and oscillatory response. However, fuzzy-GMC controller manages to track the change in set-point in less than 3000 secs and also slightly reduces the overshoot. Hence, hybrid controller exhibit better performance than GMC and PID controllers (low overshoot and fast and stable response).

Disturbance rejection

A disturbance was introduced to the system to analyse the disturbance rejection capability for each controller. Changes in superficial velocity, U_0 from its nominal value (0.35 m/s) are imposed to the system. At 60000 secs, U_0 is increased to 0.42 m/s.

Based on Figure 4, PID controller is unable to cope with the disturbance since the response did not stop oscillating. Both GMC and fuzzy-GMC hybrid controller exhibit good performance in rejecting the disturbances. The difference is fuzzy-GMC hybrid controller's response has lower overshoot and faster settling time compared to GMC controller.

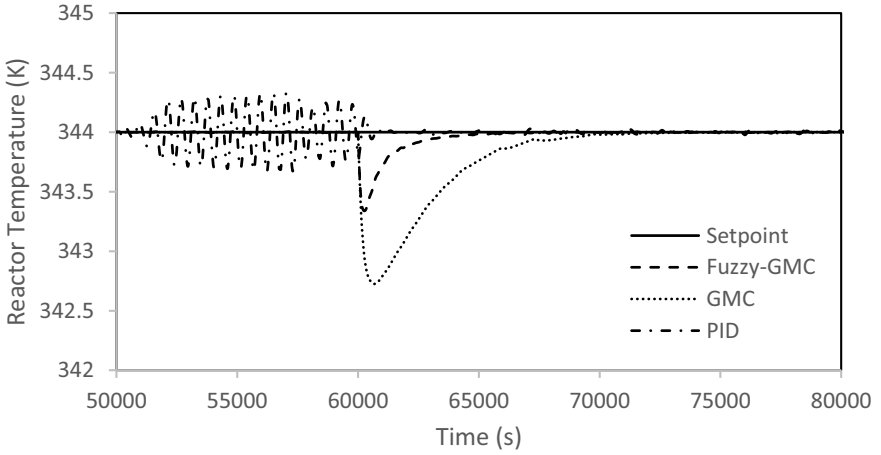


FIGURE 4. Controllers' performance in disturbance rejection.

5. Conclusion

A modified two-phase model that considered solid entrainment in the fluidized bed reactor is utilized in the study to represent the copolymerization process.

A model based controller, GMC is designed to control the temperature of the reactor. GMC gave an unstable and high overshoot response. A hybrid controller is then designed to reduce the instability. Results show that fuzzy-GMC hybrid controller is able to reduce the instability problem and provide the best response in set-point tracking as well as disturbance rejection compared to GMC and PID controller.

The authors acknowledge the funding from FRGS (FP064-2015A) University of Malaya.

References

Akbari, V., Borhani, T.N.G., Shamiri, A., Aramesh, R., Hussain, M.A., Hamid, M.K.A., 2015a. 2D CFD-PBM simulation of hydrodynamic and particle growth in an industrial gas phase fluidized bed polymerization reactor. Chemical Engineering Research and Design 104, 53-67.

Akbari, V., Nejad Ghaffar Borhani, T., Shamiri, A., Kamaruddin Abd. Hamid, M., 2015b. A CFD–PBM coupled model of hydrodynamics and mixing/segregation in an industrial gas-phase polymerization reactor. Chemical Engineering Research and Design 96, 103-120.

Alamolhoda, Fatemeh, Ahmad Shamiri, Mohd Azlan Hussain, Rahmat Sotudeh-Gharebagh, and Navid Mostoufi. 2015. 'Early detection of agglomeration in a polyethylene fluidized bed at high temperature and pressure by vibration signature analysis', Chemical Engineering Research and Design, 104: 156-63.

Ali, M. Al-Haj, B. Betlem, G. Weickert, and B. Roffel. 2007. 'Non-linear model based control of a propylene polymerization reactor', Chemical Engineering and Processing: Process Intensification, 46: 554-64.

Ghasem, N. M. 2006. 'Design of a Fuzzy Logic Controller for Regulating the Temperature in Industrial Polyethylene Fluidized Bed Reactor', Chemical Engineering Research and Design, 84: 97-106.

Ho, Yong Kuen, Ahmad Shamiri, Farouq S. Mjalli, and M. A. Hussain. 2012. 'Control of industrial gas phase propylene polymerization in fluidized bed reactors', Journal of Process Control, 22: 947-58.

Ibrehem, Ahmmed S., Mohamed Azlan Hussain, and Nayef M. Ghasem. 2008. 'Mathematical Model and Advanced Control for Gas-phase Olefin Polymerization in Fluidized-bed Catalytic Reactors', Chinese Journal of Chemical Engineering, 16: 84-89.

Lee, P. L., and G. R. Sullivan. 1988. 'Generic model control — theory and applications', IFAC Proceedings Volumes, 21: 111-19.

Mamdani, E. H. 1976. 'Advances in the linguistic synthesis of fuzzy controllers', International Journal of Man-Machine Studies, 8: 669-78.

Shamiri, Ahmad, Mohamed Azlan Hussain, Farouq Sabri Mjalli, Navid Mostoufi, and Mohammad Saleh Shafeeyan. 2011. 'Dynamic modeling of gas phase propylene homopolymerization in fluidized bed reactors', Chemical Engineering Science, 66: 1189-99.

Shamiri, Ahmad, Suk Wei Wong, Mohd Fauzi Zanil, Mohamed Azlan Hussain, and Navid Mostoufi. 2015. 'Modified two-phase model with hybrid control for gas phase propylene copolymerization in fluidized bed reactors', Chemical Engineering Journal, 264: 706-19.

Vahidi, Omid, Mohammad Shahrokhi, and Ahmad Mirzaei. 2008. 'Control of a Fluidized Bed Polyethylene Reactor', Iranian Journal of Chemistry & Chemical Engineering-International English Edition, 27: 87-101.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50121-X

Identification and Control of an Unstable SOPTD system with positive zero

Dhanya Ram V.[a*] , C. Sankar Rao[b]

[a] *Department of Chemical Engineering, Birla Institute of Technology and Science, Pilani-K.K. Birla Goa Campus, Goa 403726,India*
[b] *Department of Chemical Engineering, National Institute of Technology Karnataka, Suratkal 575025,India*
dhanyaramv@gmail.com

Abstract

The work deals with the identification and control of unstable Second Order plus Time Delay (SOPTD) system with positive zero. Presence of positive zero complicates the performance of the control system dynamics. There are many unstable systems which exhibit the second order plus time delay with positive zero such as drum boiler, distillation column. No work has been reported in the literature on identification of unstable SOPTD process with positive zero. In this work, a subspace based method and an optimization method are proposed to identify an unstable SOPTD model with positive zero followed by the PID controller design which can handle set-point changes and disturbance rejection. The subspace-based method uses input-output measurements to estimate the state space model. This method uses projections of block Hankel matrices followed by a singular value decomposition to determine the order of the system. It offers the key advantages on providing low parameter sensitivity with respect to perturbations for higher order systems. The model parameters are also identified using optimization technique by matching the closed loop responses of the process and the model. In any optimization technique, the initial guess plays an important role for proper convergence. A method is suggested to obtain the initial guess values for process gain, poles, zeros and delay. The parameters identified by subspace based method are compared with that obtained using optimization technique. For the models identified by the above two methods, controllers are designed and implemented. Simulation studies on linear and nonlinear systems are demonstrated to evaluate the performance of the proposed methodologies. The closed loop performances comparison can be made in terms of time integral errors and total variation in input variable.

Keywords: Unstable systems, SOPTD with positive zero, subspace identification, optimization method

1. Introduction

Unstable systems are common in chemical industries. Examples include isothermal CSTR (Liou and Chien, 1991), Nonlinear Bioreactor (Agarwal and Lim, 1989), Dimerization reactor (Alio and Al-humaizi, 2000), Fluidized bed reactor (Kendi and Doyle, 1996), polyolefin reactor (Seki et al., 2001). Padma Sree and Chidambaram (2006) have given an excellent review on the control of unstable systems. Unstable SOPTD systems with a zero are difficult to control due to the presence of an overshoot or inverse response. Some of the examples reported in literature include the Klien's unrideable bicycle (Klien, 1989), Jacketed CSTR (Bequette, 2003).The presence of zero

in the unstable transfer function imposes a difficulty in controlling such systems. For the purpose of designing controllers, identification of the model parameters are required which helps in improved tuning of the controllers. Open loop identification cannot be applied to unstable systems. The systems with positive zeros are slow in response because of their undershoot response at the beginning of the response. Presence of this positive zero complicates performance of control system dynamics. One of the prominent problems with this kind of system is internal stability. Ram et al. (2014) have identified unstable SOPTD systems with negative zero using optimization method. In this work, the method is extended to unstable SOPTD systems with positive zero. Recently Sankar Rao and Chidambaram (2017) reported subspace identification for unstable systems. In this work, the unstable system under consideration is stabilized by a PI/PID controller. A second order time delay model with a positive zero is identified from the closed loop response using a step change in the set point. The process model is identified using optimization method and subspace method separately. The identified models by the above two methods are used for designing controllers and the closed loop performance are evaluated for servo and regulatory response.

2. Proposed Methodology

2.1 Identification of Unstable SOPTD systems with positive zero by subspace identification method

Subspace identification is used to get a linear time invariant state space models directly from the input and output measurement data. The various forms of the subspace based identification methods have attracted much of interest. Subspace identification methods use the concepts of systems theory and linear algebra. Subspace identification methods consist of two steps. In first step, it determines the extended observability matrix and state sequences from the row and column spaces of certain matrices, which are formed from the input output data. Second step estimates the state space model using either the knowledge of extended observability matrix or state sequences.

2.2. Identification of Unstable SOPTD systems with positive zero by optimization method

Consider the unstable second order open loop transfer function with a positive zero given by $G_p(s) = \frac{k_p(1-ps)}{a_1 s^2 + a_2 s + 1} e^{-\phi s}$. In order to identify the process parameters (kp, p, a_1, a_2), system is stabilized by a PID controller and the closed loop response is noted. The system is modelled as second order process with a positive zero

$$\frac{y_{cl}(s)}{y_r(s)} = \frac{k_p(1-ps)}{(\tau_e s^2 + 2\zeta\tau s + 1)} e^{-\phi s} \tag{1}$$

For a unit step change in the setpoint the closed loop response is given by (Seborg and Millichamp, 2006)

$$y(t) = k_p\left[\frac{p}{\tau_e\sqrt{(1-\zeta^2}} e^{-\zeta t'/\tau_e} \sin(\sqrt{(1-\zeta^2}\,\frac{t'}{\tau_e})\right] + \left(1 - e^{-\zeta t/\tau_e}\left[\cos(\sqrt{(1-\zeta^2}\,\frac{t'}{\tau_e}) + \frac{\zeta}{\sqrt{(1-\zeta^2}} \sin(\sqrt{(1-\zeta^2}\,\frac{t'}{\tau_e})\right]\right) \tag{2}$$

Where $t'=t-\Phi$. Ananth and Chidambaram (1999) have defined the formulas for obtaining ζ and τ_e

$$v_1 = \frac{y_{inf} - y_{m1}}{y_{p1} - y_{inf}}; v_1 = \frac{y_{p2} - y_{inf}}{y_{p1} - y_{inf}} \tag{3}$$

$$\zeta_1 = \frac{-\ln(v_1)}{\sqrt{\Pi^2 + (\ln(v_1))^2}}; \zeta_2 = \frac{-\ln(v_2)}{\sqrt{4\Pi^2 + (\ln(v_2))^2}} \tag{4}$$

$$\zeta = \zeta_1 + \zeta_2 \tag{5}$$

$$\tau_e = \frac{\Delta T\sqrt{1-\zeta^2}}{\Pi} \tag{6}$$

$$a_1 = \tau_e^{\ 2}; a_2 = -2\zeta\tau_e^{\ 2} \tag{7}$$

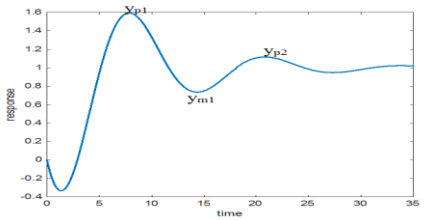


Figure 1: Closed loop response of a system to a step change in setpoint

From the closed loop response the value of y_{p1}, y_{p2}, y_{m1}, y_{inf} are noted. The sign of final steady state value of u (deviation value) is negative and the sign of k_p is positive, then there is one unstable pole and one stable pole present in the system. In order to get the initial guess values for identifying the process parameters from the closed loop response, the method given by Ananth and Chidambaram (1999) is used to obtain the initial guess values of a_1 and a_2. The initial guess value for k_p is taken as the ratio of the final steady state value of the closed loop response to the final steady state value of the manipulated variable u. Equations (3) to (6) are solved to obtain the initial guess values of a_1 and a_2. Equation (2) is solved to obtain the initial condition for the p from undershoot. To evaluate the proposed methodology a known unstable system is taken. The system is stabilized using PID controller and the closed loop response is obtained by giving a unit step change. The model to be identified is also stabilized using the same controller settings. All simulations are carried out on Matlab and Simulink. In order to solve the least square optimization problem, the Matlab routine *lsqnonlin* is used. The objective function is formulated to minimize the sum of the square of the errors between the closed loop response of the model and the process.

Using the methodology explained in Section 2.1 and 2.2, the model identification of the unstable SOPTD system with positive zero is carried out. The identified model parameters by both the methods are stabilised by designing controllers and the results compared.

3. Simulation studies

3.1. Example 1

Consider the transfer function given by $G(s) = \frac{3.5(5 - 0.7s)}{(s^2 - 6.867)}$. The system is stabilized by the PID controller k_c=0.4708; τ_I = 8; τ_D =0.58. The response of the system for unit step change in set point is given in Figure 2. The data obtained are y_{p1}=3.5090; y_{p2}=1.3478; y_{m1}=0.0658; y_{inf} =1; ΔT=12.25; u=-0.3884; ζ=0.0683; τ_e =1.9451. The system has an inverse response which shows the presence of a positive zero in the transfer function. The initial guess values for kp is taken from the closed loop response as the ratio of final steady state value to initial steady state value k_p=1/(-0.3884) = -2.575. To obtain the initial guess values of a_1, a_2 and p the methodology proposed above is applied. The initial guess values are obtained as a_1 =3.7834; a_2=-0.2658; p= -0.9121. To identify the model parameter optimization technique is carried out (Section 2.2). The converged final model parameters identified are k_p=2.5484; p=-0.14; a_1=0.1456; a_2=0. Figure 2 shows that the closed loop response of the identified model matches with the actual response for the same controller settings. Using the identified model parameters the PID controller settings are designed using IMC method. Table 1 shows the controller settings obtained and the ISE, IAE and the TV values. The transfer function model is simulated and the output data is generated. Pseudo Random Binary Signal (PRBS) is used as the exogenous input to excite the process. The order of the system is estimated by inspecting the singular values. The number of dominant singular values will give the information of the order of the system. The model parameters estimated based on subspace based method are K_p = 2.34, p = 0.12, poles are 2.618 and 2.532. From the residual analysis, it is confirmed that the identified model is capable of explaining the dynamic relationship between the cause and effect. Based on the identified model, a PID controller is designed by IMC method and the obtained settings are listed in Table 1. Using the controller settings designed based on the identified model (by optimization and subspace methods), the closed loop performances are evaluated by introducing a unit step change in the set point at time, t = 0 sec and a negative step change in disturbance variable at time, t = 25 sec which is shown in Figure 3. Figure 3 shows the comparison of the closed loop servo and regulatory performance for the identified model by optimization method and subspace identification method.

Table 1: Performance evaluation of the identified models

Examples	Models	Controller settings			Time integral errors		Total Variation in u (TV)
		Kc	τ_I	τ_D	ISE	IAE	
Ex1	Model 1	0.4914	14.6224	0.4141	214.5	42.68	48.3
	Model 2	0.5383	14.3915	0.4143	126.3	38.19	59.2
Ex2	Model 1	0.6493	30.4435	13.2848	220.6	244.3	1170
	Model 2	0.6093	30.3920	13.2903	219.2	258.4	1076

Model 1: Identified by optimization method; Model 2: Identified by subspace method

3.2. Example 2

Consider the isothermal Continuously Stirred Tank Reactor given by Rao and Chidambaram (2006). The transient nonlinear CSTR model is linearized around the

unstable operating point and the resulted transfer function relating concentration of A to to the feed concentration is given by $\frac{\Delta C_A(s)}{\Delta C_f(s)} = \frac{-0.2679(1-41.67s)}{(279.3s^2 - 2.7981s + 11)} e^{-20s}$

The system is stabilised to obtain an underdamped closed loop response of the form as shown in Figure 1 by using the controller settings given by K_c=0.3471;τ_I=-100, τ_D =0.429. From this the value of y_{p1}=1.647; y_{p2}=1.08; y_{m1}=0.7682; y_{inf} =1; ΔT=197.3; u=-3.614 are noted. The values of ζ and τ_e are obtained as ζ =0.0708; τ_e = 31.3224. Using equation (2) and (7) the initial guess values of kp, a_1, a_2 and p are obtained and these are kp=-0.2767.;p=-236.2121;a1=222.3402;a2=-9.3381.Closed loop time constant is assumed as 2-3 times of open loop time constant. The initial guess value for open loop time delay is taken same as the closed loop time delay. Using optimization the process parameters are identified and the converged parameters are obtained as k_p=-0.2679.; p=-41.6667;a_1=279.03;a_2=-2.9781, θ =20. The identified model is used for designing the controllers. In the present work, the controller settings proposed by Sree and Chidambaram (2002) by synthesis method is used. The controller settings given by them are in the form $Gc(s) = K_c(1 + \frac{1}{\tau_I s} + \tau_D s)\left(\frac{1}{\alpha s + 1}\right)$. For the present model parameters obtained by optimization, the controller settings are enlisted in Table 1. Table 1 also shows the IAE, ISE and TV values. Model parameters obtained from the subspace identification method are Kp = -0.285, p = -41.67, a_1 = 279.2, a_2 = -2.953 and θ =20, PID settings are determined by synthesis method and given in Table 1. The time integral error are also calculated and presented in Table 1. Figure 4 shows the closed loop performance for the identified models by the two proposed methodologies.

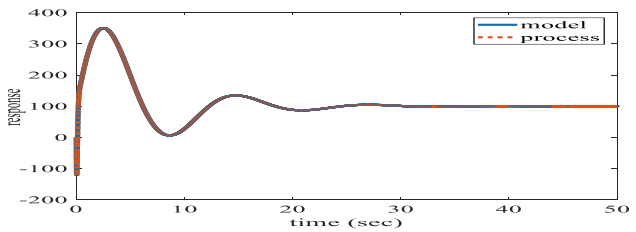


Fig 2: Closed loop servo response of the process and identified model for the same controller settings (optimization method-example 1)

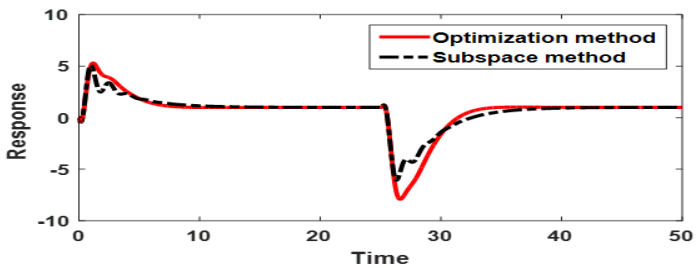


Figure 3: Closed loop response of the controllers designed based on the identified model by optimization method and subspace methods by IMC method for example 1

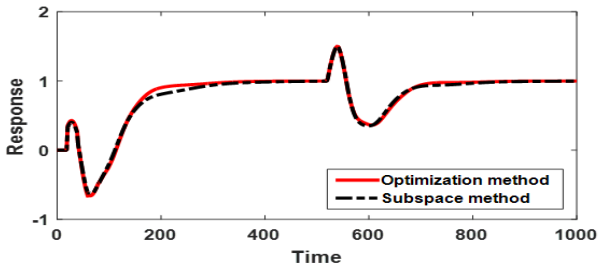


Figure 4: Closed loop response of the controllers designed based on the identified model by optimization method and subspace methods by synthesis method for example 2

4. Conclusions

A closed loop identification method based on optimization and subspace identification for unstable SOPTD systems with zero is proposed. Two case studies were demonstrated in order to evaluate the identification methods. The model parameters obtained by the proposed method were found to match with the actual values. Based on the identified model controllers were designed using methods available in literature. A good set-point tracking and disturbance rejection was obtained. The ISE, IAE and TV values are evaluated. It is found that the closed loop responses obtained by both the identified models are almost similar which emphasis that the proposed methods give good model parameters for designing suitable controller for unstable SOPTD system with zero.

References

Liou,C.T. and Y.S. Chien ,1991, The effect of nonideal mixing on input multiplicities in a CSTR, Chemical Engineering Science, 46,2113-2116.

Agarwal,P. and Lim,H.C,1989, Analysis of various control schemes for continuous bioreactor, Advances in Biochemical Engineering/Biotechnology book series,30,61-90.

Alio,E. and K. Al-humaizi,2000, Temperature control of ethylene to butene-1 dimerization reaction, Industrial & Engineering Chemistry Research,39,1320-1329.

Ananth, I. and Chidambaram, M., 1999, Closed Loop Identification of Transfer Function Model for Unstable Systems, J. Franklin Inst., 336,7, 1055–1061.

Kendi, T.A. and F.J. Doyle III, 1996, Nonlinear control of a fluidized bed reactor using approximate feedback linearization, Industrial & Engineering Chemistry Research, 35,746-757.

Seki,H.M., M. Ogawa and M. Ohshima, 2001, PID temperature control of unstable Gas-phase polyolefin reactor,34,1415-1422.

A Seshagiri Rao, M Chidambaram, 2006, Enhanced two-degrees-of-freedom control strategy for second-order unstable processes with time delay, Industrial & Engineering Chemistry Research,45, 10, 3604-3614.

RP Sree, M Chidambaram, 2006, Control of unstable systems, Alpha Science Int'l Ltd.

Klein, R.E., 1989, Using bicycles to teach system dynamics, IEEE control systems magazine, 9, 3, 4-9.

Bequette, B. W, 2003, Process Control: modelling, design and simulation, Prentice Hall, India, 644.

V Dhanya Ram, A Karlmarx, M Chidambaram, 2016, Identification of Unstable Second-order Transfer Function Model with a Zero by Optimization Method, Indian Chemical Engineer, 58,1,29-39.

C Sankar Rao, M Chidambaram, 2017, Subspace Identification of Unstable Transfer Function Models, Indian Chemical Engineer 59, 1, 1-20.

ELSEVIER

13TH INTERNATIONAL SYMPOSIUM ON PROCESS SYSTEMS ENGINEERING (PSE 2018)

PART B

Edited by
MARIO R. EDEN
MARIANTHI G. IERAPETRITOU
GAVIN P. TOWLER

COMPUTER-AIDED CHEMICAL ENGINEERING, 44

CACHE

13TH INTERNATIONAL SYMPOSIUM ON PROCESS SYSTEMS ENGINEERING (PSE 2018)

PART B

This page intentionally left blank

13TH INTERNATIONAL SYMPOSIUM ON PROCESS SYSTEMS ENGINEERING (PSE 2018)

PART B

Edited by

Mario R. Eden
Department of Chemical Engineering
Auburn University
Auburn, Alabama, USA

Marianthi G. Ierapetritou
Department of Chemical and Biochemical Engineering
Rutgers University
Piscataway, New Jersey, USA

Gavin P. Towler
Honeywell Performance Materials and Technologies
Des Plaines, Illinois, USA

ELSEVIER

Amsterdam – Boston – Heidelberg – London – New York – Oxford
Paris – San Diego – San Francisco – Singapore – Sydney – Tokyo

Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, USA

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN (Part B): 978-0-444-64244-8
ISBN (Set): 978-0-444-64241-7
ISSN: 1570-7946

Publisher: Joe Hayton
Acquisition Editor: Kostas Marinakis
Editorial Project Manager: Emily Thomson
Production Project Manager: Paul Prasad Chandramohan
Designer: Greg Harris

Typeset by SPi Global, India

Contents

Contributed Papers: Optimization Methods and Computational Tools

Contributed Papers: Supply Chain Management and Logistics

Contributed Papers: Education

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50122-1

Global Optimization of Superstructures and Decision Support via a universally applicable Optimization Tool

Kristina Zimmermann[a*], Georg Fieg[a]

[a]*Hamburg University of Technology, Institute of Process and Plant Engineering, Am Schwarzenberg-Campus 4 (C), 21073 Hamburg, Germany*
kristina.zimmermann@tuhh.de

Abstract

The simulation-based multi-objective structural and parameter optimization of continuous chemical production processes is a complex task. To provide a universally applicable optimization tool, the Advanced Process Optimizer was developed and intensively tested in cooperation with different industrial partners over several years. The superstructure can be directly built in a commercial flowsheet simulator. The resulting Generalized Disjunctive Programming formulation of the optimization problem was handled through a combination of newly developed and adapted methods. In order to validate the efficiency of the developed optimization tool the example of a dimethyl ether production process was chosen. The optimization results are compared to a reference solution achieved by conventional approaches and show a significant reduction of both the investment and operating costs.

Keywords: Superstructure, multi-objective optimization, conceptual process design, Generalized Disjunctive Programming

1. Simulation-based multi-objective optimization

Due to growing competition and environmental regulations it is no longer sufficient to fulfill the chemical and technical requirements of a process. Moreover, a holistic view of the process considering different and often opponent objectives is required. The simulation-based process synthesis and design with an overlaid multi-objective optimization is a powerful tool to rise to these challenges. The obtained result is a set of optimal process design trade-offs, illustrated as a Pareto front (PF). On this basis the user is in a position to evaluate the pros and cons of each trade-off considering higher-level qualitative considerations, e.g., the strategic focus of the company, and choose his optimal process in an early stage of process development. To do this, constructive and operative parameters of the process as well as the process structure itself have to be optimized. This is preferably done via a superstructure which is directly implemented in a commercial flowsheet simulator using mostly rigorous models of the apparatuses. Simultaneously, a set of constraints, e.g., the chemical quality of the product streams, have to be satisfied. One systematic formulation of such optimization problems is the Generalized Disjunctive Programming (GDP) representation (Navarro-Amorós et al., 2014). GDP models are known to enclose continuous and Boolean variables, algebraic equations, disjunctions and logical propositions and are often chosen by the Process System Engineering community (Trespalacios and Grossmann, 2014). In order to take advantage of the various available MINLP solvers,

GDP problems are often reformulated as MINLP problems (Trespalacios and Grossmann, 2014). This reformulation has several disadvantages, e.g. the dissimilar responses of different apparatus models in flowsheet simulators (Navarro-Amorós et al., 2014).

In this work, a universally applicable optimization tool for the solution of multi-objective structural and parameter optimization problems that use the GDP formulation is presented. In this regard, the superstructures have been directly drawn in a commercial flowsheet simulator. The developed holistic approach pursues the goal to guide the user through every stage of the optimization project. The developed optimization tool will be shortly presented in the following chapter.

2. Advanced Process Optimizer

Based on former works by Ernst et al. (2017) the Advanced Process Optimizer (Adv:PO) was further developed in order to provide a universally applicable tool for the simulation-based multi-objective structural and parameter optimization of chemical production processes using a GDP representation. The Adv:PO is implemented in Matlab and uses a bidirectional interface to several commercial flowsheet simulators. In the context of this work, the focus lies on the flowsheet simulator Aspen Plus®. A step-by-step user-interface guides the user through the different stages of the optimization project. Figure 1 gives an overview on the implemented system architecture of the Adv:PO focusing on the most important and new components.

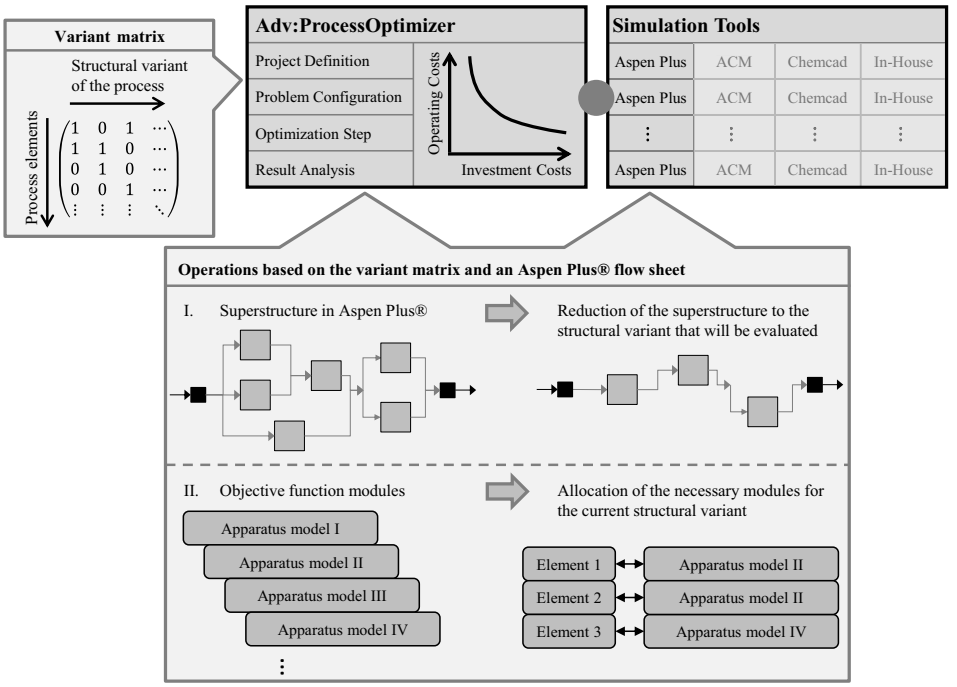


Figure 1: Schematic illustration of the system architecture of the Advanced Process Optimizer for the structural and parameter optimization of chemical production processes, adjusted considering Ernst et al. (2017). Explanatory remarks are given to the most important and new components.

The main stages of an optimization project consist of three blocks: a straightforward concept for the definition of the optimization problem, a robust solution of the multi-objective optimization problem and the provision of appropriate tools for the analysis of the optimization results and a profound decision making support.

In case of the developed Adv:PO, two user inputs are required for the problem definition. The first is the complete superstructure drawn in the flowsheet simulator. And the second is the so-called variant matrix that defines which elements of the flowsheet are part of the process alternative that is currently evaluated. These elements can be apparatus models and connecting material streams but also flowsheeting options of the simulator, like *Design Specs* or *Calculators*, which are needed to formulate special conditions and/or special convergence strategies for the whole process or single apparatuses. The variant matrix is prepared in a provided Microsoft Excel template and automatically imported via the user-interface. For the representation of the disjunctions in the GDP formulation of the problem, a value of 1 in the variant matrix is interpreted as the Boolean variable `true` and a value of 0 as 'false'. Based on this information and as illustrated in Figure 1 the superstructure is automatically cut down to the necessary components reducing the sources of errors in the simulation.

The variant matrix fulfills another important task. On the basis of the objective function modules introduced by Ernst et al. (2017) every apparatus model in the flowsheet is covered by an objective function module. It includes all necessary information about the decision variables, or in other words the apparatus parameters that are varied to find the optimum solutions, and the calculation of the objective functions. Due to this modular concept basically every process can be described. In case of a structural optimization not all apparatuses are part of the evaluation and, as a consequence, the corresponding objective function modules have to be eliminated from the calculations. Analogous to the reduction of the superstructure, this is automatically done by the introduced variant matrix as indicated by Figure 1. The same procedure is employed for the constraints.

The second block of the system architecture contains efficient optimization methods for the robust solution of the optimization problem. The solution method is based on a specially tailored genetic algorithm (GA) introduced by Ernst et al. (2017). GAs are stochastic optimization algorithms that are intensively used in academia and industry. Although GAs do not guarantee to find optimal solutions, they are able to find quasi-optimal solutions in a reasonable timespan. In addition, the adaption of GAs to the optimization problem can significantly improve the convergence velocity and behavior (Chiong, 2009). The developed specially tailored GA is characterized by an innovative elite and diversity preserving strategy (Ernst et al., 2017). Moreover, different operators of the GA have been adjusted for the consideration of the GDP formulation of the optimization problem. In addition, the two-step optimization, introduced by Zimmermann and Fieg (2017), was extended for the use in structural optimization problems. Another important component of the Adv:PO is the implemented parallelization concept for the parallel evaluation of several process design alternatives, as indicated in Figure 1. Due to this concept the hardware of modern computers is used efficiently and the computational time can be reduced significantly.

To complete the holistic view, the third block of the system architecture includes the decision support based on a detailed analysis of the obtained results from a process engineering point of view. One possibility is the export of the results to Microsoft Excel or the export of complete Aspen Plus® flowsheets. In addition, several integrated tools

for the analysis of the results are implemented, e.g. the visualization of the distribution of the process parameters or the convergence of the PF.

3. Exemplary design problem for a dimethyl ether production process

3.1. Dimethyl ether production process

For the evaluation of the developed optimization tool a dimethyl ether process (DME), which was presented in detail by Luyben (2011), was chosen. The example is especially suitable because the optimization results can be directly compared with conventionally achieved reference designs from the author.
In the proposed process, methanol is used as a feed to form DME and water in a gas phase reaction. The basic process consists of a feed preparation for the vaporization of methanol. Subsequently, methanol reacts partially via a catalyst in the reaction section of the process. The products are purified demanding a quality of 99.9 w% for all three components. Unreacted methanol is recycled back to the process. For a more detailed description of the process it is referred to Luyben (2011).

3.2. Definition of the optimization problem

The superstructure drawn in Aspen Plus® is presented in Figure 2. It was developed on the basis of the proposed three alternatives for the reaction section by Luyben (2011) and the original Aspen Plus simulation files, kindly provided by the publisher.

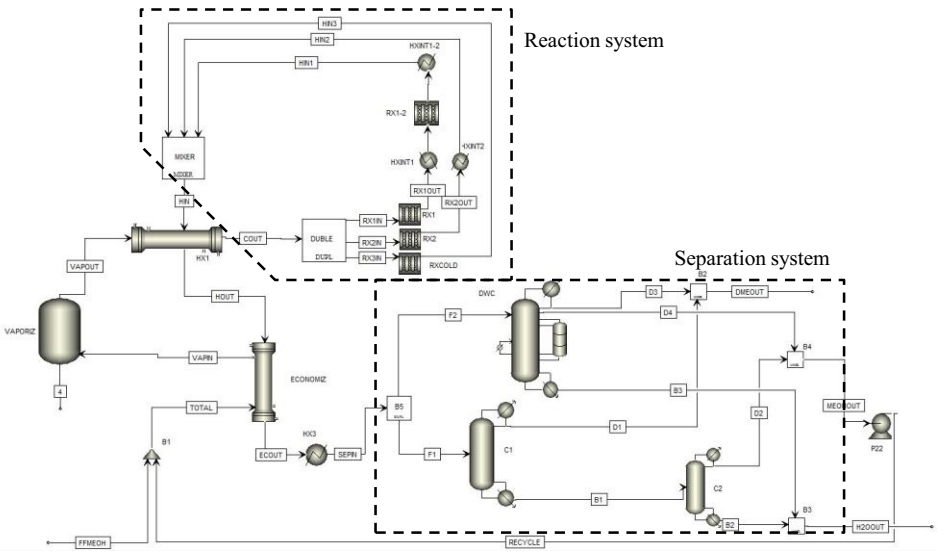


Figure 2: Aspen Plus® flowsheet of the superstructure for the multi-objective optimization of a dimethyl ether production process.

The first alternative is a cooled reactor, the second, one adiabatic reactor with a subsequent heat exchanger (HX), and the third, a series of two adiabatic reactors and HXs. For the separation system a dividing wall column (DWC) was introduced by the authors as an alternative for the direct sequence of two distillation columns proposed by Luyben (2001). In summary, six different process structure alternatives will be regarded for the optimization. For a high comparability of the results the specified objective functions, which are the investment and operating costs of all apparatuses, were

carefully reconstructed (Luyben, 2011). The costs of the newly added DWC are assumed to be similar to the costs of the distillation columns. The constraints demand a high chemical purity of 99.9 w% for all three product streams. In addition, various constraints considering the hydrodynamics of the columns are implemented in the flowsheet. The decision variables with their minimum and maximum values as well as the step size are summarized in Table 1. If an apparatus appears multiple times in the flowsheet a multiple number of the listed decision variables is used due to the modular concept described in the previous section. As the values of the decision variables depend on the process structure altogether 68 decision variables are regarded.

Table 1: Summary of the decision variables for the optimization of a dimethyl ether process.

Type of apparatus	Decision variable	Upper Bound	Lower Bound	Step size	Unit
Vaporizer	Pressure	13.5	15	0.1	atm
Adiabatic reactor	Length	5	12	0.1	m
	Length to diameter ratio	10	15	0.1	-
Heat exchanger	Temperature	550/570	650/670	1	K
Cooled reactor	Number of tubes	150	450	1	-
	Tube length	5	15	0.1	m
	Temperature	600	700	1	K
Distillation column	Number of stages	10	50	1	-
	Feed position	0.1	0.9	0.01	-
	Condenser pressure	1/9	5/10.5	0.1	atm
	Reboiler duty	1	5	0.001	MW
Dividing wall column	Total number of stages	9	65	1	-
	Number of stages for dividing wall	5	25	1	-
	Horizontal position of dividing wall	0.1	0.99	0.1	-
	Reboiler duty	2,5	5	0.001	MW
	Liquid split	0.2	1	0.001	-
	Feed position	0.1	0.9	0.05	-
	Side product position	0.1	0.9	0.05	-
	Condenser pressure	6	10.5	0.1	atm

3.3. Optimization results

For the determination of the optimization results presented in Figure 3, about 48 000 design alternatives have been simulated. The authors do not give information about the computational time intentionally because it highly depends on the independent processing units of the hardware that are used by the previously mentioned specially developed parallelization concept of the Adv:PO. The results presented in Figure 3 are normalized to one reference solution proposed by Luyben (2011) and show a broad PF with 91 process design alternatives. The alternatives consist of three different process structures: two adiabatic reactors plus two distillation columns, two adiabatic reactors plus DWC and cooled reactor plus DWC. Compared to the reference solution a maximum saving of ca. 30 % considering the investment costs is possible. The operating costs can be reduced to a maximum of 11 %. A more detailed analysis of the optimization results goes beyond the scope of this paper.

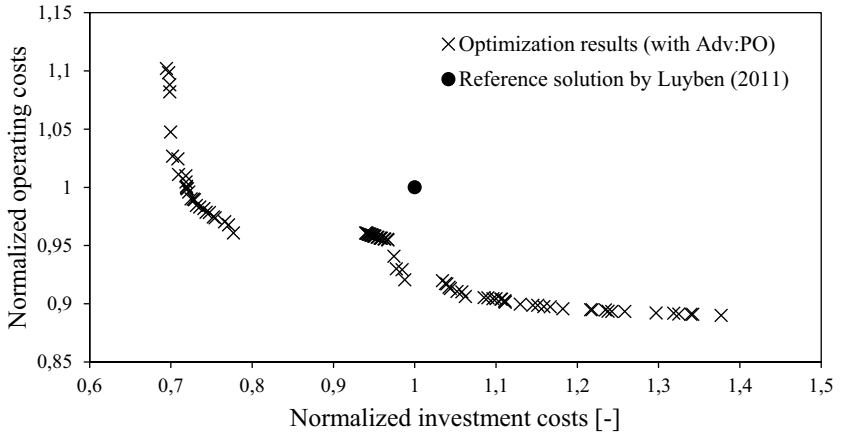


Figure 3: Optimization results determined with the Adv:PO normalized and in comparison to one reference solution calculated by Luyben (2011).

4. Conclusions

For the solution of simulation-based multi-objective structural and parameter optimization problems in a GDP formulation a robust optimization tool, the Adv:PO, was presented. The modular concept for the flexible and automatic consideration of the objective functions and constraints depending on the currently evaluated process structure allows the optimization of basically every process. Moreover, it is possible to draw the superstructure directly in a commercial flowsheet simulator. The developed optimization methods achieved very good results for the exemplary optimization of a DME process in comparison to a reference solution determined by conventional process design methods. The developed methodical approach, the open architecture of the Adv:PO combined with the encouraging results for the exemplary design problem allow the application on more complex design problems. Additionally, not only grass-root design problems can be addressed, but also retrofit problems in an early project phase.

References

P. S. Ernst, K. Zimmermann, G. Fieg, 2017, Multi-objective Optimization-Tool for the Universal Application in Chemical Process Design, Chem. Eng. Technol., 40, No. 10, 1867-1875

R. Chiong, 2009, Nature-Inspired Algorithms for Optimisation, Berlin, Germany

W. L. Luyben, 2011, Principles and Case Studies of Simultaneous Design, USA

M. A. Navarro-Amorós, R. Ruiz-Femenia, J. A. Caballero, 2014, Integration of modular process simulators under Generalized Disjunctive Programming framework for the structural flowsheet optimization, Comput. Chem. Eng., 67, 13-25

M. Shokrian, K. A. High, 2014, An efficient multi criteria process optimization framework: Sustainable improvement of the Dimethyl Ether Process, Comput. Chem. Eng, 60, 213-230

F. Trespalacios, I.E. Grossmann, 2014, Review of Mixed-Integer Nonlinear and Generalized Disjunctive Programming Methods, Chem. Ing. Tech., 86, No. 7, 991-1012

K. Zimmermann, G. Fieg, 2017, Development of a Diversity-preserving Strategy for the Pareto-optimization in Chemical Process Design, Chem. Ing. Tech., 89, No. 10, 1297-1305

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50123-3

Parsimonious Input Parameterization for Dynamic Optimization Problems

Diogo Rodrigues*, Dominique Bonvin

Laboratoire d'Automatique, EPFL, 1015 Lausanne, Switzerland
diogo.mateusrodrigues@epfl.ch

Abstract

This contribution presents an input parameterization for dynamic optimization problems, which allows reducing the number of decision variables compared to traditional direct methods. Firstly, a finite set of plausible arc sequences is postulated. Then, each arc sequence is described by a small number of parameters that include switching times and initial conditions for the sensitivity-seeking arcs. These arcs are parsimoniously described by cubic splines using only a few parameters. Finally, optimal parameter values are determined for each arc sequence. The reduction in the number of decision variables is such that the problem is often amenable to global optimization. The procedure is illustrated via the optimization of a simulated semi-batch reactor.

Keywords: global dynamic optimization, parsimonious input parameterization, cubic spline.

1. Introduction

Dynamic optimization methods that rely on local numerical optimization are well established (Biegler, 2010). Direct methods reformulate the original infinite-dimensional problem as a finite-dimensional problem by discretizing either the inputs (in sequential methods) or both the inputs and the states (in simultaneous methods). Although these methods typically attain local optimality, they could in principle be extended to global optimality. However, the complexity scales exponentially with the number of decision variables, which is proportional to the number of input parameters in the case of sequential methods (Houska and Chachuat, 2014). Hence, the use of direct methods for global optimization either calls for a rather coarse input parameterization or results in intractable problems.

An apparent way to avoid these difficulties is to use an input parameterization that can approximate the optimal inputs well with a limited number of parameters. This contribution proposes a parsimonious input parameterization, whereby the parameters correspond to (i) switching times between arcs, and (ii) a few parameters used to describe sensitivity-seeking arcs. A similar parameterization was applied to the dynamic optimization of switched systems (Xu and Antsaklis, 2004). However, to the authors' knowledge, it has not been proposed for problems with inequality constraints or in the context of global optimization.

2. Preliminaries

This section presents the formulation of the class of dynamic optimization problems that are considered in this paper and the method used to compute *adjoint-free* optimal control laws.

2.1. Formulation of the dynamic optimization problem

For the sake of clarity, but without loss of generality, this paper will deal with the single-input case. The optimization problem considered in this paper is formulated in Mayer form as

$$\min_{u(\cdot),t_f} \quad \mathscr{J}(u(\cdot),t_f) := \phi(\mathbf{x}(t_f),t_f), \tag{1a}$$

$$\text{s.t.} \quad \mathscr{T}(u(\cdot),t_f) := \boldsymbol{\psi}(\mathbf{x}(t_f),t_f) \leq \mathbf{0}_{n_\psi}, \tag{1b}$$

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}(t),u(t)), \quad \mathbf{x}(t_0) = \mathbf{x}_0, \tag{1c}$$

$$\mathbf{g}(\mathbf{x}(t),u(t)) \leq \mathbf{0}_{n_g}, \tag{1d}$$

$$\mathbf{h}(\mathbf{x}(t)) \leq \mathbf{0}_{n_h}, \tag{1e}$$

where t_f is the finite final time, $u(t)$ is the scalar piecewise-continuous input with trajectory $u(\cdot)$, $\mathbf{x}(t)$ are the piecewise-continuously differentiable states of dimension n_x, and $\phi(\mathbf{x},t)$, $\boldsymbol{\psi}(\mathbf{x},t)$, $\mathbf{f}(\mathbf{x},u)$, $\mathbf{g}(\mathbf{x},u)$, $\mathbf{h}(\mathbf{x},u)$ are smooth functions of dimensions 1, n_ψ, n_x, n_g, n_h, respectively. Moreover, it is assumed that $\mathbf{g}(\mathbf{x},u)$ and $\mathbf{h}^{(1)}(\mathbf{x},u) := \frac{\partial \mathbf{h}}{\partial \mathbf{x}}(\mathbf{x})\mathbf{f}(\mathbf{x},u)$ depend explicitly on u.

2.2. Types of input arcs

Most often, the optimal input consists of a finite number of arcs. For a given arc, the optimal input is determined by either an active path constraint or a single condition that expresses a physical compromise in the system dynamics. Moreover, the resulting optimal control laws (but not necessarily the control values) are independent of the initial conditions and the terminal constraints (Srinivasan et al., 2003).

The control laws express the optimal input or one of its time derivatives in terms of the states and possibly the input and its time derivatives, thus resulting in an *adjoint-free* optimal control law. Each input arc is of one of the following three types (Srinivasan et al., 2003):

1. If the optimal input is determined by the active mixed path constraint $g_k(\mathbf{x},u) \leq 0$ for some $k = 1,\ldots,n_g$, then the control law $u = c(\mathbf{x})$ enforces $g_k(\mathbf{x},u) = 0$.

2. If the optimal input is determined by the active pure-state path constraint $h_k(\mathbf{x}) \leq 0$ for some $k = 1,\ldots,n_h$, then the control law $u = c(\mathbf{x})$ enforces $h_k^{(1)}(\mathbf{x},u) = 0$.

3. If the optimal input is not determined by any active path constraints, then the control law $u^{(\xi)} = c(\mathbf{x},u,\ldots,u^{(\xi-1)})$ enforces $\det(\mathscr{M}) = 0$. Let $\mathbf{x}^u$ be the ρ states that can be reached by manipulating u, with dynamics $\dot{\mathbf{x}}^u = \mathbf{f}^u(\mathbf{x},u)$. Then, we can

construct the $(\rho \leftrightarrow p)$ matrix $\mathcal{M} := \left[\frac{\partial \mathbf{f}^u}{\partial u}(\mathbf{x},u) \vdots \Delta\frac{\partial \mathbf{f}^u}{\partial u}(\mathbf{x},u) \vdots \cdots \vdots \Delta^{\rho-1}\frac{\partial \mathbf{f}^u}{\partial u}(\mathbf{x},u)\right]$ using symbolic computations and the operator $\Delta\mathbf{v} := \frac{\partial \mathbf{v}}{\partial \mathbf{x}}\mathbf{f}(\mathbf{x},u) - \frac{\partial \mathbf{f}^u}{\partial \mathbf{x}^u}(\mathbf{x},u)\mathbf{v} + \sum_{k=0}^{\infty}\frac{\partial \mathbf{v}}{\partial u^{(k)}}u^{(k+1)}$, with $\Delta^l\mathbf{v} := \Delta(\Delta^{l-1}\mathbf{v})$.

The input arcs of types 1 and 2 are labeled *constraint-seeking*, while the input arcs of type 3 are *sensitivity-seeking*. Note that switching between arcs can happen any time, except for the switching to an arc of type 2 that can only occur when the states $\mathbf{x}$ satisfy $h_k(\mathbf{x}) = 0$.

The analytical approach for the computation of adjoint-free optimal control laws shown above is quite useful to get an analytical characterization of the input arcs. However, when the number of reachable states ρ grows beyond 4 or 5, the size and complexity of $\mathcal{M}$ may become intractable. In that case, a more practical approach is needed.

2.3. Approximation of optimal control laws using cubic splines

The idea is to replace the analytical characterization of sensitivity-seeking input arcs by a description using cubic splines (piecewise-cubic functions). Since the optimal profiles in these arcs are smooth functions, they can be approximated by cubic splines with only a few parameters. Piecewise-constant, piecewise-linear or piecewise-quadratic functions could also be used, but these would typically require more parameters for the same quality of approximation.

A cubic spline with π intervals starts at the point (t^0, y^0), passes through the $\pi - 1$ intermediate points $(t^1, y^1),...,(t^{\pi-1}, y^{\pi-1})$ and ends up at the point (t^π, y^π). It is typically parameterized by the $\pi + 1$ coordinates $y^0,..., y^\pi$ and 2 conditions at the endpoints t^0 and t^π (for example, on the second derivatives). Hence, $\pi + 3$ parameters are needed to describe the cubic spline.

In this paper, to enable the computation of sensitivity-seeking input arcs via forward integration of differential equations, a different interpretation is assigned to these $\pi + 3$ parameters. Three parameters correspond to the initial value $u(t^0)$, the initial first derivative $\dot{u}(t^0)$ and the initial second derivative $\ddot{u}(t^0)$, while the remaining π parameters $\mathbf{p}$ correspond to the third derivatives in the π intervals of the cubic spline. Hence, the input u is described by

$$u^{(3)}(t) = p_1, \quad \forall t \in [t^0, t^1); \quad ... \quad ; \quad u^{(3)}(t) = p_\pi, \quad \forall t \in [t^{\pi-1}, t^\pi), \tag{2}$$

with the three initial values $u(t^0)$, $\dot{u}(t^0)$, and $\ddot{u}(t^0)$.

In the remainder, these $\pi + 3$ parameters will be considered as "initial conditions" for the input arc. Note that the locations of the intermediate points are not input parameters since they are fixed at $t^m := (1 - w^m)t^0 + w^m t^\pi$ for all $m \in \{0,1,...,\pi\}$, where w^m is a constant that specifies the relative position of t^m with respect to t^0 and t^π, with $0 = w^0 < w^1 < ... < w^\pi = 1$.

3. Parsimonious input parameterization

3.1. Basic idea

This approach, labeled *parsimonious input parameterization*, was first proposed by Srinivasan et al. (2003). It consists of the following steps:

1. Generate a finite set of arc sequences from all the input arcs that can occur, and choose an initial arc sequence.
2. Use numerical optimization to compute the optimal switching times to arcs of types 1 and 3 and initial conditions for the sensitivity-seeking arcs.
3. Check if the cost for the chosen arc sequence is less than the current optimal cost and, if so, update the current optimal cost and optimal arc sequence.
4. Choose a different arc sequence and repeat Steps 2-4 until all sequences are investigated.

The optimization in Step 2 typically uses algorithms that require the computation of gradients with respect to the decision variables. At each iteration, the states are computed via forward integration of the dynamic equations, since the input can be generated without knowledge of the adjoint variables. The next subsections give more details about this implementation.

Remark. If the optimization in Step 2 can be solved to global optimality for each arc sequence, then the input trajectory with the lowest cost will be globally optimal. To enforce global optimality, one could express the terminal cost and constraints as explicit *polynomial* functions of the decision variables and compute the global solution to this polynomial optimization problem via reformulation as a semidefinite program (Lasserre, 2009). Note that the fact of having fewer decision variables facilitates the whole procedure.

3.2. Model of an arc sequence

Consider a given arc sequence that includes n_s+1 arcs of types 1 and 3, where n_s is the number of switching times to these arcs. These switching times $t_1,...,t_{n_s}$ are considered as decision variables, while the switching times to arcs of type 2 are not decision variables since they depend on the states and cannot occur arbitrarily. We show next that the input $u(t)$ for this sequence is fully defined by the switching times $t_1,...,t_{n_s}$, the final time t_f, and the initial conditions of the sensitivity-seeking arcs, which become the decision variables of the problem.

Note that, in addition to the switching times $t_1,...,t_{n_s}$ considered as decision variables, other switchings between arcs occur when a pure-state path constraint becomes active. In that case, the input $u(t)$ is adjusted such that the path constraint remains active.

Let us describe the input in the time interval $[t_{i-1},t_i)$, for some $i=1,...,n_s+1$, where $t_{n_s+1}=t_f$. There is a degree $\xi_i \geq 0$ for which $u^{(\xi_i)}(t)$ is a function of the states, the input,

the time derivatives of the input, and the vector $\mathbf{p}_i$ of constant third derivatives in the π_i intervals of the cubic spline. This differential relationship requires specifying the initial conditions $u^{(\xi_i-1)}(t_{i-1}),...,u(t_{i-1})$. The element $p_{i,m}$ of the vector $\mathbf{p}_i$ is used only for $t \in [t_i^{m-1}, t_i^m)$, with the instants $t_i^0, t_i^1, ..., t_i^{\pi_i}$ specified by the constants $w_i^0, w_i^1, ..., w_i^{\pi_i}$. Upon defining the $\pi_i + \xi_i$ states $\mathbf{z}_i(t) := [\mathbf{p}_i^{\mathrm{T}} \; u^{(\xi_i-1)}(t) \; \cdots \; u(t)]^{\mathrm{T}}$, one can describe their dynamics for $t \in [t_{i-1}, t_i)$ as

$$\dot{\mathbf{z}}_i(t) = \mathbf{q}_i(\mathbf{x}(t), \mathbf{z}_i(t)), \quad \mathbf{z}_i(t_{i-1}) = \mathbf{z}_{i,0}. \tag{3}$$

Then, for the arc sequence, one defines the extended states $\mathbf{z}(t) := [\mathbf{x}(t)^{\mathrm{T}} \; \mathbf{z}_1(t)^{\mathrm{T}} \; \cdots \; \mathbf{z}_{n_s+1}(t)^{\mathrm{T}}]^{\mathrm{T}}$ and uses the control law $u(t) = \tilde{c}(\mathbf{z}(t))$, which is continuous unless a switching occurs, to eliminate the input dependencies and rewrite the functions in Problem (1) in terms of $\mathbf{z}$, that is, $\tilde{\mathbf{f}}(\mathbf{z}(t))$, $\tilde{\phi}(\mathbf{z}(t_f), t_f)$, $\tilde{\boldsymbol{\psi}}(\mathbf{z}(t_f), t_f)$, $\tilde{\mathbf{g}}(\mathbf{z}(t))$, $\tilde{\mathbf{h}}(\mathbf{z}(t))$, $\tilde{\mathbf{h}}^{(1)}(\mathbf{z}(t))$, with the initial conditions $\mathbf{z}_0 := [\mathbf{x}_0^{\mathrm{T}} \; \mathbf{z}_{1,0}^{\mathrm{T}} \; \cdots \; \mathbf{z}_{n_s+1,0}^{\mathrm{T}}]^{\mathrm{T}}$. The system equations now read:

$$\dot{\mathbf{z}}(t) = \tilde{\mathbf{f}}(\mathbf{z}(t)), \quad \mathbf{z}(t_0) = \mathbf{z}_0. \tag{4}$$

3.3. Numerical optimization for a given arc sequence

To check whether a given arc sequence is a solution to Problem (1), one uses the fact that any input trajectory implementing that arc sequence is given by some switching times and initial conditions for the sensitivity-seeking arcs. Then, the functionals $\mathscr{J}$ and $\mathscr{T}$ reduce to functions $\bar{\phi}$ and $\bar{\boldsymbol{\psi}}$ of the switching times $t_1,...,t_{n_s}$, the final time t_f and the initial conditions $\mathbf{z}_{1,0},...,\mathbf{z}_{n_s+1,0}$, which is more convenient for numerical optimization. This implies that, for any values of the decision variables $\boldsymbol{\tau} := (t_1,...,t_{n_s}, t_f, \mathbf{z}_{1,0},...,\mathbf{z}_{n_s+1,0})$, the terminal cost and constraints $\bar{\boldsymbol{\chi}}(\boldsymbol{\tau}) := [\bar{\phi}(\boldsymbol{\tau}) \; \bar{\boldsymbol{\psi}}(\boldsymbol{\tau})^{\mathrm{T}}]^{\mathrm{T}}$ can be computed via numerical integration of Eq. (4) and evaluation of the functions $\tilde{\boldsymbol{\chi}}(\mathbf{z}(t_f), t_f) := [\tilde{\phi}(\mathbf{z}(t_f), t_f) \; \tilde{\boldsymbol{\psi}}(\mathbf{z}(t_f), t_f)^{\mathrm{T}}]^{\mathrm{T}}$.

However, the gradients of $\bar{\boldsymbol{\chi}}$ with respect to the decision variables $\boldsymbol{\tau}$ also need to be supplied to the optimization algorithm. These gradients can be computed via adjoint sensitivity analysis, which is efficient when the number of functions $\bar{\boldsymbol{\chi}}$ is small. For this, each time Eq. (4) is integrated forward from t_0 to t_f to compute the states $\mathbf{z}(t)$, the adjoint variables $\boldsymbol{\zeta}(t) := [\boldsymbol{\lambda}(t)^{\mathrm{T}} \; \boldsymbol{\zeta}_1(t)^{\mathrm{T}} \; \cdots \; \boldsymbol{\zeta}_{n_s+1}(t)^{\mathrm{T}}]^{\mathrm{T}}$ are computed by integrating backward from t_f to t_0

$$\dot{\boldsymbol{\zeta}}(t) = -\frac{\partial \tilde{\mathbf{f}}}{\partial \mathbf{z}}(\mathbf{z}(t))^{\mathrm{T}} \boldsymbol{\zeta}(t), \quad \boldsymbol{\zeta}(t_f) = \frac{\partial \tilde{\boldsymbol{\chi}}}{\partial \mathbf{z}}(\mathbf{z}(t_f), t_f)^{\mathrm{T}}. \tag{5}$$

Moreover, for each instant θ at which $\tilde{h}_k(\mathbf{z}(\theta^-)) < 0$ and $\tilde{h}_k(\mathbf{z}(\theta)) = 0$ for some $k = 1,...,n_h$, which is equivalent to saying that $\tilde{h}_k(\mathbf{z}(t)) \le 0$ becomes active at $t = \theta$, it holds that:

$$\boldsymbol{\zeta}(\theta^-) = \boldsymbol{\zeta}(\theta) - \frac{\partial \tilde{h}_k}{\partial \mathbf{z}}(\mathbf{z}(\theta^-))^{\mathrm{T}} \frac{[\tilde{\mathbf{f}}(\mathbf{z}(\theta^-)) - \tilde{\mathbf{f}}(\mathbf{z}(\theta^+))]^{\mathrm{T}} \boldsymbol{\zeta}(\theta)}{\tilde{h}_k^{(1)}(\mathbf{z}(\theta^-))}. \tag{6}$$

Finally, the gradients of $\bar{\chi}$ with respect to the decision variables $\boldsymbol{\tau}$ are obtained from the knowledge of the states $\mathbf{z}$ and adjoint variables $\boldsymbol{\zeta}$. It would also be possible to compute higher-order derivatives of $\bar{\chi}$ with respect to the decision variables, which would require the computation of derivatives of the states $\mathbf{z}$ and adjoint variables $\boldsymbol{\zeta}$ with respect to these decision variables. However, this computation is out of the scope of this paper.

4. Simulated example

The example describes the acetoacetylation of pyrrole in a homogeneous semi-batch reactor (Ruppen et al., 1998). The solution with the parsimonious parameterization results in an input trajectory described by 6 parameters. This parameterization outperforms the piecewise-constant parameterization of a direct method using the same number of parameters, which results in a rather coarse input trajectory and a worse cost; about 50 parameters are needed for the piecewise-constant parameterization to obtain an equally smooth input trajectory with similar cost. Details are available from the authors.

5. Conclusions

This paper has presented a parsimonious parameterization that describes the optimal input accurately with fewer parameters than direct methods, which reduces the number of decision variables and makes it a valid candidate for global dynamic optimization.

Future work will consider the multiple-input case, which requires more attention in the definition and selection of arc sequences. Also, it would be useful to express the terminal cost and constraints of dynamic optimization problems as explicit polynomial functions of the decision variables, which is typically required for global optimization.

References

L. T. Biegler, 2010. Nonlinear Programming: Concepts, Algorithms, and Applications to Chemical Processes. Society for Industrial and Applied Mathematics.

B. Houska, B. Chachuat, 2014, Branch-and-lift algorithm for deterministic global optimization in nonlinear optimal control, J. Optim. Theory Appl., 162, 208–248.

B. Houska, H. Ferreau, M. Diehl, 2011, ACADO toolkit – An open source framework for automatic control and dynamic optimization, Optim. Control Appl. Methods, 32, 3, 298–312.

J. B. Lasserre, 2009. Moments, Positive Polynomials and Their Applications. World Scientific.

D. Rodrigues, S. Srinivasan, J. Billeter, D. Bonvin, 2015, Variant and invariant states for reaction systems, Comput. Chem. Eng., 73, 1, 23–33.

D. Ruppen, D. Bonvin, D. W. T. Rippin, 1998, Implementation of adaptive optimal operation for a semi-batch reaction system, Comput. Chem. Eng., 22, 1, 185–189.

B. Srinivasan, S. Palanki, D. Bonvin, 2003, Dynamic optimization of batch processes: I. Characterization of the nominal solution, Comput. Chem. Eng., 44, 1–26.

X. Xu, P. J. Antsaklis, 2004, Optimal control of switched systems based on parameterization of the switching instants, IEEE Trans. Autom. Contr., 49, 1, 2–16.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50124-5

Advanced Model-Based Optimization of a Commercial Natural Gas Liquid Recovery Unit

Dara Satyadileep[a], Abdallah S. Berrouk[a*], Yasser Al Wahidi[a], S. Leyland[b], A. Andrade[b], D. Georgis[b], A.S. El Nasr[c], F. Geuzebroek[c]

[a] *Khalifa University of Science and Technology, Petroleum Institute, Abu Dhabi, PO Box 2533, Abu Dhabi, UAE*
[b] *Process Systems Enterprise, W6 7HA, London, UK*
[c] *ADNOC Gas Processing, PO Box 898, Abu Dhabi, UAE*
aberrouk@pi.ac.ae

Abstract

In this paper we use the example of a commercial NGL Plant in Abu Dhabi to explain the practical benefits of Advanced Model-based Optimization. We will describe and show how the key first step in such a study is validation: ensuring that model predictions of key results are matching the actual plant behavior. We will describe how the model is then used to understand the best operational strategies for the facility through model-based optimization. This demonstrates how the approach used overcomes the limitations of the typical workflow and models used, in particular, the ones based on the Sequential Modular approach. The trial-and-error simulations used in this approach are not only time consuming and error prone but often limited in the operating decisions and constraints that can be considered. We will show how using an Equation Oriented environment overcomes these limitations and allows an optimal solution satisfying all operational and safety constraints to be found using mathematical optimization methods. Finally we will show how this leads to the set-up of operating guidelines to allow operators to drive the plant towards optimal operation; giving them the ability to respond to changes in the plant or economic conditions quicker and with more confidence than previously possible.

Keywords: Natural Gas Liquid, Process Simulation, Equation Oriented Approach, Optimization

1. Introduction

Process simulation software has long been used as a key tool to aid in the design and operation of complex plants with the aim to identify optimized designs and operating procedures. Typically, with such tools, a trial and error methodology is adopted to "optimize" the design or operation. Rigorous optimization techniques and tools have been increasingly used, as summarized below, but these either require additional tools compared to the standard ones used in industry or expert users. A body of research on NGL unit optimization has appeared in the literature where various optimization techniques and simulators have been used for such a purpose. Diaz et al. (2002) formulated a MINLP (mixed integer non-linear programming) mathematical model to achieve the optimum operating conditions and flexibility under various feed conditions of various NGL recovery processes. Mehrpooya et al. (2006) used an advanced method known as VPGA (Variable Population size Genetic Algorithm) for parametric optimization of an existing NGL recovery plant. Chebbi et al. (2010) optimized a typical turbo-expander process to various de-methanizer pressures and feeds conditions through process simulations using HYSYS. Yusoff et al. (2011) used the Taguchi parametric

design approach to select the optimization variables for a refrigerated gas plant simulated in HYSYS. Ghorbani et al. (2012) used a combination of pinch and exergy analysis to optimize the energy recovery of the refrigeration cycle in the NGL recovery plant. Park et al. (2015) presented a novel modified form of coordinate descent methodology (MCD) for solving complex NG process optimization problems. Getu et al. (2012) developed a six chance-constrained model using HYSYS and performed the process optimization using the general algebraic modelling system (GAMS). A comprehensive analysis of existing/past technologies and the state of the art in optimization of integrated process systems through the use of process flowsheeting tools was published recently (Pantelides et al. 2013). The main technology adopted by most commercial steady-state flowsheeting tools used is based on the Sequential Modular (SM) approach, where, process units are solved in sequence. Simulation tools based on this approach are easy to use; however, in practice within this environment formal mathematical optimization is very limited due to its problematic handling of non-standard specifications, closing of recycle loops and the unavailability of gradients (which are required by numerical optimization algorithms). To overcome this difficulty, engineers resort to trial-and-error steady state simulations for their optimization studies. This can be time-consuming and error-prone without any guarantee of obtaining an optimal solution or even one that satisfies all important constraints. The need to move away from the Sequential Modular approach in process flowsheeting has long been recognized and there has been increasing development in tools using the Equation Oriented (EO) approach. In this approach, the mathematical model, built up by the representation of each unit operation in a flowsheet, is solved as a single set of equations simultaneously rather than as a unit by unit Sequential Modular approach. This technology can overcome many of the limitations of the Sequential Modular approach. In particular, because the entire system of equations is available it is possible to calculate the partial derivatives required by optimization solvers. In this work we develop a flowsheet model of an NGL plant in gPROMS ProcessBuilder® (Process Systems Enterprise Ltd.), an Equation Oriented process flowsheeting tool and demonstrate the advantages of such technology in terms of performing optimization of a commercial NGL unit.

2. NGL Recovery Process Description

2.1 Overview

The NGL recovery unit produces two main products (NGL and condensates) from the treated natural gas stream. In addition to the two product streams, the majority of the feed material (methane) exits the unit as residue gas which is then sold as sales gas with its value determined through a set of product specifications (e.g. net heating value). Removal of the NGL and condensate components from the feed gas stream results in a reduced revenue for the residue gas stream as the larger hydrocarbon molecules have a higher heating value. Depending on the market conditions a certain level of heavier hydrocarbons may be required in the residue gas to maintain the heating value specification, especially if there are inerts present such as carbon dioxide or nitrogen. Although recovering the maximum amount of liquids in NGL and condensate streams increases the revenues from these streams, a balance is required to ensure the value of the residue gas is maintained and it meets the required specifications for the customer. A simplified schematic of such a process is shown in Figure 1.

2.2 Operation

The key requirements for the operation of these units include:

- Recovery efficiency- (in this case the propane recovery is considered)
- Product purity – the NGL product stream has a maximum acceptable levels for both light (C_2 and lower) and heavy (C_{5+}) components.
- Residue gas specification –a minimum net heating value is required for distribution through the sales gas grid.

Therefore optimizing the operation of such plants is not a simple matter of producing as much NGL liquids as possible given the required product purity. Another complicating factor is variation in the composition of feed gas, which can affect the above operational requirements.

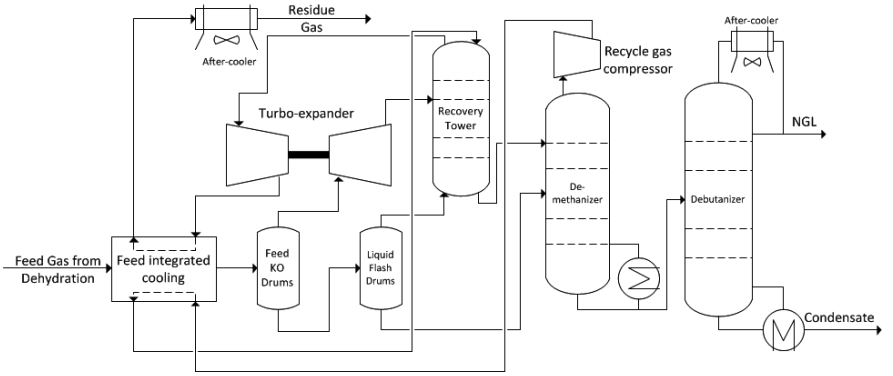


Figure 1: NGL unit simplified schematic

3. Unit Model Validation

The overall topology of the developed model is shown in Figure 2. A model of the propane refrigeration cycle was also developed and integrated with the process section via a hierarchical modeling structure. The topology of the refrigeration cycle is not shown herein. For validation the model was first configured to represent the original design and validated against the design data, with all model predictions within 1% of original design data. Plant data for an entire year was then analysed and used to validate the model at different steady state points during different seasons, using the built in parameter estimation capabilities of the software tool being used to estimate the unknown inputs to the model (not provided by plant measurements). Table 1 shows the deviations of the model predictions in key (product) results for a representative steady state. In total 90 plant measurements are considered with deviations within measurement errors considered.

4. Unit Model Optimization

4.1. Problem formulation

The main objective of the present optimization exercise is to maximize the profitability of the plant (revenue minus cost). The decision variables consist of various operating conditions that can be modified such as pressures, flow splits, steam duties into reboilers, etc. The main constraints include product specifications and equipment limitations and capacities. This optimization problem is formulated in the same Equation Oriented

package used to develop the flowsheet model. The main elements of the optimization problem formulation are listed below:

Table 1: Key result comparison with plant data

Stream	Measurement [unit]	Absolute Deviation
Residue Gas	Compressibility [-]	<0.01
	Heating Value Gross [BTU/ft3]	-2
	Heating Value Net [BTU/ft3]	-2
	Molecular Weight [g/mol]	<0.05
	Specific Gravity (relative to air)	<0.01
	Temperature [degC]	0.5
	Molar flowrate [kNm3/h]	10
NGL	Compressibility [-]	<0.01
	Heating Value Gross [BTU/ft3]	-25
	Heating Value Net [BTU/ft3]	-23
	Molecular Weight [g/mol]	-1.0
	Specific Gravity (relative to air)	-0.1
	Volumetric flowrate [m3/h]	6
Condensate	Mass density [g/mL]	0.02
	Volumetric flowrate [m3/h]	0.3
	Reid vapour pressure [psi]	-0.4
	Temperature [degC]	-5

- Objective function
  - Annualized net revenue, which is the summation of the revenue of the three product streams minus the total utility cost
- Decision variables
  - Main process
    - Reboiler steam duty in de-methanizer and debutanizer
    - Flow splits of feed gas to different heat exchangers
    - Column pressures (all columns)
  - Refrigerant loop
    - Circulating flowrate
- Constraints
  - Product specifications
    - Residue gas – minimum net heating value (BTU)
    - NGL liquid
      - C2 and lighter components maximum composition (5% by mole)
      - C5 and heavier components maximum composition (4% by mole)
    - Condensate – a maximum Reid Vapor Pressure (RVP) specification is required
  - Additional constraints
      - In addition to the product constraints, additional constraints are added to avoid undesirable operating points, such as small temperature differences at heat exchanger boundaries.

The package's built-in optimization capabilities were then used to solve the mathematical optimization problem. In total the optimization problem consists of 10 decision variables, 15 constraints and 80,000 equality constraints (flowsheet model equations).

4.2. Results and Discussion

Table 2 summarizes the results obtained during the optimization and the percentage improvement compared to the original baseline (values normalized to a product revenue of $100million/yr). An increase of 1.51% is seen in the net revenue. The increased revenue requires a higher energy utilization which is reflected in energy cost 5.5% greater than the baseline operating point. On the other hand, the economic improvement is achieved by increasing the production of both liquids, and sacrificing both the production and the residue gas specifications.

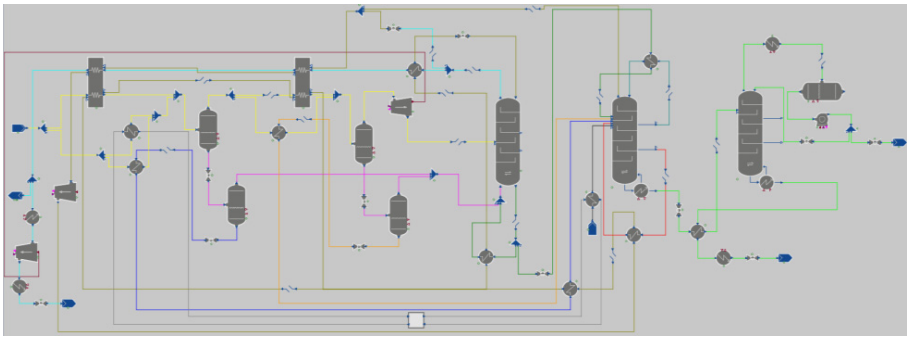


Figure 2: Overall topology of NGL flowsheet model

Table 2: Optimization results

	Baseline ($million/yr normalized)	**Optimized**
Net Revenue	99.611	+1.51%
Revenue	100	+1.53%
Costs	0.389	+5.47%

Table 3 summarizes the changes in the decision variables in the calculated optimal point compared to the original baseline values. It can be seen that the optimal point is a result of a number of factors. The recovery tower pressure is reduced, which gives a greater pressure drop and therefore refrigeration in the turbo expander. In addition flow splits of the feed gas change in a way to increase the amount of refrigeration from the propane loop and recycle heat exchangers. The pressure in the de-methanizer is also reduced compared to the baseline, lowering the steam duty required in the de-methanizer. The pressure in the de-butanizer is higher compared to the baseline, to meet the constraint on the lighter components in the NGL product stream. Due to the higher pressure in the debutanizer, the steam duty in reboiler is increased. To maximize the production of NGL and condensate, both product specifications reach their limits. It is therefore evident that imposing appropriate constraints in an optimization problem is important. The approach taken in this work, solving the optimization problem mathematically, ensures that the optimization algorithm returns a feasible (all the constraints are satisfied) solution.

5. Conclusions

In this paper, we demonstrated how a steady state flowsheet model developed in an Equation Oriented flowsheeting tool can be deployed for an optimization activity in order to improve the economics of a commercial NGL plant by modifying specified operating conditions. In the current work being carried out a series of such optimization

calculations are to be performed, with a wider range of controls, to help improve understanding of how plant operation can be changed in situations that are seen during day to day operation such as feed stock variations. Results from these calculations will be used to determine trends in plant operating conditions and to develop a set of operating guidelines to help operators drive the plant towards optimal operation.

Table 3: Optimized decision variables and product constraints

	Optimized (change from baseline)
Decision variables	
Recovery column pressure	-6.80%
De-methanizer pressure	-9.15%
De-butanizer pressure	12.53%
De-methanizer steam duty	-13.96%
De-butanizer steam duty	17.04%
Flow Split 1	-6.17%
Flow Split 2	-1.34%
Flow Split 3	-15.67%
Constraints	
Residue gas BTU	-0.6%
NGL C2-	57.2%*
NGL C5+	-28.5%
Condensate Reid Vapor Pressure	0.7%*
* these constraints are at their upper limit	

References

S. Diaz, E. Brignole, and A. Bandoni, 2002, Flexibility study on a dual mode natural gas plant in operation, Chemical Engineering Communications, 189, 623-641

M. Mehrpooya, A. Vatani, and S. Mousavian, 2010, Optimum design of integrated liquid recovery plants by variable population size genetic algorithm, The Canadian Journal of Chemical Engineering, 88, 1054-1064

R. Chebbi, N. Al-Amoodi, N. A. Jabbar, G. Husseini, and K. Al Mazroui, 2010, Optimum ethane recovery in conventional turboexpander process, Chemical engineering research and design, 88, 779-787

N. Yusoff, M. Ramasamy, and S. Yusup, 2011, Taguchi's parametric design approach for the selection of optimization variables in a refrigerated gas plant, Chemical Engineering Research and Design, 89, 665-675

B. Ghorbani, G. Salehi, H. Ghaemmaleki, M. Amidpour, and M. Hamedi, 2012, Simulation and optimization of refrigeration cycle in NGL recovery plants with exergy-pinch analysis,Journalof Natural Gas Science and Engineering, 7, 35-43

J. Park, M. Shariq, R. Andika, 2015, Techno-economic evaluation of a novel NGL recovery scheme with nine patented schemes for offshore applications, Journal of natural gas science and engineering, 27, 2-17

M. Getu, M. S. Khan, N. V. D. Long, and M. Lee, 2012, Studying the effect of feed composition variation on typical natural gas liquid (NGL) recovery process, Book Series: Computer Aided Chemical Engineering, 31, 405-409

C. Pantelides, J. Renfro, 2013, The online use of first-principles models in process operations: Review, current status and future needs, Computers & Chemical Engineering, 51, 136-148

Acknowledgements

The authors would like to acknowledge ADNOC Gas Processing and The Gas Research Center (GRC) for the support of this project.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50125-7

Surrogate Equations of State for Equation-Oriented Optimization of Polymerization Processes

John P. Eason[a], Jiayuan Kang[b], Xi Chen[b*], Lorenz T. Biegler[a*]

[a]*Chemical Engineering Dept., Carnegie Mellon University, Pittsburgh, PA 15213*
[b]*State Key Laboratory of Industrial Control Technology, Dept. of Control Science and Engineering, Zhejiang University, Hangzhou, Zhejiang 310027 P.R. China*
xichen@iipc.zju.edu.cn, lb01@andrew.cmu.edu

Abstract

In equation-oriented process optimization, thermodynamic properties are often challenging to model. Thermodynamic property models tend to be highly nonlinear and must handle complex issues such as root selection and phase identification. In this work, we propose a method for using simple surrogate equations of state (SEOS), fit locally, to derive the thermodynamic state functions of the system. This can be fit to data derived from more complex thermodynamic models or directly from experiments. The SEOS are built to complement the equation-oriented optimization paradigm, while enabling flexibility to model diverse thermodynamic behavior. While still a data-driven model, surrogate equations of state offer better extrapolation potential than general regression of thermodynamic properties. A parameter estimation problem is formulated to determine the parameters of the SEOS, and the resulting model is applied to optimize a flash model encountered in a polymerization process. Finally, we discuss the potential to incorporate SEOS in a trust region framework to adaptively manage the approximation error.

Keywords: Equation of state, Thermodynamic properties, Equation-oriented modelling, Surrogate modelling

1. Introduction

For large scale process optimization, the equation-oriented (EO) paradigm provides attractive performance in terms of computational efficiency and flexibility in problem formulation. While most unit operations are easily modelled in this approach, thermodynamic property evaluations require careful attention. Many equations of state exhibit several volume roots, which need to be either identified with their corresponding phases or ignored as physically irrelevant. Sequential modular approaches attempt to find all roots and use logical conditions to identify the phases, but the equation-oriented approach requires a customized approach for the specific form of an equation of state. For cubic equations of state, Kamath et al. (2010) and Dowling and Biegler (2015) develop phase identification conditions using the sign of the first and second derivatives of the cubic equation. These conditions can then be included in the optimization models as inequality constraints to automatically enforce correct phase identification. However, for many fluids, such as polar or large-chain fluids, cubic equations of state are not always satisfactory. The IUPAC reference (Sengers et al. 2000) reviews these methods in depth. However, when moving to more complex functional forms, the problem of phase identification becomes more serious. For example, the PC-SAFT equation can have up to five real roots for pure components (Privat et al. 2010). Even outside of the equation-oriented context phase identification is non-trivial.

Sometimes, these issues are avoided by using surrogate models of physical properties. These surrogate models may take the form of polynomial or other regressions of thermodynamic state variables as functions of temperature, pressure, and composition, which may then be incorporated in an equation-oriented framework. However, these models are inherently local in nature, and the construction of suitable approximations becomes significantly more difficult as the number of components in a mixture increases. In addition, by independently fitting thermodynamic state variables, the laws of thermodynamics relating these state variables are often ignored, so extrapolation is limited. In this work, we propose instead to use surrogate equations of state (SEOS) fit locally to data from experiment or more complex thermodynamic models. This allows the properties derived from the surrogate equation of state to follow from basic laws of thermodynamics and have better extrapolation potential. We note that the key motivation of surrogate models is not for computational efficiency, but rather to enable EO optimization technology for properties that typically are calculated procedurally.

It is desired to base the surrogate equations of state on well-known simple functional forms for equations of state. The most logical choices for the simple functional form are the virial equation and cubic equations of state. This paper considers cubic equations with the Peng-Robinson form of the attractive term. This functional form will be tuned to match thermodynamic data (either generated from a complex equation of state or experimental data) in a local space, thus providing a surrogate equation of state. Of course, the simple functional form of the SEOS is not able to accurately describe all possible states. Instead, we fit over a local region of temperature, pressure, and concentration, and rely on the trust region approach of Eason and Biegler (2016) to control the error and guarantee accurate solutions.

2. Cubic Equations of State

The basic cubic functional form first proposed by van der Waals has been modified to produce many equation of state models. Valderrama (2003) reviews these variations, showing the great flexibility of cubic equations of state to predict the behavior of many systems. In this work, we will choose the Peng-Robinson equation of state functional form, although it is simple to adjust this to other functional forms.

The Peng-Robinson equation of state for a pure component takes the form:

$$P = \frac{RT}{v-b} - \frac{ma}{v^2 + 2bv - b^2} \quad (1) \tag{1}$$

where $m = \left[1 + (0.37464 + 1.54226\omega - 0.26992\omega^2)\left(1 - \sqrt{T/T_c}\right)\right]^2$ and ω is the Pitzer acentric factor. Eq.(1) can be converted to the standard cubic form in Z by multiplying through by the denominators and rearranging terms.

The parameters a and b in the cubic equation of state can be uniquely determined by imposing criticality conditions. When we require $\left(\frac{dP}{dv}\right)_{T_c} = \left(\frac{d^2P}{dv^2}\right)_{T_c} = 0$, it is determined that $a = 0.45724(R^2\,T_c^2)/P_c$ and $b = 0.07780(RT_c)/P_c$.

From fitting the critical conditions, we can see how the cubic equation of state follows the principle of corresponding states. This principle states that if the reduced conditions T/T_c and P/P_c and the acentric factor ω are the same, then the thermodynamic properties are (nearly) the same, regardless of the fluid. This means that a fluid is parameterized by

its critical temperature, critical pressure, and acentric factor. This principle can be derived from molecular thermodynamics with assumptions including universal pairwise potential functions (see Prausnitz et al. 1999). However, real fluids can vary in behavior depending on quantum effects, polarity etc. To capture these deviations, we propose tuning the critical properties to fit the observed behavior of a fluid locally. For mixtures, it is also necessary to tune binary interaction parameters $\kappa_{i,j}$.

3. Parameter estimation formulation

In this section, we propose a parameter estimation formulation for tuning a cubic equation of state for use in vapor-liquid equilibrium calculations. Because cubic equations of state are parameterized by critical properties, we seek adjusted values for the pure component critical temperatures T_i, pure component critical pressures P_i, and binary interaction parameters $\kappa_{i,j}$ to minimize the squared error in fugacity coefficients ϕ_i for one phase. This requires VLE data as input; in this work we use data generated from the PC-SAFT equation of state. The values associated with each data point k are system temperature $\hat{T}_k$, system pressure $\hat{P}_k$, composition $\hat{x}_{i,k}$, compressibility factor $\hat{Z}_k$, and fugacity coefficient $\hat{\phi}_{i,k}$. Note that the compositions, compressibility factor, and fugacity coefficient are specific to one phase, so the parameter estimation problem is applied separately to both liquid and vapor data. The subscript i represents the components and subscript k represents the data points. All symbols with a hat, for example $\hat{Z}_k$, are parameters associated with a data point. The fitting model is presented as follows:

$$\min \sum_{\mathrm{k}} \sum_{\mathrm{i}} \left(\widehat{\phi}_{i,k} - \phi_{\mathrm{i,k}}\right)^2 + \rho \left(\mathrm{f}\left(\hat{\mathrm{Z}}_k\right)\right)^2$$

$$s.t. \quad \mathrm{a}_{i,k} = 0.45724 \frac{\mathrm{m}_{\mathrm{i,k}}^2 \mathrm{R} \mathrm{T}_{\mathrm{i}}^2}{\mathrm{P_i}}, \quad b_i = 0.07780 \frac{RT_i}{P_i}$$

$$m_{i,k} = \left[1 + (0.37464 + 1.54226\omega_i - 0.26992\omega_i^2)\left(1 - \sqrt{\frac{\hat{T}_k}{T_i}}\right)\right]^2 \tag{2}$$

$$a_k^m = \sum_i \sum_j \hat{x}_{i,k} \hat{x}_{j,k} \sqrt{a_{i,k} a_{j,k}} (1 - \kappa_{i,j}), \quad b_k^m = \sum_i \hat{x}_{i,k}\, b_i$$

$$\alpha_k = \frac{a_k^m \hat{P}}{R^2 \hat{T}^2}, \quad \beta_k = \frac{b_k^m \hat{P}}{R\hat{T}}$$

$$f\left(\hat{Z}_k\right) = \hat{Z}_k^3 + (\beta_k - 1)\hat{Z}_k^2 + (\alpha_k - 2\beta_k - 3\beta_k^2)\hat{Z}_k - \alpha_k\beta_k + \beta_k^2 + \beta_k^3$$

$$\ln\left(\phi_{i,k}\right) = \frac{b_i}{b_k^m}\left(\hat{Z}_k - 1\right) - \ln(\hat{Z}_k - \beta_k) - \frac{\alpha_k}{2\sqrt{2}\beta_k}\left(\frac{2\sqrt{a_{i,k}}}{a_k^m}\left(\sum_j \hat{x}_{j,k}\sqrt{a_{j,k}}\left(1 - \kappa_{i,j}\right)\right) - \frac{b_i}{b_k^m}\right) \times \ln\left(\frac{\hat{Z}_k + (1+\sqrt{2})\beta_k}{\hat{Z}_k + (1-\sqrt{2})\beta_k}\right)$$

The objective function consists of two terms. The first term minimizes the squared relative deviation in fugacity coefficient, while the second term enforces the root of the cubic equation using a penalty parameter ρ. Note that the compressibility factors in the cubic equation are parameters given by data; it is the cubic equation itself that changes during optimization.

An alternative formulation would define a new variable Z_k for each datum k and add constraints to require each Z_k to be a root of the cubic equation at the corresponding temperature and pressure. However, this introduces the issue of root selection and the

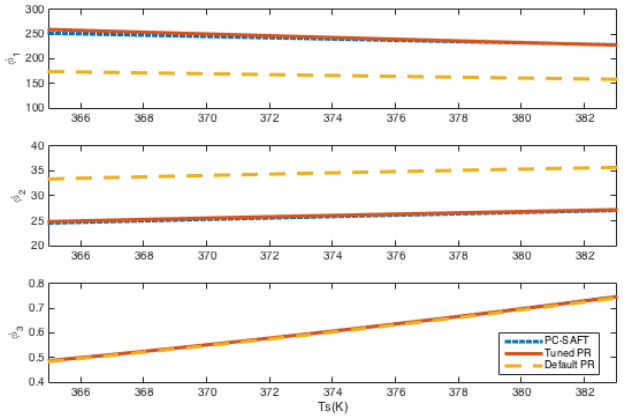


Figure 1: Fugacity coefficients with changing temperature

large number of nonlinear constraints can be difficult to converge. By contrast, the presented formulation can be solved as an unconstrained optimization problem through simple variable elimination and is found to be quite robust. Although the problem is nonconvex and the existence of several local minima has been observed, very good initialization can be derived by using the literature-reported critical temperatures, pressures, and binary interaction parameters. Formulation (2) has been implemented in Pyomo, and the parameter estimation problem is solved using IPOPT.

4. Numerical testing and flash optimization

In this section, the procedure for constructing SEOS is presented and performance is demonstrated on a flash model from a polyethylene production process. This flash model is used to separate unreacted monomer (ethylene) and chain transfer agent (hydrogen) from solvent (hexane) for recycle to the reactor. VLE data were gathered using a PC-SAFT flash model with a 3-level full factorial design on temperature, pressure, and component flowrates. The levels of each variable are shown in Table 1.

Table 1: Full factorial design

Input variable	Values	Explanation
$\hat{T}_k$	[362.15, 372.15, 382.15]	Flash temperature (K)
$\hat{P}_k$	[3.5, 4.0, 4.5]	Flash pressure (bar)
F_1	[0.093, 0.186, 0.373]	Input flowrate hydrogen
F_2	[3.01,6.02,12.05]	Input flowrate ethylene
F_3	[112.8, 141.0, 169.2]	Input flowrate hexane

Out of the 243 points in the full factorial design, two fell outside of the two-phase region and were discarded. The flash model outputs both liquid and vapor phase concentrations, compressibility factors, flowrates, and fugacity coefficients.

Formulation (2) was applied to fit SEOS for both liquid and vapor phases. Both problems were initialized with the literature values for critical properties and binary interaction coefficients. In the liquid phase, there was insufficient information in the data to obtain unique values for hydrogen's critical properties, so these were fixed at their literature values. The resulting parameters are shown in Table 2. Figure 1 compares the fugacity coefficients varying with temperature for PC-SAFT, the SEOS, and untuned Peng-

Robinson when the pressure is fixed at 3.85 bar. The SEOS matches very well over the temperature range, but the untuned PR deviates significantly.

Table 2: Results of fitting SEOS to cubic equation of state

Parameter	Literature	Liquid fit	Vapor fit
T_{H2} (K)	33.0	33.0	76.54
T_{C2H4} (K)	283	347.5	342.3
$T_{n\text{-hexane}}$ (K)	507.5	506.9	649.7
P_{H2} (bar)	13.2	13.2	10.1
P_{C2H4} (bar)	51.2	62.4	84.8
$P_{n\text{-hexane}}$ (bar)	30.1	30.0	61.9
$\kappa_{H2,C2H4}$	0.022	2.8×10^{-6}	0.509
$\kappa_{H2,n\text{-hexane}}$	0.02917	0.513	0.972
$\kappa_{C2H4,n\text{-hexane}}$	0.1144	0.179	0.158

The resulting surrogate equations of state are used in an optimization formulation of a flash drum. The objective is to minimize the amount of monomer in the liquid phase while maintaining a lower bound on mass flow of this stream such that the polymer product (not included in equilibrium calculations) remains a slurry. The model is given as:

$$\begin{aligned} \min_{T,P} \quad & x_{C2H4} L \\ \text{s.t.} \quad & 335 \le \mathrm{T} \le 370, \quad 2.5 \le \mathrm{P} \le 5 \\ & \textstyle\sum_i MW_i x_i \ge c, \quad \sum_i (y_i - x_i) = 0 \\ & F = L + V, \quad F z_i = L x_i + V y_i \\ & y_i = \frac{\phi_{L,i}}{\phi_{V,i}} x_i \end{aligned} \tag{3}$$

where MW_i is the molecular weight of component i, c is a constant specified to keep the product polymer suspended in slurry, and $\phi_{L,i}$ and $\phi_{V,i}$ are functions of T, P, x_i and y_i. Optimization problem (3) was run in three different modes: using PC-SAFT to calculate the fugacity coefficients, using un-tuned Peng-Robinson, and using the tuned SEOS approach. We note however that the PC-SAFT equation of state is difficult to use in equation-oriented optimization due to the presence of multiple volume roots, and the issue of root selection is beyond the scope of this work. Through careful initialization, a result was obtained that was validated against Aspen Plus to have found the correct roots of PC-SAFT.

Table 3: Flash optimization results. Errors are measured relative to PC-SAFT solution.

	PC-SAFT	Peng-Robinson	Error	SEOS	Error	Re-tuned SEOS	Error
x_{H2}	4.62×10^{-5}	6.3×10^{-5}	36%	4.34×10^{-5}	6%	4.62×10^{-5}	0.03%
x_{C2H4}	0.0127	0.0974	23%	0.0125	1.86%	0.0127	0.04%
x_{hexane}	0.9872	0.9902	0.3%	0.9875	0.02%	0.9873	<0.01%
T	351.67	349.68	0.6%	351.38	0.08%	351.64	<0.01%
P	2.5	2.5	0%	2.5	0%	2.5	0%

Table 3 shows that the results of using Peng-Robinson and PC-SAFT can vary by as much as 36 %. Using the SEOS parameters shown in Table 2, the maximum error relative to

PC-SAFT is reduced to 6%. Then, more data were gathered from PC-SAFT, in a full factorial design centered at the solution of the first SEOS optimization problem, with the width of data range cut in half. Then the SEOS is re-tuned using the data in the region of the solution, and the maximum error is reduced to 0.04%. This indicates the potential of using automatic update strategies in a trust region framework to control the optimization error introduced by the SEOS approximation.

5. Conclusions and future work

In this paper, we have demonstrated the potential of using a cubic equation of state to locally represent thermodynamic behavior directly from data. Tuning parameters for the cubic equation are selected based on their interpretation in terms of the principle of corresponding states. Then, an effective and reliable parameter estimation formulation is proposed. The formulation can be naturally converted to an unconstrained optimization problem for ease of initialization and solution. Finally, SEOS models are demonstrated for optimization of a flash unit in a polymerization system, and the SEOS gives accurate results with the potential for refinement by gathering more data.

The SEOS concept is not restricted to cubic functional forms and it may be worthwhile to consider virial-type equations as well. In addition, it is important to consider how to automatically and rigorously manage the error introduced by the SEOS. The use of trust region concepts, in which thermodynamic variables are constrained to lie in the region where data were obtained, is promising. However, the SEOS must be able to sufficiently approximate gradient information if it is to be valid for use in optimization. Future work will focus on verifying these properties and demonstrating the ability to adaptively adjust the surrogate equations of state in large-scale process optimization.

Acknowledgements

The authors gratefully acknowledge the financial support of National Key Research and Development Program of China (No. 2017YFE0106700), the Open Research Project of the State Key Laboratory of Industrial Control Technology, Zhejiang University, China (No. ICT170305), and the United States National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1252522.

References

A.W. Dowling, L.T. Biegler, 2015, A framework for efficient large scale equation-oriented optimization, Computers & Chemical Engineering, 72, 3-20.

J.P. Eason, L.T. Biegler, 2016, A trust region filter method for glass box/black box optimization, AIChE Journal, 62:9, 3124-3136

R.S. Kamath, L.T. Biegler., and I.E. Grossmann, 2010, An equation-oriented approach for handling thermodynamics based on cubic equation of state in process optimization. Computers & Chemical Engineering, 34, 12, 2085-2096.

J.M. Prausnitz, R.N. Lichtenthaler, E.G. Azevedo, 1999, Molecular Thermodynamics of Fluid-Phase Equilibria. Prentice Hall, Upper Saddle River, NJ, USA, 104-114.

R. Privat, R. Gani, and J.-N. Jaubert, 2010, Are safe results obtained when the PC-SAFT equation of state is applied to ordinary pure chemicals?, Fluid Phase Equilibria 295:76-92.

J.V. Sengers, R.F. Kayser, C.J. Peters, H.J. White Jr., Eds., 2000, Equations of State for Fluids and Fluid Mixtures, International Union of Pure and Applied Chemistry (IUPAC), Research Triangle Park, NC, USA.

J.O. Valderrama, 2003, The state of the cubic equation of state, Ind. Eng. Chem. Res., 42, 1603-1618

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50126-9

A General Framework for Sensitivity-Based Optimal Control and State Estimation

David Thierry[a], Bethany Nicholson[b], and Lorenz Biegler[a]*

[a]*Department of Chemical Engineering, Carnegie Mellon University*
[b]*Center for Computing Research, Sandia National Laboratories*
lb01@andrew.cmu.edu

Abstract

New modelling and optimization platforms have enabled the creation of frameworks for solution strategies that are based on solving sequences of dynamic optimization problems. This study demonstrates the application of the Python-based *Pyomo* platform as a basis for formulating and solving Nonlinear Model Predictive Control (NMPC) and Moving Horizon Estimation (MHE) problems, which enables fast on-line computations through large-scale nonlinear optimization and Nonlinear Programming (NLP) sensitivity. We describe these underlying approaches and sensitivity computations, and showcase the implementation of the framework with large DAE case studies including tray-by-tray distillation models and Bubbling Fluidized Bed Reactors (BFB).

Keywords: NMPC, state-estimation, optimal control, sensitivity, pyomo.

1. Introduction

New modelling and optimization tools have created an opportunity for the creation of frameworks for problems in design, control, etc. These recent algebraic modelling languages (AMLs) like Pyomo and JuMP are based on high-level programming languages and are much more flexible than similar tools written in proprietary languages. Combining these platforms with state-of-art solvers makes it easier to deal with more complex problems. Of particular interest are optimal control problems (OCPs) that incorporate first-principles process models as part of a mathematical optimization problem. These models are often systems of differential algebraic equations (DAEs), and can be handled directly by means of direct multiple-shooting or orthogonal collocation, combined with the optimization algorithm.

Several frameworks have been developed for model-based optimal control combining AMLs and NLP solvers, these include ACADO, CasADi and Pyomo. In Pyomo (Hart et al., 2011) the dynamic optimization environment provides a combination of accessible modelling interfaces to solution algorithms. Rather than building a whole new base programming language, Pyomo provides the necessary optimization objects in Python, which escalates the flexibility of applications for optimization-based algorithms for dynamic systems. Previously, (Santamaria and Gomez, 2016) developed NMPC frameworks based on Pyomo and its extension for dynamic optimization Pyomo.DAE. In this work, we develop a new Pyomo-based framework for NMPC and combine it with a framework MHE. Both are implemented to perform close to real time nonlinear control and state estimation by means of NLP sensitivity. We also introduce a simpler sensitivity update based on a reduced set of variables, which decreases the amount of information transfer required for on-line computation.

2. Algorithmic background

Consider a first-principles model with the semi-explicit DAE system.

$$x'(t) = f(x(t), y(t), u(t)), g(x(t), y(t), u(t)) = 0, x(t_0) = x_0 \tag{1}$$

where $x \in \mathbb{R}^{n_x}$ and $y \in \mathbb{R}^{n_y}$ are the differential and algebraic state vectors, $u \in \mathbb{R}^{n_u}$ is the control vector and $t \in \mathbb{R}^1$ is the independent variable (time). The general dynamic optimization problem includes the DAEs (1) among its constraints. One way to deal with the optimization problem is to discretize the variables through orthogonal collocation on finite elements. The resulting problem is then fully algebraic and has a large dimensionality. However, it is also sparse and highly structured, and can be solved with large-scale NLP packages like IPOPT.

Once the modelling and optimization of the dynamic problem has been established, we formulate the respective control and state-estimation problems. The NMPC approach utilizes the model of the plant to make predictions of the optimal input vectors $u_{i|k}$ for a prediction horizon of N steps, given the state vector $x(k)$ (at time step k) from the plant. This leads to the parametric optimization problem $P\big(x(k)\big)$. Without loss of generality, we assume a discretized model of the plant $x_{i+1} = f(x_i, u_i)$ (with algebraic variables and equations eliminated), so that $P(x(k))$ can be written as:

$$\begin{aligned} P\big(x(k)\big): \quad & \min_{u_k} \varphi_N\big(x_{N|k}\big) + \frac{1}{2}\sum_{i=0}^{N-1} u_{i|k}^T R_i^{-1} u_{i|k} + \frac{1}{2}\sum_{i=0}^{N} x_{i|k}^T Q_i^{-1} x_{i|k} \\ \text{s.t.} \quad & x_{l+1|k} = f\big(x_{l|k}, u_{l|k}\big), \quad x_{0|k} = x(k), \\ & x_{l|k} \in \mathbb{X}, \quad l \in \{0,1,\dots,N\}; u_{l|k} \in \mathbb{U}, \quad l \in \{0,1,\dots,N-1\} \end{aligned} \tag{2}$$

where Q and R are positive definite matrices, $x_{i|k}$ denotes the predicted state at stage i given state $x(k)$ (from the plant). Asymptotic stability of problem (2) can be established when $N \to \infty$, although this can also be established through a terminal cost $\varphi_N\big(x_{N|k}\big)$ that approximates the cost of a problem with infinite horizon. However, the NMPC computation requires the knowledge of a full state $x(k)$, which in general is not accessible. Instead, only a limited amount of output data $y(k)$ are usually available.
Reconstruction of $x(k)$ from past data $y(k-i)$ requires the solution of a state estimation problem. In this work, we have considered the Moving Horizon Estimation (MHE) to compute the states. This problem is analogous to the NMPC approach in the sense that the model for the plant is used in conjunction to a set of past data (instead of the future as posed by the NMPC). At step k, the parametric MHE problem $\mathcal{M}\big(\hat{x}_{T-N|k-1}, \Pi_{T-N|k-1}^{-1}, y(k-N), \dots, y(k)\big)$ has a horizon window of length N and depends on the prior state-estimate $\hat{x}_{T-N|k-1}$, inverse prior-covariance $\Pi_{T-N|k-1}^{-1}$, and measurements $y(k-N), \dots, y(k)$. Let $m(k)$ be short notation for all the problem parameters previously described, then the MHE problem can be written as follows:

$$\begin{aligned} \mathcal{M}(m(k)): \quad & \min_{x_{T-N}, w_k} \frac{1}{2}\big(x_{T-N|k} - \hat{x}_{T-N|k-1}\big)^T \Pi_{T-N|k-1}^{-1}\big(x_{T-N|k} - \hat{x}_{T-N|k-1}\big) \\ & + \frac{1}{2}\sum_{i=T-N}^{T} v_{i|k}^T \mathcal{R}_i^{-1} v_{i|k} + \frac{1}{2}\sum_{i=T-N}^{T-1} w_{i|k}^T \mathcal{Q}_i^{-1} w_{i|k} \end{aligned} \tag{3}$$

$$\text{s.t.} \quad x_{l+1|k} = f\left(x_{l|k}\right) + w_{l|k}, l \in \{T-N, \dots, T-1\};$$
$$y(k-l) = h\left(x_{l|k}\right) + v_{l|k},$$
$$x_{l|k} \in \mathbb{X}, \quad l \in \{T-N, \dots, T\}; w_{l|k} \in \mathbb{W}, \ l \in \{T-N, \dots, T-1\}$$

with the process and measurement noise $w_{l|k}$ and $v_{l|k}$ respectively; and positive definite covariance matrices $\mathcal{R}_i$ and $\mathcal{Q}_i$. Note that the solution of $\mathcal{M}(m(k))$ yields the sequence of process noise w_k and most likely state estimates x_k, subsequently, one can set $x(k) = x_{T|k}$ and solve the corresponding control problem. For the next sampling step $k+1$, new data will be available and for a constant horizon window, the problem would require dropping the oldest measurement (i.e. $y(k-N)$) and shift the remaining vectors. Finally, the first term in the objective function is an approximation of the arrival cost, which is analogous to the terminal cost of NMPC and accounts for all the information that is missing from the full information problem.

If the solution of each problem can be computed instantaneously, the controller will achieve stability and robustness properties (Zavala and Biegler, 2009). We denote these as the ideal-NMPC and MHE problems. However, the solution of these problems has a computational overhead, as we are dealing with large NLPs. This situation can be overcome by implementing the Advanced-Step (as-) strategy (Figure 1), in which the NLPs are solved before the sampling time for a predicted state (or measurements). Then the control input (or state-estimate) can be updated and corrected with sensitivity with respect to the state (or measurement) difference. This implies that the NLPs associated with the control and/or state-estimation can be solved in advance, thus avoiding the on-line computational delay. The only computational overhead on-line will correspond to that of the sensitivity update, which is reduced to a matrix-matrix multiplication.

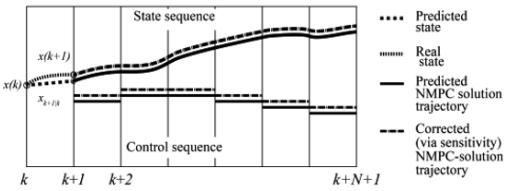


Figure 1 The Advanced-step NMPC strategy

For this framework, rather than performing the full primal-dual sensitivity update, the update is limited to a reduced set of variables of interest (i.e. for NMPC the inputs u_k and for MHE the current state-estimate $\tilde{x}_{T|k}$). This effectively reduces the amount of information that must be taken on-line, to just the sensitivity matrix $\hat{S}$. The sensitivity matrix computation is performed using the implicit function theorem on the optimality conditions of the NLP: $\min_z f(z, p_0), s.t. c(z, p_0) = 0, z \geq 0$, which will yield a linear system of equations (Pirnay et al., 2012) that depends on the parameter p. For this work, the associated linear system has the form:

$$\begin{bmatrix} \nabla_{zz}\mathcal{L}(z^*, p_0) & + Z^{*-1}V^* & \nabla_z c(z^*, p_0) \\ \nabla_z c(z^*, p_0)^T & & 0 \end{bmatrix} \hat{S}(p_0) = E \tag{4}$$

where p_0 is the nominal parameter value (i.e. predicted state for NMPC, predicted measurement vector for MHE), $\nabla_{zz}\mathcal{L}(z^*, p_0)^*$ denotes the Hessian of the Lagrange function of the corresponding NLP at its solution, $\nabla_z c(z^*, p_0)^*$ the gradients of the

constraints, E is a matrix that contains identity matrices at the positions of the variables of interest ($u_{0|k}$ for NMPC and $x_{T|k}$ for MHE) and $\hat{S}(p_0)$ is the corresponding sensitivity matrix. The KKT matrix in (4) is sparse and symmetric; therefore, the factorization can be done with sparse linear solvers like Pardiso or MA57. The subsequent sensitivity update on the parameter change p for the predicted solution $u(p_0)^*$ or $x_k(p_0)^*$ is:

$$\text{NMPC: } p = x(k) \; and \; \tilde{u}(p) = u(p_0)^* - \hat{S}(p_0)^T \begin{bmatrix} \nabla_{zp}\mathcal{L}^{*T} \\ \nabla_p c^{*T} \end{bmatrix} (p - p_0) \tag{5a}$$

$$\text{MHE: } \quad p = y(k) \; and \; \tilde{x}_{T|k}(p) = x_{T|k}(p_0)^* - \hat{S}(p_0)^T \begin{bmatrix} \nabla_{zp}\mathcal{L}^{*T} \\ \nabla_p c^{*T} \end{bmatrix} (p - p_0) \tag{5b}$$

where the term inside the brackets are the corresponding gradients with respect to the parameter. It should be noted that, equations (5a, b) are the only computations that are performed on-line, and will yield the corresponding updated control input vector $\tilde{u}_{0|k}$ (NMPC) or current state-estimate $\tilde{x}_{T|k}$ (MHE).

3. Framework

The Python based framework that formulates and solves these problems is displayed in Figure 2. The underlying idea is that we would like to provide a way to automatically set up the corresponding problems given a dynamic model, and methods that deal with algorithmic issues like initialization, sensitivity computations, etc. This framework is straightforward to extend, though most applications might require tailored algorithmic considerations.

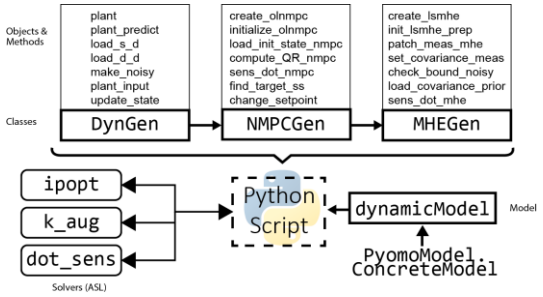


Figure 2 The Python framework for NMPC and MHE

In its current form, the framework requires a dynamic model manually discretized in time using collocation over finite elements, with expressions in place for continuity and initial conditions. Computation of the sensitivity matrices requires the solution of a symmetric sparse linear system of equation (4), that contains first and second order derivatives of the constraints and objective function. These operations are performed by combining Automatic Differentiation, sparse linear solver libraries like Pardiso and a new subprogram (`k_aug`) that performs all these operations separate from the Python framework. This computation is similar to an iteration of Ipopt with a different right-hand-side on the associated linear system. Furthermore, the update of the variables of interest is done in a second program (`dot_sens`) that updates the current solution with information from the sensitivity matrix. Both programs are compatible with Pyomo and use the Ampl Solver Library (ASL) for derivative information. These sensitivity-analysis codes were written in C instead of Python for greater computational speed.

4. Case studies

To demonstrate the application of the framework we consider the control of a Bubbling Fluidized Bed Reactor (BFB) for CO2 capture (Lee and Miller, 2012) as well as the MHE-NMPC coupling for a distillation column (Lopez Negrete et al., 2013). Both case studies were solved using Ipopt 3.12 with MA57, and an Intel(R) Core(TM) i7-6700 CPU @ 3.40 GHz processor.

BFB is a gas-solid reactor that removes CO_2 from a flue gas stream using a solid adsorbent. It can be modelled by using an extended Kunii-Levenspiel fluidized bed model. Due to the space-time discretization, the resulting problem yields 315 states and an NLP problem with 46480 variables and 46470 equations. Background computation for the control input requires solution of $P(x(k))$, which requires about 120 CPU seconds. However, the on-line cost for asNMPC requires *less than a CPU second* for the controller update (5a). Figure 3 shows the results for the NMPC of the BFB with random noise. For a change of CO2 capture set-point, the reactor is controlled by the gas output valve position. Figure 3 shows that asNMPC is not significantly different than ideal-NMPC, and is able to reach the set-point within the same time.

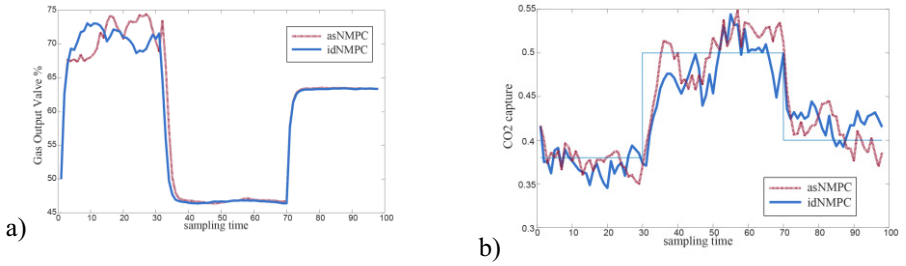


Figure 3 Results for the NMPC of the BFB; a) controller input profile, b) output profile

The full state vector for the distillation column case study is unknown, and it is estimated using MHE. This problem $\mathcal{M}(m(k))$ has 21346 variables and 20486 equations for 84 states. With this estimate, the control input is determined by NMPC. We assume that temperature and liquid molar holdup of each tray are measured, and temperature at tray 14 is the set-point. In Figure 4 both MHE results with the infinity norm of state error ($|error|_\infty$) and NMPC set-point tracking profile of the temperature are shown. Here, noise was introduced to demonstrate the performance of the as-MHE, and both advanced-step and ideal MHE and NMPC strategies reach the set-point with similar solution profiles. Background solution of problem $\mathcal{M}(m(k))$ requires 22 CPU s and the MHE sensitivity update (5b) takes 1 CPU s approximately. Both case studies show that the sensitivity update gives accurate values of control and state-estimates as noted by the values of the Root-Mean-Square Deviation of Table 1, where the deviations are relatively small for the tracked state (Figures 3-4(b)). Hence, the current Pyomo framework can handle nonlinear control effectively, with as-MHE and as-NMPC working together.

Table 1 RMSD for ideal and advanced-step Strategies

Figure	(a)	(b)
3	2.3898 (%)	0.0262 (CO2 Capture)
4	10.4334 (%)	0.0966 (K)

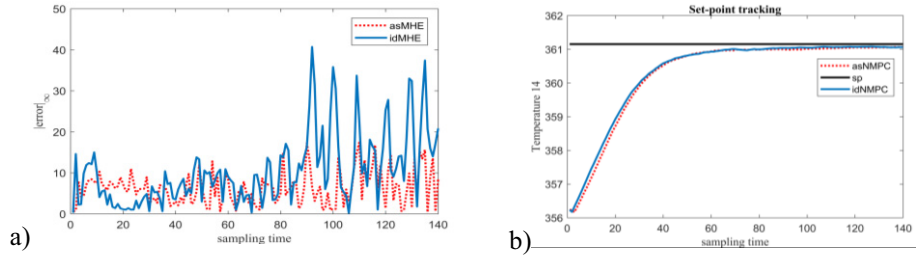


Figure 4 NMPC-MHE Results for Distillation Column; a) absolute error for MHE, b) Set-point tracking profile in NMPC

5. Conclusions

A new Python-based framework for Non-linear Model Predictive Control and Moving Horizon Estimation has been created in Pyomo. Given a discretized first principles dynamic model of a plant, the framework provides most of the common operations to perform optimal control, and it takes advantage of Pyomo to automatically generate the optimization problems. Fast control and state updates are embedded into the framework by means of NLP sensitivity with the advanced-step strategy. We also present an alternative way to compute sensitivities where the on-line update of the control is reduced to the set of input vectors (control) or state-estimates (state-estimation). The presented framework leverages state-of-art NLP and sparse linear algebra solvers to accelerate the expensive NLP computations. Moreover, capabilities currently under development include the compatibility with new Pyomo.DAE, new algorithmic features like handling active-set changes of the sensitivity updates, and structure exploitation of the underlying DAE models.

References

W. E. Hart, J.P. Watson, and D.L. Woodruff, (2011). Pyomo: modeling and solving mathematical programs in Python. Mathematical Programming Computation, 3(3), 219-260.

A. Lee and D. C. Miller, (2012). A one-dimensional (1-d) three-region model for a bubbling fluidized-bed adsorber. *Industrial & Engineering Chemistry Research*, 52(1), 469-484.

R. Lopez-Negrete, F. J. D'Amato, L. T. Biegler, and A. Kumar, (2013). Fast nonlinear model predictive control: Formulation and industrial process applications. *Computers & Chemical Engineering,* 51, 55-64.

H. Pirnay, R. López-Negrete, and L. T. Biegler, (2012). Optimal sensitivity based on IPOPT. *Mathematical Programming Computation*, 1-25.

F. L. Santamaría and J. M. Gómez, (2016). Framework in PYOMO for the assessment and implementation of (as) NMPC controllers. *Computers & Chemical Engineering*, 92, 93-111.

V. M. Zavala and L. T. Biegler, (2009). The advanced-step NMPC controller: Optimality, stability and robustness. *Automatica*, 45(1), 86-93.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50127-0

Rigorous synthesis of energy systems by relaxation and time-series aggregation to typical periods

Nils Baumgärtner[a], Matthias Leisin[a], Björn Bahl[a], Maike Hennen[a] and André Bardow[ab*]

[a]*Institute for Technical Thermodynamics, RWTH Aachen University, Schinkelstraße 8, 52062 Aachen, Germany*
[b] *Institute of Energy and Climate Research - Energy Systems Engineering (IEK-10), Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Straße, 52425 Jülich, Germany*
Andre.Bardow@ltt.rwth-aachen.de

Abstract

The synthesis of energy systems is a complex optimization task depending on multiple large time series. Time-coupling constraints, e.g., due to storage systems, complicate computation even further. To still efficiently solve time-coupled synthesis problems, we propose a rigorous synthesis method. In the proposed method, lower and upper bounds are calculated to obtain a feasible solution of the original synthesis problem with known quality. We compute the lower bound by binary relaxation. For the upper bound, we use time-series aggregation to obtain a feasible design for the system. Employing this feasible design, we solve an operational problem, which can be solved efficiently. To tighten the upper bound, we iteratively increase the time resolution of the aggregation. In a case study for an industrial energy system, we show that only few typical periods are required to obtain a solution of the original synthesis problem with excellent quality. The method has fast convergence outperforming a commercial state-of-the-art solver.

Keywords: energy systems, optimal design, typical periods, rigorous optimization, relaxation

1. Introduction

Synthesis of process or energy systems is a complex two-stage optimization task that requires the simultaneous optimization of the first design stage and the second operation stage (Yokoyama et al., 2015). On the design stage, decisions on the type and sizing of components are made. On the operational stage, the on/off status and the load allocation are defined for each energy conversion or storage unit and each time step. Operation usually depends on multiple large time series, e.g., demand profiles, electricity prices and renewable resources. These large time series lead to large-scale optimization problems, which are computationally challenging and often not solvable within reasonable computational time or memory limits. Therefore, the problem size is often reduced by model aggregation (Grossmann, 2012). Time-series aggregation is often applied in synthesis problems for the design of chemical processes (Maravelias and Sung, 2009) and energy systems (Mancarella, 2014). However, the solution of the aggregated synthesis problem is not the solution of the original synthesis problem based on the full time series. Thus, the quality of the aggregated solution is unknown and the resulting design might even be infeasible for the full time series. To obtain a feasible solution with known quality, exact solution strategies are needed. In our previous work, we proposed an exact decomposition method to ensure optimality of the resulting design (Bahl et al., 2018b).

However, the proposed method does not consider the chronology of time steps. The Chronology of time steps is necessary to model storage systems or ramping constraints. Here, we propose a rigorous synthesis method considering the chronology of time steps. We apply our method to the synthesis of an industrial energy supply system.

2. Generic synthesis problem with time-coupling constraints

In Eq. (1), we state the generic synthesis problem of energy systems as a mixed-integer linear program (MILP)

$$\min_{\dot{V}_{n,t},\dot{V}_n^N,\delta_{n,t},\gamma_n,x} TAC = OPEX + CAPEX = \sum_{t\in T}\left(\Delta t_t \sum_{n\in C\setminus C_{stor}} c_{n,t}^o \frac{\dot{V}_{n,t}}{\eta_n}\right) + \sum_{n\in C} c_n^i \dot{V}_n^N$$

$$\begin{aligned}
s.t. \quad & \sum_{n\in C\setminus C_{stor}} \dot{V}_{n,t} + \sum_{n\in C_{stor}} \left(\dot{V}_{n,t}^{out} - \dot{V}_{n,t}^{in}\right) = \dot{E}_t && \forall t \in T, \\
& V_{n,t} + \Delta t_t\left(\dot{V}_{n,t}^{in} - \dot{V}_{n,t}^{out}\right) = V_{n,t+1} && \forall t \in T, \forall n \in C_{stor}, \\
& A_1\dot{V}_{n,t} + \tilde{A}_1\delta_{n,t} \le b_1 && \forall t \in T, \forall n \in C, \\
& A_2\dot{V}_n^N + \tilde{A}_2\gamma_n \le b_2 && \forall n \in C, \\
& A_3\left(\dot{V}_{n,t}, \dot{V}_n^N, \delta_{n,t}, \gamma_n, x\right)^T \le b_3 && \forall t \in T, \forall n \in C, \\
& \gamma_n, \delta_{n,t} \in \{0,1\}; x \in \mathbb{R}^a \times \{0,1\}^{\tilde{a}}; \dot{V}_{n,t} \in \mathbb{R}; \dot{V}_n^N \in \mathbb{R}^+.
\end{aligned} \tag{1}$$

As the objective function, we employ the total annualized costs TAC. The objective function consists of two parts representing the capital and operational expenditures, CAPEX and OPEX. The CAPEX are the nominal capacity of each component $\dot{V}_n^N$ multiplied by the specific investment costs c_n^i summed up for all components C. The OPEX are defined as the sum of the output power $\dot{V}_{n,t}$ of every conversion component $n \in C\setminus C_{stor}$ in every time step t divided by the efficiency η_n and multiplied by the specific operation cost $c_{n,t}^o$ and the duration Δt_t of a time step. In contrast to CAPEX, OPEX directly depend on the set of time steps T. The TAC are minimized subject to several constraints. The sum of the components output power $\dot{V}_{n,t}$ and the net energy output of the storage units $\dot{V}_{n,t}^{out} - \dot{V}_{n,t}^{in}$ have to meet the energy demand $\dot{E}_t$ at every time step t. The future storage level $V_{n,t+1}$ is coupled to the current storage level $V_{n,t}$ by the net energy output of the storage units $\dot{V}_{n,t}^{out} - \dot{V}_{n,t}^{in}$. Further (in)equalities with the coefficient matrices $A_1; \tilde{A}_1; A_2; \tilde{A}_2$ and the vectors $b_1; b_2$ determine the binary on/off status $\delta_{n,t}$, the binary existence γ_n of components and the nominal power $\dot{V}_n^N$. Only equations including the output power $\dot{V}_{n,t}$, the storage variables or the on/off status $\delta_{n,t}$ have to be stated for each time step t and thus depend on the size of the time series T. All other variables are represented by the vector x. Additional constraints are summarized in the surrogate equation A_3. The equation A_3 includes, e.g., sizing constraints $\dot{V}_{n,t} \le \gamma_n\dot{V}_n^N$. The surrogate vector x includes, e.g., additionally variables for the linearization of the bilinear terms $\gamma_n\dot{V}_n^N$. The original synthesis problem (Eq.(1)) is often not solvable in reasonable time for large time series T. In particular, the time coupling constraints due to storage systems enhance the complexity of the synthesis problem (Rong et al., 2008).

3. Rigorous synthesis method using relaxation and time-series aggregation

In this section, we present our rigorous synthesis method for the original large-scale synthesis problem. In Section 3.1, we give a short overview of the method. In Section 3.2, the calculation of the lower bound employing relaxation is explained and in Section 3.3, the calculation of the upper bound via time-series aggregation is presented. In Section 3.4, the evaluation of the optimality gap and the potential increase in time-series resolution are described.

3.1. Method overview

The proposed method consists of six steps (Fig.1):

(i) Relaxation of all binary variables of the original problem to obtain a lower bound of the synthesis problem
(ii) Identification of the length of a typical period
(iii) Time-series aggregation to typical periods and clustering of time steps within each typical period, from our previous work (Bahl et al., 2018a)
(iv) Synthesis optimization with the aggregated time series to obtain a design candidate of the energy supply system
(v) Operational optimization of the candidate design of the energy system with the full time series to provide a feasible solution that serves as an upper bound
(vi) Reiteration of step (iii)-(v) with increasing time resolution until the resulting lower and upper bound satisfy the desired optimality gap ε.

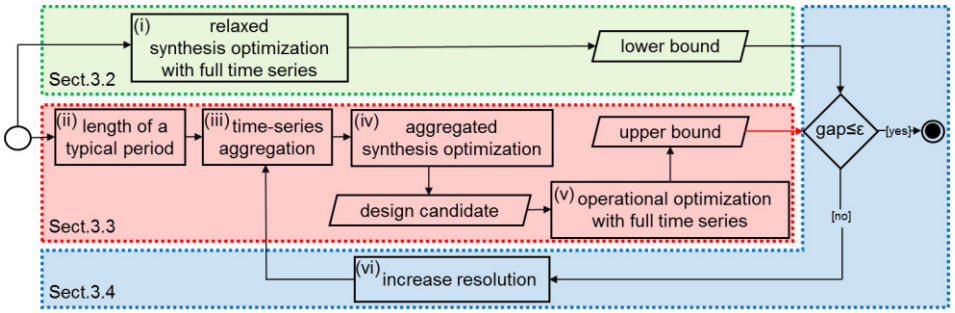


Figure 1: Rigorous synthesis method for energy systems employing time-series aggregation and relaxation to satisfy an optimality gap ε.

3.2. Step (i): Calculation of the lower bound

Step(i): First, we relax all binary variables (δ_{nt}, γ_n, $x \in \{0, 1\}$) of the original problem, thereby converting the complex MILP to an LP, which can be solved efficiently. The solution of the relaxed problem serves as a lower bound of the original synthesis problem. In this work, the lower bound is calculated only once. However, the method could also be extended by an iterative refinement of the lower bound as done in Bahl et al. (2018b) for energy systems without chronology in the time series.

3.3. Step(ii)-(v): Calculation of the upper bound

To calculate upper bounds, we employ time-series aggregation from our previous work (Bahl et al., 2018a).

Step (ii): To identify a suitable period length, we look for periodic patterns by applying autocorrelation to all time series (Box et al., 2015). Autocorrelation correlates each time

series n with itself at different time offsets resulting in a set of n autocorrelated functions ACF_n. For all time series we employ one constant period length in the time-series aggregation. Therefore, we calculate the normalized sum $\overline{ACF}$ of the resulting set:

$$\overline{ACF} = \frac{\sum_n ACF_n}{n}. \quad (2)$$

Splitting the time series into periods is meaningful if periodicity exists. An oscillation pattern in the autocorrelation indicates periodicity in the original time series. Peaks of the $\overline{ACF}$ correspond to suitable period lengths. A short period length can lead to smaller optimization problems using time-series aggregation. However, no chronology between periods is considered, and thus storage in between periods cannot be represented. Therefore, a short period length also implies to a short maximum storage load cycle.

Step (iii): Next, we start to aggregate the original time series to typical periods by first splitting the original time series into periods with the in step (ii) identified length. Typical periods require aggregation in two dimensions: the number of periods and the number of segments per period. We heuristically choose to first aggregate the number of periods and subsequently the number of segments. The aggregation of periods to typical periods is based on the k-medoids algorithm, which does not average as the k-means method does, but selects a period from the time series as a typical period (Kaufman and Rousseeuw, 2008). The k-medoids algorithm requires a mean-value correction, which we apply for each time series separately to obtain the same mean values as for the original time series as in Domínguez-Munõz et al. (2011). To aggregate segments, we iteratively calculate the average of randomly chosen consecutive time steps. The average value is used for the entire aggregated segment with the duration of the sum of the assigned time steps. The aggregated segments of the randomly chosen consecutive time steps with the lowest Euclidean distance to the original segments are used. The employed segment-aggregation method has the advantage that any time step can be used as a starting point for a typical period.

Step (iv): The time-series aggregation reduces the complexity of the synthesis problem by magnitudes. Thus, the aggregated synthesis problem can be solved efficiently. Nevertheless, the solution of the aggregated synthesis optimization cannot serve as an upper bound, as not the full time series is used. However, the optimization of the aggregated problems results in a design candidate of the energy supply system.

Step (v): The design candidate is subsequently used to solve the operational optimization problem with the full time series. For the operational optimization, the design decisions are now given. Thus, the problem can be solved much more efficiently. The feasibility of the design candidate resulting from step (iv) cannot be guaranteed for the full time series. To ensure a feasible design, we add peak demands as single time steps with duration of zero and no chronological order to the typical periods. If the resulting designs is still infeasible, we increase the time resolution in step (iii). The best known solution of the operational optimization with the full time series problem serves as an upper bound for the original synthesis problem.

3.4. Step (vi): Optimality gap ε and increase of time resolution

Step (vi): Last, we compare the resulting lower (step (i)) and upper bound (step (v)) and check if the desired optimality gap is satisfied.

$$gap\ \varepsilon = \frac{upper\ bound - lower\ bound}{upper\ bound}. \quad (3)$$

We iteratively increase the time resolution of either the number of periods or the number of segments until the gap is satisfied, as in Bahl et al. (2018a). The relaxation for the lower bound (step (i)) is not refined in this work. Hence, the method does not converge to a gap ε of 0 %.

4. Case-study

In this section, we apply the proposed method to an MILP industrial synthesis problem based on Voll et al. (2013). We use a superstructure with three units of each energy conversion technology with constant part-load efficiency (boiler, CHP engine, compression chiller, and absorption chiller) and one storage tank for hot and one for cold water. The original time series consists of one year with hourly demand data for process heat, hot and cold water, electricity, and electricity grid prices based on Bahl et. al. (2018a). The original synthesis problem with full time series contains $1.2 \cdot 10^6$ equations and $6 \cdot 10^5$ variables ($2 \cdot 10^5$ binaries) with $3 \cdot 10^6$ nonzero elements. We use a period length of 24 hours derived from step (ii). The period length of 24 hours leads to relatively minor optimization problems, but still enables daily storage cycles. All calculations are performed using 4 Intel- Xeon CPUs with 3.0 GHz and 64GB RAM. All MILP problems are solved using CPLEX 12.6.3.0 with an optimality gap of 2 %. The proposed rigorous synthesis method satisfies the gap ε of 2% in less than 1 hour (2364 s) (Fig. 2). The optimality gap ε of 2 % can be satisfied with only 2 periods and 3 segments resulting in a synthesis problem with only 4231 equations, 3054 variables (418 binaries) and 10,935 non-zero elements, reducing the complexity by about 3 orders of magnitude. Considering more periods and segments increases the accuracy only slightly (Fig. 2). For our calculation with the highest resolution of 8 periods consisting of 8 segments each, an optimality gap ε of less than 1 % is reached. The design candidate obtained from the problem with 2 periods and 3 segments consists of one boiler, 2 CHP engines, one compression and one adsorption chiller and a cold-water storage. The design changes only slightly for the solution with 8 periods and 8 segments; only a hot water storage is added and the capacity of the cold-water storage is doubled. The calculation of the lower bound (step (i)) takes about 800 s. The computational time of the upper bound consists of three parts: time-series aggregation (step (iii)), aggregated synthesis optimization (step (iv)), and operational optimization (step (v)). The computational time for the time-series aggregation and the aggregated synthesis optimization increases with a higher time resolution, but only takes few seconds for small number of time steps. In contrast, the operational optimization is independent of the aggregation of the time series, because the full time series is used. The operational optimization takes up to 5500 s, with a mean value about 1800 s. As a benchmark, we solve the original synthesis problem directly with CPLEX 12.6.3.0. The benchmark provides the first feasible solution after 3 hours (10827 s) and reaches the time limit of 120 hours with a relative gap of 10.07 %. To validate the computational results, we repeat the calculation of the original time series for 10 samples generated by statistical noise. To limit the computational effort, the time limit for each sample was set to 24h. The performance is very similar for all samples of our method (Fig. 2). The performance of the benchmark method differs: seven samples perform in a similar way to the original time series; In one sample no feasible solution is found at all; In two samples the required optimality gap is satisfied in 13h, marked as '+' (Fig. 2). The proposed method outperforms the benchmark in all samples, always satisfying the desired gap in less than an hour.

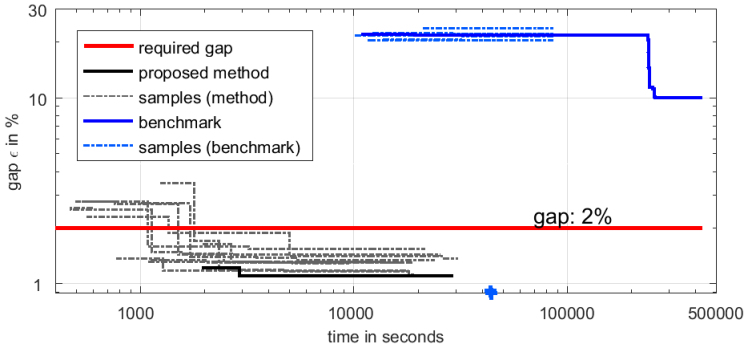


Figure 2: Logarithmic gap ε of the proposed rigorous synthesis method and the benchmark CPLEX as function of the logarithmic solution time. The required optimality gap ε of 2 % is marked in thick red.

5. Conclusions

The synthesis of energy systems results in large-scale optimization problems, which are computationally challenging and often not solvable within reasonable computational time or memory limits. In this paper, we propose a rigorous synthesis method providing feasible solutions via time-series aggregation as well as lower bounds via relaxation. The proposed method is applied to a real-world industrial synthesis problem. The results show that the lower bound provided by the relaxed problem is sufficient to reach an optimality gap of less than 2%. In our case study, only two typical periods with three segments each are required to obtain a solution of the original synthesis problem within the optimality gap. The method provides fast convergence within one hour, outperforming the state-of-the-art solver CPLEX, which does not satisfy the desired gap after 120 hours. The method is generally applicable to two-stage synthesis problems with binary decision variables, in which only the second-stage operational problem depends on the time series.

Acknowledgements: The research leading to this contribution was funded by the German Federal Ministry of Education and Research (BMBF) under grant number 03SFK3L1 and by the German Federal Ministry of Economic Affairs and Energy under grant number 03ET1259A.

References

Bahl, B., Söhler, T., Hennen, M., Bardow, A. (2018a): Typical Periods for Two-Stage Synthesis by Time-Series Aggregation with Bounded Error in Objective Function. Frontiers in Energy Research, 5, pp. 1-35.

Bahl, B., Lützow, J., Shu, D., Hollermann, D. E., Lampe, M., Hennen, M., Bardow, A. (2018b): Rigorous synthesis of energy systems by decomposition via time-series aggregation. Computers & Chemical Engineering, 112, pp. 70-81.

Box, G.E., Jenkins, G.M., Reinsel, G.C., Ljung, G.M. (2015): Time series analysis: forecasting and control.

Domínguez-Munõz, F., Cejudo-López, J. M., Carrillo-Andrés, A., and Gallardo-Salazar, M. (2011): 474 Selection of typical demand days for CHP optimization. Energ Build 43, pp. 3036 - 3043

Grossmann, I. E. (2012). Advances in mathematical programming models for enterprise-wide optimization. Computers & Chemical Engineering, 47, pp. 2-18.

Kaufman, L., Rousseeuw, P. (2008): In: Partitioning Around Mediods (Program PAM). Wiley, pp. 68-125.

Mancarella, P. (2012): Multi-energy systems): An overview of concepts and evaluation models. Energy, 65.

Maravelias, C. T., Sung, C. (2009): Integration of production planning and scheduling: Overview, challenges and opportunities. Computers & Chemical Engineering, 33, pp. 1919-1930.

Rong, A., Lahdelma, R., Luh, P. (2008): Lagrangian relaxation based algorithm for trigeneration planning with storages. European Journal of Operational Research, 188, pp. 240-257.

Voll, P.; Klaffke, C.; Hennen, M.; Bardow, A. (2013): Automated superstructure-based synthesis and optimization of distributed energy supply systems. Energy, 50, pp. 374 - 388.

Yokoyama, R., Shinano, Y., Taniguchi, S. , Ohkura, M., Wakui, T.(2015): Optimization of energy supply systems by MILP branch and bound method in consideration of hierarchical relationship between design and operation. Energy Conversion and Management, 92, pp. 92-104.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50128-2

Algorithms for Mixed-Integer Optimization Constrained by Partial Differential Equations

Christian Wesselhoeft[a], David A. Ham[b] and Ruth Misener[a*]

[a]*Department of Computing, Imperial College London, SW7 2AZ, UK*
[b]*Department of Mathematics, Imperial College London, SW7 2AZ, UK*
r.misener@imperial.ac.uk

Abstract

Mixed-integer, partial differential equation (PDE) constrained optimization (MIPDECO) is a flexible modeling framework with many engineering applications, e.g. tidal and wind turbine siting, pharmaceutical business operations, disaster recovery, and solid product creation. But solving MIPDECO is daunting because it combines both integer programming and partial differential equations in a single optimization model. We consider a range of optimization algorithms for addressing MIPDECO from two applications: (i) the Source Inversion and (ii) the Tidal Stream Turbine Optimization problems. We report on the relevant merits of these approaches and make our results available as an open source extension to OpenTidalFarm.

Keywords: Mixed integer optimization, partial-differential equation constraints, tidal stream turbines

1. Introduction

This paper discusses mixed-integer, partial differential equation (PDE) constrained optimization (MIPDECO) through two applications: (i) the Source Inversion problem with applications in well drilling and (ii) the Tidal Stream Turbine Optimization problem with applications in renewable tidal energy (Cay et al. 2016; Funke et al. 2016). To solve these two MIPDECO problem classes, we introduce several novel optimization algorithms including two methods based on branch-and-bound, i.e. relaxation of the PDE constraints, and four methods based on gradient descent, i.e. relaxation of the integer restrictions. We analyze the efficacy of each algorithm with respect to time complexity and performance on our MIPDECO test set.

As an application, we focus on the tidal turbines and tidal stream turbine MIPDECO previously proposed by Funke et al. (2016). We reproduce the Funke et al. (2016) results, i.e. derive results maximizing power from a field of turbines. We also formulate a new objective function incorporating revenue and the time-value of money. We show that our novel MIPDECO algorithms significantly outperform the Funke et al. (2016) approach. Our results are available (https://github.com/cog-imperial/tidal_turbines-source_inversion) as an extended version of the open-source software OpenTidalFarm using FEniCS (Logg et al. 2012), CasADi, dolfin-adjoint (Farrell et al. 2013), and IPOPT (Wächter and Biegler, 2006).

2. Background

2.1. Mixed-integer partial differential equation constrained optimization

The following equations formalize a mixed-integer PDE-constrained optimization (MIPDECO) problem. MIPDECO generalizes mixed-integer optimal control, e.g. as

considered by Mohideen et al. (1997), Allgor & Barton (1999) and Sager et al. (2009). In mathematical terms, it is important to note that the PDE states $\boldsymbol{u}$ and the decision

$$\min_{\boldsymbol{u},\boldsymbol{x}} \; J(\boldsymbol{u},\boldsymbol{x}) \quad s.t. \quad \begin{cases} \mathcal{F}(\boldsymbol{u},\boldsymbol{x}) = 0 \\ \boldsymbol{g}(\boldsymbol{x}) = 0 \\ \boldsymbol{h}(\boldsymbol{x}) \leq 0 \\ x_i \in \mathbb{Z} \quad \forall i \in 1,...,K \\ x_j \in \mathbb{R} \quad \forall j \in K+1,...,N \\ \boldsymbol{x},\boldsymbol{u} \in \mathcal{H}_1, \mathcal{H}_2 \end{cases}$$

variables $\boldsymbol{x}$ exist in a possibly infinite-dimensional Hilbert space, as they represent arbitrary functions varying over the entire domain. The resulting MIPDECO takes the form (where *F* denotes the PDE system, and *K* and *N* are potentially infinite):

2.2. Discretize then optimize versus optimize then discretize

The two primary approaches to solving MIPDECO are illustrated in Figure 1. This paper assumes that the MIPDECO is too big for global optimization, e.g. as in Houska & Chachuat (2017), so the illustrated solutions are local only. Given the initial MIPDECO problem with a continuous PDE constraint, we can either (i) find points satisfying the necessary optimality conditions (NC) and then discretize the PDE, yielding Solution 1 (Sol. 1) or (ii) first discretize the PDE constraint and then solve using nonlinear programming (NLP), yielding Solution 2 (Sol. 2). These two techniques may not yield equivalent answers, i.e. ¬ (Sol.1 = Sol.2).

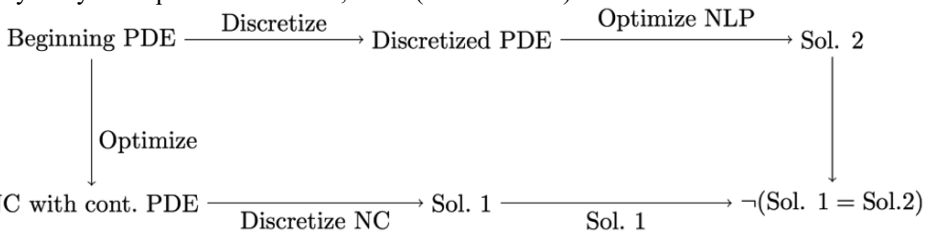


Figure 1 Discretization order matters: discretize then optimize vs. optimize then discretize

Before discretizing/optimizing MIPDECO, observe that MIPDECO can be classified into problems with (i) mesh-independent and (ii) mesh-dependent decision variables. Consider the MIPDECO problem after continuous PDE *F* has been discretized. Denoting the discrete PDE as $\mathbb{F}$, and transforming our infinite dimensional Hilbert spaces into finite vector spaces V_1 and V_2:

$$\min_{\boldsymbol{u},\boldsymbol{x}} \; J(\boldsymbol{u},\boldsymbol{x}) \quad s.t. \quad \begin{cases} \mathbb{F}(\boldsymbol{u},\boldsymbol{x}) = 0 \\ \boldsymbol{g}(\boldsymbol{x}) = 0 \\ \boldsymbol{h}(\boldsymbol{x}) \leq 0 \\ x_i \in \mathbb{Z} \quad \forall i \in 1,...,K \\ x_j \in \mathbb{R} \quad \forall j \in K+1,...,N \\ \boldsymbol{x},\boldsymbol{u} \in \mathcal{V}_1, \mathcal{V}_2 \end{cases}$$

If the problem is mesh-independent, then the decision variables are independent of the PDE discretization, e.g. *O(n)* memory complexity for *n* mesh-independent decision variables. But the number of mesh-dependent decision variables increases as the discretized PDE is refined. Obviously, this distinction has significant ramifications on the scalability of a MIPDECO problem, e.g. with a *m x m* mesh any optimization algorithm will have memory complexity $O(m^2)$ with regards to the decision variables.

3. Test Instances

We consider two optimization problems: (i) the Source Inversion problem with applications in well drilling and (ii) the Tidal Stream Turbine Optimization problem with applications in renewable tidal energy (Cay et al. 2016; Funke et al. 2016).

3.1. Source Inversion Problem

Following Cay et al. (2016), we consider the Source Inversion MIPDECO problem, though with the use of adjoint-based gradient optimization methods (implemented with the FEniCS PDE-solving framework and dolfin-adjoint). The Source Inversion problem is a function-tracking model minimizes the distance between a reference state $\bar{\boldsymbol{u}}$ and the solution to a source PDE $\boldsymbol{u}$ (which is a Poisson equation with the source term equalling the sum of a host of source functions) over the simple box domain Ω, paired with a Dirichlet boundary condition. This minimization problem also requires that the maximum number of activated source functions does not exceed a user-defined integer S, and that each source function at location {k,l} is modelled as a Gaussian function f_{kl} with binary coefficient w_{kl} (which controls the activation/inactivation of each source function). In other words, we want to find the most accurate way to recreate the solution to the reference PDE with a combination of up to S source functions:

$$\min_{\boldsymbol{u},\boldsymbol{w}} \ \mathcal{J}(\boldsymbol{u},\boldsymbol{w}) = \frac{1}{2}\int_{\Omega}(\boldsymbol{u}-\bar{\boldsymbol{u}})^2 d\Omega \quad s.t. \quad \begin{cases} -\Delta \boldsymbol{u} = \sum_{k,l} w_{kl} f_{kl} & \in \Omega \\ \boldsymbol{u} = 0 & \in \delta\Omega \\ \sum_{k,l} w_{kl} \leq S & \\ w_{kl} \in \{0,1\} & \forall k,l \end{cases}$$

The Source Inversion problem has mesh-independent binary variables w_{kl}. We considered 4 methods for solving this problem:

1. **Heuristic Integer Algorithm** This method solves the continuous relaxation of the Source Inversion problem, i.e. $w_{kl} \in [0,1]$, and then rounds each continuous w_{kl} depending on its optimal continuous magnitude and the value of S.
2. **Penalty Algorithm** This method is equivalent to Algorithm 1 except that it penalizes non-binary w_{kl} by adding terms $w_{kl} \times (1 - w_{kl})$ in the objective.
3. **Two-Step Algorithm** This method takes a two-stage approach by beginning with solving a continuous relaxation of the Source Inversion problem and then trying to enforce binary by adding terms $w_{kl} \times (1 - w_{kl})$ to the objective.
4. **Branch & Bound** This method enforces binary w_{kl} using divide and conquer. It converges to the global optimum if all continuous relaxations are convex, but we do not solve each (possibly nonconvex) subproblem to global optimality.

3.2. Tidal Stream Turbine Optimization

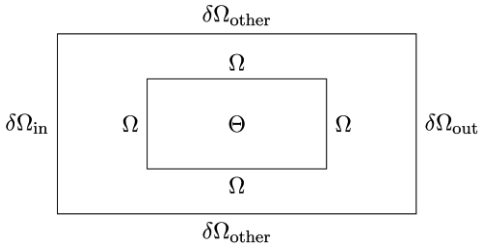


Figure 2 Problem Topology

In tidal stream turbine optimization (TSTO), there is a two- or three-dimensional domain Ω in which the tidal turbines are placed. There is also a sub-domain $\Theta \in \Omega$ such that the tidal turbines are all placed in Θ, with tidal modelling also occurring around the farm area in Ω. In this paper, we consider unidirectional tidal currents in a 2-D rectangular domain. That is, the tidal

velocity enters Ω from boundary $\delta\Omega_{in}$ and leaves from $\delta\Omega_{out}$. We assume no tidal activity on the other two boundaries.

Our model formulation is very similar to the formulation of Funke et al. (2016) and incorporates the following modeling considerations:

- **Constant water velocity.** Alternative approaches may consider a sinusoidal model, e.g. to represent tidal inflow.
- **Shallow water equations.** This approximation of the Navier-Stokes equation models fluid motion around the turbines.
- **Energy captured by the turbine farm.** This will feed into the profit function.
- **Proximity constraints.** Turbines cannot be built too close to one another.

The original optimization formulation of Funke et al. (2016) also required that we specify the number of turbines, but we make this a decision variable. To make the objective function more realistic, we also incorporate the cost of the installed turbines and the revenue from the produced energy into the objective function.

We compute the cost of the turbines using the levelized cost of energy (*LCOE*) to model the lifetime cost of a single turbine with discount rate r = 10%, time period *(0,T)*, value of energy output O_t, and cost C_t at time t as:

$$LCOE = \frac{\sum_{t=0}^{T} C_t/(1+r_t)^t}{\sum_{t=0}^{T} O_t/(1+r_t)^t}$$

From this, given a predefined constant *LCOE*, we can easily calculate:

$$\text{Total Discounted Cost} = \sum_{t=0}^{T} C_t/(1+r_t)^t = LCOE \cdot O_t/(1+r_t)^t$$

Modeling revenue is more difficult because it cannot be modeled per turbine: power extraction (and therefore revenue) depends on the entire turbine field configuration. Consequently, we create a revenue function primarily dependent on: (i) the 10% discount rate, (ii) the power produced by the field at each time step, (iii) a constant k which corrects for efficiency losses, and (iv) the income per unit of energy *I(t)* at time t. To penalize the energy optimism, we set $k = 0.5$, so that only 50% of theoretically extractable energy is actually captured by the farm. Furthermore, the income per energy unit *I(t)*, is based upon the UK government's Electricity Market Reform program which encourages (and subsidies) renewable electricity generation.

$$\text{Total Discounted Revenue} = R(\boldsymbol{u}, d) = k \sum_{t=0}^{T} \frac{IE_t(\boldsymbol{u}, d)}{(1+r_t)^t}$$

Our complete model formulation is available in the MSc thesis of C. Wesselhoeft. We considered several different methods for solving tidal stream turbine optimization:

1. **Continuous**. This method of Funke et al. (2016) calculates the optimal continuous density field, profit, power extraction/cost, and number of turbines.
2. **Two-Step**. This algorithm uses the continuous solution and a method to place N turbines to set up an initial layout for the discrete problem with the number of turbines pre-specified by a continuous relaxation approximation.
3. **MIPDECO**. This algorithm uses the continuous solution and chooses an initial lay-out for a MIPDECO problem without a pre-specified number of turbines.

4. Results

4.1. Case Study 1: Source Inversion Problem

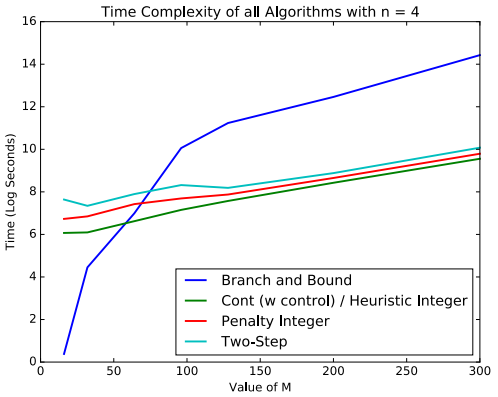


Figure 3 Time complexity versus mesh size

The MIPDECO algorithms for Source Inversion are effectively: (i) solving a relaxed nonlinear optimization problem and then rounding and (ii) solving relaxed nonlinear optimization problems in a branch and bound framework. Branch and bound is fast and accurate when the mesh m is coarse. But the key difference between rounding and branch and bound algorithms lies in their different scaling. Gradient-based algorithms create a control for each source function and only optimize based on these controls; consequently they scale well in mesh size, yet poorly in number of source functions. Branch and bound must explore constraints based on the total set of N + M variables, and thus performs well when $n^2 + m^2$ is small, but performs poorly as the total number of variables increases. In the Source Inversion problem, $n << m$ by construction, and thus mesh refinement will often be the bottleneck whilst scaling.

4.2. Case Study 2: Tidal Stream Turbine Optimization

Table 1 Results for tidal stream turbine maximization (the continuous solution is infeasible). The runtime includes the initialization step. "Initial Two Step" is the first step of the MIPDECO row.

Algorithm	Profit (GBP)	Number of Turbines	Runtime (seconds)
Continuous	9.57e7	122.5	7978
Two-Step Discrete	5.04e7	123	38363
MIPDECO	4.021e7	123	9622
Initial Two-Step	3.11e7	123	7979

Solving the problem posed in Section 3.2 (see Wesselhoeft's MSc https://goo.gl/Srx1Zr and our code, https://github.com/cog-imperial/tidal_turbines-source_inversion), we obtain the results in Table 1 from the optimization algorithms. We chose a mesh size of 4m in the farm area and 100m outside the farm area, i.e. a computational mesh of 90,500 elements. The first three optimization methods are the ones outlined in Section 3.2 and the *Initial Two-Step* method places turbines randomly from a continuous relaxation. Since this is a maximization problem, randomly placing turbines clearly has the weakest performance of the 4 approaches.

The results in Table 1 lie on a spectrum. The continuous formulation automatically yields the optimal number of turbines and supposedly yields the highest profit, though as it does not individually resolve turbines, the profit comparison between the continuous and discrete situations is spurious. Crucially, however, optimizing the continuous case only requires ~133 minutes, despite the large number of elements in the

mesh. After running the continuous example, we placed 123 turbines at random. This initial placement from the two-step algorithm yielded 3.11 x 10^7 GBP profit.

To improve the baseline result, we implemented the second step of the two-step algorithm from Section 3.2 and ran a full discrete optimization randomly-generated layout as a starting position. While the optimization resulted in 5.04 x 10^7 GBP profit, i.e. a large increase over the baseline, the increase in profit came at the cost of a severe increase in time complexity as the second step of the two-step algorithm individually resolves the location of each turbine.

As an intermediate solution, we used the MIPDECO algorithm to fix the possible turbine locations to a subset of the 200 locations, and then optimize binary variables assigned to each fixed turbine. Consequently, the total MIPDECO runtime is roughly one fifth of the Two-Step algorithm's runtime – in this case 2.65 hours instead of 12.87 hours. Additionally, the MIPDECO turbine layout improved profit as compared with the initial random heuristic layout by 33.5%, though still yielded far less profit than the discrete case, which improved upon the initial layout by 62.1%. In short, the two-step discrete method returned a significantly higher profit than the baseline, at the expense of a significant increase in algorithm runtime. The MIPDECO formulation improved moderately upon the baseline, with comparatively little time increase.

5. Conclusions

The methods described in this paper apply to test cases with mesh-dependent (source inversion) and mesh-independent (tidal turbine) optimization problems. We see the trade-off, e.g. with respect to quality of solution or time invested, with respect to a number of approaches. In particular, notice that exact methods for resolving the integer variables are often difficult in these large-scale problems.

This work was funded by an Engineering & Physical Sciences Research Council Research Fellowship to RM [grant number EP/P016871/1] and a Natural Environment Research Council Independent Research Fellowship to DH [grant number NE/K008951/1].

References

RJ Allgor, PI Barton. **(1999)** Mixed-integer dynamic optimization I: problem formulation. *Comput Chem Eng*, 23(4-5), 567-584.

P Cay, S Leyffer, B van Bloemen Waanders, D Kouri, A Thuenen. **(2016)** SAMSI Optimization Workshop. Research Triangle Park. pages 24

PE Farrell, DA Ham, SW Funke, ME Rognes. **(2013)** Automated derivation of the adjoint of high-level transient finite element programs *SIAM J Sci Comput* 35(4), C369-C393

S Funke, S Kramer, M Piggott. **(2016)** Design optimisation & resource assessment for tidal-stream renewable energy farms using a new continuous turbine approach. *Renew Energy*, 99, 1046–1061.

B Houska, B Chachuat. **(2017)** Global optimization in Hilbert space. *Math Program* https://doi.org/10.1007/s10107-017-1215-7

A Logg, KA Mardal, GN Wells et al. **(2012)** Automated Solution of Differential Equations by the Finite Element Method, Springer.

MJ Mohideen, JD Perkins, EN Pistikopoulos. **(1997)** Towards an efficient numerical procedure for mixed integer optimal control. *Comput Chem Eng*, *21*, S457-S462.

S. Sager, HG Bock, G Reinelt. **(2009)** Direct methods with maximal lower bound for mixed-integer optimal control problems. *Math Program*, 118(1), 109-149.

A Wächter, LT Biegler **(2006)** On the Implementation of a Primal-Dual Interior Point Filter Line Search Algorithm for Large-Scale Nonlinear Programming, *Math Program* 106(1), 25-57

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50129-4

Parallel cooperative optimization through hyperheuristics

Paola P. Oteiza[a], Diego A. Rodríguez[b, c] , Nélida B. Brignole[a, b*]

[a]*Planta Piloto de Ingeniería Química (Universidad Nacional del Sur - CONICET) Camino La Carrindanga km. 7 - 8000 Bahía Blanca – Argentina*
[b]*Laboratorio de Investigación y Desarrollo en Computación Científica (LIDECC)- Departamento de Ciencias e Ingeniería de la Computación (DCIC), Universidad Nacional del Sur (UNS), Bahía Blanca, Argentina.*
[c]*Facultad de Ciencias Exactas, Universidad Nacional de Salta (UNSa), Salta, Argentina.*
dybrigno@criba.edu.ar

Abstract

A hyperheuristics that coordinates the interaction between various metaheuristic techniques is presented. The proposed algorithm, which we called Parallel Optimizer With Hyperheuristics (POWH), includes a Genetic Algorithm, Simulated Annealing, and Ant Colony Optimization. In view of the need to escape from local optima, information exchanges take place between these metaheuristics. In this way, it is possible to take advantage of each metaheuristics' particular strengths during the search process. Testing related to the hyperheuristic approach was carried out by using the following real-life case studies: I. the optimal design of a subsea pipeline network and II. the urban bus-transit optimal planning. In both cases, a satisfactory reduction of the computational time was achieved due to the parallel implementation that allowed several metaheuristics to run simultaneously. Moreover, better results were also obtained thanks to the parallel cooperative combination of metaheuristics compared with serial executions.

Keywords: parallel programming, metaheuristics, pipelines, LRP, hyperheuristics.

1. Introduction

Nowadays, metaheuristic optimization algorithms have turned out to be quite attractive because of their distinct advantages over traditional algorithms (Gupta and Ramteke, 2014). Since metaheuristics can solve multiple-objective multiple-solution and nonlinear formulations, they are employed to find high-quality solutions to an ever-growing number of complex real-world problems, such as the combinatorial ones. Evolutionary techniques, like Genetic Algorithms (GA) and Ant Colony Optimization (ACO), have successfully been employed in solving many general optimal problems, such as the optimization of fed-batch fermentation. Optimization frameworks combining GA and Simulated Annealing (SA) have also been applied for the design and synthesis of heat exchanger networks (Shelokar *et al*., 2014), as well as for model predictive control (Venkateswarlu and Reddy, 2008). In particular, some variants of hybrid-GA techniques have been reported as novel prediction techniques (Sukumar *et al.,* 2014).

In the present study we have proposed a hyperheuristic approach by incorporating GA, SA and ACO. A hyperheuristics (Burke et al., 2013) is a general search method that has a set of solvers (low-level methods) and manages the execution of the most convenient

technique at a given time during the search process. When solving difficult computational search problems, hyperheuristics have proved to be advantageous. In operational research, there are some hyperheuristic-based algorithms that have been developed for single kinds of classical metaheuristics (Koulinas et al, 2014), but the combination of several types of metaheuristics has not been thoroughly analysed. Branke et al. (2016) summarized the state-of-the-art alternatives for hyperheuristic design choices and identified the main issues that future work should focus on. They pinpointed that their high computational requirements constitute a major disadvantage of hyperheuristics. In this context, parallelism emerges as an advanced strategy to reduce computational times significantly.

Both the demand for fast solutions of optimization problems, together with the advent of novel concurrent computing architectures constitute incentives for the development of parallel algorithms (Alba, 2005) in order to solve challenging problems efficiently. Regarding computational efficiency in urban planning strategies, Agrawal and Mathew (2004) have implemented a parallel algorithm based on a GA model for a large-scale problem, reporting satisfactory results in computational time, speedup and efficiency.

In view of the widespread need to solve challenging large-scale PSE problems, a parallel implementation was adopted and the proposed algorithm was called Parallel Optimizer With Hyperheuristics (POWH). The parallelization allows several metaheuristics to run simultaneously in threads, thus achieving a satisfactory decrease of the computational time. POWH design is briefly described in Section 2. Then, in Section 3 both pipelining and bus planning are presented as test problems, whose main features are pointed out before the discussion of some computational results, which are given in Section 4. Finally, some conclusions and hints for future work are summarized in Section 5.

2. Methodology

A cooperative strategy applicable to optimization problems is presented. The proposed algorithm is a hyper-heuristics, which comprises 3 well-known canonical metaheuristics with various strengths. The algorithm for the Parallel Optimizer With Hyperheuristics (POWH) is based on the master–slave paradigm (Fig. 1) and has multiple metaheuristics that cooperate periodically and work concurrently on distributed computing environments. Its organizational framework is an A-team architecture (Talukdar et al., 1998), where the autonomous agents are the following metaheuristics: Simulated Annealing (SA), Genetic Algorithm (GA) and Ant Colony Optimization (ACO).

The partnership is useful because it has been conceived by combining different primary procedures: SA is a trajectory-based technique, GA is a population-based method and ACO is a constructive approach.SA's strategy avoids local minima by allowing with a certain probability to choose a solution whose fitness value is worse than the current solution. Therefore, the SA algorithm converges relatively slowly towards the final solution. This drawback can be remediated by both implementing parallel programming and hybridizing the algorithm with ACO, which is particularly advantageous due to the remarkable ability of artificial ants to construct solutions guided by the pheromone trails. Moreover, SA tends to find local optima, thus delaying convergence. This weakness can also be surpassed thanks to a combination of SA with other metaheuristics that may contribute to diversify the search by enlarging the search space by means of efficient explorations. For this purpose, it was necessary to incorporate another exploratory method, such as GA.

In GA the initial population is chosen randomly, while the best solution designated by the master is always incorporated as an elite solution. In general, the termination criteria are all based on the amount of executions. These limits are: 50 generations for GA, 100 iterations for SA and 50 colonies for ACO.

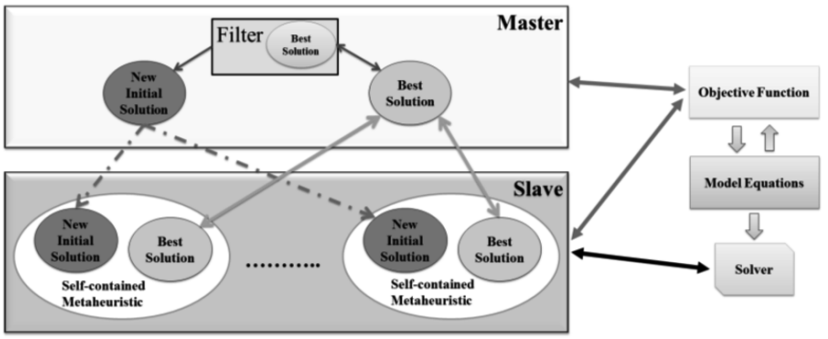


Figure 1. Hyperheuristic Model.

As to the control system, the choice of a suitable metaheuristics is guided by a ranking index (*Ind*) that assesses each metaheuristics' performance through Eq. (1-6), whenever a processor becomes idle. The technique with the highest rank will thus be eligible for the next execution. Besides, the particular problem is not only incorporated in modules for the model equations and the objective function, but also in the filter module because filtering is necessary to detect unfeasible solutions. The filter is a problem-dependent selector that indicates which solutions should be erased.

$$Ind = Fitness + Evaluation + Time + Optimality + Repetition \quad (1)$$

$$Fitness = 1 - \left({BestFitness}/{SumFitness}\right) \quad (2)$$

$$Evaluation = 1 - {Eval}/{SumEval} \quad (3)$$

$$Time = 1 - {BestTime}/{SumTime} \quad (4)$$

$$Optimality = 1 - \left({BestFitness}/{(SA_{bf} + GA_{bf} + ACO_{bf})}\right) \quad (5)$$

$$Repetition = 1 - \left({e^{Rep}}/{Q_{runs}}\right) \quad (6)$$

3. Test Cases

A couple of PSE-representative case studies were solved. Aiming at algorithmic generalization, Case II was chosen since it also serves to create very large challenging instances that may demand long computational times.

CASE I: The optimal design of a subsea pipeline network. It was modelled based on the problem formulation in Oteiza *et al.* (2015). In contrast, the well platforms were represented by concentrating nodes, which are located by the optimizer in the adapted model. Besides, the objective function (Eq. (7)) was adapted in this work so as to comprise offshore pipelining by contemplating specific terms related to installation and operating costs. The Net Present Value (NPV) of the project was calculated from revenue

flows and terms related to the construction and operating costs for each well platform WP_k. In Eq (7), the building costs were taken into account by considering both pipeline construction $CC_{ij,t}$ and the installation costs of every well platform $\delta_{k,t}$. The operating costs include the maintenance cost $MC_{ij,t}$ and the operative labor cost LC_t. The transport tariff P_t is useful to determine whether the project is economically viable. Since fitness values were measured in terms of NPV, this is a maximization problem.

$$NPV = \sum_{t=1}^{o} \frac{P_t \text{ x } Q_t}{(1+r)^{t+d+e}} - \sum_{t=1}^{e} \sum_{i=1}^{f+q} \sum_{j=1}^{q+p} \frac{CC_{ij,t}}{(1+r)^{t-1}} FP_{ij} - \sum_{t=1}^{d} \sum_{k=1}^{q} \frac{\delta_{k,t}}{(1+r)^{t+e-1}} WP_k - \sum_{t=1}^{d} \sum_{k=1}^{q} \frac{MC_{k,t}}{(1+r)^{t+e-1}} WP_k - \sum_{t=1}^{o} \frac{LC_t}{(1+r)^{t+d+e}} \quad (7)$$

CASE II: Bus planning. It was modelled by means of an LRP (Location Routing Problem), which constitutes an NP-complete problem (Wolsey, 2000). The stops and routes were determined in order to meet the demand with minimal cost *Z*. The location and routing model for the urban bus transport was based on the formulation of classical LRPs (Ceder, 2016) with suitable modifications in fitness evaluation, whose function was formulated in terms of network costs Z. It includes three terms to cover the offer, while the fourth one contemplates the demand in more detail through a matrix, whose elements m_{ij} represent walking-distance penalties. As to the offer, the following terms were included: the costs of traversed edges, the fixed costs of opening bus stops and the fixed vehicular costs associated with employed buses. Then, the lower costs become, the better.

4. Performance Results

All the algorithms were implemented in Java and run on a PC with 8 GB of RAM and Intel Core i5-3330. For all instances, each test was repeated independently 30 times. As to the "best average fitness", it was calculated as the average across the 30 runs of the best value identified in each run. In the first place, serial algorithms were programmed to evaluate the behaviour of the individual metaheuristics whose computational times (costs) are employed as the upper bounds for this analysis. An incremental variation of the amount of nodes was considered on Table 1. It shows the best average fitness and the corresponding average time required for each metaheuristics running sequentially for different sample problems. The solution quality (fitness) was always satisfactory, but computational time also increased as problem size grew up. It can be concluded that POWH´s organization (GA-SA-ACO) is effective because the addition of various agents has provided benefits in terms of solution quality (fitness) for large instances.

Table 1 The effect of metaheuristics on the computational time (ms)

Algorithms		**GA**		**SA**		**ACO**		**GA-SA-ACO**	
Case	**Size (Nodes)**	**Fitness (U$S)**	**Time (ms)**	**Fitness (U$S)**	**Time (ms)**	**Fitness (U$S)**	**Time (ms)**	**Fitness (U$S)**	**Time (ms)**
I Small	5	$9.69E^{+07}$	217	$9.69E^{+07}$	328	$8.60E^{+07}$	527	$9.69E^{+07}$	677
I Large	18	$3.19E^{+08}$	3,233	$1.97E^{+08}$	928	$2.00E^{+08}$	923	$1.14E^{+11}$	1,370

Since substantially smaller times are required, a parallel combination of metaheuristics (POWH) was implemented. Parallelization by threading allows several metaheuristics to run simultaneously in threads. On Table 2 the impact of parallelism was corroborated by

comparing serial and parallel times.The average deviation between the best lower bounds (POWH) and the upper bound (SOWH) is around 50% on smaller instances and around 30% on some large instances, whenever solutions of the same quality are reached.

Table 2. Contrasting Parallel (POWH) and Serial (SOWH) algorithms for Case I

Algorithm		**SOWH**		**POWH**	
Case	**Size (Nodes)**	**Fitness (U$S)**	**Time (ms)**	**Fitness (U$S)**	**Time (ms)**
I Small	5	$9.69E^{+07}$	1,458	$9.69E^{+07}$	677
I Large	18	$1.14E^{+11}$	2,097	$1.14E^{+11}$	1,370

As to solution quality and for Case II, Table 3 shows that as the number of processors increases, POWH finds solutions of better quality (i.e., lower Z values) in shorter times.

Table 3. Contrasting Parallel (POWH) and Serial (SOWH) cooperative algorithms.

Algorithm	**# Processes**	**Average Z (U$S)**	**Average Time (ms)**	**Speed-up**	**Efficiency (%)**
SOWH	1	6,354	2,764		
POWH	4	5,804	1,897	1.46	36.5
POWH	7	2,473	1,793	1.54	22.0

5. Conclusions and Future Work

Some real-life PSE optimization problems proved to be benefited from quick solutions that could be achieved with the help of parallel programming by means of the proposed hyperheuristic method (POWH) presented here. The solution method coordinates interactions among metaheuristics to create a fast strategy that is able to escape from local optima, thus performing an effective search of a solution space. Three metaheuristics have been combined: Simulated Annealing, Genetic Algorithm, and Ant Colony Optimization. The solution strategy worked out satisfactorily in practical problems. For different sizes, POWH showed promising performance in the search for solutions since it was able to obtain high-quality results in reasonably shorter computing times. For the continuation of this research work, it is necessary to verify in general terms that POWH is always both rigorous and efficient. Then, further testing that involves a wide choice of sample problems oriented to parallel computing running in threads is vital to complete thoroughly the computational design.

Notation

ACO_{bf}= The best solution yielded by the ACO technique
BestFitness = The best fitness yielded for the best attained solution
BestTime = The shortest time demanded by an execution
$CC_{ij,t}$: construction cost from the *i*-th to the *j*-th point, in the *t*-th period
$\delta_{k,t}$: installation cost of the *k*-th Well Platform in the *t*-th period
d: total number of time periods related to WP construction
e: total number of pipeline construction time periods
Eval = The minimum number of evaluations
f: number of Wells

FP_{ij}: feasible path from the i-th to the j-th point
GA_{bf} = The best solution yielded by the GA technique
Ind = Ranking index
LC_t: operative labor cost in the t-th period
$MC_{k,t}$: maintenance cost of the WP structure
o: project lifetime
p: number of plants onshore
P_t: sale price of gas in the t-th period
q: number of Well Platforms that have been activated
Q_t: amount of gas transported in the t-th period
Q_{runs} = Number of performed executions
r: discount rate
Rep = Counter
SA_{bf} = The best solution yielded by the SA technique
$SumEval$ = The sum of all evaluations
$SumFitness$ = The sum of all BestFitness values
$SumTime$= Runtimes required by the past Ne executions
WP_k: for the k-th Well Platform: WP_k=1 if active, otherwise WP_k=0
Z = Total system cost per trip

References

E. Alba, 2005, Parallel Metaheuristics: A New Class of Algorithms, Wiley.

J. Agrawal, T.V. Mathew, 2004, Transit route network design using parallel genetic algorithm, Journal of Computing in Civil Engineering, 18, 3, 248-256.

E. K. Burke, M. Gendreau, M. Hyde, G. Kendall, G. Ochoa, E. Özcan, R. Qu, 2013, Hyper-heuristics: A survey of the state of the art, J. Operational Research Society, 64 ,12, 1695-1724.

J. Branke, S. Nguyen, C.W. Pickardt, M. Zhang, 2016, Automated design of production scheduling heuristics: A review, IEEE Transactions on Evolutionary Computation, 20, 1, 110-124.

A. Ceder, 2016, Public Transit Planning and Operation: Modeling, Practice and Behavior, 2nd Ed., Boca Raton, FL, USA, CRC Press.

S. K. Gupta, M. Ramteke, 2014, Applications of genetic algorithms in chemical engineering I: Methodology, In: Applications of Metaheuristics in Process Eng., Switzerland, Springer, 39-59.

G. Koulinas, L. Kotsikas, K. Anagnostopoulos, 2014, A PSO based hyper-heuristic algorithm for the classic resource constrained project scheduling problem, Inf. Sci., 277, 680-693.

P.P. Oteiza, M. De Meio, D.A. Rodriguez, V. Viego, N.B. Brignole, 2015, Metaheuristic Techniques for the Optimal Design of NGL Pipelining, Computer-Aided Chem. Eng., 37, 785-790.

P. Shelokar, A. Kulkarni, V.K. Jayaraman, P. Siarry, 2014, Metaheuristics in Process Eng.: A Historical Perspective, In: Applications of Metaheuristics in Process Eng., Switzerland, 1-38.

N. Sukumar, G. Prabhu, P. Saha, 2014, Applications of genetic algorithms in QSAR/QSPR modeling, In: Applications of Metaheuristics in Process Eng., Switzerland, Springer, 315-324.

S. Talukdar, L. Baerentzen, A. Gove, P. De Souza, P, 1998, Asynchronous teams: Cooperation schemes for autonomous agents, Journal of Heuristics, 4, 4, 295-321.

C. Venkateswarlu, A.D. Reddy, 2008, Nonlinear model predictive control of reactive distillation based on stochastic optimization, Ind. Eng. Chem. Res., 47(18), 6949–6960

L.A. Wolsey, 2000, Integer programming, IIE Transactions, 32, 273-285, 2-58.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50130-0

Optimization of polymer molecular weight distribution with non-ideal reactors using surrogate models

Jing Kong[a], Xi Chen[a*], Zhijiang Shao[a], Lingyu Zhu[b]

[a]*College of Control Science and Engineering, Zhejiang University, Hangzhou, Zhejiang, 310027, P.R. China*
[b] *College of Chemical Engineering, Zhejiang University of Technology, Hangzhou, Zhejiang, 310014, P.R. China*
xichen @iipc.zju.edu.cn

Abstract

Molecular weight distribution (MWD) is very important to characterize polymer quality. To reveal non-ideal mixing conditions in a polymerization reactor, computational fluid dynamics (CFD) methods are essential. However, the combination of CFD and MWD calculation is time-consuming. Moreover, it is difficult to obtain gradient information in the CFD simulation. All these challenges make optimization process inefficient. In this project, optimization with embedded polymer MWD for a low-density polyethylene (LDPE) tubular reactor is conducted. To overcome the time-consuming and insufficient gradient information problem in calling a CFD simulation, a surrogate model method is proposed. Instead of building an accurate and global surrogate model with a large number of simulation data, an iterative surrogate model with limited data is preferred to save simulation runs. The result shows that the surrogate method can greatly reduce the time cost and achieve the optimal result.

Keywords: computational fluid dynamics; optimization; molecular weight distribution; surrogate model.

1. Introduction

Molecular weight distribution (MWD) describes the percentage of different-chain-length molecule in a polymer, which reflects more detailed information than macro quality indices, such as average molecular weight and melt index. A number of methods have been reported on MWD computation, e.g., Flory method, discrete Galerkin method Monte Carlo method, et al. But most of these studies focus on ideal reactors, which can be very different from a practical process. To deal with this problem, computational fluid dynamics (CFD) technology has been widely developed in chemical engineering. By adding reactant source terms and reaction heat to transport equations, CFD can extend to non-ideal polymerization reactors. Zhou et al. (2001) applied CFD to model LDPE tubular and autoclave reactors with the method of moments and discussed the influence of initiator concentration and the inlet temperature on the reaction. Roudsari et al. (2013) developed a CFD model to study methyl methacrylate (MMA) solution polymerization and investigated the influence of operating conditions, impeller speed, reaction temperature, residence time and inlet monomer concentration on the conversion and homogeneity of the reaction mixture. However, none of the aforementioned

research considered the MWD in their work. Recently, Zhang et al. (2017) applied the simplex method to solve the optimization of LDPE with MWD constrains. To improve the calculation efficiency, an iterative surrogate model method is proposed in this study to deal with the optimization of CFD models for polymerization reactor.

2. CFD Simulation with MWD for Non-Ideal Reactor

Free radical polymerization involves initiator decomposition, chain initiator, chain propagation, chain transfer to monomer, termination by combination, and termination by disproportionation. The reaction mechanism is shown in Table 1, where *I* denotes the initiator, *A* the initial radical, *M* the monomer, R_n the free radical, P_n the dead polymer. Subscript *n* stands for the chain length, which changes from 1 to a large number. The method of moment is introduced as

$$\lambda_i = \sum_{n=1}^{\infty} n^i [R_n], \mu_i = \sum_{n=1}^{\infty} n^i [P_n], i = 0,1,2 \tag{1}$$

where *i* represents the order of moments.

Denote k_x as the reaction rate coefficient of reaction *x*,

$$k_x = k_{x0} e^{[\frac{-(E_a + V_a p)}{R_g T}]} \tag{2}$$

where k_{x0} is initial reaction rate coefficient, E_a activation energy, and V_a is activation volume. These parameters change with different reaction systems.

The calculation equation of viscosity used in this paper is described as

$$\eta = \eta_{ethy} e^{[2.00+0.017(\frac{\mu_1}{\mu_0})^{0.556} \mu_1 + \frac{E_v}{R_g}(\frac{1}{T} - \frac{1}{423})]} \tag{3}$$

$$E_v = -500 + 560\mu_1 \tag{4}$$

where η_{ethy} is the viscosity of ethylene valued by 0.0016 *kg/(m•s)*.

According to the Flory distribution method (Flory, 1953) with the long chain and quasi-steady assumptions, MWD can be calculated by Eq.(5)-(6).

$$W(n) = (\tau + \beta)[\tau + \frac{\beta}{2}(\tau + \beta)(n-1)]n(\frac{1}{1+\tau+\beta})^{n+1} \tag{5}$$

$$\tau = \frac{k_{td}\lambda_0 + k_{trm}[M]}{k_p[M]}, \beta = \frac{k_{tc}\lambda_0}{k_p[M]} \tag{6}$$

The diameter and length of the tubular reactor are 0.038 m and 10 m, respectively. Gambit 2.4.6 is used to generate the geometry structure and the mesh of the reactor. After deriving the mesh file, Fluent 15.0 is used to establish the CFD model of LDPE based on the reaction mechanism in Table 1 and Eq.(1)-(6). Set the inlet initiator and

monomer mass fractions as 0.00051 and 0.99949, respectively. The inlet temperature is 500 K. Other variable settings are the same as Zhang et al. (2017). The CFD Simulation takes 457.34 s on a desktop with 2.50GHz Intel Core i5-7200U CPU. Fig. 1 shows the conversion distribution. To better illustrate the results, the radial direction is scaled up for 20 times. Fig. 2 shows the distribution of MWD along the centre axial direction.

Table 1. Reaction mechanism of free radical polymerization

Reaction Type	Reaction Mechanism
initiator decomposition	$I \xrightarrow{k_d} 2A$
chain initiator	$A + M \xrightarrow{k_{ini}} R_1$
chain propagation	$R_n + M \xrightarrow{k_p} R_{n+1}$
chain transfer to monomer	$R_n + M \xrightarrow{k_{trm}} P_n + R_1$
termination by combination	$R_n + R_m \xrightarrow{k_{td}} P_n + P_m$
termination by disproportionation	$R_n + R_m \xrightarrow{k_{tc}} P_{n+m}$

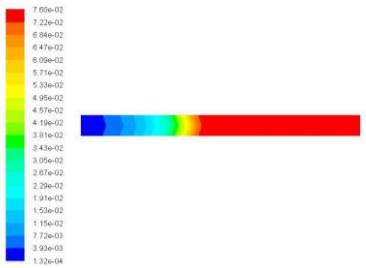


Figure 1. Conversion distribution in the tubular reactor.

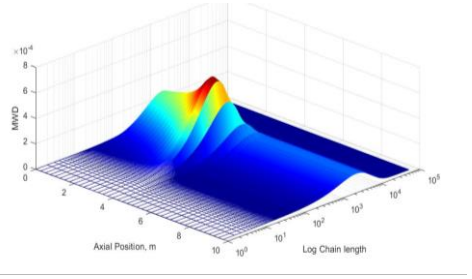


Figure 2. MWD distribution along the axial direction.

3. Optimization with Surrogate Model

The optimization problem studied in this paper is formulated as:

$$\begin{aligned} &\max\ conversion = CFD(M_I, T) \\ &s.t.\ MWD = CFD(M_I, T) \\ &\frac{1}{m}\sum_{i=1}^{m}((MWD_i - MWD_i^*)/MWD_i^*)^2 \le 0.0025 \\ &M_{Il} \le M_I \le M_{Iu} \\ &T_l \le T \le T_u \end{aligned} \tag{7}$$

It is to maximize the monomer conversion subject to an MWD constraint. m denotes the number of points sampled to represent the curve of MWD. MWD_i^* denotes the i^{th} sample point on the target MWD curve. The decision variables include M_I and T, which are the inlet initiator mass fraction and the inlet temperature, respectively. The *CFD* function stands for the CFD simulation presented in section 2. However, as it is time-

consuming and difficult to derive gradient information during the CFD simulation, the *CFD* function is replaced by a surrogate model during the solving process. We propose an iterative surrogate model method, which combines global coarse surrogate model and local accurate surrogate model, to solve the optimization problem (7). Flowchart of the method is shown in Fig. 3.

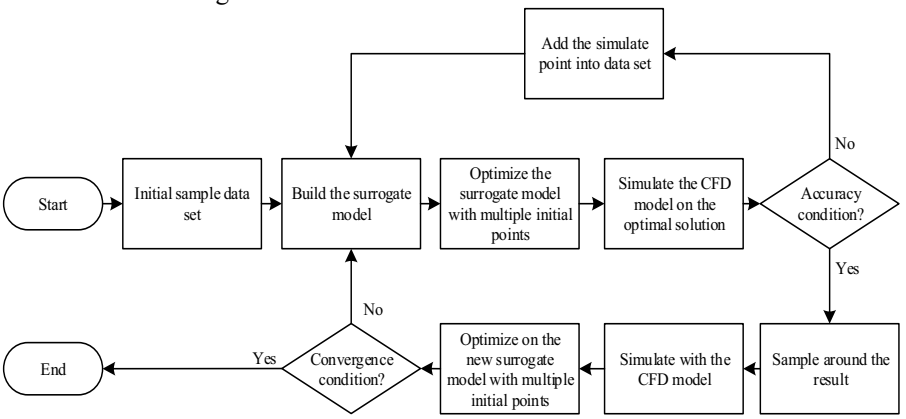


Figure 3. Flowchart of the iterative surrogate model method.

First, this method samples only a few points on the global space to form an initial sample set, S_0. The sampling method is Latin Hypercube Sampling (LHS) method (McKay et al., 1979). Then we establish a coarse global surrogate model M_0 using Kriging method (Lophaven et al., 2002) based on the sample set S_0. Set i=1 and conduct the optimization with the surrogated CFD model (M_{i-1}). To avoid of being trapped into local optimal points, multiple initial points are used for the optimization. The optimal solution is marked as $(x_i, \hat{y}_i)$. Then a CFD simulation is conducted on point x_i, and we obtain the result y_i. The error between y_i and $\hat{y}_i$ reveals the accuracy of M_i. If the accuracy condition is not satisfied, this new point (x_i, y_i) will be added into S_{i-1} to form S_i. Then a new surrogate model M_i will be established based on S_i. Else if the accuracy condition is satisfied, *2*n* (*n* is the dimension of decision variables) sample points will be generated around the point x_i. Each of the sample points is a little bit different from point x_i in one dimension. After conducting simulations on these sample points, a new surrogate model M_i' is built based on the results obtained, followed by another optimization. If the optimal solution x_i' locates in the box formed by the *2*n* sample points, the calculation process stops. As the kriging model based on the *2*n* sample points satisfies κ-fully linear condition in the box region, this convergence condition means that a local optimal point is obtained. Otherwise, this process will go back into the inner loop until the convergence condition is satisfied.

4. Results and Discussion

There are two decision variables in problem (7). Ten points are generated to build the initial sample set. By following the proposed methods shown in the previous section, the optimization process is shown in Table 2.

Table 2 Optimization process of the iterative surrogate model method

Step	Temp.	Initiator	Conversion			MWD Error		
			Surrogate	CFD	Error	Surrogate	CFD	Error
1	518.959	5.50000E-4	0.0766385	0.0678933	1.28808E-1	7.27822E-4	2.89482E-2	1.03979E+0
2	496.465	4.02772E-4	0.0769198	0.0742393	3.61062E-2	2.49889E-3	1.87446E-2	5.06697E-1
3	493.392	4.07052E-4	0.0756905	0.0759785	3.79039E-3	2.48287E-3	2.89946E-2	6.94137E-1
4	492.646	4.15457E-4	0.0758894	0.0766765	1.02646E-2	2.42709E-3	2.87467E-2	6.96441E-1
5	491.319	4.34871E-4	0.0781145	0.0780700	5.70499E-4	2.49726E-3	2.67297E-2	6.54499E-1
6	490.153	4.53017E-4	0.0786476	0.0793158	8.42449E-3	2.30165E-3	2.48172E-2	6.43336E-1
7	489.210	4.98483E-4	0.0813522	0.0813246	3.39683E-4	2.50237E-3	1.54637E-2	4.36829E-1
8	488.483	5.21489E-4	0.0817789	0.0824330	7.93526E-3	2.48035E-3	1.25993E-2	3.71559E-1
9	487.638	5.50000E-4	0.0833683	0.0837449	4.49735E-3	2.01603E-3	9.63005E-3	3.36809E-1
10	487.408	5.50000E-4	0.0837594	0.0838677	1.29087E-3	6.35053E-4	1.00709E-2	6.01052E-1
11	486.614	5.50000E-4	0.0840585	0.0842906	2.75356E-3	1.13593E-6	1.15827E-2	2.07028E+0
12	485.334	5.50000E-4	0.0846699	0.0849661	3.48642E-3	5.79694E-5	1.35634E-2	1.26854E+0
13	482.308	5.50000E-4	0.0859728	0.0862823	3.58673E-3	2.49721E-3	4.81021E-3	1.22834E-1
14	482.428	5.50000E-4	0.0862480	0.0862678	2.29844E-4	2.49372E-3	1.23038E-3	1.05435E-1
15	482.375	5.50000E-4	0.0862750	0.0862755	5.63399E-6	2.48683E-3	2.45487E-3	2.15263E-3
16	482.373	5.50000E-4	0.0862758	0.0862758	5.61808E-7	2.51138E-3	2.51044E-3	6.23818E-5
16a	482.368	5.50000E-4	0.0862764	0.0862764	3.75222E-7	2.65615E-3	2.65520E-3	5.98688E-5
16b	482.378	5.50000E-4	0.0862751	0.0862752	9.59119E-7	2.37159E-3	2.37074E-3	5.90140E-5
16c	482.373	5.49995E-4	0.0862751	0.0862756	5.36766E-6	2.53740E-3	2.51351E-3	1.58014E-3
16d	482.373	5.50006E-4	0.0862765	0.0862760	5.66551E-6	2.48031E-3	2.50721E-3	1.80136E-3

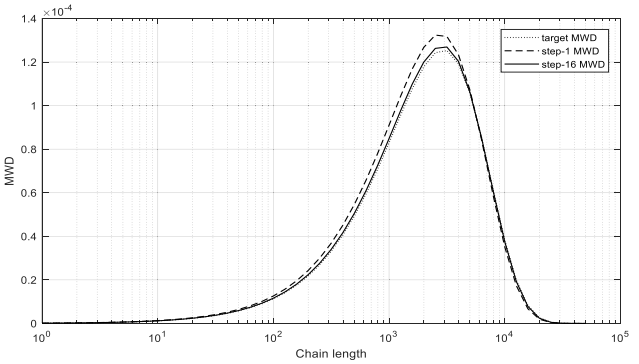


Figure 4. MWD curves of the target, step-1 and step-16 (optimum) points.

From Table 2, the surrogate model satisfies the accuracy condition at the point x_{16}= (482.373,0.00055) after 16 steps, where the absolute errors of τ, β, MWD and conversion are less than 0.001. Then another four points around x_{16} are sampled as listed in Table 2 from 16a to 16d. Each of the four points is slightly different from x_{16}, e.g., x_{16a} = (482.373*(1-1e-5), 5.5e-4). The optimization result of the new surrogate model based on the sample set including the four points is still equal to x_{16} = (482.373,0.00055), indicating that x_{16} is a local optimal point. Together with the initial 10 sample points, the iterative surrogate model method uses only 30 CFD simulations to find the optimal solution. Compared with the results in Zhang et al. (2017), which solved a similar problem using simplex method with 83 simulation runs, this method reduces more than a half calculation burden. As shown in the fifth and eighth columns in Table 2, the conversion is increased from 0.0678933 to 0.0862758, while the MWD error is decreased from 2.89482e-2 to 2.51044e-3. The MWD curves of the first and the last iterations, as well as the target MWD curve, are shown in Fig. 4. The initial MWD (step 1) is quite different from the target MWD, whereas the optimal result (step 16) is very close to the target. According to Wild et al. (2008), surrogate model in the neighbor of x_{16} formed by 4 sample points is satisfied κ-fully linear condition, which ensures local accuracy of function value and derivatives. To further prove the reliability of the

optimization result obtained by the iterative surrogate model method, we sampled another 1000 points and established a global accurate model. Optimization result of the 1000-point surrogate model is (482.327, 0.00055), which is very close to the point x_{16}. The results prove that the propose method, which combines the coarse global model and accurate local surrogate model, is efficient to reduce calculation burden and ensure an optimal solution.

5. Conclusion

The CFD model of LDPE polymerization in the tubular reactor is established to calculate the MWD. Optimization with embedded MWD is proposed for the non-ideal reactor. To overcome the time-consuming and insufficient gradient information problem in calling a CFD simulation, an iterative surrogate model method is proposed to solve the optimization with the CFD model, through which, the optimal solution can be found efficiently and accurately.

Acknowledgements

The authors gratefully acknowledge the financial support of National Key Research and Development Program of China (No. 2017YFE0106700) and NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization (No. U1509209).

References

P.J. Flory, 1953, Principles of Polymer Chemistry, New York: Cornell University Press.

S.N. Lophaven, H.B. Nielsen, J. Søndergaard, 2002, DACE: a Matlab kriging toolbox (version 2), IMM, Informatics and Mathematical Modelling, The Technical University of Denmark.

M.D. McKay, R.J. Beckman, W.J. Conover, 1979, Comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics, 21, 2, 239-245.

S.F. Roudsari, F. Ein-Mozaffari, R. Dhib, 2013, Use of CFD in modeling MMA solution polymerization in a CSTR, Chemical engineering journal, 219, 429-442.

S.M. Wild, R.G. Regis, C.A. Shoemaker, 2008, ORBIT: Optimization by radial basis function interpolation in trust-regions, SIAM Journal on Scientific Computing, 30, 6, 3197-3219.

C. Zhang, Z. Shao, X. Chen, C. Xu, J. Wang, L. Feng, 2017, Simulation and optimization of polymer molecular weight distribution with nonideal reactors, Computers & Chemical Engineering, 106, 744-757.

W. Zhou, E. Marshall, L.O. Shinowo, 2001, Modeling LDPE Tubular and Autoclave Reactors, Industrial & Engineering Chemistry Research, 40, 23, 5533-5542.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50131-2

Hydrodynamic Behaviour of Helical Rings Random Packing Using CFD Simulation

Jia-Lin Kang[a]*,Ya-Cih Ciou[b], Dong-Yang Lin[b], Ching-Hung Cheng[b], David Shan-Hill Wong[b]*, Shi-Shang Jang[b]*

[a] *Department of Chemical and Material Engineering, Tamkang University, New Taipei City, 25137, Taiwan, ROC*
[b] *Department of Chemical Engineering, National Tsing Hua University, Hsinchu 30013, Taiwan, ROC my.mail@mail.server*

Abstract

This study presented computational fluid dynamics (CFD) and experimental validation of hydrodynamic behavior of helical rings random packing. Random packings have been widely used in chemical industries in absorbers, strippers etc. Up till now, development of novel random packing structure has essentially been empirical. Even with the increased computing power, CFD simulations of random packings are hard to find in the literature. The random nature of the packing structures, and the stacked geometry make acceptable grid generation and convergence very difficult because the structure of random packing is complicated when large amounts of it are stacked in a column. In this work, a CFD model was first time developed to simulate countercurrent gas-liquid flow in random packings formed by helical rings. Gravity simulation was used to generate stacking structures. A simple feedback control scheme was applied to control the gas inlet flow rate so that a particular pressure. Multiphase model was employed to compute the gas and liquid interaction in which the surface tension and wall contact angle were found as key factors. The predictions of the CFD model were validated with a lab-scale packed-bed absorber. It was found that the helical structure did increase the interfacial area, liquid hold-up when compared to Raschig rings, and such predictions can be validated by our in-house experiment. The model also showed that helical rings will have lower pressure drop and can sustain a larger liquid-gas ratios compared to Raschig rings . In summary, our study found that CFD simulations can obtain reasonable predictions of hydrodynamic behaviour of packings, and the inclusion of microstructures into a packing element will improve its hydrodynamic properties. These results showed that CFD can be used as a basis for rational packing design.

Keywords: computational fluid dynamics, random packing, helical Rings, Hydrodynamic Behaviour

1. Introduction

Traditional absorbers require huge volume due to the limitation of mass transfer rate. The direct way of intensification process of traditional absorbers is improving the specific surface area utilization of packings. Developing novel packing structures can imporve the mass transfer efficiency to increase gas-liquid contact area. Computational fluid dynamic (CFD) is widely used to investigate hydrodynamic behavior among packings. Klöker et al. presented that using CFD constructed a visual platform of experiment for structured packing.(Klöker et al., 2003; Klöker et al., 2005) The study successfully established the hydraulic and mass transfer correlations of reactive distillation column in

the visual platform. The correlations from the visual platform were validated by real experiments. In order to simplify complicated structure, Petre et al. (2003) and Larachi et al. (2003) presented representative units (REUs) to simulate single phase steam flow for different structure packings, respectively. Raynal et al. (2004) validated that 2D and 3D geometric structure were able to predict liquid holdup and dry pressure. These studies are sufficient to support that using CFD is able to solve complicated issues of flow fields. Generally, traditional absorber columns contain two or more fluids such as gas and liquid fluids. Multiphase model is used to solve the interaction among multi-fluids. (Valluri et al., 2005; CHEN et al., 2009; Haroun et al., 2012) Iliuta and Larachi (2001) developed a one-dimensional two-zone two-fluids mechanistic model for the prediction of structured packing by using multiphase model. The results showed that the multiphase model indeed is able to calculate gas and liquid contact area as well as liquid holdup for structured packing. Haroun et al. (2014) used volume of fluid (VOF) model which is one of multiphase models, to predict the gas and liquid contact area of Mellapak 250X. Simulated results showed that when liquid velocity increased, the gas and liquid contact area increased but liquid holdup decreased. The simulated results had good agreement with experiments. For CFD application for CO_2 capture process, Pham et al. (2015) established a porous media Eulerian CFD model to simplify complicated geometry of structured packing in an absorber as well as to predict mass transfer and pressure drop. The study showed that the model was able to estimate wetted pressure drop, liquid holdup and CO_2 removal efficiency. Based on the aforementioned studies, most studies of CFD focused on regular packing rather than random packing and lacked the investigation of novel structures of packings. The purpose of this study is to present computational fluid dynamics (CFD) simulation and experimental validation of hydrodynamic behavior of helical rings random packing. Using volume of fluid (VOF) method simulates the comprehensive counter-current flow field of randomly stacking of random packing in an absorber. This study evaluates the performance of helical rings random packing by using hydraulic behavior including the gas and liquid contact area liquid holdup, and pressure drop.

2. Methods and Materials

2.1 Process modeling in Fluent®

Fluent®, a commercial CFD simulator, solves the flow field going through the packing. Finite volume element is used for analysing diversity of phenomenon in various flow fields. To solve complicated multiphase fluids, the governing equations include mass conservation equations, momentum conservation equations, and volume of fluid equations. This study is going to compute and investigate gas and liquid contact area and liquid holdup in 3D flow field using VOF for 3 cases of irregular arrangements of random packing. Transient state is used in this study because the multiphase model adopted. To reduce computation time, a supercomputer having 72 cores of Intel ®CPU is used for calculating flow fields. The time step set as 0.0001 s a case took 35 h for integrating 20 sec. of simulation time.

2.2 Volume of Fluid (VOF) Model

This study chooses VOF model to investigate the interface between gas and liquid phases because the gas and liquid are miscible phases. The model going through surface-tracking technique is able to solve out interfaces between two insoluble fluids. Using this model

can clearly define the interface of gas and liquid in the flow field. In a two-phase system as an example, the continuity equations are shown below:

$$\frac{1}{\rho_q}\left[\frac{\partial(\alpha_q\rho_q)}{\partial t} + \nabla\cdot\left(\alpha_q\rho_q\boldsymbol{v}_q\right)\right] = \frac{1}{\rho_q}\left[\dot{S}_q + \dot{m}_{pq} - \dot{m}_{qp}\right] \quad (1)$$

Subscripts p and q indicate two phase indices. α_q is the volume fraction, ρ_q is density, and $\boldsymbol{v}_q$ is velocity vector of phase q. $\dot{S}_q$ is the rate of mass added to phase q. $\dot{m}_{pq}, \dot{m}_{qp}$ are the mass transfer rates to and from phase q to phase p. Continuum surface force (CSF) model was solved to calculation of surface tension force. Wall contact angle of packings represents the material hydrophobic, which set as 70° for plastic packings in this study.

2.3 Helical ring Random Packing Specification and Stacking Method

Helical rings are used to investigate in this case. The height of the ring is around 10 mm, the diameter of rings is 0.6 mm, the interval between the rings is 0.18 mm, and the helical packing diameter is 3.5 mm. To create a structure of irregular arrangement, every Helical ring has to be randomly rotated in 3D space, and then they are randomly stacked in the flow field. In order to prevent making excessively tiny mesh grid, a short distance is required between rings.

2.4 Geometry of Flow Field and Boundary Conditions

A geometry of flow field is created, which is an upright cylinder simplified based on realistic stacking situation of random packing in an absorber, shown in Figure 1. There are 4 circular areas at the top and bottom of the cylinder, respectively. The diameter of the circular areas is 5 mm. The circulars area on the top cylinder is liquid inlet, and the rest area of the top of the cylinder is gas outlet. The circulars area on the bottom is gas inlet and the rest area is liquid outlet. Specifications of the flow field and Helical rings are listed in Table 1. The flow direction of liquid is from top to bottom while the flow direction of gas is from bottom to the top. The inlets and outlets of gas and liquid are set as boundaries. The operating condition for inlet streams is shown in Table 2. The wall boundary is set as "no-slip wall". This study provided three different cases of stacking types in the same flow field.

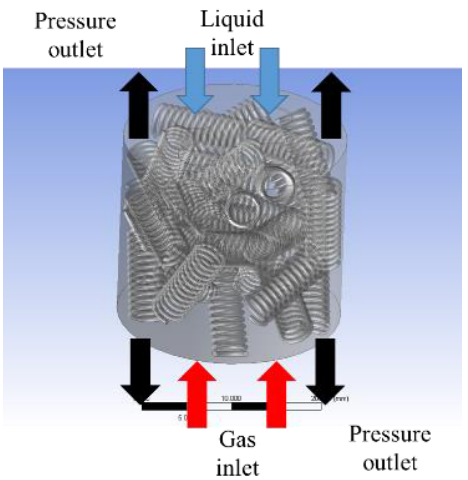


Figure 1 Schematic diagram of helical rings stacking

Table 1 Specifications of flow field and packing

Flow Field	
Diameter	25 mm
Height	22mm
Gas and liquid inlet ports	
Diameter	5 mm

Table 2 Operating conditions

Gas inlet velocity (m/s)	0.1
Liquid inlet velocity (m/s)	0.1, 0.2 ,0.3 ,0.4 , 0.5
Outlet gauge pressure of gas and liquid (pa)	0
Turbulent intensity	1%
hydraulic diameter (m)	0.005

3. Results and Discussions

3.1 Effect of Liquid Flow Rate for Gas-Liquid Contact Area

Figure 2 showed that the agreements of the interfacial area between helix rings and Raschig rings with the experimental data. The simulated interfacial areas for helix rings and Raschig rings were close to the experimental results. The interfacial area of helix rings had a better performance than the interfacial area of Raschig rings. The helix ring experimental data of the interfacial area was hard to measure at low L/G, leading to the larger deviation so that the model cannot match well, but the model can catch the trend of the interfacial area with L/G increasing.

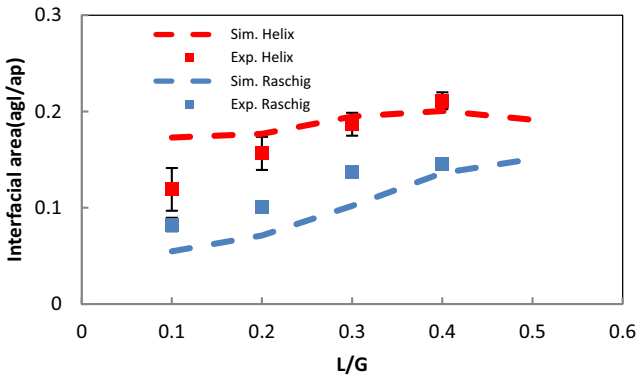


Figure 2 The validation of the interfacial area of helical rings and Raschig rings

3.2 Effect of Liquid Flow Rate for Liquid Hold Up

Figure 3 showed that validations of the liquid holdup between helix rings and Raschig rings with the experimental data. The simulated liquid holdups matched the experimental liquid holdups very well. The liquid hold-up of helix rings was higher than the liquid hold-up of Raschig rings. A slight difference between the helix-simulated result and the helix-experimental data may come from experiment.

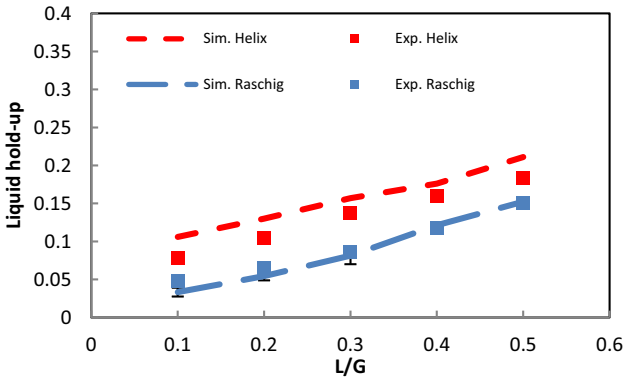


Figure 3 The validation of the liquid holdup of helical rings and Raschig rings

3.3 Effect of Liquid Flow Rate for Pressure Drop

Figure 4 showed that validations of the liquid holdup between helix rings and Raschig rings with the experimental data. The model also was able to predict the pressure drop of helical and Raschig rings. The results showed that the helical rings has a lower pressure drop than the Raschig rings. When the L/G achieved to 0.5, the pressure drop of Raschig rings rapidly increased to flooding but the helical rings still stayed at a lower pressure drop.

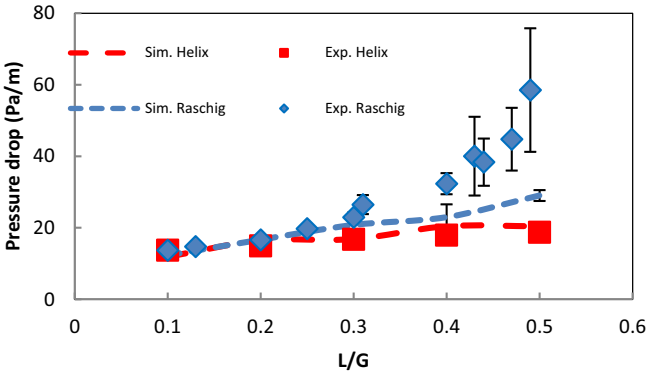


Figure 4 The validation of the pressure drop of helical rings and Raschig rings

4. Conclusions

This study presented CFD simulations and experimental confirmation of the hydrodynamic properties of a helical ring random packing. The results showed that the interfacial area and liquid holdup of helical rings were higher than those of Raschig rings, and the pressure drop of Helical rings is lower than the that of Raschig rings. This study showed that CFD can be implemented to predict hydrodynamics in random packings and serve as a basis for rational packing design and evaluation.

5. Acknowledgement

This research was supported by Grant MOST 106-2221-E-007-093 –MY2 from the Ministry of Science and Technology of the Republic of China.

6. References

J. Chen,C. Liu, X. Yuan, and G. Yu, 2009. CFD simulation of flow and mass transfer in structured packing distillation columns. Chinese Journal of Chemical Engineering 17, 381-388.

Y. Haroun, L. Raynal, and P. Alix, 2014. Prediction of effective area and liquid hold-up in structured packings by CFD. Chemical Engineering Research and Design 92, 2247-2254.

Y. Haroun, L. Raynal, and D. Legendre, 2012. Mass transfer and liquid hold-up determination in structured packing by CFD. Chemical Engineering Science 75, 342-348.

I. Iliuta and F. Larachi, 2001. Mechanistic model for structured-packing-containing columns: irrigated pressure drop, liquid holdup, and packing fractional wetted area. Industrial & engineering chemistry research 40, 5140-5146.

M. Klöker, E. Kenig, and A. Górak, 2003. On the development of new column internals for reactive separations via integration of CFD and process simulation. Catalysis today 79, 479-485.

M. Klöker, E. Y. Kenig, R. Piechota, S. Burghoff, and Y. Egorov, 2005. CFD- based Study on Hydrodynamics and Mass Transfer in Fixed Catalyst Beds. Chemical engineering & technology 28, 31-36.

F.ç. Larachi, C. F. Petre, I. Iliuta, and B. Grandjean, 2003. Tailoring the pressure drop of structured packings through CFD simulations. Chemical Engineering and Processing: Process Intensification 42, 535-541.

C.F. Petre, F. Larachi, I. Iliuta, and B. Grandjean, 2003. Pressure drop through structured packings: Breakdown into the contributing mechanisms by CFD modeling. Chemical Engineering Science 58, 163-177.

D.A. Pham, Y. Lim, H. Jee, E. Ahn, and Y. Jung, 2015. Porous media Eulerian computional fluid dynamics(CFD) model of amine absorber with structured-packing for CO2 removal. Chemical Engineering Science 132, 259-270.

L. Raynal, C. Boyer, and J.P. Ballaguet, 2004. Liquid holdup and pressure drop determination in structured packing with CFD simulations. The Canadian Journal of Chemical Engineering 82, 871-879.

P. Valluri, O.K. Matat, G.F. Hewitt, and M.A. Mendes, 2005. Thin film flow over structured packings at moderate Reynolds numbers. Chemical Engineering Science 60, 1965-1975.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50132-4

Sustainability assessment using local lazy learning: The case of post-combustion CO_2 capture solvents

Gulnara Shavalieva[a*], Athanasios I. Papadopoulos[b], Sara Badr[a], Panos Seferlis[b], Stavros Papadokonstantakis[a]

[a] *Chalmers University of Technology, Department of Space, Earth and Environment, Division of Energy Technology, SE-412 96 Gothenburg, Sweden*
[b] *Chemical Process and Energy Resources Institute, Centre for Research and Technology-Hellas, 57001 Thermi, Greece*
gulnara.shavalieva@chalmers.se

Abstract

The consideration of sustainability is very important for the assessment of life cycle, environmental, health and safety properties of chemicals used in various applications. The screening of wide ranges of molecular structures, prior to the identification of the optimum and most sustainable options, requires the use of efficient and inclusive predictive models. Group contribution (GC) models are popular for the evaluation of numerous molecular options, but they support the calculation of few properties related to sustainability, while their predictive capabilities are often limited by significant data gaps. To address such challenges, we propose the use of a local learning approach as a means of evaluating sustainability properties for a wide range of molecular structures. Supplementing GC methods with data mining ones, such as local lazy learning approaches and exploiting molecular similarities has a potential to improve the predictive capacity of sustainability indices and offers an alternative when GC methods or empirical models are not available for life cycle assessment (LCA) and environmental, health and safety (EHS) hazard assessment indicators. The proposed approach is applied to predict a set of properties (bioaccumulation, persistence and acute aquatic toxicity) of 101 commercial solvents for post-combustion CO_2 capture.

Keywords: Local-lazy learning, Sustainability, Post-combustion CO_2 capture, EHS, data mining

1. Introduction

GC methods have been used very widely for the evaluation and selection of molecular structures with desired properties out of numerous options. Their exploitation in computer-aided molecular design (CAMD) approaches has resulted in numerous prior implementations in the design of solvents, polymers, biofuel additives etc. (Papadopoulos et al., 2016a).

In such implementations, it was realized that it is important to also account for sustainability in the context of environmental, health and safety aspects. Early evaluation of issues such as hazards associated with toxic releases of chemicals could help prevent incidents and minimize harmful impact to human health and environment. One of the main characteristics shared among studies that address the integration of sustainability assessment to automated selection of chemicals is that they focus on only few sustainability indicators. This is because the available GC and/or empirical models only

support few EHS and LCA endpoints. Recent efforts to incorporate sustainability assessment in GC-based CAMD are reported in Ten et. al (2016) who integrates inherent safety and occupational health to the framework. The parameters required for application in CAMD are calculated via property prediction methods available for selected indicators. Another example is the study by Khor et. al (2017) who use a CAMD approach to select sustainable solvents for palm oil extraction. The sustainability indices are obtained by GC and empirical models available for the required parameters. Both studies include safety and health indices, whereas potential impact to the environment is either not assessed or considered to a limited extent. However, a holistic sustainability assessment requires diverse sustainability indicators to achieve a reliable differentiation over a certain, typically large, number of molecules. Data mining methods can be used to analyse, interpret and exploit molecular similarities to supplement GC models to predict sustainability properties, to potentially improve their predictive capabilities and to expand the range of sustainability indicators that can be predicted. Local lazy learning is one such method that was proposed in recent years by Lu et al (2014) to predict the chemical toxicity profile of drug compounds, with encouraging results. To this end, we investigate the implementation of this method in the evaluation of the sustainability performance of numerous, diverse molecular structures used as solvents in CO_2 capture.

2. Methodology

2.1 Motivation and approach

This work is motivated by Papadopoulos et al. (2016b), where CAMD was used for the design of CO_2 capture solvents. The solvent design and selection was based on thermodynamic and reactivity property criteria predicted by GC and empirical models. Sustainability assessment was also performed but only at the post-design stage, where the required information to perform the assessment for few selected solvents was gathered manually from various databases, material data sheets and estimated using available software tools based on Quantitative Structure Activity Relationships. Even though the method provided reliable assessment results, it could not be applied at the design stage hence limiting the evaluated molecular structures.

To estimate safety, health and environmental scores of the metrics, at least one indicator value for dangerous property per aspect should be found for every molecule. An example of dangerous properties and indicators to calculate a score for the environment within EHS is presented in Table 1. To automate the assessment, the dangerous properties should be calculated or estimated by techniques which can be performed computationally and for the desired (wide) range of sustainability indicators. While some of the indicators can be calculated by GC and empirical models, it is troublesome to obtain information for other endpoints which cannot be estimated by such methods. In this work, the local lazy learning (LLL) technique is applied to evaluate its capability in closing data gaps and its potential to be integrated in a CAMD. LLL is a data mining technique which approximates a molecular target function (e.g. a sustainability indicator) locally, based on the known function value of the most similar molecule(s) available in a database (closest neighbours). Closest neighbours can be identified using the Tanimoto coefficient and the extended connectivity fingerprints (ECFPs). ECFPs encode molecular features (e.g. atoms, bonds, stereochemistry) and store them in a bit form. ECFPs are further used to calculate the Tanimoto index to examine similarity between compounds. The most similar compounds have the Tanimoto coefficient close to 1. The aim is to develop LLL

Table 1. Dangerous properties and indicators for Safety score calculation, based on Badr et al. (2017).

Environment	
Dangerous Property	Indicator
Water Mediated Effects	*LC50*aquatic
	Hazard statements
Degradation	Persistency (days)
	Halflife in water (days)
Air Mediated Effects	Chronic toxicity index
Solid waste	Amount of inorganic material
Accumulation	Bioconcentration factor (*BCF*)
	Octanol/water partitioning coefficient (*KOW*)

models and compare their prediction accuracy with GC approaches for three important property indicators mentioned in Table 1 (LC50aquatic, persistence and bioconcentration) using a wide range of molecular structures for which available GC methods can provide predictions.

2.2 Dataset

The comprehensive database of Strempel (2012) was used to create a dataset to generate the models. This data was sorted and only molecules with the existing value for all three indicators of bioaccumulation, persistence and acute toxicity were left in the database. The database was further reduced to a dataset containing compounds consisting of sixteen functional groups relevant to the CO_2 capture case. The selection of the groups was determined by the structure of the molecules identified in the work of Papadopoulos et al. (2016). The influence of functional group selection should be studied further; but it is assumed that such a reduction creates a dataset with more similar compounds that, according to Lu et.al (2014), is needed to enhance the prediction performance of LLL models. This led to a final dataset containing 2106 molecules. Acute toxicity and persistence were converted to logarithmic values.

2.3 LLL Model

For a given compound with an absent indicator value the LLL method identifies *n* nearest neighbours, i.e. most similar molecules in a dataset, and then uses these molecules to predict a property of the compound as follows (Lu et al, 2014):

$$y_{pre} = \sum_{i=1}^{n} \frac{S_i}{\sum_{j=1}^{n} S_j} * y_{i,db} \tag{1}$$

where y_{pre} is the predicted value of the query compound, $y_{i,db}$ is the property value of the *i*-th nearest neighbour from the dataset, S_i is the Tanimoto coefficient indicating the similarity value between the given compound and the *i*-th neighbour, and *n* is the optimized number of nearest neighbours used in the prediction. The Tanimoto coefficient can be calculated as follows:

$$S = \frac{c}{a + b + c} \tag{2}$$

where *c* is the number of atom pairs common for the molecules and *a, b*-the number of molecules' unique atom pairs.

Thus, developing an LLL model consists in finding the optimal value of *n*, over a dataset

and a target property. This optimal value is found by monitoring the LLL model performance over test compounds which are not included in the dataset used to search for the nearest neighbours. In this study this is achieved by leave-one-out cross-validation (LOO-CV). The procedure was applied as follows (Zheng and Tropsha, 2000):
1) One compound (*i*) of the database of *N* compounds was eliminated and the rest of the dataset molecules were used to predict the compound's property (Eq. 3) using *k* (initially set to 1) most similar compounds from the dataset:

$$y_{i,pre,k} = \sum_{i=1}^{k} \frac{S_k}{\sum_{j=1}^{k} S_j} * y_{k,db} \quad (3)$$

where $y_{i,pre,k}$ is the predicted value of the query compound using *k* closest neighbours, $y_{k,db}$ is the property value of the *k*-th nearest neighbour from the database, S_k-the Tanimoto similarity value between the eliminated compound and the *k*-th neighbour, $\sum_{j=1}^{k} S_j$ is the sum of Tanimoto coefficients of the eliminated compound and all its *k* nearest neighbours. The step was repeated for all the compounds in the database (*i*=1 to *N*).
2) Cross-validated Q^2 (Eq.4) was calculated using all the database molecules and their predicted $y_{i,pre}$ and actual values y_i, where y_{avr} is the average property value of all molecules in the dataset.

$$Q^2 = 1 - \frac{\sum(y_i - y_{i,pre})^2}{\sum(y_i - y_{avr})^2} \quad (4)$$

3) Steps 1 and 2 are repeated for *k* in range 1-50. The maximum possible *k* equals to the number of molecules in the dataset, however further increase of the range was unnecessary since the optimal number of neighbours was found to lay in range from 3 to 16 (see Fig. 1). This range remained the same even when the dataset was reduced to 700 and 1400 molecules to analyze the dependence of a dataset size on *k*. The value of *k* which gave the highest Q^2 was set to *n* and used for the final model (Eq. 1).
The optimal *n* number depends on the quality of the database. For instance, a larger database with more similar compounds could possibly result in a lower number of *n*. The obtained *n* number was used to predict properties of solvents for post-combustion CO_2 capture. The assumption here is that a new optimal number of *n* should not be required if this family of molecules has been indirectly included in the initial LLL development (i.e., the initial dataset was obtained by using the same functional groups as those used in the CO_2 capture molecules). Q^2 was calculated in the same way as stated in Eq. (4), and an

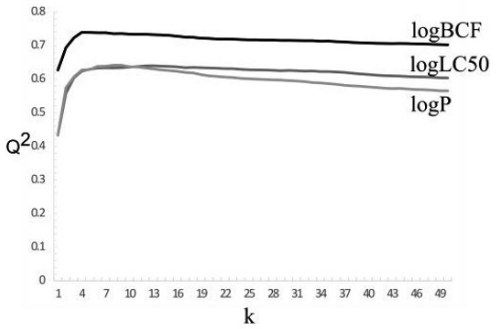


Figure 1. Q^2 vs number of neighbours *k*

average relative prediction error was determined by Eq. (5):

$$RE = \frac{1}{N}\sum_{i}\left|\frac{y_i - y_{i,pre}}{y_i}\right| * 100 \quad (5)$$

3. Results

To demonstrate the performance of the model, the results are presented both for molecules from the dataset of 2106 molecules and for a subset previously identified as potential post-combustion CO_2 capture solvents (Papadopoulos et al., 2016b).

3.1. LLL model performance

The results of LOO-CV showed that the number of closest neighbours required to get the highest Q^2 values are different for every property indicator (Table 2 and Figure 1). After reaching the optimal values, the graphs level up and adding more neighbours into the model does not improve its predictive power. The work carried out by Lu et al. (2014) acquired similar Q^2 values for predicting *LD50* indicator by LLL and their reference library consisting of 2271 compounds. Table 2 also contains Q^2 and *RE* values for the GC+ method as reported by Hukkerikar et al. (2012) for the case of the stepwise approach. It can be seen that published Q^2 numbers for the GC+ method are higher than those obtained by the LLL method for the dataset used.

3.2. Post-combustion CO_2 capture case

The best *n* values found by LOO-CV were used to predict the property indicators for 101 CO_2 capture commercial solvents from Papadopoulos et al. (2016), which all were also present in the dataset of the 2106 molecules. The results were compared to numbers obtained by the GC+ method, as shown in Table 3. The LLL model has a reasonable performance. Higher accuracy might be achieved by adding more molecules that are relevant for CO_2 capture and more structurally similar. Negative Q^2 numbers for the GC+ model show that the model fits the data for CO_2 capture solvents poorly.

Table 2. LLL method results compared to results reported for GC+

	LLL			GC+	
Property indicator	Optimal *n*	Q^2 for the optimal *n* value	Average *RE*	Q^2*	Average *RE**
Log*BCF*	13	0.64	33	0.78	-
Log*LC50*	4	0.74	69	0.82	21
Log *P*	9	0.64	8	-	-

**RE* of Log*BCF* is not given and log *P* is not studied by Hukkerikar, et.al (2012)

Table 3. Results for CO_2 capture solvents obtained by LLL and GC+ methods

	LLL				GC+*			
Property indicator	Q^2 for the optimal n	R^2	Average *RE*	Median *RE*	Q^2	R^2	Average *RE*	Median *RE*
Log*BCF*	0.18	0.55	41	32	-0.71	0.21	71	55
Log*LC50*	0.38	0.75	105	32	-0.17	0.51	150	70
Log *P*	0.68	0.71	9	8	-	-	-	-

4. Discussion and Conclusions

LLL gives different Q^2 values leading to different prediction capacity for particular indicators, which can probably be influenced by the quality of the dataset or character of molecular structure-dangerous property relationship for every indicator. This aspect needs to be studied further to identify which structural fragments are responsible for the highest contribution to further improve the model. An advantage of the LLL method is that it does not require extensive work prior to implementation and can be applied to any set of molecules, whereas GC methods are restricted to molecules that have groups for which contributions have been calculated. Up to now, there is quite a limited data on GC for sustainability assessment indicators which makes it even more difficult to use GC methods for such an analysis. Depending on the number of missing values and the database size, the LLL method can be computationally expensive. This effort can be reduced with a different *n* optimization method.

LLL can be integrated into CAMD to predict information when values are missing, providing an alternative when GC methods or empirical models are not available for some EHS and LCA indices.

References

1. Badr, S., Frutiger, J., Hungerbuehler, K., Papadokonstantakis, S., 2017,A framework for the environmental, health and safety hazard assessment for amine-based post combustion CO_2 capture, Int. J. Greenh. Gas Control, 56.
2. Hukkerikar, A. S., Kalakul, S., Sarup, B., Young, D.M., Sin,G., Gani, R., 2012, Estimation of environment-related properties of chemicals for design of sustainable processes: Development of group-contribution+ (GC+) property models and uncertainty analysis, J. Chem. Inf. Model., 52 (11), 2823-2839.
3. Khor, S. Y., Liam, K. Y., Loh, W. X., Tan, C. Y., Ng, L. Y., Hassim, M. H., Ng, D. K. S., Chemmangattuvalappil, N. G., 2017, 'Computer aided molecular design for alternative sustainable solvent to extract oil from palm pressed fibre', Process Saf. Environ. Prot., 106
4. Lu, J., Peng, J., Wang, J., Shen, Q., Bi, Y., Gong, L., Zheng, M., Luo, X., Zhu, W., Jiang, H. and Chen, K., 2014, 'Estimation of acute oral toxicity in rat using local lazy learning', J Cheminf, 6(1).
5. Papadopoulos, A.I, Linke, P, Seferlis, P. 2016a, Integrated multi-objective molecular and process design: Operational and computational frontiers, in Tools for chemical product design: From consumer products to biomedicine, M. Martin, M. Eden and N. Chemmangattuvalappil (Eds.), Computer Aided Chemical Engineering, 39, Elsevier
6. Papadopoulos, A.I., Badr, S., Chremos, A., Forte, E.,Zarogiannis, T., Seferlis, P., Papadokonstantakis,S., Galindo,A., Jackson, G., and Adjiman, C..S., 2016b, Computer-aided molecular design and selection of CO_2 capture solvents based on thermodynamics, reactivity and sustainability, Mol. Syst. Des. Eng.,1.
7. Strempel, S., Scheringer, M., Carla A. Ng, Hungerbühler, K., 2012, Screening for PBT chemicals among the "Existing" and "New" chemicals of the EU, Environ. Sci. Technol., 46 (11), 5680-5687.
8. Ten, J. Y.; Hassim, M. H.; Chemmangattuvalappil, N., Ng, D. K. S., 2016, A novel chemical product design framework with the integration of safety and health aspects J. Loss Prev. Process Ind., 40.
9. Zheng, W., Tropsha, A., 2000. Novel variable selection quantitative structure-property relationship approach based on the k-nearest-neighbor principle, J. Chem. Inf. Comput. Sci., 40, 185-194.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50133-6

Experimental implementation of dynamic real-time optimization in a graft polymerization reactor

Ryad Bousbia-Salah, François Lesage, Guo-Hua Hu, Abderrazak Latifi

Laboratoire Réactions et Génie des Procédés, CNRS-ENSIC, Université de Lorraine, 1 rue Grandville, BP20451, 54001, Nancy Cedex, France
Abderrazak.Latifi@univ-lorraine.fr

Abstract

This paper deals with experimental implementation of dynamic real-time optimization (D-RTO) in a batch reactor where polymer grafting reactions take place. The objective is to determine the online reactor temperature profile that minimizes the batch period while meeting terminal constraints on the overall conversion rate and grafting efficiency. The optimal temperature profile previously determined in (Bousbia-Salah et al, 2017) was implemented in open loop control and the reactor temperature, the conversion rate and the grafting efficiency were measured. The computed and measured optimal profiles exhibit a quite good agreement and show also that the constraints are satisfied. Moreover, the resulting batch period is reduced by more than 20% with respect to the usual one.

Keywords: Dynamic real-time optimization, Polymer grafting batch reactor, Experimental implementation

1. Introduction

In chemical industry, many plants (especially those that are never in steady-state) deal more and more with uncertainty and changing conditions in their operation. This is due to different factors such as volatile market demands, tougher competition, restrictive environmental standards, changing product specifications, etc.
Their optimal operation is therefore not only a need but also a necessity, and dynamic real-time optimization (D-RTO) is the most suitable technology to succeed (Rawlings et al, 2011; François and Bonvin, 2013; Ellis et al, 2013 and 2014). It makes use of the online available measurements to maximize a process performance index while meeting environmental and operating constraints (e.g. terminal specifications, temperature limits…). The results from the optimization problem can then be sent directly to a control system.
In a previous simulation study (Bousbia-Salah et al, 2017), we considered a batch reactor where polymer grafting reactions take place. The goal was to value the used ground tire rubber (GTR). The latter results from grinding of the rubber part of used tires which retains excellent elasticity. The idea was to take advantage of its elasticity to toughen brittle polymers such as polystyrene upon incorporating GTR into them. D-RTO methodology was applied to the reactor and the objective was to minimize the batch period subjected to some terminal industrial specifications (i.e. conversion rate and grafting efficiency) with the reactor temperature and batch period as decision variables. The simulations showed a good behavior of D-RTO to adapt to parameters mismatch and to generate the optimal temperature trajectory with respect to final constraints and

temperature bounds. A moving horizon observer was successfully implemented to estimate the state of the system.

In the present work, the D-RTO methodology was experimentally implemented within the same batch polymerization reactor. Open-loop and closed-loop D-RTO were implemented and the experimental measurements were compared to the simulation profiles previously computed.

2. Grafting polymerization

When styrene is added to GTR particles, it can be located inside and/or outside them, depending on the styrene/GTR ratio. The latter is kept below 2 so that styrene is completely located inside GTR particles where it polymerizes in two types of polymerization. The first type is the polymerization of styrene itself leading to polystyrene (PS) chains which are not linked to GTR particles. These PS chains are called free PS. The second type is the polymerization of styrene from the rubber chains and the resulting PS chains are attached (grafted) to the rubber ones. These PS chains are designated as GTR-g-PS. The resulting material is GTR particles inside which there are free PS and GTR-g-PS. The main objective is to minimize the amount of free PS in the PS/GTR-g-PS while maximizing the amount of GTR-g-PS in order to allow recycling a maximum of GTR with the best properties, especially the impact strength (Yu, 2015).

This objective is taken into account here through constraints on the grafting efficiency (GE) and the styrene conversion rate (X). GE is defined as the ratio between the amount of grafted PS and that of the total PS (grafted PS + free PS).

3. Formulation of the optimization problem

The dynamic optimization problem of the polymerization reactor considered here may be formulated as

$$\min_{\theta, t_f} J = t_f \tag{1}$$

Subject to

$$\dot{x} = f(x, \theta) \text{ with } x(t_k) = \hat{x}_k \tag{2}$$

$$GE(t_f) \geq GE_f \tag{3}$$

$$X(t_f) \geq X_f \tag{4}$$

$$\theta \leq \theta_{max} \tag{5}$$

where $\theta = \left(T_1, T_2, \dots, T_{n_\theta}\right)^T$ is the vector of time-independent parameters used to approximate the reactor temperature and x the state variables. GE and X are the measured process outputs and GE_f and X_f are their desired final values. $(\theta_{max})_i = T_{max}\,; i = 1,2, \dots, n_\theta$; is the temperature upper bound.
The process model equations (2) consist of 20 ODEs involving 24 unknown kinetic parameters which were previously estimated from experimental measurements (Yu, 2015).

The optimization problem is solved over a shrinking horizon using the control vector parametrization (CVP) method (Goh and Teo, 1988). The latter is based on the approximation of decision variables by means of piece-wise constant functions over the optimization horizon.

The resulting non-linear programming problem is solved by means of a gradient-based method where the gradients are computed through the integration of sensitivities at each iteration of the optimizer. The equations of sensitivities are expressed as

$$\dot{s} = \frac{\partial f}{\partial x} s + \frac{\partial f}{\partial \theta} \text{ with } s(t_k) = 0 \quad (6)$$

where $s(t) = \frac{\partial x(t)}{\partial \theta}$ are the sensitivities of the vector of state variables with respect to the vector of parameters. Note that Eqs.(6) are integrated from t_k to t_f at each iteration of the optimization solver.

The D-RTO proceeds then as follows. Starting from the real process to which the optimal decision variables are applied at a sampling time t_k, the outputs (i.e. GE) as well as the input (i.e. temperature) are used to estimate the state vector $\hat{x}_k$ that will be used as the initial condition for the next optimization at the sampling time t_{k+1}. The corresponding initial condition for the sensitivities is always taken equal to zero. The estimation $\hat{x}$ of x is carried out by means of a moving horizon estimator (MHE) in a closed-loop control (Michalska and Mayne, 1995; Robertson and Lee, 1995) using input and output variables.

4. Experimental rig and measurements

4.1. Experimental rig

The polymerization reactions take place in a stirred batch reactor equipped with a condenser, a temperature control device and a sample collection system. The substances involved in the reactions are GTR particles of 800 μm in diameter, styrene (monomer) and two initiators, i.e. benzoyl peroxide (BPO) and dicumyl peroxide (DCP). Hydroquinone (HQ) and chloroform are used for the measurements of the conversion rate and the grafting efficiency.

4.2. Conversion measurement

The conversion of monomer is measured by a gravimetric method. Samples are taken from the reactor and weighed. A known amount of hydroquinone is added in to stop the polymerization. Then they are put into a Halogen Moisture Analyzer and the temperature is raised to 170°C and kept at this temperature for 10 min. The residual monomer could completely evaporate and the remaining material is composed of free PS, GTR-g-PS and HQ.

4.3. Grafting efficiency measurement

Samples taken from the reactor are put in chloroform to measure the grafting efficiency of the polymerization of styrene inside GTR particles. The free PS and HQ are soluble in chloroform, while GTR and/or GTR-g-PS are not. After at least four days during which the samples are shaken many times, the solvent is removed using filter paper. The free PS and HQ are in the solvent, and the non-soluble portion in the filer paper is GTR or GTR-g-PS.

5. D-RTO implementation

The optimal temperature profiles previously obtained in (Bousbia-Salah et al, 2017) are experimentally implemented in both open loop and closed loop control. However, for the paper length constraint, only open loop control results are presented.

The implementation procedure is as follows. The optimal temperature profile computed for $GE_f = 60\%$ and $X_f = 80\%$ is applied to the reactor using a control temperature device. Since the dynamics of the polymerization reaction is relatively slow, manual temperature control is found to be sufficient to achieve reasonable accuracy. Samples of the reaction medium are taken every 15 minutes and analyzed offline. It is important to note that the conversion rate is measured in 30 minutes whereas the measurement of GE needs much more time and takes about 6 hours.

Figures 1, 2 and 3 show the temperature, monomer conversion and GE profiles, respectively. They compare the computed optimal profiles and the measured ones in the reactor. The temperature profile exhibits a regular increase in order to fulfill the required monomer conversion rate and grafting efficiency. This regular increase is meaningful since at constant temperature, the conversion rate increases with time whereas the grafting efficiency decreases. Therefore the temperature should increase with time in order to achieve the desired conversion rate but not too much in order to guarantee the specified terminal value of grafting efficiency.

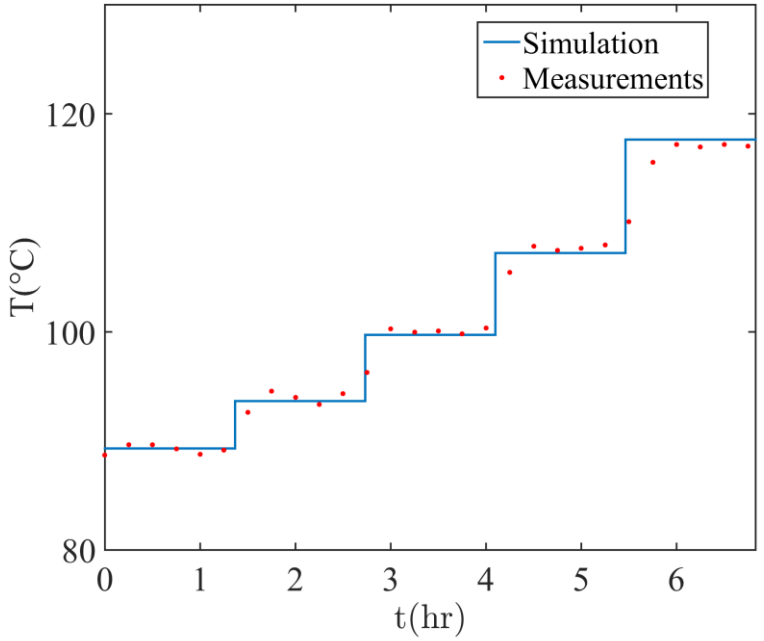


Figure 1: Optimal profiles of reactor temperature for $GE_f = 60\%$ and $X_f = 80\%$.

Figures 2 and 3 show the time-varying process outputs, i.e. conversion rate and grafting efficiency, respectively, with two different profiles. The computed profile given by the integration of the process model equations (Eqs.(2)) using the computed optimal temperature profile, and the measured one. It can be seen that the computed terminal inequality constraints on both process outputs are satisfied and are almost the same as the measured ones. This very good agreement between the computed and measured profiles can be explained by the absence of disturbances and a very good quality of the process model. Furthermore, the resulting batch period is at least 20% less than the usual one

based on two (non-optimized) levels of temperature without terminal constraints on conversion and grafting efficiency (Yu, 2015).

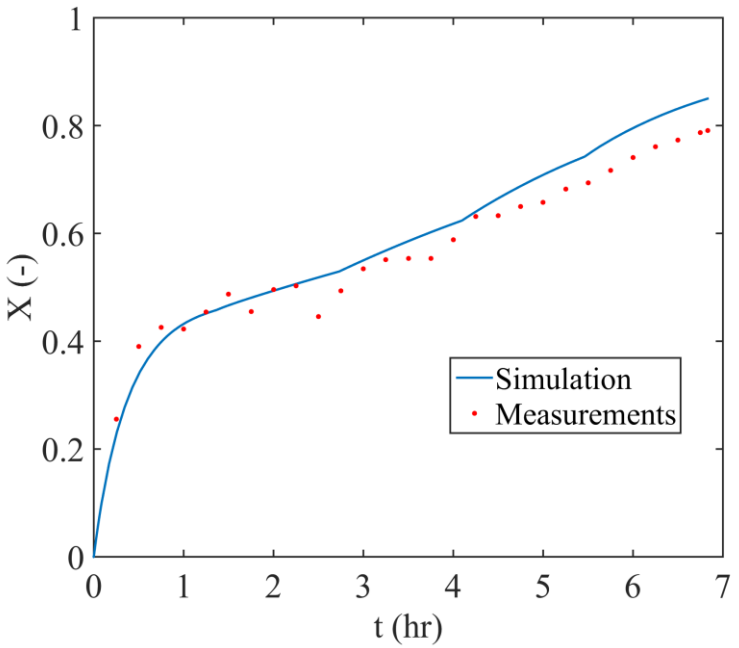


Figure 2: Time-varying profiles of the conversion rate for $GE_f = 60\%$ and X_f =80%.

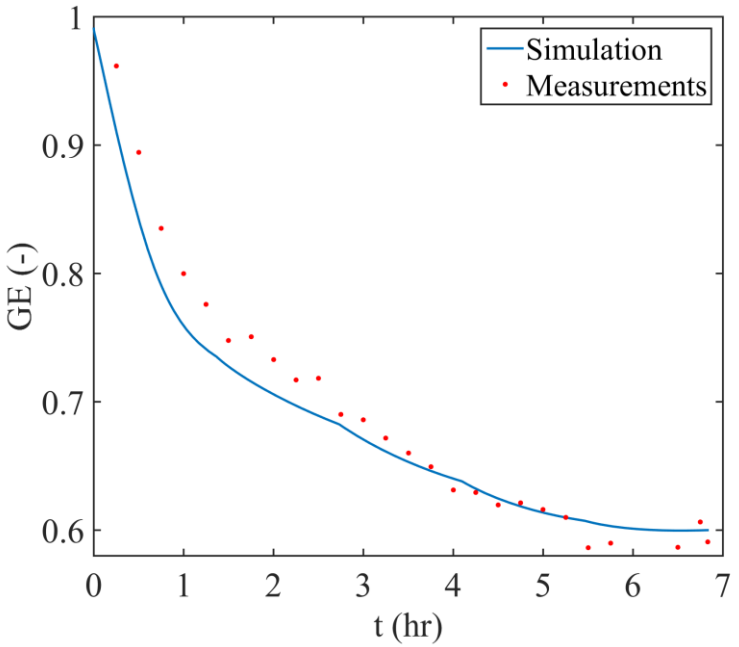


Figure 3: Time-varying profile of the grafting efficiency for $GE_f = 60\%$ and $X_f = 80\%$.

6. Conclusions

The results obtained in this work showed again that real-time optimization is the most suitable technology for online dynamic optimization of the grafting polymerization reactor considered here. It allowed to determine the temperature profile that minimizes the batch period under terminal constraints on monomer conversion and grafting efficiency. The computed profile is then successfully implemented within the reactor in open loop control. The results show very good agreement between the computed and measured profiles of temperature, monomer conversion and grafting efficiency. Moreover, the computed and measured terminal constraints are satisfied and are almost the same. The stability of the computed optimal temperature is guaranteed since the process considered is of the batch type and the shrinking optimization horizon covers the whole operational time. The experimental results obtained in closed loop control will be presented in a paper in preparation.

References

R. Bousbia-Salah, F.Lesage, G.H. Hu, M.A.Latifi, 2017, Dynamic real-time optimization of a batch polymerization process, Computer - Aided Chemical Engineering, 40, 1741 – 1746 , 2017.

M.Ellis, H.Durand, P.D. Christofides, 2014, A tutorial review of economic model predictive control methods, Journal of Process Control, 24, 1156 – 1178.

M.Ellis, M.Heidarinejad, P.D.Christofides, 2013, Economic model predictive of nonlinear singularly perturbedsystems, Journal of Process Control, 23, 743 – 754.

G.Francois, D.Bonvin, 2013, Measurement-Based Real-Time Optimization of Chemical Processes, In: S. Pushpavanam, Editor(s), Advances in Chemical Engineering, Academic Press, 2013, 43, 1-50.

C.J. Goh, K.L. Teo, 1988, Control parametrization: A unified approach to optimal control problems with general constraints, Automatica, 24, 3-18.

H. Michalska, D. Q. Mayne, 1995, Moving horizon observers and observer-based control, IEEE Transactions on Automatic Control},40,995-1006.

J.B.Rawlings, D.Angeli, C.N. Bates, 2012, Fundamentals of economic model predictive control, In Proceedings of the 51st IEEE Conference on Decision and Control, 3851 – 3861, Manui, Hawaii, USA.

D.G. Robertson, J.H. Lee, 1995, A least squares formulation for state estimation, Journal of Process Control, 5, 291-299.

N.Yu, 2015, Étude de la cinétique de polymérisation radicalaire du styrène dans un réseau tridimensionnel et application à la valorisation de pneus usagés, PhD Thesis, Université de Lorraine, Nancy, France.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50134-8

Data-driven Modeling and Optimization of Complex Chemical Processes Using a Novel HDMR Methodology

Qingsong Xie, Hua Liu, Di Bo, Chang He, Ming Pan*

School of Chemical Engineering and Technology / Guangdong Engineering Center for Petrochemical Energy Conservation, *Sun Yat-Sen University*, *Guangzhou 510275, China*
panm5@mail.sysu.edu.cn

Abstract

This paper presents a novel data-driven modeling strategy for highly accurate prediction and optimization of complex chemical processes. The material balance, equilibrium, and heat balance equations addressed in a chemical process are nonlinear, thus making it very difficult to optimize. To overcome the above difficulty, a high dimensional model representation (HDMR) method was developed to represent a complex process (Pan et al., 2016), and a novel linear programming (LP) model was then proposed to find the HDMR parameters. Finally, the complexity of a chemical process model can be reduced significantly as its mechanism formulations were replaced with a simple nonlinear HDMR model. The resulting simple nonlinear optimization problem can be solved efficiently by using the iterative linear programming (LP) method proposed in the earlier work (Pan et al., 2013). To validate the proposed approach, a propane dehydrogenation (PDH) process was studied.

Keywords: Complex Chemical Processes; Data-driven Modeling; High Dimensional Model Representation (HDMR); Highly Accurate Prediction; Optimization

1. Introduction

In general, the trade-off between computational efficiency and accuracy must be considered for process modeling, especially for process optimization. Chemical process models can be divided into mechanism-based models and surrogate models. A large number of unit operations and complex physical-chemical reactions are addressed in the mechanism-based models. It is very difficult to optimize such complex nonlinear programming problems by using conventional optimization methods. Compared to the mechanism models, surrogate models can be seen as “black boxes”. The challenge of using surrogate models is to remain high accuracy of the process performance. There is a growing interest in modeling and optimizing process system using complex mathematical models, such as Polynomial (Draper and Smith, 1981), Kriging (Krige, 1951), Artificial Neural Networks (Himmelblau, 2000), and Support Vector Regression (P. Jain et al., 2007). However, some of them need an expensive CPU computation time, or cannot guarantee the model accuracy because of the limit of their approximation algorithms. Recently, high dimensional model representation (HDMR) method has been proposed as an efficient method for chemical process optimization (Pan et al., 2016). It can also be applied to analyze the economic viability of a chemical process under technical and economic uncertainties (Brownbrige et al., 2014) and global sensitivity analysis (Azadi

et al., 2015). Surrogate modeling can be generally classified into two categories: model order reduction and data-driven modeling (Biegler et al., 2014). Model order reduction aims to lower the computational complexity of a high fidelity model, meanwhile, retains most of the structure of the primitive equations. Data-driven modeling only makes use of the data generated with complex models or experiments.

In this work, we study a complex chemical process of propane dehydrogenation (PDH) by date-driven surrogate models based on a novel high dimensional model representation (HDMR) methodology, and then present the highly accurate prediction of the HDMR model and the optimal solutions obtained for the PDH process. A case study is carried out to demonstrate that the proposed HDMR method is computationally efficient and highly accurate.

2. An HDMR model for describing complex chemical process

In practical modeling work, most of chemical process systems are complex to be coded. There are many specialised simulation software packages having been used in the existing studies (e.g., Aspen Plus^{TM}, PRO/II, gPROMS). However, in most situation, these simulation software packages appear to the users as black boxes. To tackle the complexity of building very detailed models based on the explicit knowledge of the physical behaviour of the system, surrogate models are adopted to find the connections between the system state variables (input, internal and output variables). Data generated with applying commercial software packages (e.g., Aspen Plus was applied in this work) or experiments are used to construct simpler but accurate models which include empirical approximations describing the relation between input variables and response values of a process.

In order to build the surrogate models for representing complex chemical processes, HDMR method is used to generate surrogate models which have been demonstrated as the most efficient and actual models in industrial applications (Pan et al., 2016). Moreover, the HDMR method can take into account the inherent uncertainties in input parameters, and also presents the potential nonlinearities and contributions due to the interactions between input parameters. The main feature of HDMR is shown as follows:

$$y = f_0 + \sum_{i=1}^{N} f_i(x_i) + \sum_{i=1}^{N}\sum_{j=i+1}^{N} f_{ij}(x_i, x_j) + \cdots + f_{12\ldots N}(x_1, x_2, \cdots x_N) \tag{1}$$

Where N is the number of input parameters, i and j index the input parameters, and f_0 is the mean value of $f(x)$.

The expansion given in Eq. (1) has a finite number of terms and $f(x)$, however for most practical applications, the functions containing more than two input parameters can often be ignored due to their less contributions compared to the former terms (Rabitz and Als, 1999, Li et al., 2002). Therefore, Eq. (1) can be simplified as (Pan et al., 2016):

$$\begin{aligned} &y = C + \sum_{i=1}^{N}\sum_{k=1}^{K} A_{i,k} \times x_i^k + \sum_{i=1}^{N}\sum_{j=i+1}^{N}\sum_{k=1}^{K}\sum_{n=1}^{K} B_{i,j,k,n} \times x_i^k \times x_j^n \\ &i, j \in \{1, 2, 3, \cdots, N\} \qquad k, n \in \{1, 2, 3, \cdots, K\} \end{aligned} \tag{2}$$

Where C is a constant term, $A_{i,k}$ and $B_{i,j,k,n}$ are the first and second order coefficients, K is the highest degree of input variables , subscript i and j denote the input parameters, and y is function value.

In order to generate accurate surrogate models, the HDMR parameter values of C, $A_{i,k}$ and $B_{i,j,k,n}$ are found by using a novel linear programming (LP) model in GAMS. The objective function of the LP model is shown as follows:

$$\min \sum_{d}^{D} \left| y_d - y_d^* \right|, \quad d \in \{1,2,3,\cdots,D\} \tag{3}$$

In addition, there is a constraint of model maximum allowable error:

$$\left| y_d - y_d^* \right| \le \delta, \quad d \in \{1,2,3,\cdots,D\} \tag{4}$$

D represents the number of data sets, y_d is model calculated value, y_d^* is actual value, δ is the maximum allowable error.

3. Case study

3.1. Process simulation (Aspen Plus™)

To validate the proposed approach, a propane dehydrogenation (PDH) process is simulated by using Aspen Plus software. Figure 1 presents the detailed flowsheet of PDH process developed in Aspen Plus simulation environment. The process feed, LPG, is pre-processed to make propane volume fraction meet the feed requirements for reaction at first, then is reacted with H_2 in a moving bed reactor with four parallel stages and radial flow. The reaction product enters refining section after the rapid cooling, high pressure dehydration and cryogenic dehydrogenation. In the refining section, light end (C1 and C2) are removed from an dethanizer column, and then the propylene and propane are separated via a propylene rectification column. Propylene is sold as a product, while propane enters the reactor as a cyclic feed. To demonstrate the validity of our Aspen Plus models, the Aspen Plus results are compared with the literature data, and the errors of the propane conversion, propylene selectivity, and propylene yield are 0.271%, 0.127% and 0.063%, respectively.

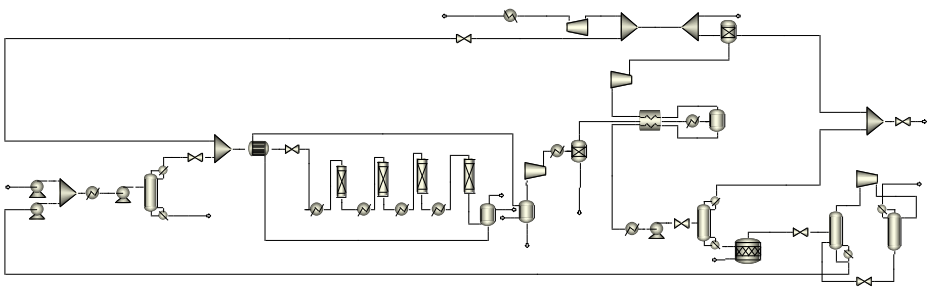


Figure 1. Flowsheet of the PDH process built in Aspen Plus™.

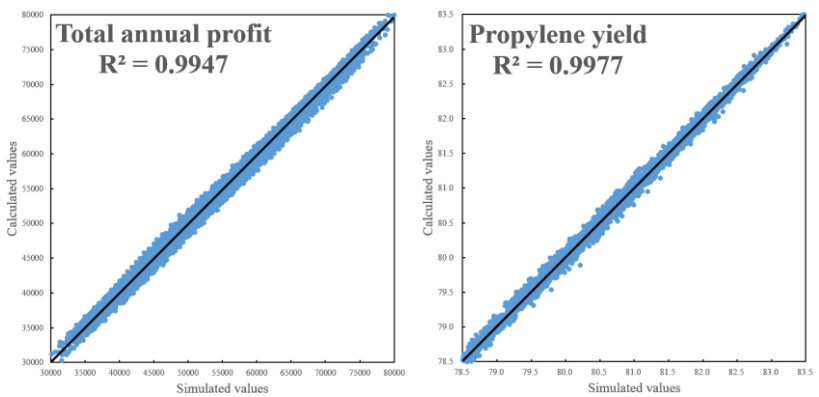
3.2. Process modeling and prediction

The key equipment of PDH process includes pumps, heat exchangers, moving bed reactors, compressors and rectifying columns. For the HDMR modeling of PDH process, the total annual profit and the propylene yield are selected as the outputs, and the five major operating parameters affecting the outputs are set as the inputs, including: the feed flow of LPG, the flow ratio of hydrogen-to-propane, the temperature and the pressure of reactor, and the pressure of propylene rectifier. The simulation data are obtained with varying the five input parameters. The ranges of these parameters are shown in Table 1.

Table 1. Ranges of input parameters for surrogate modeling of the PDH process.

Parameters	Lower bound	Upper bound
x_1: LPG feed flow (kmol/h)	560	840
x_2: Hydrogen to propane ratio	1.60	2.40
x_3: Reactor temperature (℃)	570	610
x_4: Reactor pressure (MPa)	0.18	0.20
x_5: Propylene rectifier pressure (MPa)	0.65	1.00

The HDMR model parameters of the PDH process are obtained by feeding the 8750 groups of simulated results to the proposed LP model, which is performed on a desktop PC with an Intel Core i7 (2.4 GHz) and 16 GB RAM using GAMS 24.4. The CPU time of model training of the total annual profit and propylene yield are 134 and 18 seconds, which are much less than kriging metamodel, in which the time consumption increases exponentially as the data volume picks up. For HDMR model training, the R^2 values of the total annual profit model and propylene yield model are 0.9947 and 0.9977 respectively, which presents a high goodness of fit between the HDMR and Aspen Plus models, and it also can be seen from Figure 2, most of data points of total annual profit or propylene yield are nearby diagonal line.

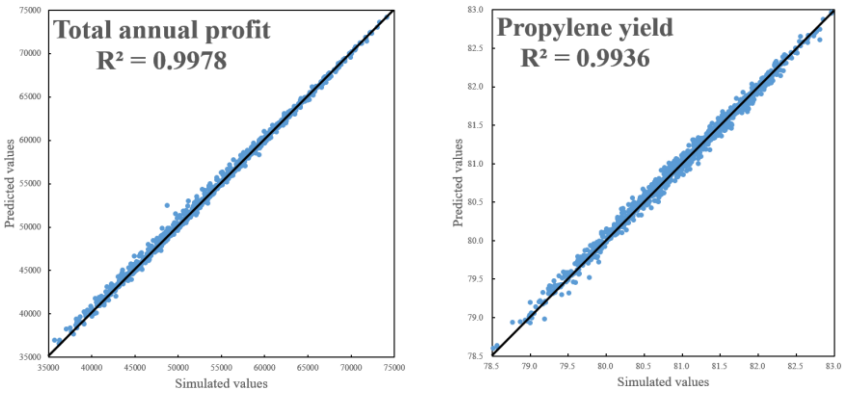


(a) Total annual profit HDMR model (b) Propylene yield HDMR model

Figure 2. Calculated versus simulated total annual profit/propylene yield values

For model prediction, 1024 new groups of simulation data are tested. Highly accurate prediction of the PDH process is shown in Figure 3, and the R^2 values of the HDMR

model prediction of total annual profit and propylene yield are 0.9978 and 0.9936 respectively.



(a) Total annual profit HDMR model (b) Propylene yield HDMR model

Figure 3. Predicted versus simulated total annual profit/propylene yield values

3.3. Process optimization

Regarding the process operation optimization, the resulting simple nonlinear optimization problem can be solved efficiently by using the iterative linear programming based method proposed in the earlier work (Pan et al., 2013). The best operating solutions are obtained (Table 2): (1) the total annual profit reaches the maximum 120.98 million dollars when the feed flow of LPG is 840 kmol/h, the flow ratio of hydrogen-to-propane is 1.6, the reactor's temperature is 610 ℃, the pressure of reactor is 0.18 MPa, and the pressure of propylene rectifier is 0.9727 MPa; (2) the propylene yield reaches the maximum 83.56% when the feed flow of LPG is 840 kmol/h, the flow ratio of hydrogen-to-propane is 1.6, the reactor's temperature is 610 ℃, the pressure of reactor is 0.18 MPa, and the pressure of propylene rectifier is 0.6665 MPa. Our optimal results are identical to the Aspen Plus simulation results under the same input parameters.

Table 2. Optimal solutions in two scenarios (maximisation of total annual profit and propylene yield) based on the obtained surrogate models for PDH process.

	HDMR models						Aspen Plus results	Error (%)
	x_1	x_2	x_3	x_4	x_5	Objective value		
Max total annual profit (M$/year)	840	1.6	610	0.18	0.9727	120.98	120.70	0.16
Max propylene yield (%)	840	1.6	610	0.18	0.6665	83.56	83.54	0.02

x_1: LPG feed flow (kmol/h); x_2: hydrogen to propane ratio; x_3: reactor temperature (℃); x_4: reactor pressure (MPa); x_5: propylene rectifier pressure (MPa).

4. Conclusion

This paper investigates data based modeling of complex nonlinear processes, a novel data-driven modeling strategy based on high dimensional model representation (HDMR) method is proposed to represent a complex chemical process of propane dehydrogenation,

and a novel linear programming model is then used to find the HDMR parameters. Finally, the complexity of a chemical process model can be reduced significantly as its mechanism formulations are replaced with a simple nonlinear HDMR model and the resulting simple nonlinear optimization problem can be solved efficiently by using an iterative linear programming (LP) method. According to the results of case study, the proposed HDMR method can achieve a highly accurate prediction and optimization of the PDH process.

Acknowledgements

Financial support from 100 Top Talents Program of Sun Yat-Sen University, the National Natural Science Foundation of China (No. U1462113) and the Science and Technology Planning Project of Guangzhou city (No. 201704030136) are gratefully acknowledged.

References

P. Azadi, G. Brownbridge, S. Mosbach, O.R. Inderwildi, and M. Kraft, 2015, Simulation and life cycle assessment of algae gasification process in dual fluidized bed gasifiers, Green Chem 2015, 17(3):1793–801.

G. Brownbridge, P. Azadi, A. Smallbone, A. Bhave, B. Taylor, and M. Kraft, 2014, The future viability of algae-derived biodiesel under economic and technical uncertainties. Bioresource Technol, 151:166–73

L.T. Biegler, Y. Lang, and W. Lin, 2014, Multi-Scale Optimization for Process Systems Engineering. Comput Chem Eng, 60(10):17–30.

N.R. Draper and H. Smith, 1981, Applied Regression Analysis, New York: Wiley.

D.M. Himmelblau, 2000, Applications of artificial neural networks in chemical engineering. Korean J Chem Eng, 17(4):373–392.

P. Jain, I. Rahman, and B.D. Kulkarni, 2007, Development of a Soft Sensor for a Batch Distillation Column Using Support Vector Regression Techniques, Chem.Eng.Res.Des., 85, 283–287.

D.G. Krige, 1951, A Statistical Approach to Some Mine Valuations and Allied Problems at the Witwatersrand, South Africa: University of Witwatersrand.

M. Pan, I. Bulatov, and R. Smith, 2013, New MILP-based iterative approach for retrofitting heat exchanger networks with conventional network structure modifications, Chemical Engineering Science, 104: 498–524.

M. Pan, J. Sikorski, J. Akroyd, S. Mosbach, R. Lau, and M. Kraft, 2016, Design technologies for multi-level innovation and integration in eco-industrial parks: from unit operations to processes, plants and industrial networks, Applied Energy, 175, 305–323.

H. Rabitz, and O.F. Als, 1999, General foundations of high-dimensional model representations. J Math Chem, 25:197–233.

G. Li, S.W. Wang, and H. Rabitz, 2002, Practical approaches to construct RS-HDMR component functions. J Phys Chem A, 106:8721–33.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50135-X

Modeling Ethylene Cracking Process by Learning Convolutional Neural Networks

Feng Hua, Zhou Fang, Tong Qiu*

Department of Chemical Engineering, Tsinghua University, Beijing 100084, China
qiutong@tsinghua.edu.cn

Abstract

Naphtha pyrolysis is the major source of ethylene and other olefin feedstock for petrochemical industry. Due to the complex reaction kinetics and mass transfer process, the traditional kinetic models are often computationally expensive. Neural nets have been used to overcome the difficulty, but the so-called black-box models ignore reaction mechanism in the training process. They typically need a large training set and often perform poorly when the test data is out of the training range. To improve this, the article puts forward a model that uses the topology of the reaction network to guide the construction of the neural network. After undertaking a graph theoretical analysis of the reaction network, we find that similar to images, the architecture of the naphtha pyrolysis network has modular features, which indicates its deep learning might use the convolutional neural networks (CNNs) to extract the features of its topology. Here, a novel model of naphtha pyrolysis is built based on CNN, which reduces the computational cost of previous kinetic models by up to 700 times. Since the CNN architecture learns the network representations, the proposed model generalizes very well and predicts the yields of key products with high accuracy. The outlined framework is not specific to ethylene cracking process and can be applied to other similar chemical reaction network problems.

Keywords: convolutional neural network; chemical reaction network; naphtha pyrolysis

1. Introduction

The modeling of ethylene cracking process is of high importance due to growing interest in olefins as essential feedstock in the petrochemical industry (Zhang et al., 2012). Recently, comprehensive kinetic models are used in the pyrolysis of hydrocarbons to predict the product yields and other information (Green et al., 2007). Researchers have developed a model to simulate the steam cracking process (Van et al., 2002). Its kinetic scheme, having progressed across the years (Ranzi et al.,2005) is based on equations from the standpoint of free radicals, which enables a more detailed description in the heavy feedstock components. To reduce the network scale, it is asumed that in the reaction process, radicals of C5 and above conform to the quasi-steady-state approximate assumption (QSSA) and can be directly removed (Ranzi et al.,2001), thereby greatly reducing the amount of components in the reaction network. As a result, computational complexity of the lumped model is greatly reduced. For the purpose of moving beyond single particles and modeling the entire cracking reactor, the chemical kinetic mechanisms need to be coupled with the simulation of furnace. the three dimensional heat radiation model of the furnace is established, and the corresponding flow patterns is simulated (Detemmerman et al.,2006). The impact of the flue gas flow is further analyzed for the thermal cracking of hydrocarbons. The detailed pyrolysis network is combined

with the heat transfer in furnace simulation (Niaei et al.,2004). The results show good agreement with experimental data in industrial cracking furnaces.

Nevertheless, such models describe the chemical kinetic mechanisms by conservation equations for mass, heat and momentum transport. Numerical solutions to those models, especially in three-dimensions, are computationally expensive.

After undertaking a graph theoretical analysis (Wagner et al., 2001) of the previous reaction network of naphtha pyrolysis, we find that it has evident modular features. The high modularity can be considered as the result of the existence of local communities in the network. In this paper, a framework based on convolutional neural network (CNN) is used to extract the community structure of the naphtha pyrolysis network.

In the last few years, CNNs have been extensively studied and developed by researchers and has led to good performance on a variety of problems, especially in the field of computer vision (Fukushima et al., 1982; LeCun et al., 1998;2015). CNN is a deep learning model inspired by the visual cortex of living creatures. Its architecture usually consists of five parts: an input layer, convolutional layers, pooling layers, fully connected layers and an output layer. For image-based CNNs, the input is the matrix of image pixels. Convolutional layers are the layers where convolutional kernels extract locally connected regions from the input. Each convolutional layer usually contains several kernels, making sure that different patterns can be filtered. Pooling layers aim to the resolution of the feature maps while retaining important information. Fully connected layers perform high-level reasoning, take all data in the previous layer to generate output results.

Currently, CNNs are mainly applied in image-based problems. Distinct from images whose pixels have an implicit spatial order, the connection pattern of the nodes in networks is quite irregular. Furthermore, large-scale chemical reaction networks usually involve large amounts of components and reactions. Thus, chemical networks are more difficult to be described. The major challenge relies in extending the concepts of CNNs to chemical reaction networks. In the paper, a deep-learning model of ethylene cracking is developed by learning from convolutional neural networks.

2. Graph theoretical analysis of the naphtha pyrolysis network

2.1. Graph Construction

Based on publicly available information of naphtha pyrolysis (Zhang et al., 2012), we assemble a list of 4694 reactions with 142 components. Then, a substrate graph (Wagner et al, 2008) is constructed as a representation derived from the stoichiometric equations and is visualized in Fig. 1.

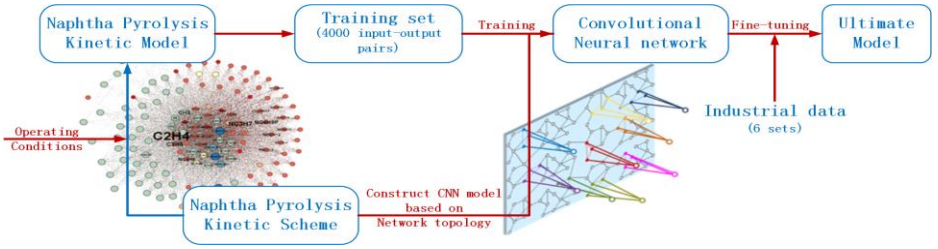


Figure 1: Framework of the modelling process

Figure 1 also gives an overall framework of the modelling process. The substrate graph is then used to guide the construction of the neural net. Its local features are extracted by kernels in the convolutional structure. Each kernel produces 4694 values of output, representing 4694 reactions in the kinetic scheme. The number of outputs is decreased to 142 in the max-pooling layer, representing 142 components. Since the industrial data (8 samples) is not sufficient, the training set (containing about 4000 samples) is generated by the previous naphtha pyrolysis kinetic model. 6 out of 8 of the industrial samples are chosen to fine-tune the model, while the other 2 to test the model.

2.2. Modularity of the network

The modularity of the substrate graph is measured using the community detection method (Blondel et al., 2008). The modularity is defined as a scalar value between -1 and 1 that measures the density of links inside the community compared to links between communities. If the number of inside-community edges is the same as random, we will get Q=0. Q increases with stronger community structure. Through analyse, the modularity of the substrate graph is 0.194. The high modularity can be considered as the result of the similarity of reactions. Reactions of the same mechanism connect the relevant components in the same pattern, resulting in the clustering 'cliques' in this network. The wide existence of communities suggests that the local features of the network topology might be extracted by learning CNN architecture.

3. The neural network architecture

One common type of artificial neural network models is the so-called "black-box" models, where the prior knowledge of the chemical mechanism is completely neglected. This kind of pure data-driven model can simulate much faster than traditional kinetic models and with high predictive accuracy given a large training set. However, even with sufficient training data, the model might still have a poor performance when the input data is out the range of the training set.

In order to improve the ductility of the neural network, we make use of the prior knowledge of the reaction network to guide the design of the neural network architecture. By combining the structural features of naphtha pyrolysis network with the designed neural net, we transform the data-driven model to a hybrid model of both mechanism and data.

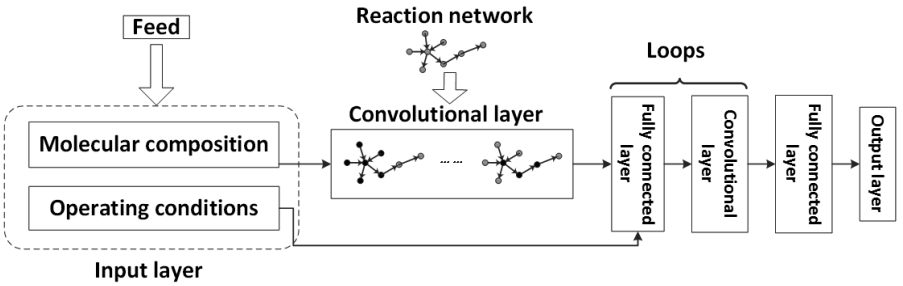


Figure 2: A schematic view of the proposed CNN architecture

Our proposed CNN architecture is shown in Fig.2. The input consists of detailed molecular composition of feedstock and operating conditions. The reaction network is embedded in the convolution layer where the features of the network is learned. The output of the convolutional layer goes into loops. Here, we set 2 loops in the structure.

The output layer exports the yields of 9 key products: H2, CH4, C2H4, C2H6, C3H6, C3H8, C4H6, NC4H8 and IC4H8.

3.1. Convolutional and pooling layer

The design of convolutional layer is presented in Fig3. Each convolutional layer consists of 16 kernels. Each kernel contains 4694 sets of parameters representing 4694 reactions and each set of parameters operates on the relevant feedstock components. After activation function, the convolutional layer gives out a 16*4694 matrix. The max pooling operation is equivalent to summarizing the reaction information by components. The output of the max-pooling layer is 142 values, representing 142 components in the reaction network.

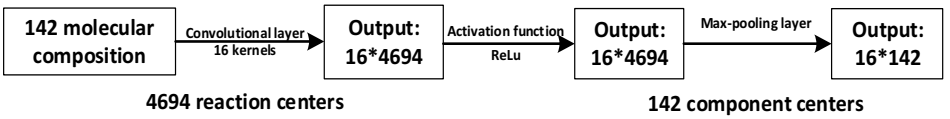


Figure 3: A schematic view of the convolutional and max-pooling layer

3.2. Loop

Each loop consists of a fully connected layer, a convolutional and a pooling layer. Other than the output of the previous layer, the input of the first fully connected layer involves 5 operating conditions including coil inlet temperature (CIT), coil outlet temperature (COT), coil inlet pressure (CIP), federate and water/oil ratio. Thus, the number of input is 16*142+5=2277. We set the number of neurons in the fully-connected layer 142. Each neuron in the fully-connected layer adds the weighted input together plus a certain bias. The result is then transferred by certain activation function and delivered to the next layer. From the mathematic point of view, the operation can be regarded as a row vector (1*2277) multiplied by a matrix of 2277*142 and added to a column vector of 1*142. Thus, the output is decreased to 142, in line with the input of the next layer. A schematic view of the loop is given in Fig4.

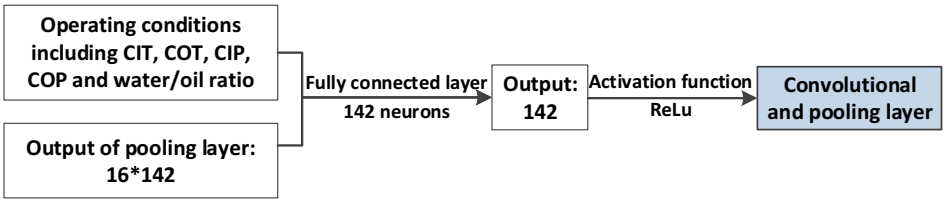


Figure 4: A schematic view of the loop

4. Case study

An industrial case of naphtha pyrolysis is used to testify the proposed CNN model (Fang et al., 2017). Since the experimental data is not enough to form a sufficiently large training set, we train the CNN model by two steps. In the first step, based on the published industrial data of naphtha pyrolysis in the cracking reactor KBR SC-1, we generate 4000 sets of input data by adding a random perturbation and the kinetic model gives out the corresponding output data. The range of input variation is around ±10%. The training set range is thus relatively small because the working conditions don't generally change much for the same furnace in the same factory. In order to apply the neural network to a larger scope, more data will be needed. However, only one cracking reactor is studied

here. The initial network is trained with these input-output pairs. In the second step, we take 6 out of 8 sets of the real industrial data and use them to further fine-tune the neural network.

The ultimate model is used to predict the product yields of three input sets. The operating conditions of the three sets are shown in Table 1. Among them, the first sample is in the training set, while the second and third ones are new to the model.

Table 1: Simulation output of the CNN model

		Sample1 (Trained)		Sample2 (Untrained)		Sample3 (Untrained)	
Feedrate	t/h	6.5686		6.2355		6.2476	
Water/oil ratio	t/t	0.4958		0.5059		0.4837	
CIT	℃	645.68		662.82		617.15	
COT	℃	888.16		868.29		838.83	
CIP	Mpa	0.292		0.294		0.303	
		Industrial	Simulation	Industrial	Simulation	Industrial	Simulation
H2	wt%	1.00	0.96	1.03	0.96	1.06	0.96
CH4	wt%	16.28	16.64	15.88	16.54	17.10	16.80
C2H4	wt%	30.03	30.37	29.66	30.10	30.72	30.26
C2H6	wt%	3.87	3.89	3.79	3.89	3.67	3.89
C3H6	wt%	16.91	16.89	16.27	16.78	16.41	16.84
C3H8	wt%	0.78	0.81	0.72	0.81	0.76	0.81
C4H6	wt%	6.21	6.23	6.14	6.23	5.78	6.23
NC4H8	wt%	1.18	1.16	1.14	1.16	1.03	1.15
IC4H8	wt%	2.78	2.74	2.72	2.74	2.65	2.74
Relative Error Average		1.73%		3.71%		5.55%	

The results show that after the fine-tuning stage, the model can predict the product yields with excellent accuracy. All of the three samples show good consistency with the industrial data. Among them, the trained sample performs better than the untrained ones, with the average relative error of the former equals to 1.73% and that of the latter ones around 5%. Though the phenomenon is quite common for the neural network models and the error is still acceptable for industrial use, it should be mentioned that the application of this model need to be retrained and rebuilt when the input has large differences with the trained ones. To assess the reduction of computational cost compared with the kinetic model, we run 1000 calls of both models and count the average execution time. Each call of the kinetic model takes 31.112s while the CNN model takes 0.044s, which demonstrates the neural set can reduce the computational cost by up to 700 times.

5. Conclusion

In this paper, we build a deep learning model of ethylene cracking process by learning from convolutional neural networks. Different from traditional neural nets that are used as black boxes, the proposed model uses the kinetic scheme to guide the construction of the neural network. Its architecture reads the detailed molecular compositions as well as the operating conditions. The convolutional layers and pooling layers in the network are designed based on the features of the reaction network. As demonstrated above, the CNN model predicts the product yields with excellent agreement with the values in a real industrial case and is about 700 times faster than the previous kinetic model.

In future work, since the performance of neural networks depends largely on the range of training set, industrial data of various furnaces with various working conditions should be added to the training set. Besides, we plan to develop the model to predict the temporal or periodical changes of the product yields. A logical next step is to include time as an additional input of the model. This development will lead to large file size of the training data, which demands the increase of the network depth and width. To overcome this obstacle, various optimization techniques need to be applied to construct deep CNN architectures.

Acknowledgements

The authors gratefully acknowledge the National Natural Science Foundation of China for its financial support (Grant No. U1462206).

References

V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, 2008, Fast Unfolding of Communities in Large Networks, *Journal of Statistical Mechanics: Theory and Experiment* 2008 (10):P10008.

Y.-L. Boureau, J. Ponce, and Y. LeCun, 2010, A Theoretical Analysis of Feature Pooling in Visual Recognition, *In Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 111–118.

Z. Fang, T. Qiu, and W. Zhou, 2017, Coupled Simulation of Recirculation Zonal Firebox Model and Detailed Kinetic Reactor Model in an Industrial Ethylene Cracking Furnace, *Chinese Journal of Chemical Engineering*.

K. Fukushima and S. Miyake, 1982. Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Visual Pattern Recognition, *In Competition and Cooperation in Neural Nets*, 267–285. Springer.

W. H. Green, 2007, Predictive Kinetics: A New Approach for the 21st Century, *Advances in Chemical Engineering* 32:1–313.

Y. LeCun, Y. Bengio, and G. Hinton, 2015, Deep Learning, *Nature* 521 (7553):436–444.

Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, 1998, Gradient-Based Learning Applied to Document Recognition, *In IEEE*, 2278–2324.

Van Goethem, MWM, FI Kleinendorst, N van Velzen, M Dente, and E Ranzi, 2002, Equation Based Spyro® Model and Optimiser for the Modelling of the Steam Cracking Process, *Escape-12 Supplementary Proceedings*, 26–29.

A. Wagner and D. A. Fell, 2001, The Small World inside Large Metabolic Networks, *Proceedings of the Royal Society of London B: Biological Sciences* 268 (1478):1803–1810.

L. Zhang, B. Chen, and T. Qiu, 2012, A Multi-Scale Model of Naphtha Pyrolysis Process, *In 11th International Symposium on Process Systems Engineering-PSE2012*, 15:120. Elsevier.

E. Ranzi, A. Frassoldati, G. Silvia, T. Faravelli, 2005, Wide-Range Kinetic Modeling Study of the Pyrolysis, Partial Oxidation, and Combustion of Heavy n-Alkanes, *Industrial & Engineering Chemistry Research,* 44: 5170–5183.

E. Ranzi, M. Dente, A. Goldaniga, G. Bozzano, T. Faravelli, 2001, Lumping procedures in detailed kinetic modeling of gasification, pyrolysis, partial oxidation and combustion of hydrocarbon mixtures, *Progress in Energy and Combustion Science*, 27, 99–139.

T. Detemmerman, G. F. Froment, 2006, Three Dimensional Coupled Simulation of Furnaces and Reactor Tubes for the Thermal Cracking of Hydrocarbons, *Oil & Gas Science & Technology*, 53:81–194.

A. Niaei, J. Towfighi, S. M. Sadrameli, R. Karimzadeh, 2004, The combined simulation of heat transfer and pyrolysis reactions in industrial cracking furnaces, *Applied Thermal Engineering*, 24: 2251–2265.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50136-1

Design of Experiments for Model Discrimination using Gaussian Process Surrogate Models

Simon Olofsson, Marc Peter Deisenroth and Ruth Misener*

Department of Computing, Imperial College London, SW7 2AZ, UK
r.misener@imperial.ac.uk

Abstract

Given rival mathematical models and an initial experimental data set, optimal design of experiments for model discrimination discards inaccurate models. Model discrimination is fundamentally about finding out how systems work. Not knowing how a particular system works, or having several rivalling models to predict the behaviour of the system, makes controlling and optimising the system more difficult. The most common way to perform model discrimination is by maximising the pairwise squared difference between model predictions, weighted by measurement noise and model uncertainty resulting from uncertainty in the fitted model parameters. The model uncertainty for analytical model functions is computed using gradient information. We develop a novel method where we replace the black-box models with Gaussian process surrogate models. Using the surrogate models, we are able to approximately marginalise out the model parameters, yielding the model uncertainty. Results show the surrogate model method working for model discrimination for classical test instances.

Keywords: Design of experiments, model discrimination, Gaussian processes

1. Introduction

Biological engineering deals with noisy and uncertain processes. Modelling these processes is often difficult, and exacerbated by the difficulty of observing mechanisms and reactions on the molecular level. We can hypothesise several different mathematical models to explain a system's behaviour, and run experiments to discriminate between the models. The idea is that there is a real, expensive-to-evaluate system, and M rival models predicting the system behaviour. These models are effectively different hypotheses about some underlying system mechanism. We seek the most accurate model with as few experiments as possible. We will assume the experimental measurement noise is Gaussian distributed with zero mean and a given covariance $\boldsymbol{\Sigma}$.

For analytical models, where the functional relationship can be written down in closed form, extensive literature exists for design of experiments for model discrimination, e.g. Asprey and Macchietto (2000) and Michalik *et al.* (2010). The challenge is that most mathematical models for industrially relevant biological and chemical processes are neither simple nor analytical. They are often, from an optimisation point-of-view, complex black boxes, e.g. legacy codes representing large systems of partial differential equations. For these models, we can simulate the process at discrete locations, but gradient information with respect to model parameters is not readily available.

This paper develops a novel method of using Gaussian process (GP) surrogate models from which model output distributions can be computed and used with existing design criteria from literature, e.g. Box and Hill (1967) and Buzzi-Ferraris *et al*. (1990).

2. Background

2.1. Design Criteria for Model Discrimination

The fundamental principle of experimental design for model discrimination is to select the next experimental point where the model predictions differ the most (Hunter and Reiner, 1965). The measure of how much the models differ is called the *design criterion* and should be maximised. We compare four different design criteria. The first three can be written on the general form $D(\mathbf{x}) = \sum_{i=1}^{M}\sum_{j=1}^{M} D_*(i,j)$, where

- $D_{\mathrm{HR}}(i,j) = \left\|\Delta_{i,j}\right\|_2^2$ (Hunter and Reiner, 1965),
- $D_{\mathrm{BH}}(i,j) = \pi_i\pi_j\left\{\mathrm{tr}\left(\boldsymbol{\Sigma}_i\boldsymbol{\Sigma}_j^{-1} + \boldsymbol{\Sigma}_j\boldsymbol{\Sigma}_i^{-1} - 2\mathbf{I}\right) + \Delta_{i,j}^{\mathrm{T}}\left(\boldsymbol{\Sigma}_j^{-1} + \boldsymbol{\Sigma}_i^{-1}\right)\Delta_{i,j}\right\}$ (Box and Hill, 1967; Prasad and Someswara Rao, 1977),
- $D_{\mathrm{BF}}(i,j) = \mathrm{tr}\left(2\boldsymbol{\Sigma}\left(\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}_i\right)^{-1} + \Delta_{i,j}^{\mathrm{T}}\left(\boldsymbol{\Sigma}_j + \boldsymbol{\Sigma}_i\right)^{-1}\Delta_{i,j}\right.$ (Buzzi-Ferraris *et al*., 1990},

$\pi_i \in [0,1]$ is the prior likelihood of model i, $\Delta_{i,j} = \boldsymbol{f}_i(\mathbf{x},\boldsymbol{\theta}_i) - \boldsymbol{f}_j(\mathbf{x},\boldsymbol{\theta}_j)$ with $\boldsymbol{f}_i(\mathbf{x},\boldsymbol{\theta}_i)$ the prediction of model i given design $\mathbf{x}$ and model parameters $\boldsymbol{\theta}_i$, and $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} + \boldsymbol{\Sigma}_{\mathcal{M},i}(\mathbf{x})$ is the covariance of model i, with $\boldsymbol{\Sigma}_{\mathcal{M},i}(\mathbf{x})$ the model uncertainty due to parameter estimation from noisy observations.

The Michalik *et al.* (2010) design criterion uses Akaike's information criterion: $D_{\mathrm{AW}}(\mathbf{x}) = \sum_{i=1}^{M}\pi_i\left\{\sum_{j=1}^{M}\exp\left(\frac{1}{2}\Delta_{i,j}^{\mathrm{T}}\boldsymbol{\Sigma}_i^{-1}\Delta_{i,j} + D_i - D_j\right)\right\}$, where D_i is the number of model parameters for model i.

We let $\phi_{(e),\ell}(\mathbf{x}) = \partial f_{i,(e)}(\mathbf{x},\boldsymbol{\theta}_i)/\partial\theta_{i,\ell}\big|_{\boldsymbol{\theta}_i=\boldsymbol{\theta}_i^*}$ and use it to define $\left[\boldsymbol{\Phi}_{(e)}\right]_{n,\ell} = \phi_{(e),\ell}(\mathbf{x}_n)$, where $\mathbf{x}_n$ is the location of the n^{th} observation, and $[\boldsymbol{\Phi}(\mathbf{x})]_{(e),\ell} = \phi_{(e),\ell}(\mathbf{x})$. Based on a first-order Taylor expansion of the model functions around $\boldsymbol{\theta}^*$ (Prasad and Someswara Rao, 1977) the model parameters are approximated as Gaussian distributed $\mathcal{N}(\boldsymbol{\theta}^*, \boldsymbol{\Sigma}_\theta)$, with $\boldsymbol{\Sigma}_\theta^{-1} = \sum_{e_1=1}^{E}\sum_{e_2=1}^{E}[\boldsymbol{\Sigma}]_{(e_1),(e_2)}\boldsymbol{\Phi}_{(e_1)}^{\mathrm{T}}\boldsymbol{\Phi}_{(e_2)}$, where E is the number of observable states. The model uncertainty is then approximated by $\boldsymbol{\Sigma}_{\mathcal{M},i}(\mathbf{x}) = \boldsymbol{\Phi}(\mathbf{x})\boldsymbol{\Sigma}_\theta\boldsymbol{\Phi}^{\mathbf{T}}(\mathbf{x})$.

2.2. Gaussian Process Regression

A Gaussian process (GP) is a collection of random variables, any finite subset of which is jointly Gaussian distributed (Rasmussen and Williams, 2006). We can place a GP prior $\mathcal{GP}(m(\mathbf{x}), k(\mathbf{x},\mathbf{x}'))$ on an unknown function f, where m and k are the mean function and covariance function, respectively. Given observations $\mathbf{y} = y_{1:n} \sim \mathcal{N}(f(\mathbf{x_{1:n}}), \sigma_\eta^2)$ at locations $\mathbf{X} = \mathbf{x}_{\mathbf{1}:\boldsymbol{n}}^{\mathrm{T}}$ and a test point $\mathbf{x}_*$, we can compute the posterior predictive distribution $f(\mathbf{x}_*)|\,\mathbf{y},\mathbf{X} \sim \mathcal{N}(\mu(\mathbf{x}_*), \sigma^2(\mathbf{x}_*))$, where $\mu(\mathbf{x}_*) = m(\mathbf{x}_*) + k(\mathbf{x}_*,\mathbf{X})\,\mathbf{K}^{-1}(\mathbf{y}-\mathbf{m})$ and $\sigma^2(\mathbf{x}_*) = k(\mathbf{x}_*,\mathbf{x}_*) - k(\mathbf{x}_*,\mathbf{X})\mathbf{K}^{-1}k(\mathbf{X},\mathbf{x}_*)$, with, in turn, $\mathbf{K} = k(\mathbf{X},\mathbf{X}) + \sigma_\eta^2\mathbf{I}$ the kernel matrix and $\mathbf{m} = m(\mathbf{x_{1:n}})$. The covariance between predictions $f(\mathbf{Z})$ and $f(\mathbf{Z}')$ for the input locations $\mathbf{Z} = \mathbf{z}_{\mathbf{1}:\boldsymbol{H}}^{\mathrm{T}} \in \mathbb{R}^{H\times h}$ and $\mathbf{Z}' = \mathbf{z}'^{\mathrm{T}}_{\mathbf{1}:\boldsymbol{H}'} \in \mathbb{R}^{H'\times h}$ is denoted $k(\mathbf{Z},\mathbf{Z}') \in \mathbb{R}^{H\times H'}$ with elements $[k(\mathbf{Z},\mathbf{Z}')]_{i,j} = k(\mathbf{z}_i,\mathbf{z}_j')$.

A common covariance function is the radial basis function (RBF) kernel $k(\mathbf{x},\mathbf{x}') = \rho^2\exp(-r^2/2)$ with automatic relevance determination (ARD), where ρ^2 is the signal variance, $r^2 = \|\mathbf{x}-\mathbf{x}'\|_{\boldsymbol{\Lambda}^{-1}}^2$, and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_{1:D}^2)$, $\mathbf{x} \in \mathbb{R}^D$, are the length scales.

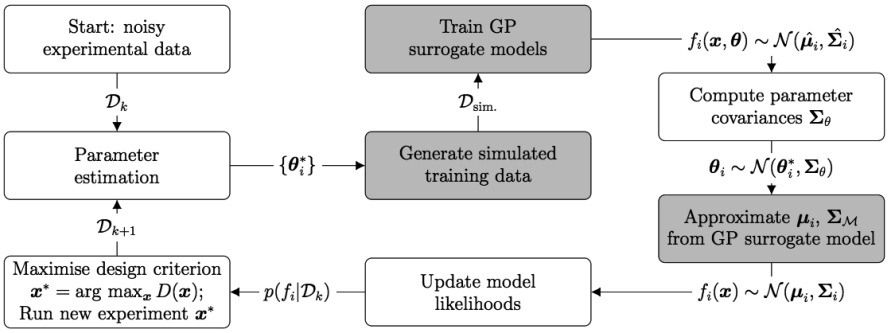


Figure 1: Flowchart description of design of experiments for model discrimination, with novel steps in our method highlighted.

3. Method

We begin by studying the output $\boldsymbol{f}(\mathbf{x}, \boldsymbol{\theta})$ of a single model. We wish to compute the predictive output distribution of the model, taking the uncertainty in the model parameters $\boldsymbol{\theta}$ into account. We place GP priors $\mathcal{GP}(m_{(e)}(\mathbf{x}), k_{x,(e)}(\mathbf{x}, \mathbf{x}')k_{\theta,(e)}(\boldsymbol{\theta}, \boldsymbol{\theta}'))$ on each output $f_{(e)}(\mathbf{x}, \boldsymbol{\theta})$, $e = 1, \dots, E$, of $\boldsymbol{f}$. The noise variance for model output is denoted $\sigma_{\eta,(e)}^2$. The GPs have been trained on the data set $\mathcal{D}_{\text{sim}} = \{\mathbf{X}, \boldsymbol{\Theta}, \mathbf{Y}\} = \{\mathbf{x}_{1:n}, \boldsymbol{\theta}_{1:n}, y_{1:n,1:E}\}$, from n model evaluations.

The predictive distribution $p\big(\boldsymbol{f}(\mathbf{x}, \boldsymbol{\theta})\big) \sim \mathcal{N}\big(\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta}), \boldsymbol{\Sigma}_{\mathcal{M}}(\mathbf{x}, \boldsymbol{\theta})\big)$ is Gaussian, where $\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta}) = \mu_{1:E}(\mathbf{x}, \boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_{\mathcal{M}}(\mathbf{x}, \boldsymbol{\theta}) = \text{diag}(\sigma_{\mathcal{M},1:E}^2(\mathbf{x}, \boldsymbol{\theta}))$ with $\mu_{(e)}(\mathbf{x}, \boldsymbol{\theta}) = m_{(e)}(\mathbf{x}) + \mathbf{k}_*^{\mathrm{T}}\mathbf{K}_{(e)}^{-1}(\mathbf{y}_{(e)} - \mathbf{m}_{(e)})$ and $\sigma_{\mathcal{M},(e)}^2(\mathbf{x}, \boldsymbol{\theta}) = k_{x,(e)}(\mathbf{x}, \mathbf{x})k_{\theta,(e)}(\boldsymbol{\theta}, \boldsymbol{\theta}) - \mathbf{k}_*^{\mathrm{T}}\mathbf{K}_{(e)}^{-1}\mathbf{k}_*$. The vector $\mathbf{k}_* = k_{x,(e)}(\mathbf{x}, \mathbf{X}) \circ k_{\theta,(e)}(\boldsymbol{\theta}, \boldsymbol{\Theta})$ and the kernel matrix $\mathbf{K}_{(e)} = k_{x,(e)}(\mathbf{X}, \mathbf{X}) \circ k_{\theta,(e)}(\boldsymbol{\Theta}, \boldsymbol{\Theta}) + \sigma_{\eta,(e)}^2\mathbf{I}$, and $\circ$ denotes element-wise multiplication.

Given a distribution $\boldsymbol{\theta} \mid \mathcal{D} \sim \mathcal{N}(\boldsymbol{\theta}^*, \boldsymbol{\Sigma}_\theta)$ over the model parameters, where $\mathcal{D}$ is the experimental data set, we wish to determine the resulting model uncertainty. The marginalised model output distribution $p(\boldsymbol{f}(\mathbf{x})|\mathcal{D}) = \int p\big(\boldsymbol{f}(\mathbf{x}, \boldsymbol{\theta})\big)p(\boldsymbol{\theta}|\mathcal{D})\mathrm{d}\boldsymbol{\theta}$, which is intractable and has to be approximated. We choose to approximate the marginalised distribution $p(\boldsymbol{f}(\mathbf{x})|\mathcal{D})$ with a Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}(\mathbf{x}), \boldsymbol{\Sigma}_{\mathcal{M}}(\mathbf{x}))$, where $\boldsymbol{\mu}(\mathbf{x}) \approx \mathbb{E}_{\boldsymbol{\theta}|\mathcal{D}}[\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta})]$ and $\boldsymbol{\Sigma}_{\mathcal{M}}(\mathbf{x}) \approx \mathbb{E}_{\boldsymbol{\theta}|\mathcal{D}}[\boldsymbol{\Sigma}_{\mathcal{M}}(\mathbf{x}, \boldsymbol{\theta})] + \mathbb{V}_{\boldsymbol{\theta}|\mathcal{D}}[\boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta})]$.

A design of experiments for model discrimination flowchart is shown in Figure 1, with novel steps in our method highlighted. We will compare first- and second-order Taylor approximations of $\boldsymbol{\mu}(\mathbf{x})$ and $\boldsymbol{\Sigma}_{\mathcal{M}}(\mathbf{x})$. We let $\Delta\boldsymbol{\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}^*$, $\nabla_{\boldsymbol{\theta}}\mu_{(e)} = \nabla_{\boldsymbol{\theta}}\mu_{(e)}(\mathbf{x}, \boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$ and $\nabla_{\boldsymbol{\theta}}\sigma_{\mathcal{M},(e)}^2 = \nabla_{\boldsymbol{\theta}}\sigma_{\mathcal{M},(e)}^2(\mathbf{x}, \boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$.

3.1. First-Order Taylor Approximation

The first-order Taylor approximation around a model parameter value $\boldsymbol{\theta}^*$ is given by

$$\mu_{(e)}(\mathbf{x}, \boldsymbol{\theta}) \approx \mu_{(e)}(\mathbf{x}, \boldsymbol{\theta}^*) + \nabla_{\boldsymbol{\theta}}\mu_{(e)}\Delta\boldsymbol{\theta} \tag{1}$$

$$\sigma_{\mathcal{M},(e)}^2(\mathbf{x}, \boldsymbol{\theta}) \approx \sigma_{\mathcal{M},(e)}^2(\mathbf{x}, \boldsymbol{\theta}^*) + \nabla_{\boldsymbol{\theta}}\sigma_{\mathcal{M},(e)}^2\Delta\boldsymbol{\theta} \tag{2}$$

If the model parameter distribution is given by $\boldsymbol{\theta} \mid \mathcal{D} \sim \mathcal{N}(\boldsymbol{\theta}^*, \boldsymbol{\Sigma}_\theta)$, using the first-order Taylor approximations in Eq. (1)-(2), the mean $\boldsymbol{\mu}(\mathbf{x})$ of the approximate marginalised

model output distribution becomes $\boldsymbol{\mu}(\mathbf{x}) \approx \boldsymbol{\mu}(\mathbf{x}, \boldsymbol{\theta}^*)$, and the corresponding covariance matrix has elements $[\boldsymbol{\Sigma}_{\mathcal{M}}(\mathbf{x})]_{(e_1),(e_2)} \approx [\boldsymbol{\Sigma}_{\mathcal{M}}(\mathbf{x}, \boldsymbol{\theta}^*)]_{(e_1),(e_2)} + \nabla_{\boldsymbol{\theta}} \mu_{(e_1)} \boldsymbol{\Sigma}_{\theta} \nabla_{\boldsymbol{\theta}} \mu_{(e_2)}^{\mathrm{T}}$.

3.2. Second-Order Taylor Approximation

The second-order Taylor approximation around a model parameter value $\boldsymbol{\theta}^*$ is given by

$$\mu_{(e)}(\mathbf{x}, \boldsymbol{\theta}) \approx \mu_{(e)}(\mathbf{x}, \boldsymbol{\theta}^*) + \nabla_{\boldsymbol{\theta}} \mu_{(e)} \Delta\boldsymbol{\theta} + \frac{1}{2} \Delta\boldsymbol{\theta}^{\mathrm{T}} \nabla_{\boldsymbol{\theta}}^2 \mu_{(e)} \Delta\boldsymbol{\theta} \tag{3}$$

$$\sigma_{\mathcal{M},(e)}^2(\mathbf{x}, \boldsymbol{\theta}) \approx \sigma_{\mathcal{M},(e)}^2(\mathbf{x}, \boldsymbol{\theta}^*) + \nabla_{\boldsymbol{\theta}} \sigma_{\mathcal{M},(e)}^2 \Delta\boldsymbol{\theta} + \frac{1}{2} \Delta\boldsymbol{\theta}^{\mathrm{T}} \nabla_{\boldsymbol{\theta}}^2 \sigma_{\mathcal{M},(e)}^2 \Delta\boldsymbol{\theta} \tag{4}$$

If the model parameter distribution is given by $\boldsymbol{\theta} \mid \mathcal{D} \sim \mathcal{N}(\boldsymbol{\theta}^*, \boldsymbol{\Sigma}_{\theta})$, using the second-order Taylor approximations in Eq. (3)-(4), the mean $\boldsymbol{\mu}(\mathbf{x})$ of the approximate marginalised model output distribution has elements $\mu_{(e)}(\mathbf{x}) \approx \mu_{(e)}(\mathbf{x}, \boldsymbol{\theta}^*) + \frac{1}{2} \mathrm{tr}(\nabla_{\boldsymbol{\theta}}^2 \mu_{(e)} \boldsymbol{\Sigma}_{\theta})$. The corresponding covariance $\boldsymbol{\Sigma}_{\mathcal{M}}(\mathbf{x})$ has elements

$$\begin{aligned} [\boldsymbol{\Sigma}_{\mathcal{M}}(\mathbf{x})]_{(e_1),(e_2)} \approx \delta_{(e_1),(e_2)} & \underbrace{\left(\sigma_{\mathcal{M},(e_1)}^2(\mathbf{x}, \boldsymbol{\theta}^*) + \frac{1}{2} \mathrm{tr}\left(\nabla_{\boldsymbol{\theta}}^2 \sigma_{\mathcal{M},(e_1)}^2 \boldsymbol{\Sigma}_{\theta} \right) \right)}_{\mathbb{E}_{\boldsymbol{\theta}|\mathcal{D}}[\boldsymbol{\Sigma}_{\mathcal{M}}(\mathbf{x},\boldsymbol{\theta})]} \\ & + \underbrace{\nabla_{\boldsymbol{\theta}} \mu_{(e_1)} \boldsymbol{\Sigma}_{\theta} \nabla_{\boldsymbol{\theta}} \mu_{(e_2)}^{\mathrm{T}} + \frac{1}{2} \mathrm{tr}\left(\nabla_{\boldsymbol{\theta}}^2 \mu_{(e_1)} \boldsymbol{\Sigma}_{\theta} \nabla_{\boldsymbol{\theta}}^2 \mu_{(e_2)} \boldsymbol{\Sigma}_{\theta} \right)}_{\mathbb{V}_{\boldsymbol{\theta}|\mathcal{D}}[\boldsymbol{\mu}(\mathbf{x},\boldsymbol{\theta})]}, \end{aligned} \tag{5}$$

where $\delta_{i,j}$ is the Kronecker delta.

4. Results

We compare the performance of (i) the first- and second-order Taylor approximations, and (ii) the four different design criteria in Sec. 2.1 as well as random design, for two different case studies. For each experiment we uniformly sample a new set of N_0 initial data points. Each iteration we compute the normalised likelihood $\hat{\pi}_{i,N}$ of model $\boldsymbol{f}_i$ given dataset $\mathcal{D}_N$ of size N as

$$\hat{\pi}_{i,N} = \frac{p(\boldsymbol{f}_i|\mathcal{D}_N)}{\sum_{j=1}^{M} p(\boldsymbol{f}_j|\mathcal{D}_N)}, \qquad p(\boldsymbol{f}_i|\mathcal{D}_N) = \prod_{j=1}^{N} \mathcal{N}\left(\mathbf{y}_j | \boldsymbol{\mu}_i(\mathbf{x}_j), \boldsymbol{\Sigma} + \boldsymbol{\Sigma}_{\mathcal{M},i}(\mathbf{x}_j) \right) \tag{7}$$

The normalised likelihoods $\hat{\pi}_{i,N}$ fluctuate significantly in the experiments due to measurement noise and refitting of model parameters each iteration. Therefore, we choose to update the model likelihoods using a moving average, such that given the $(N+1)^{\text{th}}$ observation the likelihood for model $\boldsymbol{f}_i$ becomes $\pi_{i,N+1} = (\pi_{i,N} + \hat{\pi}_{i,N})/2$, with $\pi_{i,0} = 1/M$.

In each case study, we choose one model to be the true model, from which we generate observations with added noise. We sample all models in a black-box fashion to construct the corresponding GP surrogate models. Performance is measured in the average number (A) of additional experiments $N - N_0$ required for the data-generating model i to reach likelihood threshold $\pi_{i,N} \geq \tau$. (SE) denotes the standard error of (A).

In both case studies, the RBF-ARD covariance function is used for both k_x and k_θ, because we expect the model functions to be smooth.

4.1. Case Study 1

Case study from Buzzi-Ferraris *et al*. (1990) of a reaction with two observables y_1, y_2 and two design variables x_1, x_2. We want to discriminate between four chemical kinetic models $\boldsymbol{f}_i$, each with four model parameters $\theta_{i,j}$. The observed data are generated from model 1 with $N_0 = 5$ initial data points. The results are shown in Table 1.

Table 1. Results after 100 experiments in case study 1, using first- (T_1) and second-order (T_2) Taylor approximation. Averages within one SE of best result in bold font.

T_1		Design criterion				
τ		D_{HR}	D_{BH}	D_{BF}	D_{AW}	Random
0.90	A	3.79	3.90	4.41	**3.64**	5.42
	SE	0.12	0.14	0.17	0.10	0.26
0.95	A	5.16	5.37	5.94	**4.86**	7.05
	SE	0.17	0.18	0.22	0.12	0.28
0.99	A	8.25	8.26	9.13	**7.50**	10.67
	SE	0.24	0.21	0.25	0.16	0.34

T_2		Design criterion				
τ		D_{HR}	D_{BH}	D_{BF}	D_{AW}	Random
0.90	A	**4.00**	**4.08**	5.00	**3.97**	5.70
	SE	0.13	0.13	0.23	0.14	0.24
0.95	A	5.34	5.43	6.41	**5.13**	7.32
	SE	0.19	0.14	0.25	0.15	0.27
0.99	A	8.54	8.46	9.59	**7.89**	11.49
	SE	0.23	0.17	0.26	0.19	0.39

4.2. Case Study 2

Case study of the reaction $\mathrm{A} + \mathrm{B} + \mathrm{E} \rightarrow \mathrm{P} + \mathrm{D}$ (Tandogan *et al.*, 2017). There are three different models to describe this process, with intermediary, non-observable reactions. We generate the observed data from model 3, with $N_0 = 3$ initial data points for each experiment. The model parameters $\theta_{i,j}$ are the kinetic coefficients. There are 5, 4 and 7 parameters for model 1, 2 and 3, respectively. The design variables are (1) the ratio $r = V_{\mathrm{A},0}/750$ ml of the initial volume $V_{\mathrm{A},0} + V_{\mathrm{B},0}$ made up by solvent A, (2) the initial amount of $\mathrm{E} \in [10, 200]$ mg added to the solution, and (3) the time point $\mathrm{t} \in [0, 1]$ s at which we measure the concentrations of A, B, E, P and D. We assume that A and B have a molar volume 55 mol/l, and that E has molar weight 440 g/mol. The initial concentrations of all other components are set to 0. The results are shown in Table 2.

Table 2. Results after 100 experiments in case study 2, using first- (T_1) and second-order (T_2) Taylor approximation. Averages within one SE of best result in bold font.

T_1		Design criterion				
τ		D_{HR}	D_{BH}	D_{BF}	D_{AW}	Random
0.90	A	2.81	**2.40**	**2.42**	**2.49**	**2.36**
	SE	0.31	0.14	0.17	0.22	0.13
0.95	A	4.29	**3.40**	**3.42**	**3.49**	**3.38**
	SE	0.48	0.14	0.17	0.22	0.14
0.99	A	6.76	**6.21**	**6.23**	**6.33**	**6.44**
	SE	0.36	0.11	0.12	0.19	0.31

T_2		Design criterion				
τ		D_{HR}	D_{BH}	D_{BF}	D_{AW}	Random
0.90	A	4.13	3.54	3.79	**3.14**	5.17
	SE	0.67	0.30	0.33	0.21	0.61
0.95	A	5.77	4.59	4.79	**4.14**	6.53
	SE	0.84	0.31	0.34	0.22	0.68
0.99	A	8.81	7.07	7.33	**6.69**	8.56
	SE	1.09	0.32	0.33	0.21	0.73

5. Discussion

Tables 1-2 show that replacing the original models with GP surrogate models is outperforming the random strategy. For these test instances, our method performs better using the first-order than the second-order Taylor approximation for marginalising out the model parameters, i.e. fewer additional experiments are required to find the data-generating model. This could be because a first-order Taylor expansion is used for computing the parameter covariance $\boldsymbol{\Sigma}_{\theta}$. Alternatively, the first-order (second-order) approximation might underestimate (overestimate) the model uncertainty, resulting in faster discrimination in these simulated case studies. In future research we will study this more, and explore additional approximation methods.

For case study 1, D_{AW} significantly outperforms the other design criteria. For case study 2, the performances are more similar. The D_{AW} design criterion might have an advantage over the D_{HR} and D_{BF} design criteria because it incorporates the model probabilities used for the subsequent model discrimination.

In Case Study 2, the concentrations can only be measured at a single time point for each experiment. Measurements at multiple time points can be implemented by modelling each output and fixed time point with a separate GP surrogate model, i.e. treat the measurement at each time point as a separate model output.

6. Conclusions

Design of experiments for black-box model discrimination is a difficult but important problem. We have shown that our novel method, hybridising the classical analytical approach with a statistical sampling-based approach using GP surrogate models, can work well. It allows us to be flexible with regards to the software implementations of the underlying models. In future research we will apply our method to more case studies and compare our method to competing black-box design and discrimination methods. We also hope to apply our method in a control setting using a multi-objective approach (Olofsson *et al.,* 2017). The GP surrogate model-based method presented in this paper has been implemented in Python and made available on GitHub (Olofsson, 2017).

Acknowledgement

This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no.675251, and an EPSRC Early Career Fellowship (EP/P016871/1) to R.M.

References

S.P. Asprey, and S. Macchietto, 2000. Statistical tools for optimal dynamic model building. In *Comput Chem Eng* 24, 1261–1267.

G.E. Box, and W.J. Hill, 1967. Discrimination among mechanistic models. *Technometrics* 9 (1).

G. Buzzi-Ferraris, P. Forzatti, G. Emig, and H. Hofmann, 1984. Sequential experimental design for model discrimination in the case of multiple responses. *Chem Eng Sci* 39 (1), 81–85.

G. Buzzi-Ferraris, P. Forzatti, and P. Canu, 1990. An improved version of a sequential design criterion for discriminating among rival multiresponse models. *Chem Eng Sci* 45 (2), 477–481.

W.G. Hunter, and A.M. Reiner, 1965. Designs for discriminating between two rival models. *Technometrics* 7 (3), 307–323.

C. Michalik, M. Stuckert, and W. Marquardt, 2010. Optimal experimental design for discriminating numerous model candidates: The AWDC criterion. *Ind Eng Chem Res* 49, 913–919.

S. Olofsson, 2017. GPdoemd. https://github.com/scwolof/GPdoemd.

S. Olofsson, M. Mehrian, L. Geris, R. Calandra, M.P. Deisenroth, and R. Misener, R., 2017. Bayesian multi-objective optimisation of neotissue growth in a perfusion bioreactor set-up. In *ESCAPE-27*.

K.B. Prasad, and M. Someswara Rao, 1977. Use of expected likelihood in sequential model discrimination in multiresponse systems. *Chem Eng Sci* 32, 1411–1418.

C.E. Rasmussen, and C.K. Williams, 2006. Gaussian processes for machine learning. MIT Press.

N. Tandogan, S. Garcia-Muñoz, M. Sen, T.M. Wilson, J.Y. Buser, S.P. Kolis, I.V. Borkar, and C.A. Alt, 2017. Use of model discrimination method in drug substance process development. In *Proceedings of the 2017 AIChE Annual Meeting*.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50137-3

Efficient Implicit-Explicit Time Stepping for Accurate and Rapid Simulation of Chromatographic Models

Kristian Meyer[a*], Jakob K. Huusom[a], Jens Abildskov[a]

[a]*Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engieering, Technical University of Denmark, Building 229, 2800 Kgs. Lyngby, Denmark*
krismey@kt.dtu.dk

Abstract

In this paper, we give a summary of recent advances (K. Meyer et al, 2018) on the use of a nodal discontinuous Galerkin finite element (DG-FE) method for spatial discretization of chromatographic models. Explicit Runge-Kutta (ERK) methods are popular for integrating the semi-discrete systems of equations resulting from DG space discretization. However, ERK methods suffer from stability-based time step restrictions for stiff problems. Therefore, we implement a high order implicit-explicit additive Runge-Kutta (IMEX-ARK) method (Kennedy and Carpenter, 2003, 2007) to overcome system stiffness. The IMEX-RK method advances the non-stiff parts of the model using explicit methods and solve the more expensive stiff parts using an L-stable stiffly-accurate explicit, singly diagonally implicit Runge-Kutta method (ESDIRK). We show that for a multicomponent nonlinear chromatographic system, the IMEX-ARK scheme becomes more efficient than explicit methods for increasingly stiff systems. We recommend integrating the convective term using explicit methods and to integrate both the diffusive and reactive terms using implicit methods.

Keywords: Langmuir isotherm, Nonlinear chromatography, high-order, discontinuous Galerkin finite element method, implicit-explicit.

1. Introduction

Biopharmaceuticals are becoming increasingly important in recent years in treating chronic diseases such as cancer and arthritis. The manufacturing processes are inherently complex due to their non-linear bioprocess dynamics and batch-to-batch variability leading to high production costs of which 50-80 % are attributed to downstream processing (Guiochon and Beaver, 2011). Here, chromatographic separations are essential since they achieve high purity as required for biopharmaceutical drugs. The current trend in the biopharmaceutical industries is the digitalization of the manufacturing processes to switch from a recipe-driven to a data-driven operational approach. This digitalization has an enormous potential to facilitate rapid cost-effective process development for accelerating speed-to-market and near-optimal operation using on-line monitoring strategies to control production costs. This has been successfully used for decades in other industries such as the chemical and petrochemical industries. Biopharmaceutical industries can reap similar benefits from digitalization. Presently, the utility of computers for process design and operation is limited. The current challenges are computational efforts, quality of models, accessibility to experimental data and

component-specific on-line concentration measurements in complex biological feedstocks.

This contribution presents a novel tool for rapid and accurate simulation of chromatographic models to reduce computational efforts and hence, promote the digitalization of the biopharmaceutical industry. The parabolic and convection-dominated transport-dispersive model of chromatography is considered and is discretized using a method-of-lines approach in which the spatial and temporal domains are considered separately. A discontinuous Galerkin finite element (DG-FE) method was presented by Meyer et al. (2018) for spatial discretization resulting in a large system of ordinary differential equations (ODE) in time. Despite the popularity of high-order explicit Runge-Kutta (ERK) methods for time marching of the ODE system resulting from DG space discretization, ERK methods suffer from stability-based time step restrictions for stiff problems. Therefore, we implement a high-order implicit-explicit additive Runge-Kutta (IMEX-ARK) scheme to overcome system stiffness. The IMEX-ARK scheme uses an explicit method to solve non-stiff parts of the model and isolate and solve the more expensive stiff parts using an Explicit, Singly Diagonally Implicit Runge-Kutta method (ESDIRK). In future, we want to consider dynamic optimization problems for optimal closed loop control of chromatographic columns. Therefore, it is especially beneficial to choose an ESDIRK-based method for time integration of the stiff parts over the often applied BDF methods (Schlegel et al., 2004). In general, ESDIRK methods in combination with DG space discretization is superior over BDF methods (Bijl et al., 2002). We show that for increasingly stiff systems, the IMEX-ARK scheme becomes more efficient than explicit Runge-Kutta (ERK) methods. We recommend integrating the convective term with explicit methods and both the diffusive and reactive terms using implicit methods.

2. Transport-dispersive model of chromatography

2.1. Mass transport

The transport-dispersive model of chromatography is used to describe the effects of mass transport (Guiochon et al., 2006). It is for the ith component given, on dimensionless form, by

$$\frac{\partial \boldsymbol{u}}{\partial \tau} + \frac{\partial \boldsymbol{f_c}}{\partial z} - \frac{\partial \boldsymbol{f_d}}{\partial z} = \boldsymbol{p} \tag{1}$$

where

$$\boldsymbol{u} = \begin{bmatrix} c_i + F q_i \\ q_i \end{bmatrix}, \boldsymbol{f_c} = \begin{bmatrix} c_i \\ 0 \end{bmatrix}, \quad \boldsymbol{f_d} = \begin{bmatrix} 1/Pe_i \cdot \partial c_i / \partial z \\ 0 \end{bmatrix}, \quad \boldsymbol{p} = \begin{bmatrix} 0 \\ St_i (q_i^* - q_i) \end{bmatrix} \tag{2}$$

Here, c_i and q_i are the mobile phase and stationary phase concentrations, respectively, $F = (\epsilon_c - 1)/\epsilon_c$ is the column phase ratio, ϵ_c is the column porosity, q_i^* is the equilibrium stationary phase concentration and τ and z are the dimensionless time and axial position, respectively. The Peclet number $Pe_i = vL/D_i$ and the Stanton number $St_i = k_i L/v$ are important dimensionless groups that determine the stiffness of the

system, where v is the interstitial velocity, L is the column length, D_i is the axial dispersion coefficient, k_i is the lumped mass transfer coefficient and L is the column length. As the Peclet number increases, the system stiffness due to diffusion vanishes. For increasingly large Stanton numbers, the reactive term contributes increasingly to system stiffness. Eq.(1) is for the ith component complemented with Danckwerts (1954) boundary conditions and initial conditions corresponding to an initially equilibrated column are used.

2.2. Adsorption

The column adsorption equilibria between bound q_i and unbound c_i states are described by a competitive nonlinear Langmuir isotherm. The model is derived by Guiochon et al. (2006) and is for the ith component given by

$$q_i^* = \frac{a_i c_i}{1 + \sum_{s=1}^{N_c} b_s c_s} \tag{3}$$

where b_s is the ratio of the rate constant of adsorption and desorption and the ratio a_i/b_i is the column saturation capacity.

3. Time integration schemes

The numerical solution of the parabolic and convection dominated system described by Eq.(1) is solved using a methods-of-lines approach. Here, the spatial and temporal discretization's are considered separately. A nodal discontinuous Galerkin finite element (DG-FE) method is presented by Meyer et al. (2018). The resulting semi-discrete equation system is of the form

$$\frac{\partial \boldsymbol{u_h}}{\partial t} = \boldsymbol{F}\big(t, \boldsymbol{u}_h(t)\big) = \sum_{v=1}^{2} \boldsymbol{F}^{[v]}(t, \boldsymbol{u}_h(t)), \quad \boldsymbol{u}_h(t_0) = \boldsymbol{u}_{h,0} \tag{4}$$

where $\boldsymbol{u_h}$ is the numerical approximation to the exact solution $\boldsymbol{u}$, $\boldsymbol{F}^{[1]}(t, \boldsymbol{u_h(t)})$ is the spatial DG approximation to the convective term which we integrate with an explicit method and $\boldsymbol{F}^{[2]}(t, \boldsymbol{u}_h(t))$ is the DG approximation to both the diffusive and reactive terms which we integrate with an implicit method.

3.1. Implicit-Explicit additive Runge-Kutta (IMEX-ARK) method

A 6 stage 4$^{\text{th}}$–order IMEX-ARK$_2$ method (Kennedy and Carpenter, 2003) is used to advance Eq.(2) from time step $t^{(n)}$ until $t^{(n+1)} = t + \Delta t$ by

$$\begin{aligned} \boldsymbol{u}_h^{(i)} &= \boldsymbol{u}_h^{(n)} + \Delta t \sum_{v=1}^{2} \sum_{j=1}^{6} a_{ij}^{[v]} \boldsymbol{F}^{[v]}\left(t^{(n)} + c_j \Delta t, \boldsymbol{u}_h^{(j)}\right), \quad 1 \le i \le 6 \\ \boldsymbol{u}_h^{(n+1)} &= \boldsymbol{u}_h^{(n)} + \Delta t \sum_{v=1}^{2} \sum_{j=1}^{6} b_j^{[v]} \boldsymbol{F}^{[v]}\left(t^{(n)} + c_j \Delta t, \boldsymbol{u}_h^{(j)}\right) \end{aligned} \tag{5}$$

$$\widehat{\boldsymbol{u}}_h^{(n+1)} = \boldsymbol{u}_h^{(n)} + \Delta t \sum_{v=1}^{2} \sum_{j=1}^{6} \hat{b}_j^{[v]} \boldsymbol{F}^{[v]} \left(t^{(n)} + c_j \Delta t, \boldsymbol{u}_h^{(j)} \right)$$

where $a_{ij}^{[v]}$, $b_i^{[v]}$ and $c_i^{[v]}$ are fixed constants. The $\boldsymbol{u}_h^{(n+1)}$-vector is of order 4 and the $\widehat{\boldsymbol{u}}_h^{(n+1)}$-vector associated with the embedded scheme is of order 3. Using the embedded scheme, a local error estimate $\delta = u_h^{n+1} - \hat{u}_h^{n+1}$ of order 3 is obtained. The local error estimate is fed into an I-controller to adaptively change the time step, i.e.

$$(\Delta t)_I^{n+1} = \kappa (\Delta t)^{(n)} \left(\frac{\epsilon}{\| \delta^{(n+1)} \|_2} \right)^{\left(\frac{1}{3}\right)} \tag{6}$$

where $\kappa \approx 0.9$ is a safety factor and ϵ is a user-specified temporal integration tolerance. The ith non-linear stage given by Eq.(5) is solved iteratively using a modified Newton's method, that is

$$\begin{aligned} \boldsymbol{M} \Delta \boldsymbol{u}_h^{(i)} &= -\boldsymbol{R}^{(i)} \left(\boldsymbol{u}_h^{(i,l)} \right) \\ \boldsymbol{u}_h^{(i,l+1)} &= \boldsymbol{u}_h^{(i,l)} + \Delta \boldsymbol{u}_h^{(i)} \end{aligned} \tag{7}$$

where $\boldsymbol{M} = \boldsymbol{I} - \Delta t a_{ii} J$ is the iteration matrix, $\boldsymbol{J} = \partial \boldsymbol{F} / \partial \boldsymbol{u}_h^{(i)}$ is the Jacobian matrix and $\boldsymbol{R}^{(i)}$ is the stage residual. The iteration is termined using a residual test rather than a displacement test for robustness (Houbak, Nørsett, & Thomsen, 1985). Thus, the iteration is terminated when $\boldsymbol{R}^{(i)} \leq c\epsilon$ where $c \approx 0.01$ is chosen such that the temporal integration error dominates. The iteration matrix is factored using Matlab's *lu* function whenever the step size changes. To reduce the number of factorizations, a constant step size is enforced when the step size computed by the I-controller is small compared to the current step size, e.g. if $1 \leq \Delta t^{(n+1)} / \Delta t^n \leq 1.2$, then $\Delta t^{(n+1)} = \Delta t^{(n)}$ (E. Hairer and G. Wanner, 1991). The Jacobian matrix is formed adaptively using similar logic as Shampine (1980). To speed up the Newton iterations and avoid convergence failures a good stage predictor is required. We use stage values from previous time steps to extrapolate into the current step using a dense output formula (Kennedy and Carpenter, 2007; Sand, 1988; Nørsett and Thomsen, 1986). For large time steps, e.g. if $(\Delta t)^{(n+1)} / \Delta t^{(n)} > 3$, we use the trivial guess instead (the most recent stage value).

3.2. Explicit Runge-Kutta (ERK) method

We implement an efficient and accurate 5 stage 4th-order low storage ERK scheme (M.H. Carpenter, and C.A. Kennedy, 1994) which requires storage and overwriting of only 2 vectors $\boldsymbol{k}^{(j)}$ and $\boldsymbol{p}^{(j)}$ by

$$\begin{aligned} &\boldsymbol{p}^{(0)} = \boldsymbol{u}_h^{(n)} \\ &j \in [1, \ldots, 5]: \begin{cases} \boldsymbol{k}^{(j)} = a_j \boldsymbol{k}^{(j-1)} + \Delta t \boldsymbol{F} \left(t^{(n)} + c_j \Delta t, \boldsymbol{p}^{(j-1)} \right) \\ \boldsymbol{p}^{(j)} = \boldsymbol{p}^{(j-1)} + b_j \boldsymbol{k}^{(j)} \end{cases} \\ &\boldsymbol{u}_h^{(n+1)} = \boldsymbol{p}^{(5)} \end{aligned} \tag{8}$$

where a_j, b_j, and c_j are fixed constants. Here, we use a fixed time step size which is restricted by a lower Courant-Friedrichs-Lewy (CFL) condition.

4. Numerical results

The case study considered is concerned with the separation of a two-component mixture under non-linear chromatographic conditions where the two components compete for available sites in the stationary phase. The detailed model parameters are: column length $L = 1$ m, porosity $\epsilon = 0.4$, interstitial velocity $v = 0.1$, isotherm parameters $(a_1, a_2) = (0.5,1)$ and $(b_1, b_2) = (0.05,0.1)$ L/mol. A rectangular pulse is injected at the inlet of the column for 12 s at a feed concentration of $(c_1, c_2) = (10,10)$ mol/L. The spatial domain is discretized using a 5th-order DGFE scheme with 200 elements.

The simulations are performed with varying Peclet numbers at a fixed Stanton number of 10^{-12} to isolate the effect of diffusion with respect to system stiffness. Similarly, the Stanton number is varied at a fixed Peclet number of $Pe = 10^{12}$ to isolate the effect of the reaction term. The results for varying Peclet numbers and varying Stanton numbers are given in Figs. 1A and B, respectively. In Fig. 1A it is seen that decreasing the Peclet number increases the stiffness of the system. Below a certain critical Peclet number, $Pe \approx 2 \cdot 10^4$, the IMEX-RK scheme starts to perform better than the ERK scheme. Under liquid chromatography conditions, Peclet numbers below $Pe = 2 \cdot 10^4$ are common. Therefore, we recommend integrating the diffusive term using an implicit method, although this might be unnecessary under high-performance liquid chromatography conditions where Peclet numbers can be expected to be very high. In Fig. 1B it is seen that increasing the Stanton number increases the stiffness of the system. Above a critical Stanton number of $St \approx 7 \cdot 10^3$ the IMEX-ARK scheme becomes more efficient compared to the ERK scheme. Again, we recommend integrating the reaction term using implicit time integration since Stanton numbers of $St = 7 \cdot 10^3$ are common under liquid chromatography conditions, although lower Stanton numbers can be expected under high-performance liquid chromatography conditions.

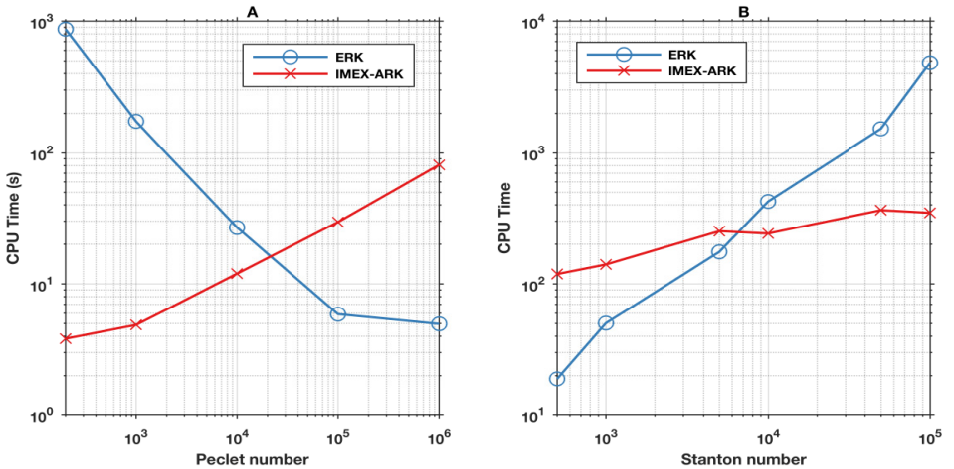


Figure 1. Comparison of IMEX-ARK and ERK results for a two component, non-linear chromatographic system. In A, the Peclet number is varied with a fixed Stanton number of 10^{-12} and in figure B the Stanton number is varied with a fixed Peclet number of 10^{12}.

5. Conclusions

An IMEX-ARK scheme was implemented and compared to an efficient ERK scheme. It was shown that for increasingly stiff systems, characterized by the dimensionless Peclet and Stanton numbers, the IMEX-ARK scheme becomes more efficient than the ERK scheme. In the future, we will improve our IMEX-ARK implementation by introducing a customized linear solver module. For larger systems, we will consider inexact Newton methods with efficient preconditioners. The code will be tested for the most widely used isotherm functions such as the Steric Mass Action isotherm (Brooks and Cramer, 1992) and Mollerup's (2007) isotherms and will be used to study optimal closed loop control of chromatographic systems.

References

H. Birken, G. Gassner, M. Haas, and C.-D. Munz, 2002, Implicit time integration schemes for the unsteady compressible navier-stokes equations: Laminar flow. J. Comp. Phys., 179, 313-329

C.A. Brooks, and S.M. Cramer, 1992, Steric-mass-action ion exchange: displacement profiles and induced salt gradients, AiChE J., 38, 12, 1969-1978

M.H. Carpenter, and C.A. Kennedy, 1994, Fourth-order 2N-storage Runge-Kutta schemes, NASA-TM-109112, 1-24

P.V. Danckwerts, 1953, Continuous flow systems, Chem. Eng. Sci., 2, 1, 1-13

G. Guiochon, A. Felinger, and D.G. Shirazi, 2006, Fundamentals of Preparative and Nonlinear Chromatography. 2nd ed., Elesiver

G. Guiochon, and L.A. Beaver, 2011, Seperation Science is the Key to Successful Biopharmaceuticals, J. Chrom. A, 1218, 49, 8836-8858

E. Hairer, and G. Wanner, 1991, Solving ordinary differential equations II, Springer, 2nd revised ed.

J.S. Hesthaven, and T. Warburton, 2002, High-order nodal discontinuous Galerkin methods for the Maxwell eigenvalue problem, Philos. Trans. R. Soc. A, 362, 1816, 493-524

J.S Hesthaven, and T. Warburton, 2008, Nodal discontinuous Galerkin methods: Algorithms, Analysis, and Applications, Springer, New York

N. Houbak, S.P. Nørsett, and P.G. Thomsen, 1985, Displacement or Residual Test in the application of implicit methods for stiff problems, IMA J. Numer. Anal., 5,3,297-305

A. Kanevsky, M.H. Carpenter, D. Gottlieb, and J.S. Hesthaven, 2007, Application of implicit-explicit high order Runge-Kutta methods to discontinuous-Galerkin schemes, J. Comput. Phys., 225, 1753-1781

C.A. Kennedy, and M.H. Carpenter, 2003, Additive Runge-Kutta schemes for convection-diffusion-reaction equations, Appl. Numer. Math., 44, 1-2, 139-181

K. Meyer, J.K. Huusom, and J. Abildskov, 2018, High-order approximation of chromatographic models using a nodal discontinuous Galerkin approach, Comput. Chem. Eng., 109, 68-76

J.M. Mollerup, 2007, The thermodynamic principles of ligand binding in chromatography and biology, J. Biotechnol., 132, 2, 187-195

S.P. Nørsett, and P.G. Thomsen, 1986, Local error control in SDIRK-methods, BIT, 26, 1, 100-113

J. Sand, 1988, RK-predictors: Extrapolation methods for implicit Runge-Kutta formulae, DIKU Report Nr. 88/18, Datalogisk Institut, Københavns Universitet, Copenhagen,

M. Schlegel, W. Marquardt, R. Ehrig, and U. Nowak, 2004, Sensitivity analysis of linearly-implicit differential-algebraic systems by one-step extrapolation, Appl. Numer. Math., 48, 1, 83-102

L.F. Shampine, 1980, Implementation of implicit formulas for the solution of ODEs, SIAM J. Sci. Stat. Comput., 1,1,103-118

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50138-5

On the efficiency of PT Flash calculations with equations of state

Xiaodong Liang

Center for Energy Resources Engineering, Department of Chemical and Biochemical Engineering, Technical University of Denmark,Kgs. Lyngby, Denmark
xlia@kt.dtu.dk

Abstract

The isobaric-isothermal equilibrium (PT Flash) calculation has been an active research topic in thermodynamics community for decades, of which the conventional framework consists of two subproblems: stability test and phase split calculation. In this work, various aspects on the efficiency of the conventional PT Flash calculation procedures have been investigated, which includes the significances of a few successive substitution steps before starting the second-order method and volume based methods for both PT stability test and PT Flash problems, as well as the use of advanced equations of state.

Keywords: PT Flash, Stability Test, Phase Split, Successive Substitution, Second-order

1. Introduction

The isobaric-isothermal equilibrium (PT Flash) calculation is one of the most important one-stage equilibrium problems. In principle, the PT Flash calculation is a minimization problem, as the objective is to find the number of phases and their compositions that represent the minimum Gibbs free energy of the system. It has been an active research topic for decades. There are different approaches/frameworks for the PT Flash calculation. This work follows the approach proposed by Michelsen (1982a, 1982b, 2007), which consists of two equally important subproblems: stability test, checking if a new phase shall be added, and phase split calculation, determining equilibrium compositions. The PT Flash calculation could start from either of them. In this work, the calculation starts from the stability test, as it is assumed that there is no advanced knowledge available for the equilibrium factors. The output from the stability test is used to generate initial estimates for the phase split calculation, if it indicates that the systems is unstable. The phase split calculation then determines the equilibrium compositions and also take responsible for removing phases if necessary. When the phase split calculation finishes successfully, the stability of the equilibrium system is tested again, and the procedure iterates until the system is stable. It is possible to use either local global minimization methods for stability test, phase split calculation or both. It is generally considered that local minimization methods are more efficient, while the global minimization ones are safer. According to the previous experience, a well-organized PT Flash calculation procedure with a globally convergent local minimization algorithm is very robust, and it is the strategy used in this work. This study focuses on answering (1) if a few successive substitution (*SS*) and/or (2) volume-based methods can improve the efficiency for stability test and/or phase split calculation; (3) how much efficiency cost people need to pay when advanced equations of state (EOS) are used.

2. Methodology

The objective function for stability test, when temperature and pressure are specified, can be formulated (Michelsen, 1982a, 2007)

$$tm(\boldsymbol{n}) = 1 + \sum_i n_i[\ln n_i + \ln\varphi_i(T,P,\boldsymbol{n}) - \ln z_i - \ln\varphi_i(T,P,\boldsymbol{z}) - 1] \tag{1}$$

where n_i and z_i are mole numbers and mole fraction of component *i* of the trial phase and the tested phase, respectively, and $\boldsymbol{n}$ and $\boldsymbol{z}$ are the corresponding vectors. φ_i is the fugacity coefficient of component *i*. Equivalently, it can be written

$$tpd(T,V,\boldsymbol{n}) = \sum_i n_i\left[\frac{\mu_i(T,V,\boldsymbol{n})}{RT} - \frac{\mu_i(T,v_0,\boldsymbol{z})}{RT}\right] - V\left[\frac{P(T,V,\boldsymbol{n})}{RT} - \frac{P_0}{RT}\right] \tag{2}$$

where V is the total volume, μ_i is the chemical potential of component i, v_0 and P_0 are respectively the molar volume and pressure of the tested phase.

The objective function for phase split calculation can be written

$$Q = \sum_{j=1}^{\Pi}\sum_{i=1}^{C} n_{i,j}\ln f_{i,j}\left(T,P,\boldsymbol{n}_j\right) \tag{3}$$

where $f_{i,j}$ is the fugacity of component *i* in phase *j*. Π and C are number of phases and number of components, respectively. The objective function can also be built on the volume-composition space,

$$Q = \sum_{j=1}^{\Pi}\left[\sum_{i=1}^{C} n_{i,j}\frac{\mu_{i,j}\left(T,V_j,\boldsymbol{n}_j\right)}{RT} - \frac{P_j - P_0}{RT}V_j\right] \tag{4}$$

The general ideas of the solution methods for stability test and phase split calculation are from literature as listed in Table 1. There is no volume based (V-based) *SS* method available for stability test and phase split calculation, so whenever V-based and P-based (pressure based) are mentioned hereafter, they refer to the *SO* methods. The two *SO* methods for stability test and the P-based method for phase split calculation are treated as unconstrained minimization problems, solved by the combination of a modified Murry's method for Hessian matrix decomposition and a line search approach developed by Fletcher (2000). The V-based method for phase split calculation is solved by a Newton-Raphson method instead of an optimization method. Whenever the minimum phase volume becomes negative during iteration, the procedure switches to the P-based *SO* method.

Table 1. Solution methods for stability test and phase split calculation

Orders of method	Variable	Stability test	Phase split calculation
First-order (*SS*)	P-based	Michelsen [2007]	Michelsen [2007]
Second-Order (*SO*)	P-based	Michelsen [2007]	Michelsen [2007]
Second-Order (*SO*)	V-based	Nichita [2017]	Paterson [2017]

3. Models and test cases

The Soave-Redlick-Kwong (SRK) (Soave, 1972), the Peng-Robinson (PR) (Peng et al., 1976) and the Perturbed-Chain Statistical Associating Fluid Theory (PC-SAFT) EOS (Gross et al., 2001; von Solms et al., 2003) are the models used in this work. The investigated systems are listed in Table 2.

Table 2. Systems studied in this work

System	Description
S1	Example Ex5 from Nichita (2017)
S2	Example Ex7 from Nichita (2017)
S3	C_1-H_2S-CO_2 with composition [0.5, 0.4, 0.1]. The k_{ij} for PR from Li et al. (2012), while the k_{ij} for SRK and PC-SAFT are from Yan et al. (2015)
S4	An asphaltene fluid (Fluid-2) and the PC-SAFT parameters from Arya et al. (2016) The parameters for SRK from Arya et al. (2017)

4. Results and discussions

In this section, we present the results on various aspects of the computational efficiency. The investigations focus on: V-based versus P-based for stability test alone or the entire PT Flash calculation, number of *SS* steps before the *SO* methods and different EOS models. The core codes are implemented in FORTRAN and compiled by Intel Fortran Compiler X2011 into a DLL. Interfaces to MATLAB are created to manage data input and output, plot and analysis, with possibilities for secondary developments. All the calculations have been conducted in a laptop with Intel Core i5-5300U CPU@2.3GHz. Demonstration cases may be obtained from the author upon request.

4.1. PT Stability test

The average iteration numbers of the V-based and P-based *SO* methods, no *SS* steps conducted, for the stability test for systems S1 and S2 are given in Figure 1 (a) and (b), respectively. The temperature and density ranges for these two systems are taken from Nichita (2017), and the pressure is then calculated from the given temperature and density. The calculations are conducted only if the pressure is larger than 100Pa. Vapour and Liquid in the *x*-axis represent if the feed is considered as a vapour-like or liquid-like fluid. The results of the V-based *SO* method are similar to those presented by Nichita (2017). To a large extent, it can be seen that the P-based *SO* method uses less iterations to converge. As shown in Figure 1 (c) and (d), however, the V-based *SO* method costs less in terms of running time when no *SS* steps are taken. It has to point out that multiple initial estimates are used (Michelsen, 1982a, Castier, 2014) when counting the running time. The running time is an average time for one point (a given temperature and pressure), for which 1000 calculations have been repeated. It can be seen from Figure 1 (c) and (d) that a few *SS* steps can significantly reduce the running cost, while how many number of *SS* steps that shall be taken depends the system of interest. It is recommended 4-7 *SS* steps taken before the *SO* method, in which manner V-based and P-based approaches show comparable performance.

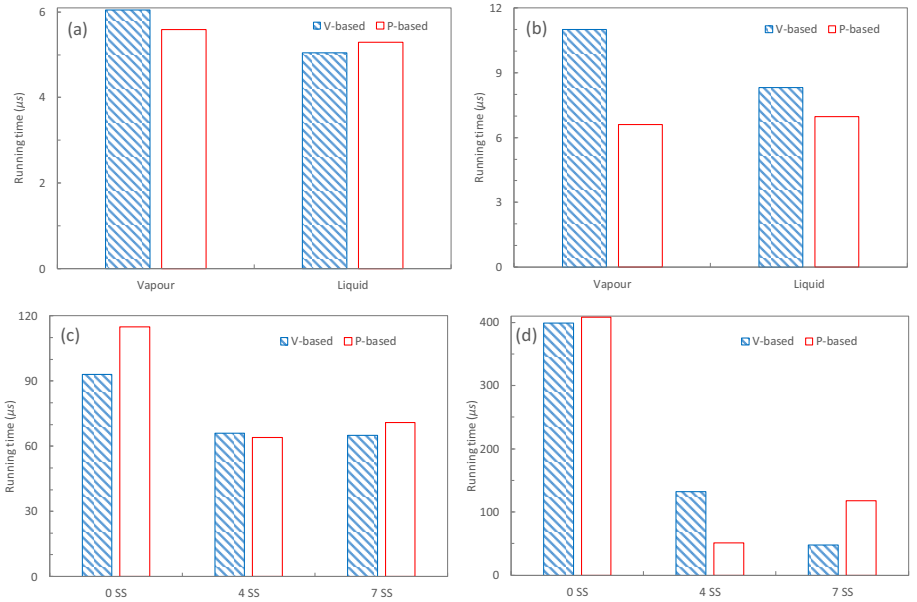


Figure 1. Average iterations for V-based and P-based *SO* methods for PT stability test for systems S1 (a) and S2 (b); The stability test running time of one point by using V-based and P-based *SO* methods following different number of *SS* steps for systems S1 (c) and S2 (d).

4.2. PT Flash (stability test + phase split)

The PT phase envelops of the systems S3 and S4 shown in Figure 2 are constructed by running the PT flash, consisting of stability test and phase split, calculations over wide ranges of temperature and pressure. Three models, PR, SRK and PC-SAFT, are applied for the system S3, while SRK, PC-SAFT (WOA) and PC-SAFT (WA) are used for the system S4. PC-SAFT (WOA) represents the modelling approach without association for asphaltene, and PC-SAFT (WA) is for that with association. More details are referred to Alay et al. (2016, 2017). As shown in Figure 2, PR, SRK and PC-SAFT might given different (Left) or similar (Right) results, which depend on the system and conditions under investigation as well as number of parameters and their tunings.

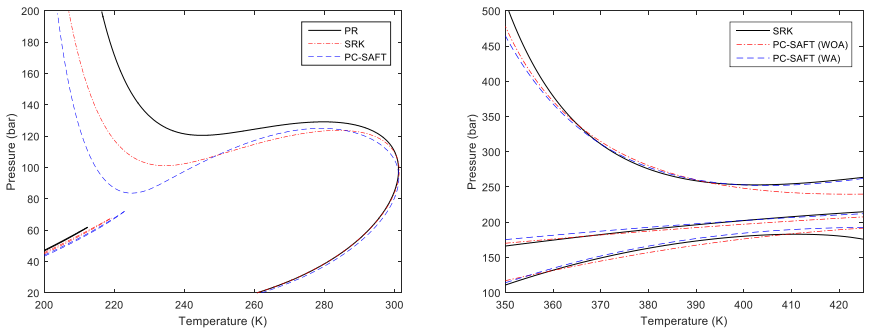


Figure 2. PT phase envelops of the systems S3 (Left) and S4 (Right).

The average PT Flash running times of one point are presented in Figure 3 (a) and (b) for systems S3 and S4, respectively. The ranges of temperature and pressure are: T(S3) = [200 to 300]K and P(S3) = 0.8(T-200) + 20bar, corresponding to a two-phase region shown in Figure 2 (Left); T(S4) = [350 to 425]K and P(S4) = 0.8(T-350)+140bar, corresponding to a three-phase region shown in Figure 2 (Right). As used by Paterson et al. (2017), 2 *SS* steps are taken before entering the *SO* methods in these calculations. The running time is averaged from 1000 repeated calculations. There is no doubt that PC-SAFT in general takes more time (with the same number of compounds), while the efficiency deterioration is acceptable (less than one time). The simplified PC-SAFT (von Solms et al. 2003) is used for the system S3. The efficiency deterioration for highly asymmetric systems may become less even with the original PC-SAFT (Gross et al. 2001) as given in Figure 3 (b). It is surprising that the introduction of association for asphaltene does not deteriorate the efficiency much, which means that association shall not be avoided on purpose because of its superficial complexity. A robust and efficient algorithm (Michelsen, 2006, 2007) is recommended for solving the association term.

It can be seen from Figure 3 that the V-based *SO* method can improve the efficiency by 10-30%, more pronounced with PC-SAFT. The effect of the number of *SS* steps before entering the *SO* methods (both V-based and P-based) on the overall efficiency has been presented for the systems S3 and S4 in Figure 4. It is clear that it is not necessary to take up to 15 *SS* steps, especially for PC-SAFT. The results indicate that 2-7 *SS* steps, preferably 4 *SS* steps, could be recommended in real applications.

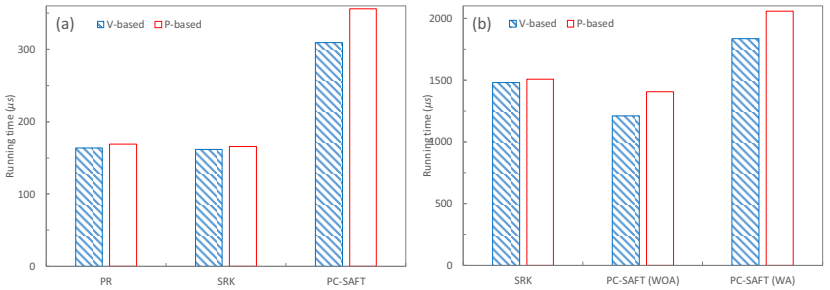


Figure 3. Comparison of the PT Flash running times of the V-based and P-based *SO* methods for phase split calculations and different models: systems S3 (a) and S4 (b).

5. Conclusions

In this work, the various aspects on the efficiency of PT Flash calculations with EOS are investigated. First, in order to study if V-based methods can improve the efficiency, the objective functions of the subproblems of a PT Flash calculation, i.e. stability test and phase split calculation, are formulated in both P-based and V-based methods. The results reveal that V-based and P-based methods perform similarly for a PT stability test in terms of running time after a few *SS* steps, which are necessary. Second, the study on the significance of a few *SS* steps before the *SO* method shows that 4-7 *SS* steps are good choices for stability test and 2-7 *SS* steps, preferably 4 *SS* steps, are recommended for phase split calculations. Third, the efficiency deterioration of using PC-SAFT compared to cubic EOS is acceptable, less than one time, and may become even less as the system becomes more asymmetric. Forth, the association term shall not be avoided on purpose due to a worry of efficiency deterioration.

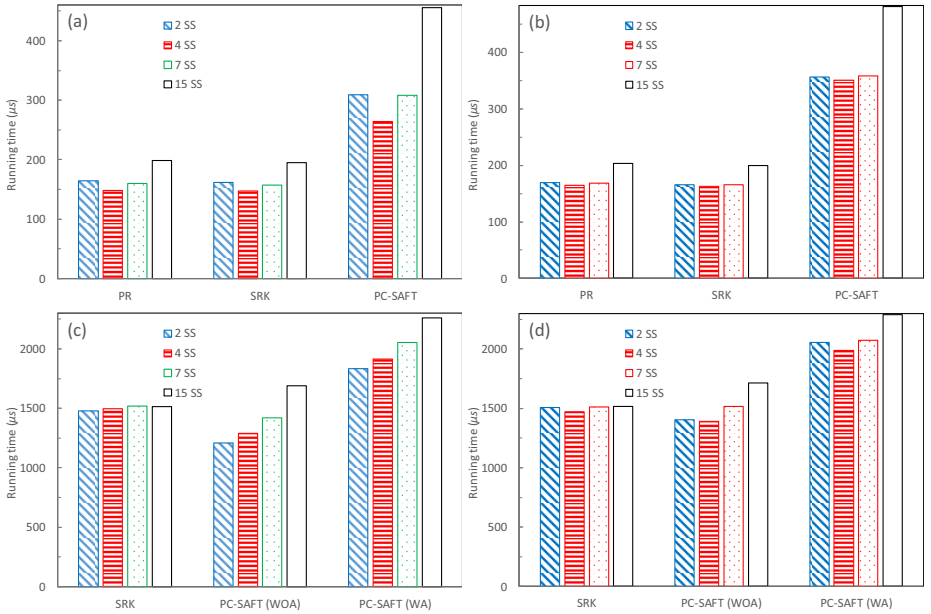


Figure 4. Comparison of the PT Flash running times of one point starting with different number of *SS* steps before using V-based and P-based *SO* methods for phase split calculations and different models. System S3: (a) V-based *SO* method; (b) P-based *SO* method; System S4: (c) V-based *SO* method; (d) P-based *SO* method.

References

A. Arya, X.D. Liang, N. von Solms, G.M. Kontogeorgis, 2016, Energy & Fuels, 30, 6835-6852.

A. Arya, X.D. Liang, N. von Solms, G.M. Kontogeorgis, 2017, Energy & Fuels, 31, 3313-3328.

R. Fletcher, 2000, Practical methods of optimization. John Wiley and Sons: New York.

J. Gross, G. Sadowski, 2001, Ind. Eng. Chem. Res. 40, 1244-1260.

Z.D. Li, A. Firoozabadi, 2012, SPE J. 17, 1096-1107.

M.L. Michelsen, 1982a, Fluid Phase Equilib. 9, 1-19.

M.L. Michelsen, 1982b, Fluid Phase Equilib. 9, 21-40.

M.L. Michelsen, J.M. Mollerup, 2007, Thermodynamic Models: Fundamentals & Computational Aspects; Tie-Line Technology: Holte, Denmark.

M.L. Michelsen, 2006, Ind. Eng. Chem. Res., 45, 8449-8453.

D.V. Nichita, 2017, Fluid Phase Equilib. 447, 107-124.

D. Paterson, M.L. Michelsen, E.H. Stenby, W. Yan, 2017, SPE-182706-MS.

D.Y. Peng, D.B. Robinson, 1976, Ind. Eng. Chem. Fundam. 15, 59-65.

G. Soave, 1972, Chem. Eng. Sci. 27, 1197-1203.

N. von Solms, M.L. Michelsen, G.M. Kontogeorgis, 2003, Ind. Eng. Chem. Res. 42, 1098-1105.

W. Yan, F. Varzandeh, E.H. Stenby, 2015, Fluid Phase Equilib. 386, 96-124.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50139-7

A comparative study between GDP and NLP formulations for conceptual design of distillation columns

Jaffer H. Ghouse[a], Qi Chen[b], Miguel A. Zamarripa[a], Andrew Lee[a], Anthony P. Burgard[a], Ignacio E. Grossmann[b], David C. Miller[a*]

[a]*National Energy Technology Laboratory, Pittsburgh, PA 15236*
[b]*Dept. of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213*
David.Miller@netl.doe.gov

Abstract

In this work, a rigorous tray-by-tray distillation column model is presented based on a new open-source modeling framework built on Pyomo that is designed to specifically support optimization of steady state and dynamic processes. The modeling framework allows for process synthesis and design of conventional tray columns either as an NLP or as a GDP problem. The model was used to simulate a distillation column for a binary mixture using three different property packages (Ideal, NRTL and Peng-Robinson), and the results are compared to results from existing commercial tools. Furthermore, a comparative case study is presented for an optimal column design problem using both frameworks.

Keywords: conceptual design, optimization, modelling framework, distillation

1. Introduction

In the past two decades, the advent of inexpensive computing power and improved numerical algorithms has enabled the increased use of equation oriented (EO) process models in process systems engineering. Though the current commercial equation oriented simulators can be used to model complex unit operations, the platforms do not provide the flexibility required for solving large-scale conceptual design optimization problems. Typically, such design problems are either solved with simplifying assumptions that lead to sub-optimal solutions or with rigorous models that are developed from the ground up, which is both time and labor intensive. The optimal design of a distillation column is an open and extensively researched problem in conceptual design owing to its importance in the process industries. Optimization problems concerning distillation columns can be broadly classified into two categories: (i) minimizing the operating cost subject to purity constraints for a fixed column design and (ii) minimizing the capital and operating cost subject to purity constraints. The former is relatively easy to solve with existing commercial simulation tools, but the latter commonly requires workarounds to overcome challenges such as limited compatibility for deterministic optimization and the lack of initialization routines with existing tools. This is computationally difficult for two reasons: (1) the presence of discrete decision variables for the number of trays and the feed tray location and (2) the non-linearity of the problem when using rigorous models.

Both a Generalized Disjunctive Programming (GDP) and NLP formulations have been proposed to model the discrete decision variables in this problem. The GDP approach

may be solved with a logic-based outer-approximation (LOA) algorithm, where the trays are activated/deactivated using Boolean variables (Yeomans and Grossmann, 2000). In the NLP approach, integer variables are avoided and trays are deactivated using bypass streams (Dowling and Biegler, 2015). The GDP approach simplifies the problem by eliminating the equations associated with deactivated trays. However, the solution to the GDP problem may require several major iterations (Barttfeld et al., 2004).

On the other hand, the NLP approach avoids the combinatorial problem associated with binary variables, but requires solving a larger NLP problem. In general, both approaches provide robust methods that can be solved with existing solvers; however, (i) a comparative study between the two methods is lacking (Dowling and Biegler, 2015) and (ii) distillation models available in existing commercial simulation tools are not compatible with either approach. The objective of the study is to present a comparison of the methodologies, which have been implemented within the IDAES conceptual design framework. A rigorous tray-by-tray unit model that has been developed to optimize the size, feed tray location and operating conditions of the distillation column is also presented.

2. Unit Operations Modelling and Optimization – the IDAES approach

The goal of the Institute for the Design of Advanced Energy Systems (IDAES), a U.S. Department of Energy initiative, is to build an open-source, next generation process systems engineering framework that will help accelerate the development of advanced energy systems. The IDAES modeling framework is based on Pyomo (Hart et al. 2017) – a Python-based algebraic modeling language that leverages the capabilities of a high level programming language to support optimization. One of the major objectives of the IDAES framework is to simplify the formulation and solution of conceptual design problems encountered in process systems engineering, thus addressing limitations with existing commercial tools. To this end, the models that are developed within this framework will be specifically designed for optimization of steady state and dynamic processes.

3. Distillation Tray Column Model

The unit model for the conventional tray-by-tray column in IDAES consists of the standard MESH equations for a distillation column. The model can be used to simulate a tray column with or without a condenser (total or partial) and a reboiler. The trays are numbered from top to bottom where the condenser and reboiler are not included as trays i.e. a distillation column with ten trays has ten trays plus a condenser and reboiler. For every tray there is an option for a feed inlet, outlet side draws for the liquid and vapor streams and heat addition/removal. The IDAES PSE framework is modular such that the user has flexibility to add or remove objects. For example, the property methods are instantiated as separate objects at the flowsheet level keeping them independent of the unit models. This provides the flexibility to easily change property calculations without making any changes to the model equations. At every tray, there is a property block for the feed stream, a property block for the mixer outlet and a property block for the equilibrium separator. The property blocks contain the equations (or constraints) for a flash calculation and also to compute properties like the specific enthalpy, vapor pressure and equilibrium coefficient as a function of the local state variables.

A major limitation for distillation column models that employ the MESH equations is the lack of consistent and robust initialization schemes. The IDAES framework ensures that sufficient flexibility is available to simplify the initialization procedure. For the column model in IDAES, the following three methods have been built to initialize the column:

- **Method 1:** The trays are initialized one by one starting from the top tray.
- **Method 2:** All trays are initialized together by sequentially activating and solving the mass, pressure, and energy balances.
- **Method 3:** Similar to method 1 but the feed stage is solved first followed by the rectification section and then the stripping section.

4. Optimization Framework for Distillation Columns in IDAES

4.1 NLP framework

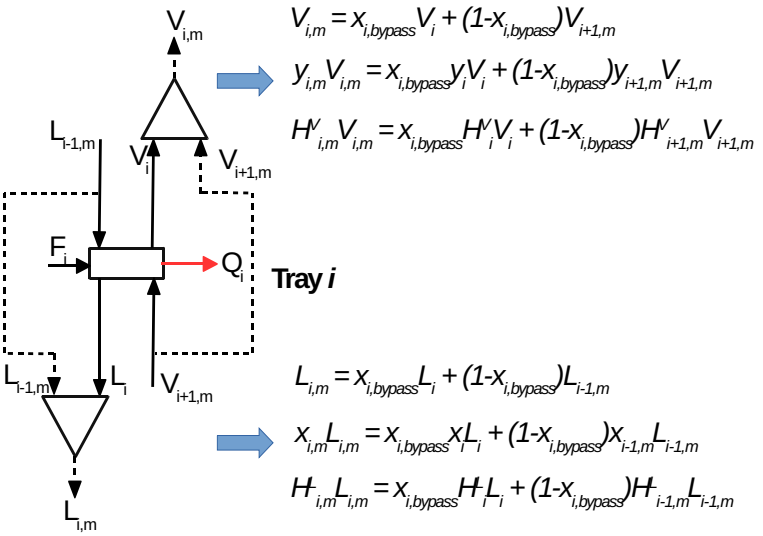


Figure 1: NLP framework with bypass streams

The NLP framework that has been implemented for solving conceptual design problems concerning distillation columns in IDAES is based on Dowling and Biegler (2015). The formulation allows trays to be bypassed using bypass variables ($x_{i,bypass}$) as shown in Figure 1. Typically, these variables should be binary where 1 denotes active and 0 denotes inactive trays. However, in the NLP formulation these variables are continuous variables bounded between 0 and 1, i.e., a relaxed MINLP problem. The general consensus while using bypass variables in distillation columns is to split the tray outlet streams (liquid/vapor) to two streams: a bypassed stream and a stream input to the tray above or below. While this formulation allows trays to be bypassed, using a conventional splitter results in a zero flow if the tray is bypassed which will affect convergence of the flash calculations on the tray. Instead, this framework duplicates the bypass stream and employs an "unconventional mixer" for the vapor and liquid outlets such that zero flows are avoided through bypassed trays. The mass/energy balances for these mixers are shown in Figure 1. When a tray is bypassed, the bypass variable $x_{i,bypass}$

should be 0. Then from the mass and energy balances for the mixer, it can be observed that the inputs to next tray (above/below) are appropriately accounted for depending on whether the tray is active or inactive. Furthermore, engineering knowledge tells us that partial bypass will be avoided as it is inefficient to mix two streams of different purities and separate them again. Another feature with this framework is that it indirectly, concurrently optimizes for the optimal feed tray location. Although the feed tray is fixed, the bypass variables are restricted to trays in the stripping and rectification sections. Thus, the feed tray location is determined depending upon the number of trays eliminated in each section.

4.2 GDP framework

The GDP framework for conceptual design of distillation columns is based upon the formulation given in Barttfeld et al. (2003). Disjunctions at each candidate tray describe the presence or absence of the tray. The general form of the disjunctions is given in Equation 1.

$$\begin{bmatrix} Y_t \\ g_1(x) \leq 0 \end{bmatrix} \vee \begin{bmatrix} \neg Y_t \\ Ax \leq b \end{bmatrix} \tag{1}$$

Y_t is the Boolean variable associated with tray existence. Constraints denoted by g_1 contain the MESH equations for each stage. When the tray does not exist, linear bypass constraints $Ax \leq b$ ensure that flowrates and enthalpies are transmitted unchanged to the next stage. An analogous disjunction governs selection of the feed tray: a tray either is the feed stage and has the relevant inlet material and energy flows OR it is not and the flows are set to zero. A logical proposition enforces that if a tray is selected as the feed, then it implies that the tray is also active. The first and final stages correspond to the condenser and reboiler; the Boolean existence variables associated with these trays are fixed to True.

5. Results and Discussion

The IDAES tray-by-tray column model was used to simulate a distillation column with ten trays (+ condenser + reboiler) to separate an equimolar mixture of benzene and toluene using three different property methods: Ideal, Ideal-NRTL and the Peng-Robinson cubic equation of state. The feed specifications are as follows: Feed = 100 mol/s, T_{feed} = 368 K, P_{feed} = 101 kPa (no pressure drop), feed tray = 5. The molar reflux and reboil ratio were set to 1.4 and 1.3 respectively. The results were compared to a "RadFrac" model in Aspen Plus and are presented in Table 1.

The results using the EO IDAES model show an excellent match with Aspen Plus models solved sequentially. The model was successfully initialized using "Method 2" for all the three simulations and demonstrates robustness even when using a rigorous property method like the cubic equation of state.

5.1 Optimal design problem

The problem for the optimal design of a distillation column has been formulated as follows:

$$\min_{RF,RB,\, feed\ tray, N_{trays}} 10^3 \left(\sum_{i=1}^{N_{trays}} x_{bypass,i}\right) + (Q_R + Q_C)$$

$$s.t.: x_{D,benzene} \geq 0.95\ ;\ x_{B,toluene} \geq 0.95;\ RF \leq 4;\ RB \leq 4 \tag{2}$$

The objective function minimizes the capital cost (number of active trays) and operating cost (reboiler and condenser duty). The number of active trays is multiplied by a cost coefficient of 1000 to scale it to the same range as that of the condenser and reboiler duties in kW so as to ensure equal weights to the capital and operating cost in the objective function. The feed conditions are the same as those used for simulation studies outlined in Table 1 (Ideal case). For the NLP framework, the model was initialized with 35 active trays (feed tray = 10) and initial values for the decision variables were unchanged from the simulation case study. The solver used was IPOPT with MA 27 as the linear solver. The optimal solution using the NLP framework is outlined in Table 2,

Table 1: Comparison of model predictions using IDAES and Aspen Plus models

Variable	Ideal		Ideal – NRTL		Peng-Robinson	
	IDAES	Aspen	IDAES	Aspen	IDAES	Aspen
Distillate (mols/)	47.76	47.71	47.94	47.82	45.31	45.33
$x_{D,benzene}$	0.886	0.887	0.885	0.887	0.900	0.900
$x_{D,toluene}$	0.114	0.113	0.115	0.113	0.100	0.100
Bottoms (mol/s)	52.24	52.29	52.06	52.18	54.69	54.67
$x_{B,benzene}$	0.147	0.147	0.145	0.145	0.168	0.168
$x_{B,toluene}$	0.853	0.853	0.855	0.855	0.832	0.832
$Q_{condenser}$ (kW)	-3626.4	-3594.5	-3639.9	-3591	-3375.2	-3378.7
$Q_{reboiler}$ (kW)	2307.9	2272.2	2300.6	2245.7	2366.8	2366.4

and it can be seen that the total number of trays is an integer value (11 in this case) even though the bypass variables were declared as continuous variables. Also, the optimal feed location is tray 6. The same solution was obtained when the model was initialized with 15/20/50/60 active trays or when the feed location was changed to 10/15/25/30 in the superstructure respectively. The GDP model was solved using the GDPopt solver in Pyomo, which implements LOA and uses both Gurobi and IPOPT as sub-solvers. The GDP model was initialized using the column profile results from a 10-stage square problem; to accommodate additional conditional trays, a linear interpolation of the profile was used. Even though the formulation with the bypass framework consists of a larger NLP model, it required 597 function evaluations to determine the optimal solution. In comparison, the GDPopt terminated after 20 master iterations requiring 2057 NLP function evaluations in total, mainly due to weak linear approximations. However, this may be avoided with better reformulation techniques.

Table 2: Optimal solutions from NLP and GDP frameworks

Optimal solution	NLP framework	GDP framework
Objective	19430	19450
No. of trays	11	10
Feed tray	6	5
Reflux ratio	2.07	2.45
Reboil ratio	2.13	2.39
$x_{D,benzene}$/ $x_{B,toluene}$	0.95/0.95	0.95/0.95

Both the NLP and GDP frameworks yield similar solutions. The NLP framework solution includes one extra tray to reduce column duties, making a slightly different capital vs. operating cost trade-off compared to the GDP framework.

6. Conclusions

This work presents a tray-by-tray distillation column model using MESH equations within the IDAES modeling framework that is suitable for both simulation and deterministic optimization. For the conceptual design of distillation columns, both NLP and GDP frameworks are available, and a comparative case study was presented in this work. In the case study, the NLP and GDP frameworks yield similar solutions; however, the NLP framework using bypass streams requires fewer nonlinear function evaluations compared to the GDP solution, but a robust initialization scheme was necessary. On the other hand, the main limitation of the GDP framework was the quality of the linear approximation, leading to slower convergence. This may be addressed by adding support for stronger GDP to MILP reformulation techniques in GDPopt, improving the bounds via automatic bound-strengthening tools and adding logical propositions to screen out structurally redundant configurations. At the same time, the impact of the reduced space sub-problems in LOA can be seen, as only a few function evaluations are required for each sub-problem, even without a complex initialization scheme. A thorough comparison between the two frameworks will be considered in the future and will incorporate more rigorous property models such as the cubic equation of state.

Disclaimer

 This research was supported in part by an appointment to the National Energy Technology Laboratory Research Participation Program, sponsored by the U.S. Department of Energy and administered by the Oak Ridge Institute for Science and Education. KeyLogic Systems, Inc.'s contributions to this work were funded by the National Energy Technology Laboratory under the Mission Execution and Strategic Analysis contract (DE-FE0025912) for support services. This material is based upon work supported by the U.S. Department of Energy, Office of Science, under Contract No. DEAC02-05CH11231. This material is based upon work supported by the U.S. Department of Energy, Office of Science, under Contract No. DEAC02-05CH11231.

References

M. Barttfeld, P.A. Aguirre, I.E. and Grossmann, 2003. Alternative representations and formulations for the economic optimization of multicomponent distillation columns. Computers and Chemical Engineering 27: 363-383.

M. Barttfeld, P.A. Aguirre, I.E. and Grossmann, 2004. A decomposition method for synthesizing complex column configurations using tray-by-tray GDP models. Computers and Chemical Engineering 28: 2165-2188.

A.W. Dowling, and L.T. Biegler, 2015. A framework for efficient large scale equation-oriented flowsheet optimization. Computers and Chemical Engineering 72: 3-20.

W.E. Hart, C.D. Laird, J.-P. Watson, D.L. Woodruff, G.A. Hackebeil, B.L.Nicholson, and J.D. Siirola, 2017, Pyomo – Optimization Modeling in Python. Second Edition. Vol. 67. Springer.

H. Yeomans, and I.E. Grossmann, 2000. Optimal Design of Complex Distillation Columns Using Rigorous Tray-by-Tray Disjunctive Programming Models. Industrial and Engineering Chemistry Research 39(11):4326-4335.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50140-3

A Smooth, Square Flash Formulation for Equation-Oriented Flowsheet Optimization

Anthony P. Burgard[a,*], John P. Eason[c], John C. Eslick[a], Jaffer H. Ghouse[a], Andrew Lee[b], Lorenz T. Biegler[c], David C. Miller[a]

[a] *National Energy Technology Laboratory (NETL), 626 Cochrans Mill Road, Pittsburgh, PA, 15236, U.S.A.*
[b] *KeyLogic Systems, NETL, 626 Cochrans Mill Road, Pittsburgh, PA, 15236, U.S.A.*
[c] *Department of Chemical Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA, 15213, U.S.A.*
Anthony.Burgard@netl.doe.gov

Abstract

This work introduces a novel flash formulation to calculate vapor-liquid equilibria (VLE) based on a general cubic equation of state (cEOS). The approach leverages smooth max operators to implement complementarity conditions leading to a model with an equal number of variables and constraints (i.e., a square problem). The model correctly calculates equilibrium phase fractions and compositions over a broad range of temperatures and pressures by ensuring that the VLE calculation is only enforced below the critical pressure and between the bubble point and dew point temperatures. We demonstrate the robustness of the flash model by embedding it directly into optimization problems that require moving from initial temperatures and pressures across phase transition boundaries as well as between subcritical and supercritical regions. The flash model is implemented in the open source, Python-based, algebraic modeling language, Pyomo. The roots of the cEOS are calculated analytically using an external function call that returns exact 1st and 2nd derivatives and is embedded directly via the AMPL Solver Library. The model will be publicly available as part of the Institute for the Design of Advanced Energy Systems (IDAES) process systems engineering framework.

Keywords: vapor-liquid equilibrium, equation-oriented flowsheet optimization, cubic equation of state, mathematical programs with complementarity constraints

1. Introduction

Flash calculations that require non-ideal vapor-liquid equilibrium (VLE) relationships are notoriously challenging to implement in an equation-oriented (EO) flowsheet optimization environment. Traditional solution algorithms are susceptible to failure in regions where only one phase exists since VLE relationships are only valid in the two-phase region between the bubble point and dew point conditions. Robust EO implementations have been achieved by formulating flash calculations as mathematical programs with complementarity constraints. For example, complementarities that relax the VLE when either the vapor or liquid phase is absent can be transferred to the objective function of an optimization problem via penalty terms (Kamath et al., 2010; Dowling and Biegler, 2015; Dowling et al., 2015). Choosing the correct weightings for the penalty terms, however, may require extensive trial-and-error to ensure that the VLE conditions are properly enforced but do not dominate the optimization leading to suboptimal solutions.

The square flash model presented here can be solved without optimization or adding penalty terms to an objective function. It greatly facilitates a modular approach to EO flowsheet construction as unit operations involving flash calculations do not have to be initialized by solving sets of optimization problems or by implementing complex solution algorithms. The following sections describe the methodology and demonstrate the robustness of the formulation when implemented in its square form (i.e., zero degrees of freedom) or as part of optimization problems (i.e., ≥ 1 degree of freedom) that require moving across phase boundaries and between subcritical and supercritical regions.

2. Problem Formulation

2.1. General Flash Formulation

The standard form for calculating the liquid and vapor phase flow rates and compositions for a mixture of *i* components is:

$$F = L + V \quad , \quad Fz_i = Lx_i + Vy_i \;\; \forall \;\; i \quad , \; 0 = \sum_i y_i - \sum_i x_i \tag{1-3}$$

$$y_i = K_i(x, Z^L(x,T,P), y, Z^V(y,T,P), T, P)x_i \;\; \forall \;\; i \tag{4}$$

$$Fh_F = Lh_L(x, Z^L(x,T,P), T, P) + Vh_V(y, Z^V(y,T,P), T, P) - Q \tag{5}$$

where F, L and V are the flow rates of the inlet, outlet liquid, and outlet vapor, respectively, while z, x, and y are their mole fractions. The equilibrium ratios, K_i, are defined by an equation of state and are a function of the temperature, pressure, and the liquid and vapor phase compositions. h_F, h_L, and h_V are the specific enthalpies of the inlet, outlet liquid, and outlet vapor streams, while Q is the heat duty. The general flash formulation is necessary but not sufficient to predict the proper number of phases. Thus, it must be supplemented with additional constraints to ensure thermodynamic stability. In this work, we implement the Peng-Robinson Equation of State described in Dowling et al. (2015), though the framework described herein is general enough to handle essentially any cEOS.

2.2. Calculation of Z^L and Z^V

The liquid and vapor phase compressibility factors, Z^L and Z^V, correspond to roots of the cubic equation where A and B are functions of the temperature, pressure and composition, while u and w are equal to 2 and -1, respectively, for the Peng-Robinson EOS.

$$Z^3 - (1 + B - uB)Z^2 + (A + wB^2 - uB - uB^2)Z - AB - wB^2 - wB^3 = 0 \tag{6}$$

One particularly effective method to embed Eq. (6) into EO flowsheet optimizations entailed using inequality constraints on its first and second derivatives to properly assign the smallest root as Z^L and the largest root as Z^V (Kamath et al., 2010; Dowling and Biegler, 2015). Slack variables were employed to relax the constraints for conditions where the liquid or vapor root disappears. More recent approaches have proposed modifying the classical inside-out algorithms using non-smooth equations to relax equilibrium conditions when appropriate (Watson, et al., 2017) or applying a modified root discrimination procedure based on Cardano's method (Glass and Mitsos, 2017). In this work, rather than adding Eq. (6) directly, Z^L and Z^V are calculated analytically using an AMPL Solver Library user-defined function that takes A and B as arguments, and returns the exact solution, using the cubic formula, along with exact 1st and 2nd derivatives. When the cubic equation has more than one root, it is assumed the smallest root is Z^L and the largest is Z^V. If only one real root exists, the root calculating function provides two options. The first option, employed for Eq. (5) only, returns the same root for both the liquid and vapor. This is a discontinuous function with respect to A and B. Also, the derivatives are unde-

fined where there is a double or triple root. Nevertheless, because Z^L and Z^V are not variables in the model, their discontinuity is essentially smoothed by the continuous nature of Eq. (5). For example, the vapor root disappears in the liquid region. Thus in Eq. (5), h_V will be multiplied by zero since V=0. This option becomes necessary in the supercritical region where it is not possible to classify a root as vapor or liquid. The second option returns a fictitious, though continuous, solution for the non-existing root (Z^L or Z^V) while providing the exact solution for the one root that exists. The fictitious solution is found by generating two distinct curves: one for the liquid and one for the vapor. When the cubic function $f(Z)$ (left-hand side of Eq. 6) has two points where the $f'(Z) = 0$, the curves are extended symmetrically from those points to create monotonic functions. When there are no points where $f'(Z) = 0$, the liquid and vapor curve are the same and match the original cubic. Thus, the same Z is returned for both liquid and vapor. This option is employed in all equilibrium equations since our approach ensures that they are applied only between the bubble point and dew point temperatures. Thus, the true solution will always be in the two-phase region where the fictitious root does not exist. The benefits of calculating Z^L and Z^V using an external function are that it reduces the number of variables and constraints per flash calculation, eliminates the need to initialize Z^L and Z^V since they are not variables, keeps the problem square by avoiding inequality constraints, and provides continuous values for Z^L and Z^V for equilibrium equations even when one of the roots disappears. One potential disadvantage is that the algebraic modeling language must support the ability to embed user-defined function calls.

2.3. Flash Model for Subcritical Pressures

A major challenge with applying cEOS thermodynamic models in large-scale flowsheet optimizations is that the problem can be nonconvex and attain nonglobal solutions that represent nonphysical equilibria. In particular, convergence to solutions where $Z^V = Z^L$, $K = 1$, and $x = y$ is a known problem (Gunderson, 1982) that can lead to arbitrary phase assignments when the mixture is not at the critical point or in the supercritical region. When solving square flash models, careful initialization can often avoid convergence to these unwanted "trivial" flash solutions; however, during optimization with ≥ 1 degree(s) of freedom, the modeler has less control over the search path chosen by the optimizer motivating the active elimination of trivial solutions via additional constraints. This is accomplished by calculating the bubble point and dew point temperatures for the mixture Eqs. (7-10) and constraining the temperature at which Eq. (4) is evaluated to be between these two temperatures Eq. (11-12). This differs from the approach proposed by Watson, et al. (2017) which relaxes the equilibrium relationship outside of the two-phase region.

$$y_i^s = K_i(y^s, Z^V(y^s, T_{Bub}, P), z, Z^L(z, T_{Bub}, P), T_{Bub}, P)z_i \;\forall\; i \tag{7}$$

$$z_i = K_i(z, Z^V(z, T_{Dew}, P), x^s, Z^L(x^s, T_{Dew}, P), T_{Dew}, P)x_i^s \;\forall\; i \tag{8}$$

$$0 = \sum_i y_i^s - \sum_i z_i \;\;,\;\; 0 = \sum_i z_i - \sum_i x_i^s \;,\; T_{Eq} = \min(\max(\mathrm{T}, T_{Bub}), T_{Dew}) \tag{9-11}$$

$$y_i = K_i\big(y, Z^V(y, T_{Eq}, P), x, Z^L(x, T_{Eq}, P), T_{Eq}, P\big)x_i \;\forall\; i \tag{12}$$

Here y_i^s and x_i^s are the vapor and liquid mole fractions, respectively, for "shadow" streams used to calculate the bubble and dew point temperatures of the incoming mixture at the flash pressure, P. In this work, the non-smooth Eq. (11) is replaced by Eqs. (13-14) which ensure that T_{Eq} is smooth and continuous for all values of T.

$$T_1 = 0.5 \cdot \left(T + T_{Bub} + \sqrt{(T - T_{Bub})^2 + \varepsilon_1^2}\right) \cong \max(T, T_{Bub}) \tag{13}$$

$$T_{Eq} = 0.5 \cdot \left(T_1 + T_{Dew} - \sqrt{(T_1 - T_{Dew})^2 + \varepsilon_2^2}\right) \cong \min(T_1, T_{Dew}) \tag{14}$$

In general, ε_1 must be greater than ε_2. We chose $\varepsilon_1 = 0.01$ and $\varepsilon_2 = 0.0005$. Eqs. (1-3, 5, 7-10, and 12-14) thus comprise a smooth square flash model with $7 + 4n$ equations (where n is the number of components) that avoids convergence to trivial flash solutions, requires initialization of only physically intuitive variables, and is applicable for all $P < P_c$.

2.4. Flash Model Extension for Supercritical Conditions

The flash model can be extended to handle cases for which the modeler does not know *a priori* if a given stream will be below or above its critical point. To this end, we propose partitioning the phase diagram into two halves and assigning phases as follows: $P > P_c$ (Liquid only), $P < P_c$ (Liquid and/or Vapor). The rationale is that fluid densities above the critical pressure will be liquid-like and more suited to pumps than compressors. We note that this approach will lead to inaccuracies within the small range of pressures between the critical pressure, P_c, and the cricondenbar, P_{cc} (i.e., the maximum pressure at which vapor can exist for a mixture), when $P_{cc} > P_c$. Nevertheless, since few processes are designed to operate in this range, the approach proposed here can be applied with confidence to large-scale flowsheet optimization problems assuming post-solution checks are made to ensure that pressures are either below P_c or above P_{cc}. The supercritical extension first entails adding Eqs. (15-16) to enable calculation of the mixture's critical temperature and pressure using the Peng-Robinson EOS.

$$0.45724 = a^m(z)P_c/(T_c^2R^2)\ ,\ 0.077796 = b^m(z)P_c/(T_cR) \tag{15-16}$$

Both a^m and b^m are functions of the inlet composition as described by Dowling et al. (2015). Next, the variables P_m and P_{mr}, are introduced and assigned such that they can be used to determine whether the flash pressure, P, is above or below P_c.

$$P_c \cdot P_m = 0.5 \cdot \left(P - P_c + \sqrt{(P - P_c)^2 + \varepsilon_3^2}\right) \cong \max(0, P - P_c) \tag{17}$$

$$P_c \cdot P_{mr} = 0.5 \cdot \left(P_c - \mathrm{P} + \sqrt{(P_c - \mathrm{P})^2 + \varepsilon_4^2}\right) \cong \max(0, P_c - P) \tag{18}$$

In this work, ε_3 and ε_4 were chosen to be 0.0005. We then replace P in the VLE equations 7, 8, and 12 with $P - P_cP_m - \varepsilon_4/4 \cong \min(P, P_c)$. This ensures that the VLE equations are never evaluated at pressures above the P_c. It also prevents T_{Eq}, T_{Bub}, and T_{Dew} from assuming arbitrary values when $P > P_c$. Finally, Eq. (14) is replaced by Eq. (19) while Eq. (20) ensures that $T_{Eq} = T_2$ when $P < P_c$ (i.e., $P_m = 0$ and $P_{mr} > 0$) thus enforcing the VLE equations only at subcritical pressures. Eq. (20) also forces the vapor flow rate, V, to be equal to zero when $P > P_c$ (i.e., $P_{mr} = 0$ and $P_m > 0$).

$$T_2 = 0.5 \cdot \left(T_1 + T_{Dew} - \sqrt{(T_1 - T_{Dew})^2 + \varepsilon_2^2}\right) \cong \min(T_1, T_{Dew}) \tag{19}$$

$$\left(T_2 - T_{Eq}\right) \cdot P_{mr} + V \cdot P_m = 0 \tag{20}$$

These modifications lead to a final smooth square flash formulation comprising (Eqs. 1-3, 5, 7-10, 12-13, and 15-20) that contains $12 + 4n$ variables and equations. When applied towards optimization problems, it is sometimes beneficial to replace the right-hand side of Eq. (20) with a small value (e.g., 10^{-7}) to avoid converging to locally optimal solutions where $P = P_c$ and thus both P_m and P_{mr} are near zero. We, however, recommend replacing Eq. (20) with the inequalities below during optimization as this approach has proven most effective at avoiding $P = P_c$ local solutions.

$$V \cdot P_m \leq \left(T_2 - T_{Eq}\right) \cdot P_{mr} \leq V \cdot P_m \tag{21}$$

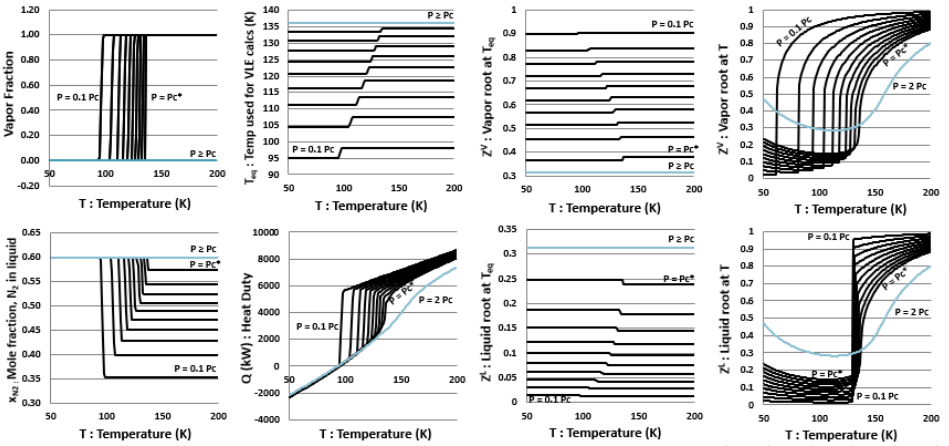


Figure 1: Flash results for the N_2, O_2, and Ar mixture. $P_c^* = 0.99998\ P_c$.

3. Results

3.1. Flash Model Implementation

The flash formulation comprising (Eqs. 1-3, 5, 7-10, 12-13, and 15-20) was applied to an inlet liquid mixture at 95 K and 7 bar containing 60 % N_2, 35 % O_2, and 5 % Ar. The thermodynamics of the two phases were modeled using the Peng-Robinson EOS. Parameters for this study can be found in Dowling and Biegler (2015). The critical temperature and pressure for the mixture were found to be 136.5 K and 3,947,426 Pa, respectively. The equilibrium was calculated at 1 K increments from 50 to 200 K at several pressures: 0.1 P_c, 0.2 P_c, 0.3 P_c, 0.4 P_c, 0.5 P_c, 0.6 P_c, 0.7 P_c, 0.8 P_c, 0.9 P_c, 0.99998 P_c, and 2 P_c. Results are shown in Figure 1. All calculations were performed using IPOPT with the following settings: linear_system_scaling = mc19, linear_scaling_on_demand = no, recalc_y = yes. Notably, correct phase transitions and heat duties were obtained at all temperatures and pressures. The outlet is classified as all liquid for all $P > P_c$. VLE calculations were performed at $T = T_{Eq}$ and P = "smooth" min (P, P_c) resulting in continuous and smooth Z^L and Z^V values regardless of the flash temperature or pressure. Only the energy balance given by Eq. (5), which is inherently smooth, requires the non-smooth Z^L and Z^V values evaluated at the actual flash temperature, T.

3.2. Model Performance within Optimizations requiring PT or PQ Flashes

To demonstrate robustness of the flash model during optimization, the model (Eqs. 1-3, 5, 7-10, 12-13, and 15-19, 21) was first initialized at 4 different starting points: 1) 50 K, 0.1 Pc; 2) 200 K, 0.1 Pc; 3) 50 K, 2 Pc; and 4) 200 K, 2 Pc. Next, P and T were unfixed and calculated by solving several optimization problems employing the objective function below where T_{target} and P_{target} correspond to T's and P's explored in section 3.1.

$$\text{Minimize} \qquad \left(T - T_{Target}\right)^2 / T_{Target} + (P - P_{Target})^2 / P_{Target} \qquad (22)$$

In all cases, an objective function of zero and identical results to Figure 1 were obtained. This clearly demonstrates the ability of the model to move across phase transition boundaries and through both sub and supercritical regions to obtain an optimal solution. Identical results were also obtained using the same strategy when P and Q were unfixed.

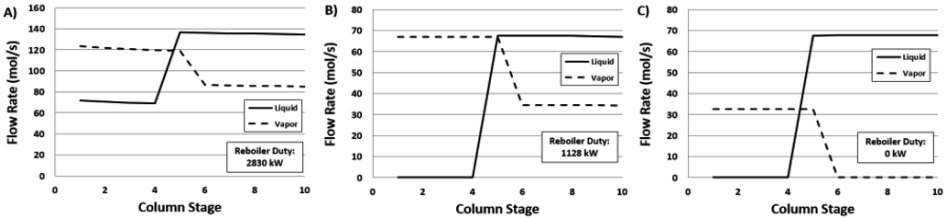


Figure 2: Distillation optimization results for A) $x_{benzene,top}$, $x_{toluene,bottom} \geq 0.90$, B) $x_{toluene,bottom} \geq 0.80$, and C) $x_{toluene,bottom} \geq 0.57$.

Lastly, the model was tested as part of a distillation optimization for a 50% benzene/50% toluene mixture. A 10-stage column at a uniform pressure of 1 bar was used to separate a 100 mol/s inlet stream entering at stage 5. Reboiler duty was minimized subject to three sets of purity constraints. Figure 2 shows results for cases with non-zero flows on each stage (case A) and vanishing phases when the purity constraints are relaxed (cases B, C).

4. Conclusions

The flash model enables reliable VLE calculations over a broad range of operation that spans both subcritical and supercritical conditions. All variables are state variables greatly simplifying the initialization procedure compared to previous work (e.g., Kamath et al. (2010), Dowling et al. (2015)) that required reliable initial guesses and bounds on several intermediate thermodynamic quantities and non-physical slack variables. This model requires no slack variables to relax the VLE constraints outside of the two-phase region and no non-physical terms must be added to the objective function of optimizations. The model moves reliably from initial starting points to distant optimal solutions suggesting that the formulation itself does not introduce non-convexities that might cause optimization solvers to get trapped in local optima. It is thus well-suited for EO flowsheet optimization and will be released as part of the IDAES process systems engineering framework.

Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not state or reflect those of the United States Government or any agency thereof. KeyLogic Systems, Inc.'s contributions to this work were funded by the National Energy Technology Laboratory under the Mission Execution and Strategic Analysis contract (DE-FE0025912) for support services. This material is based upon work supported by the U.S. Department of Energy, Office of Science, under Contract No. DEAC02-05CH11231.

References

R.S. Kamath, L.T. Biegler, and I.E. Grossmann, 2010, "An equation-oriented approach for handling thermodynamics based on cubic equation of state in process optimization," Computers & Chemical Engineering, 34(12), 2085-2096.

A.W. Dowling and L.T. Biegler, 2015, "A framework for efficient large scale equation-oriented flowsheet optimization," Computers & Chemical Engineering, 72, 3-20.

A.W. Dowling, C. Balwani, Q. Gao, and L.T. Biegler, 2015, "Optimization of sub-ambient separation systems with embedded cubic equation of state thermodynamic models and complementarity constraints," Computers & Chemical Engineering, 81, 323-343.

T. Gunderson, 1982, "Numerical aspects of the implementation of cubic equations of state in flash calculation subroutines," Computers & Chemical Engineering, 6, 245-255.

H.A.J. Watson, M. Vikse, T. Gundersen, and P.I. Barton, 2017, "Reliable Flash Calculations: Part 1. Non-smooth Inside-Out Algorithms," I&E Chemistry Research, 56(4), 960-973.

M. Glass and A. Mitsos, 2017, "Thermodynamic analysis of formulations to discriminate multiple roots of cubic equations of state in process models," Computers and Chemical Engineering, 106, 407-420.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50141-5

Price-based coordination of interconnected systems with access to external markets

Lukas S. Maxeiner[a*], Simon Wenzel[a], and Sebastian Engell[a]
[a] *Process Dynamics and Operations Group, Department of Biochemical and Chemical Engineering, TU Dortmund, Emil-Figge Straße 70, 44227 Dortmund, Germany*
sebastian.engell@tu-dortmund.de

Abstract

Many industrial processes are coupled via multiple networks of energy and materials to achieve a resource and energy efficient production. In many cases however, setting up an integrated optimization problem for all units or plants that are directly connected to the networks is not possible, especially when not all information can be shared. In such cases, dual decomposition or price-based coordination can be used, where optimal transfer prices are iteratively determined at which the networks are balanced and the resources are allocated optimally between the participants.

In this contribution, price-based coordination is extended to include the situation where limited resources can be bought or sold at predefined prices from external markets (e.g. via pipelines) and the resulting algorithms are demonstrated for a realistic example.

Keywords: Distributed optimization, Price-based coordination, Dual decomposition, External resources, Shared-resource allocation.

1. Introduction

Many industrial production sites consist of individual plants or units, which are tightly coupled by streams of energy and material. Especially in the chemical industry, these streams, e.g., energy in the forms of steam and electricity as well as raw materials and intermediate products, are exchanged via networks that link consuming and producing systems, where the holdups usually are very small. In order to optimize the overall economic performance as well as the resource efficiency of the site, it is not sufficient to optimize each plant or unit individually, but rather a joint optimization of the production plants or site is necessary taking the balancing of the resource networks into account.

However, finding the solution to the overall optimization problem in a centralized manner may not be realizable. Due to various reasons, such as robustness and local autonomy or confidentiality, when systems belong to different business units or companies, optimizing each plant independently is preferred. One method to coordinate the individual optimization problems is dual decomposition, which uses transfer prices to balance the interconnecting streams, instead of solving the whole problem at once. An optimal solution is determined by iterating between setting the transfer prices (dual variables) and each system optimizing its response, until equilibrium transfer prices are found (Everett III, 1963). This method has various applications and can be used, for instance, to manage the energy flow in electrical micro grids (Zhang et al., 2013) or to optimally allocate resources in chemical production sites (Wenzel et al., 2016, 2017).

Considering all the networks of such a production process as closed systems is not realistic, since usually at least some of the networks are connected to external sources or

sinks, where, on the spot market or according to contracts, limited amounts can be purchased or sold. The topic of optimal procurement for chemical process networks was already discussed by Calfa and Grossmann (2015) for a centralized optimization. In this contribution, the distributed optimization framework of dual decomposition is extended such that it solves this type of problem in a distributed manner by a modified update of the dual variables. We derive this update mechanism and provide a graphical interpretation. Furthermore, we demonstrate the proposed update scheme using an example.

The following syntax is used: Italic lower case letters represent scalar functions or values. Bold lower case letters represent column vectors or vector-valued functions and bold upper case letters represent matrices. The transpose of vector $\mathbf{a}$ is denoted by $\mathbf{a}^T$. We use [k] to identify the k-th row of vectors and matrices. Absolute values and norms are denoted by $|a|$ and $\|\mathbf{a}\|$. To ease the notation, we here assume that all systems have the same number of variables and constraints, which however is not required in general.

2. Problem formulation

The joint optimization of N systems with $n_{Networks}$ network constraints can be written as Eqs.(1a) – (1d):

$$\min_{\mathbf{x}_i\,\forall i} \quad \sum_{i=1}^{N} f_i(\mathbf{x}_i), \tag{1a}$$

$$\text{s.t.:} \quad \mathbf{g}_i(\mathbf{x}_i) \le \mathbf{0}, \tag{1b}$$

$$\mathbf{x}_{LB,i} \le \mathbf{x}_i \le \mathbf{x}_{UB,i}, \tag{1c}$$

$$\sum_{i=1}^{N} \mathbf{A}_i \mathbf{x}_i = \mathbf{0}. \tag{1d}$$

$$\min_{\mathbf{x}_i\,\forall i,\,\mathbf{r}_j\,\forall j} \quad \sum_{i=1}^{N} f_i(\mathbf{x}_i) + \sum_{j=1}^{M} \mathbf{p}_j^T \mathbf{r}_j, \tag{2a}$$

$$\text{s.t.:} \quad \mathbf{g}_i(\mathbf{x}_i) \le \mathbf{0}, \tag{2b}$$

$$\mathbf{x}_{LB,i} \le \mathbf{x}_i \le \mathbf{x}_{UB,i}, \tag{2c}$$

$$\sum_{i=1}^{N} \mathbf{A}_i \mathbf{x}_i - \sum_{j=1}^{M} \mathbf{r}_j = \mathbf{0}, \tag{2d}$$

$$\mathbf{r}_{LB,j} \le \mathbf{r}_j \le \mathbf{r}_{UB,j}. \tag{2e}$$

Each sub-problem i has an objective f_i, is constrained by g_i, and the states $\mathbf{x}_i$ are bounded via Eq.(1c). Additionally, the systems interact with the different networks via their contribution $\mathbf{A}_i\mathbf{x}_i$ and overall, the network constraint Eq.(1d) has to be satisfied for each network l.

In Eqs.(2a) – (2e), the network constraint is extended to include access to M external resources or sinks j, where the amounts $\mathbf{r}_j$, limited by $\mathbf{r}_{LB,j}$ and $\mathbf{r}_{UB,j}$, can be bought or sold at predefined prices $\mathbf{p}_j$.

3. Distributed solution

The problem shown in Eqs.(1a) – (1d) can be solved using dual decomposition, splitting the problem into N independent sub-problems and an overarching coordination problem by relaxation of the network constraint Eq.(1d). The local contributions to the network constraint are multiplied by the corresponding dual variable $\boldsymbol{\lambda}$ and added to the objective of the sub-problem. Locally, Eq.(3) is solved for a given value of $\boldsymbol{\lambda}$:

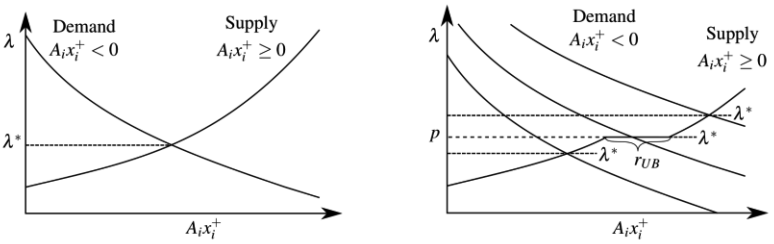
$$\mathbf{A}_i\mathbf{x}_i^+ = \mathbf{A}_i \arg\min_{\mathbf{x}_i} \quad f_i(\mathbf{x}_i) + \boldsymbol{\lambda}^T \mathbf{A}_i \mathbf{x}_i, \tag{3a}$$

$$\text{s.t.:} \quad \mathbf{g}_i(\mathbf{x}_i) = \mathbf{0}, \tag{3b}$$

$$\mathbf{x}_{LB,i} \le \mathbf{x}_i \le \mathbf{x}_{UB,i}. \tag{3c}$$

It is assumed that the subsystems do not share their objectives or constraints due to confidentiality reasons, only the responses to transfer prices $\mathbf{A}_i\mathbf{x}_i^+$, are communicated.

In Figure 1a, a one-dimensional example of aggregated supply, which is defined as the sum of all positive contributions to the network constraint, i.e., $\sum_i\{A_i x_i^+ | A_i x_i^+ \geq 0\}$, and aggregated demand, defined as $\sum_i\{A_i x_i^+ | A_i x_i^+ < 0\}$, depending on the value of λ are shown.



(a) No external resources and outlets corresponding to Eqs.(1a) – (1d).

(b) Possibility to buy external resources at price p for three different demand scenarios corresponding to Eqs.(2a) – (2e).

Figure 1: Optimal response of the systems to the transfer prices in one dimension.

Feasibility, i.e., satisfaction of the network constraint Eq.(1d), is achieved by adjusting the dual variables $\boldsymbol{\lambda}$ in an iterative procedure. Since only the responses to the transfer prices are known on the coordinator level, this cannot be done explicitly. Instead the sub-gradient method is used, where transfer prices $\boldsymbol{\lambda}$ are adjusted proportionally to the difference between supply and demand

$$\boldsymbol{\lambda}^+ = \boldsymbol{\lambda} + \alpha \sum_{i=1}^{N} \mathbf{A}_i \mathbf{x}_i^+. \tag{4}$$

Iterating between the sub-problems Eq.(3) and this update of the transfer prices is done until a stopping criterion is met, e.g., that the residua $\boldsymbol{\rho} = \sum_i^N \mathbf{A}_i\mathbf{x}_i^+$ are all less than a predefined tolerance ϵ. Economically, this can be interpreted as finding the intersection between supply and demand, cf. Figure 1a. The update of the transfer prices is done using the step size parameter $\alpha \in (0, 1]$, which has to be chosen carefully in order to ensure convergence, especially when local constraints $\mathbf{g}_i(\mathbf{x}_i)$ are active. If all objectives and constraints are convex, the solution found upon convergence is the global optimum.

The augmented problem from Eqs.(2a) – (2e) can also be solved using dual decomposition. On the coordination layer, additionally to finding the optimal transfer prices, the cost for interaction with external markets, $\sum_{j=1}^{M} \mathbf{p}_j^T \mathbf{r}_j$, has to be minimized. Finding the optimal values for $\mathbf{r}_j$ can either be done separately from the price adjustment or in a combined step. In Eqs.(5) and (6), the new transfer prices $\boldsymbol{\lambda}$ and the amounts to be bought or sold $\mathbf{r}_j$ are calculated separately using the sub-gradient method:

$$\boldsymbol{\lambda}^+ = \boldsymbol{\lambda} + \alpha \left[\sum_{i=1}^{N} \mathbf{A}_i \mathbf{x}_i^+ - \sum_{j=1}^{M} \mathbf{r}_j \right], \tag{5}$$

$$\mathbf{r}_j^+ = \min\left(\max\left(\mathbf{r}_{LB,j}, \mathbf{r}_j + \operatorname{diag}(\mathbf{p}_j)^{-1} \operatorname{diag}(\mathbf{p}_j - \boldsymbol{\lambda}^+)(\mathbf{r}_{UB,j} - \mathbf{r}_{LB,j})\right)\mathbf{r}_{UB,j}\right). \tag{6}$$

The other approach is to update both variables at the same time, where the coordinator update is shown in Algorithm 1.

Algorithm 1 Coordinator level update step to determine $\boldsymbol{\lambda}^+$ and $\mathbf{r}_j^+$

$\mathbf{r}_j^+ \leftarrow \mathbf{r}_{LB,j}\ \forall j$
for all $\boldsymbol{\lambda}[k]$ in $\boldsymbol{\lambda}$ **do**
  $\boldsymbol{\pi} \leftarrow [-\infty, \mathbf{p}_1[k], \mathbf{p}_2[k], \ldots, \mathbf{p}_M[k], \infty]^T$
  $\underline{\Lambda} \leftarrow \boldsymbol{\lambda}[k] + \alpha \left(\sum_{i=1}^{N} \mathbf{A}_i[k]\mathbf{x}_i - \sum_{j=1}^{M} \mathbf{r}_{LB,j}[k] \right)$
  for $l = 1$ **to** $M+1$ **do**
    $\overline{\Lambda} \leftarrow \boldsymbol{\lambda}[k] + \alpha \left(\sum_{i=1}^{N} \mathbf{A}_i[k]\mathbf{x}_i - \sum_{j=1}^{l} \mathbf{r}_{UB,j}[k] - \sum_{j=l+1}^{M} \mathbf{r}_{LB,j}[k] \right)$
    if $\boldsymbol{\pi}[l] \leq \underline{\Lambda} < \boldsymbol{\pi}[l+1]$ **then**
      $\boldsymbol{\lambda}^+[k],\ \mathbf{r}_j^+[k] \leftarrow \underline{\Lambda},\ \mathbf{r}_{UB,j}^+[k]$ Case 1
      Break
    else if $\overline{\Lambda} \leq \boldsymbol{\pi}[l+1] < \underline{\Lambda}$ **then**
      if $\sum_{j=1}^{l-1} \mathbf{r}_{j,max}[k] \leq \sum_{i=1}^{N} \mathbf{A}_i[k]\mathbf{x}_i < \sum_{j=1}^{l} \mathbf{r}_{j,max}[k]$ **then**
        $\boldsymbol{\lambda}^+[k],\ \mathbf{r}_j^+[k] \leftarrow \boldsymbol{\pi}[l+1],\ \left(\sum_{i=1}^{N} \mathbf{A}_i[k]\mathbf{x}_i - \sum_{j=1, j\neq l}^{M} \mathbf{r}_j^+[k] \right)$ Case 2
      else
        $\boldsymbol{\lambda}^+[k],\ \mathbf{r}_j^+[k] \leftarrow \underline{\Lambda},\ \mathbf{r}_{UB,j}^+[k]$ Case 3
      end if
      Break
    end if
    $\underline{\Lambda} \leftarrow \overline{\Lambda}$
  end for
end for

In the example in Figure 1b, this corresponds to evaluating different realizations for λ^+ using Eq.(5), based on the minimum and maximum values for r. If the price range given by $\underline{\Lambda}$ and $\overline{\Lambda}$ is below the price for the external resource p, then r will be at the lower bound, and the maximum value for λ^+ is used in the next iteration, cf. Case 1 in Algorithm 1. If the price range includes p, Case 2, the difference between supply and demand can be covered by the external resources and therefore $\lambda^+ = p$. If the price range is above the price for the external resources, Case 1 also applies and the external resource r is at the upper bound. Case 3 is only required, if the step size α is too large. The residuals $\boldsymbol{\rho} = \sum_{i=1}^{N} \mathbf{A}_i\mathbf{x}_i^+ - \sum_{j=1}^{M} \mathbf{r}_j^+$ are used for the evaluation of the stopping criterion in both approaches.

4. Example

In the following, an example is used to demonstrate the proposed extension of the dual decomposition approach. Quadratic objective functions with positive diagonal scaling matrices and affine constraints of the following form are assumed:

$$f_i(\mathbf{x}_i) = \left(\mathbf{x}_i - \mathbf{x}_{i,Target}\right)^T \mathbf{B}_i \left(\mathbf{x}_i - \mathbf{x}_{i,Target}\right), \tag{7a}$$

$$\mathbf{g}_i(\mathbf{x}_i) = \mathbf{C}_i\mathbf{x}_i - \mathbf{d}_i = \mathbf{0}, \tag{7b}$$

Five subsystems ($N = 5$) are connected via three shared resources ($n_{Networks} = 3$). For each resource, there are three different sources from which it can be bought ($M = 3$). Each subsystem has four independent variables $\mathbf{x}_i \in \mathbb{R}^4$ and is subject to two local con-

straints, i.e., $\mathbf{g}_i : \mathbb{R}^4 \rightarrow \mathbb{R}^2$. The transfer prices $\boldsymbol{\lambda} \in \mathbb{R}^3$ are initialized with $\mathbf{0}$, the step size is chosen as $\alpha = 0.03$, and the tolerance is set to $\epsilon = 10^{-6}$. The matrices and vectors $\mathbf{x}_{i,Target}$, $\mathbf{B}_i$, $\mathbf{C}_i$, $\mathbf{d}_i$, $\mathbf{A}_i$, $\mathbf{p}_j$, and $\mathbf{r}_{UB,j}$ are generated from a random seed and the lower bound for flows of external resources is $\mathbf{r}_{LB,j} = \mathbf{0}$.

$$\mathbf{x}_{Target,i} = \begin{bmatrix} -2 & -7 & 4 & 6 \end{bmatrix}; \begin{bmatrix} -3 & 9 & -10 & 5 \end{bmatrix}; \begin{bmatrix} -7 & 2 & 8 & -7 \end{bmatrix}; \begin{bmatrix} -8 & 1 & 4 & 9 \end{bmatrix}; \begin{bmatrix} 1 & 2 & -7 & -6 \end{bmatrix},$$

$$\mathbf{B}_i = diag\left(\begin{bmatrix} 3 & 5 & 7 & 8 \end{bmatrix}\right); diag\left(\begin{bmatrix} 6 & 4 & 8 & 9 \end{bmatrix}\right); diag\left(\begin{bmatrix} 1 & 3 & 4 & 5 \end{bmatrix}\right); diag\left(\begin{bmatrix} 4 & 7 & 5 & 9 \end{bmatrix}\right); diag\left(\begin{bmatrix} 3 & 4 & 7 & 7 \end{bmatrix}\right),$$

$$\mathbf{C}_i = \begin{bmatrix} 9 & 7 & 9 & 7 \\ 0 & 5 & 4 & 4 \end{bmatrix}; \begin{bmatrix} 4 & 1 & 8 & 5 \\ 0 & 0 & 5 & 4 \end{bmatrix}; \begin{bmatrix} 8 & 7 & 2 & 3 \\ 0 & 0 & 8 & 1 \end{bmatrix}; \begin{bmatrix} 0 & 2 & 1 & 8 \\ 1 & 6 & 2 & 8 \end{bmatrix}; \begin{bmatrix} 5 & 7 & 6 & 6 \\ 9 & 3 & 9 & 9 \end{bmatrix},$$

$$\mathbf{d}_i = \begin{bmatrix} 9 & 6 \end{bmatrix}^T; \begin{bmatrix} 1 & 3 \end{bmatrix}^T; \begin{bmatrix} 6 & 1 \end{bmatrix}^T; \begin{bmatrix} 3 & 6 \end{bmatrix}^T; \begin{bmatrix} 5 & 6 \end{bmatrix}^T,$$

$$\mathbf{A}_i = \begin{bmatrix} -8 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 6 & 0 \end{bmatrix}; \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & -9 & 0 \end{bmatrix}; \begin{bmatrix} -9 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix}; \begin{bmatrix} 9 & 0 & 0 & 0 \\ 0 & -5 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}; \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & -7 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

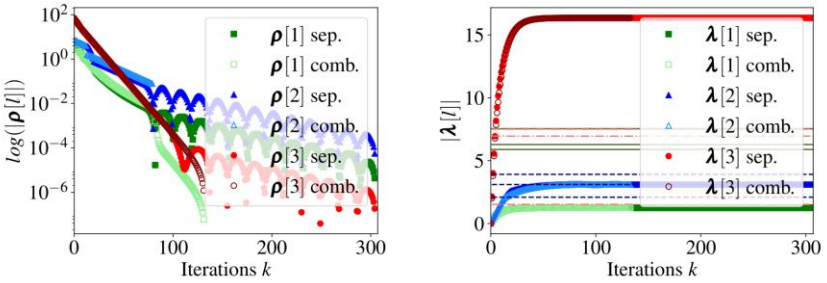
$$\mathbf{p}_j = \begin{bmatrix} 5.89 & 2.09 & 1.52 \end{bmatrix}^T; \begin{bmatrix} 6.28 & 3.1 & 6.95 \end{bmatrix}^T; \begin{bmatrix} 7.54 & 3.91 & 7.5 \end{bmatrix}^T,$$

$$\mathbf{r}_{UB,j} = \begin{bmatrix} 3.1 & 4.5 & 3 \end{bmatrix}^T; \begin{bmatrix} 2.2 & 2.9 & 1.4 \end{bmatrix}^T; \begin{bmatrix} 4.8 & 1.9 & 4 \end{bmatrix}^T,$$

In Figure 2, the evolution of the residuals $\boldsymbol{\rho}$ and of the transfer prices $\boldsymbol{\lambda}$ over the iterations can be seen for both, the separate (filled) and combined (empty), update rules. The distributed solutions converge towards the optimum of the centralized solution. Furthermore, it can be seen that upon convergence $\boldsymbol{\lambda}[1]$ (square) is less than, $\boldsymbol{\lambda}[2]$ (triangle) is equal to, and $\boldsymbol{\lambda}[3]$ (circle) is larger than the prices at which external resources can be bought (indicated by the thin lines), i.e., this example covers all three scenarios from Figure 1b. Comparing the two different methods, it becomes evident that the method with separate steps requires more iterations and oscillates towards the optimum.

5. Discussion and Conclusion

The significantly larger number of iterations in the approach with separate steps is due to $\boldsymbol{\lambda}[2]$ being equal to the external price. Since the variables are set consecutively, an overshoot is necessary because otherwise the sub-gradient for the usage of the external resource does not change. For the combined approach, after 90 iterations the network constraint of resource 2 is balanced by the external market and hence the residuum is 0. The prices can be initialized more efficiently, e.g., at one of the prices for external resources, however the goal was to show convergence regardless of initialization strategy.



(a) Evolution of the residuals.

(b) Evolution of the prices. The lines show $\boldsymbol{p}_j[l]$.

Figure 2: Resulting trajectories of characteristic quantities for the separated (sep.) and the combined (comb.) approach with the iterations.

In general, it can be said that distributed optimization can be used, also when there is access to external markets. What the authors observed running a suite of test cases with 1000 additional randomly generated examples is that the number of iterations of the combined approach is in a similar range as for a price-based optimization without external markets. Additionally, using this method bounds on the transfer prices of resources can be imposed by setting $\mathbf{r}_{LB,j}$ and $\mathbf{r}_{UB,j}$ to low and high values, respectively. The authors found this to be necessary to avoid excessive prices in real applications.

6. Outlook

Future work includes investigating the influence of a dynamic adaptation of the step size parameter α, in order to enable an arbitrary starting value for α and to speed up convergence. From a practical perspective, the distribution of the gains from the usage of external markets needs to be investigated, since here these are only realized at the coordinator level. Another topic of interest is the implementation of other contracts, such as take or pay, minimum required amounts, discounts for large amounts, etc.

Acknowledgment

The project leading to this publication has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 723575 (CoPro, spire2030.eu/copro) in the framework of the SPIRE PPP.

References

Calfa, B., Grossmann, I., 2015. Optimal procurement contract selection with price optimization under uncertainty for process networks. Computers & Chemical Engineering 82, 330–343.

Everett III, H., 1963. Generalized lagrange multiplier method for solving problems of optimum allocation ofresources. Operations research 11 (3), 399–417.

Wenzel, S., Paulen, R., Stojanovski, G., Krämer, S., Beisheim, B., Engell, S., 2016. Optimal resource allocation in industrial complexes by distributed optimization and dynamic pricing. at - Automatisierungstechnik 64 (6), 428–442.

Wenzel, S., Paulen, R., Beisheim, B., Krämer, S., Engell, S., 2017. Adaptive pricing for optimal resource allocation in industrial production sites. IFAC-PapersOnLine 50 (1), 12446–12451.

Zhang, Y., Gatsis, N., Giannakis, G. B., 2013. Robust Energy Management for Microgrids With High Penetration Renewables. IEEE Transactions on Sustainable Energy 4 (4), 944–953.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50142-7

CFD simulation of a full scale LNG storage tank

Abdullah Saleem [a], Shamsuzzaman Farooq [a*], Iftekhar A. Karimi [a], Raja Banerjee [b]

[a] *Department of Chemical & Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, 117585, Singapore*
[b] *Department of Mechanical Engineering, Indian Institute of Technology Hyderabad, Yeddumailaram, AP 502 205, India*
chesf@nus.edu.sg

Abstract

A well-insulated LNG (Liquefied Natural Gas) tank is a key facility in a regasification terminal. The management of Boil off Gas (BOG) produced from a LNG storage tank due to heat leak from the surrounding is an important issue. Many small-scale experimental studies have been reported on evaporation-condensation of different volatile liquids. However, these studies cannot mimic the behaviour of an industrial process. Computational Fluid Dynamics (CFD) simulation is the preferred alternative to study boiling phenomena at a large scale. However, no such study exists for LNG. A comprehensive dynamic CFD model presented in this study for an onshore full-scale LNG storage tank is a maiden attempt to fill that gap. We carry out a thorough investigation of internal flow dynamics and heat transfer characteristics owing to free convection. Our work provides insights on the boiling phenomena and allows reliable quantification of BOG production.

Keywords: LNG, full-scale tank, BOG, surface evaporation, nucleate boiling.

1. Introduction

Natural gas (NG) use is on the rise due to its lowest CO_2 emission per unit energy among the fossil fuels. After transportation by pipeline, the next preferred mode of NG transport is as Liquefied Natural Gas (LNG). Hence, the demand for LNG is also on the rise. In order to import LNG, a country must have one or more LNG receiving terminals. A LNG terminal consists of several facilities such as unloading, storing, regasification and transmission. A properly insulated LNG storage tank is a key unit in a receiving terminal. No amount of insulation can completely stop heat ingress into the LNG storage tank from the surroundings. Thus, production of boil off gas (BOG) is unavoidable. If not removed, BOG can elevate the tank pressure. Hence, continuous removal of BOG is essential to keep the tank pressure at a safe level. The management of BOG is a key issue for which a good understanding of the mechanism of its generation and at what rate it is generated are important.

LNG storage tank is a multiphase system that encounters liquid-vapor phase change. Many laboratory scale experimental studies have been performed on volatile liquids to understand evaporation and condensation phenomenon. However, laboratory studies are limited by the size and cannot mimic the large static pressure encountered in a full scale industrial system. Furthermore, the variation in surface area to volume ratio during

scaling significantly impacts heat transfer characteristics and flow dynamics of the system. Thus, the alternative to experimental studies for analyzing large scale industrial systems is the use of CFD simulation. In computational phase change studies, governing equations for mass, momentum and energy conservation are solved for entire computational domain (Juric and Tryggvason, 1998). The volume of fluid (VOF) method has been widely used in the literature to track vapor and liquid phases (Banerjee, 2008). CFD tank simulations have been reported to investigate different fuel systems (Hassanvand et al., 2010). However, there are few reported studies on LNG storage. A simplified isobaric and well-mixed 1D model was previously developed by our research group for BOG prediction (Effendy et al., 2017). The CFD model in the present study considers static pressure and hence the associated sensible heat gain. The boiling mechanism inside a properly insulated full-scale LNG storage tank is unknown in the open literature. The rigorous dynamic CFD model presented in this study is the maiden attempt to fill that gap. It examines the validity of the well-mixed assumption of the 1D model and sheds light on the likelihood of nucleate boiling in a typical LNG storage tank in a regasification terminal.

2. Model equations

2.1. VOF model

The volume of fluid (VOF) model is used, which allows tracking of the vapor-liquid interface. It is a multiphase model and employed for two or more phases which are immiscible. The interphase tracking between the two phases is executed by solving the equation of continuity, given by Eq. (1), for one or more phases.

$$\frac{1}{\rho_q}\left[\frac{\partial}{\partial t}\left(\alpha_q \rho_q\right) + \nabla.\left(\alpha_q \rho_q \vec{v}_q\right) = \sum_{p=1}^{n}(\dot{m}_{pq} - \dot{m}_{qp})\right] \quad (1)$$

In the VOF method, a single set of momentum equations in each control volume is solved and volume fraction (α) of each of the comprising phases in each cell is evaluated. Either of three possibilities can exist for a control volume: (a) If the control volume is completely vacant of q^{th} fluid i.e. $\alpha_q = 0$; (b) If control volume is fully filled with q^{th} fluid i.e. $\alpha_q = 1$; and (c) If control volume is partially filled of q^{th} fluid i.e. $0 < \alpha_q < 1$. Inside each computational cell, the summation of volume fractions of all phases must be unity. In the present study, number of phases (n) is two i.e. liquid and vapor.

2.2. Physical properties

If a computational cell contains more than one phase, then the cell properties are evaluated based on volume-average. The volume-average assumption is an integral part of VOF methodology. It provides accurate interface tracking within a cell. Hence, transport properties such as viscosity and density appearing in transport equation are computed based on volume fraction of each phase as indicated by Eq. (2).

$$\rho = \alpha_q \rho_q + (1 - \alpha_q)\rho_q \quad (2)$$

2.3. Lee model

The Lee model is employed to account for evaporation-condensation (Lee, 1980). The mass transfer from liquid to vapor and vapor to liquid during evaporation and

condensation respectively is described by the following Lee model equations i.e. Eq. (3) and Eq. (4).

$$m_{lv} = coeff * \alpha_l \rho_l \frac{(T_l - T_{sat})}{T_{sat}} \quad (3)$$

$$m_{vl} = coeff * \alpha_v \rho_v \frac{(T_{sat} - T_v)}{T_{sat}} \quad (4)$$

The Lee model defines a positive mass transfer from the liquid phase to the vapor phase for the evaporation-condensation problems. The coefficient appearing as first term on right hand side of above equations is called as relaxation time or mass transfer intensity factor. A default value of 0.1 for the intensity factor is used in the present simulation. To account for static pressure effect, a piecewise-polynomial correlation is defined to consider variation of saturation temperature (T_{sat}) with static head of liquid. Thus, Lee model uses Eq. (5) to evaluate the saturation temperature.

$$T_{sat} = 0.3709\, e^{-15}\, P^3 - 3.8013\, e^{-10}\, P^2 + 17.956\, e^{-5}\, P + 96.859 \quad (5)$$

2.4. Boussinesq equation

The Boussinesq approximation given by Eq. (6) treats density as constant in all the terms of solved equations except the buoyancy term in the momentum equation. Thus, a faster convergence is achieved by using this approximation for buoyancy driven flows.

$$(\rho - \rho_o) g = -\rho_o \beta (T - T_o) g \quad (6)$$

In this study, the Boussinesq approximation, valid for small density changes, is used for liquid density.

3. Results and discussion

ANSYS Fluent is used to simulate a tank 80 m in diameter and 40 m high. The axisymmetric CFD model accounts for the effect of gravity or static pressure. Liquid methane fills up to 30 m of the tank height and is initially at a temperature of 111.6 K (-161.5 °C) and pressure of 1 atm. The pressure is maintained constant at the tank top. Using U=0.025 W/m^{2}K, representative of a well-insulated full-scale LNG storage tank, it is seen that evaporation occurs only at the vapor-liquid interface and no nucleate boiling is observed. At the level of heat flux possible in the properly insulated full scale tank, the internal mixing due to natural convection is able to continuously remove heat from the bottom. Hence, the temperature at the tank bottom never reaches the saturation temperature of 122.6 K at the pressure of 2.26 bar (1.26 bar is the static pressure). The well mixed behavior results in a uniform temperature distribution throughout bulk liquid phase.

3.1. BOG profile for a full-scale properly insulated tank

The BOG formed is removed continuously through the vent provided. The BOG flowrate with time is shown in Figure 1. The rate of BOG generation increases with time for about 20 hours before attaining a steady state. Initially, when the heat leaks into tank, a portion of it is used as sensible heat to raise the temperature of subcooled liquid and to superheat the vapor. This explains the transient part until, as explained earlier, the system becomes well mixed due the natural convection current and no further sensible heat gain occurs. Therefore, the steady state BOG generation rate corresponds to entire heat leak supplying the latent heat of vaporization. Since our CFD model

incorporates sensible heat gain in both liquid and vapor phases, thus it accurately captures the initial transient BOG characteristics. The BOG generation rate given by a published isobaric and well-mixed 1D model (Effendy et al., 2017) is also shown in Figure 1. They differ in the transient behaviour but practically converge at steady state.

The initial BOG flowrates were overestimated by the 1D model. In the 1D model, only sensible heat gain is in the vapor phase and steady state is reached in about 4 hours. We also simulated a scenario without static pressure effect by appropriately adjusting the Lee model and the profile is also included in Figure 1. Without the static pressure effect, the CFD and 1D simulation results are nearly identical. This not only confirms the effect of the liquid height in the storage tank on the initial BOG rate, but also shows that its impact on the steady state behaviour is not significant. The CPU time required to simulate 40 hours of BOG production is 118 hours. The computational cost is quite reasonable considering the accuracy by which CFD simulation captures the transient dynamics of BOG production. In addition to validating the adequacy of the 1D model at steady state, this study has investigated the boiling mechanism which would not be possible with the 1D model.

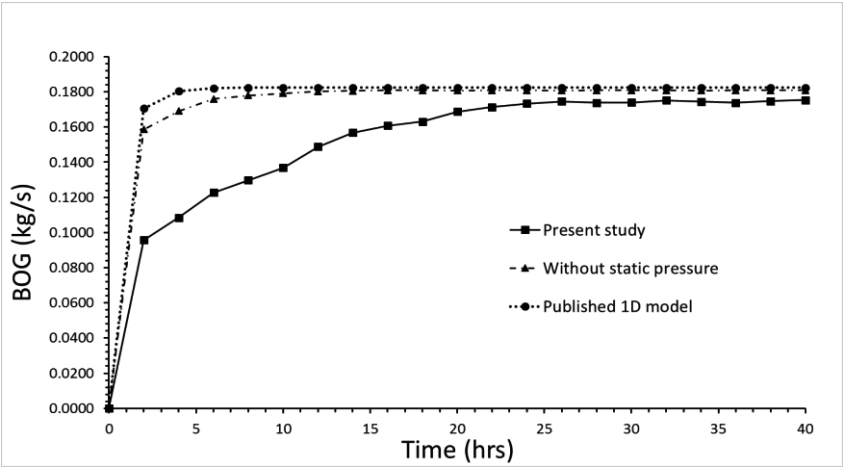


Figure 1. Boil-off-gas (BOG) generation rate in a full-scale LNG storage tank.

3.2. Boiling mechanism in a full-scale tank

To comprehensively understand the boiling mechanism inside a full-scale tank, simulations are performed at U = 0.025, 2.5 and 25 W/m^{2}K. U = 0.025 W/m^{2}K represents a properly insulated tank, whereas U = 25 W/m^{2}K corresponds to a tank which is completely stripped of its insulation. The plot of average bottom wall temperature with time is shown in Figure 2. For U = 0.025 and 2.5 W/m^{2}K, the wall temperature is always less than the saturation temperature at the bottom (122.6 K). This is because at this low heat flux, the heat ingress through the walls is carried to the top by the natural convection before the fluid near the walls can be heated to the local saturation temperature. A steady state is reached well below the local saturation temperature / bubble point. In contrast, at U = 25 W/m^{2}K the rate of heat ingress exceeds the rate at which it is removed to the top by the natural convection current and the wall temperature continuously increases. Eventually, wall temperature reaches the threshold value at which bubble formation occurs. Therefore, it is concluded that surface evaporation is the governing boiling mechanism for a full-scale properly

insulated tank. Thus, advent of nucleate boiling (bubble formation) in a full-size LNG storage tank is unlikely unless the insulation is completely removed.

3.3. Comparison of full-scale and centimetre-scale tanks at U = 2.5 W/m²K

To investigate differences in dynamics of an experimental scale and industrial scale tank, a centimetre-scale tank is simulated and compared with full-scale tank at U = 2.5 W/m²K. In the centimetre scale tank, the numerical values of tank dimensions are kept the same as in the large-scale tank, but the dimension is changed from meter to centimetre. The volume fraction contours for two cases are shown in Figure 3. (Note- white denotes vapor phase and black indicates liquid phase). It is observed that surface evaporation occurs at the vapor-liquid interface for full-scale tank, whereas, bubble formation is witnessed near bottom wall for centimetre-scale tank. The contrasting boiling mechanisms arise due to two factors; (a) static pressure, and (b) surface area to volume ratio. The static pressure for full-scale tank is 1.25 barg at bottom and the corresponding saturation temperature is 122.6 K. In contrast, the static pressure at the bottom of the centimetre-scale tank is almost negligible (0.0125 barg) and the corresponding saturation temperature is 111.6 K.

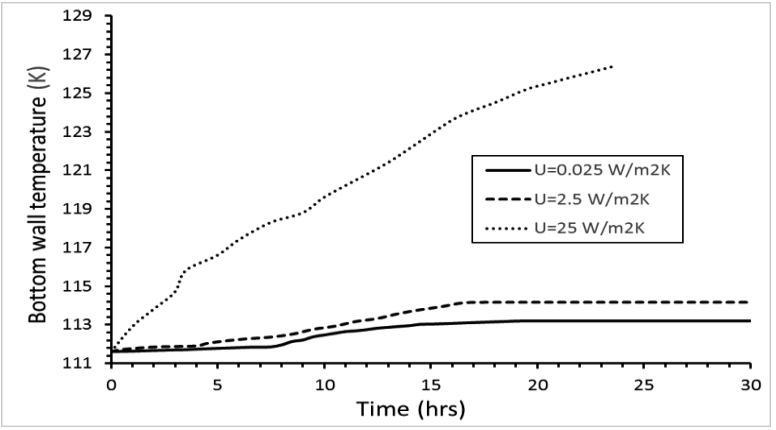


Figure 2. Plot of average bottom wall temperature with time at different values of U.

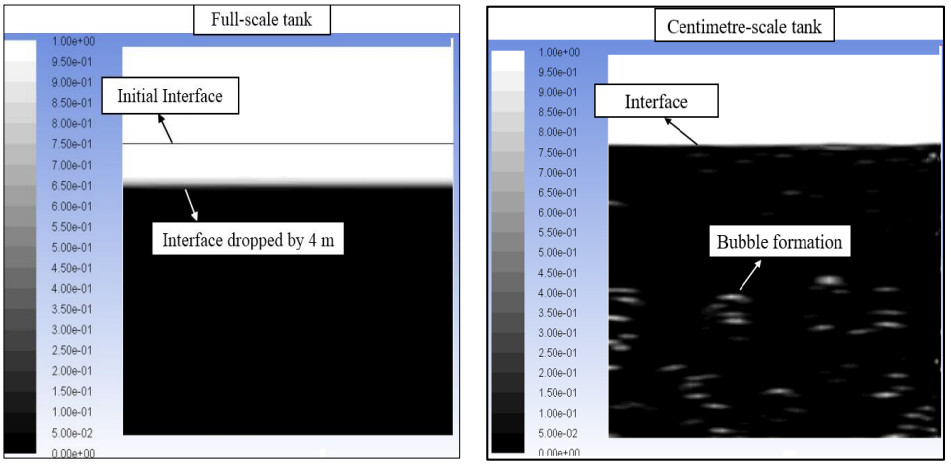


Figure 3. Phase contours for full-scale and centimetre-scale tanks at U = 2.5 W/m²K.

When the scale of the tank is reduced from metre to centimetre, the surface area to volume ratio is increased by 100 times. A higher surface area/volume allows heat to penetrate more aggressively in the bulk liquid phase, which the natural convection current is unable to remove at a comparable rate. Therefore, the temperature near the bottom wall rises to a level that induces bubble formation leading to nucleate boiling.

3.4. Onset of nucleate boiling

In order to investigate the onset of nucleate of boiling in LNG, the heat transfer coefficient is increased by over three orders of magnitude (up to 100 W/m^2K) and the dimensions of the storage tank are also scaled down and the hydrostatic pressure is varied from 1 to 2.26 bar in the scaled down tanks. For liquid methane, nucleate boiling occurs when critical wall superheat (wall temperature above the boiling point near the wall at which the first bubble is seen) is circa 3 degrees, which is unattainable in a well-insulated large scale LNG storage tank.

4. Conclusions

A comprehensive CFD simulation of a well-insulated full-scale LNG storage tank is presented. The BOG generation rate increases with time for about 20 hours before reaching a steady state. The transient is attributed to the static pressure due to the liquid height, which increases the bubble point above the initial temperature and consumes the initial heat ingress to supply the sensible heat. By virtue of the natural convection current, the bulk liquid temperature also becomes nearly uniform except very near the walls. Thus, the 1D model (well-mixed assumption) and CFD simulation results for BOG generation rate deviate in the early part but show good agreement at steady state. The onset of nucleate boiling in LNG requires a critical wall superheat of circa 3 degrees. This is not achieved in a well-insulated full-scale LNG storage tank.

References

S. Effendy, M.S. Khan, S. Farooq, I.A. Karimi, 2017, Dynamic modelling and optimization of an LNG storage tank in a regasification terminal with semi-analytical solutions for N2-free LNG, Computers & Chemical Engineering, 99, 40-50.

D. Juric, G. Tryggvason, 1998, Computations of boiling flows, International Journal of Multiphase Flow, 24, 387-410.

R. Banerjee, 2008, Turbulent conjugate heat and mass transfer from the surface of a binary mixture of ethanol/iso-octane in a countercurrent stratified two-phase flow system, International Journal of Heat and Mass Transfer, 51, 5958-5974.

A. Hassanvand, S.H. Hashemabadi, M. Bayat, 2010, Evaluation of gasoline evaporation during the tank splash loading by CFD techniques, International Communications in Heat and Mass Transfer, 37, 907-913.

W.H. Lee, 1980, Pressure iteration scheme for two-phase flow modeling, Multiphase transport: fundamentals, reactor safety, applications, 4, 407-432.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50143-9

Pyomo.GDP: Disjunctive Models in Python

Qi Chen[a,*], Emma S. Johnson[b,c], John D. Siirola[c], Ignacio E. Grossmann[a]

[a]*Center for Advanced Process Decision Making, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA*
[b]*H. Milton Stewart School of Industrial & Systems Engineering, Georgia Institute of Technology, 755 Ferst Drive, Atlanta, GA 30318, USA*
[c]*Sandia National Laboratories, Albuquerque, NM 87185, USA*
qichen@andrew.cmu.edu

Abstract

In this work, we describe new capabilities for the Pyomo.GDP modeling environment, moving beyond classical reformulation approaches to include non-standard reformulations and a new logic-based solver, GDPopt. Generalized Disjunctive Programs (GDPs) address optimization problems involving both discrete and continuous decision variables. For difficult problems, advanced reformulations such as the disjunctive "basic step" to intersect multiple disjunctions or the use of procedural reformulations may be necessary. Complex nonlinear GDP models may also be tackled using logic-based outer approximation. These expanded capabilities highlight the flexibility that Pyomo.GDP offers modelers in applying novel strategies to solve difficult optimization problems.

Keywords: python, Pyomo, modeling, optimization, disjunctive programming

1. Introduction

Optimization problems involving both discrete and continuous variables are commonplace in Process Systems Engineering (PSE) applications, including process design, planning, scheduling, and operations. Modelers typically formulate these problems as MILPs, or if nonlinear, MINLPs. However, there is no unique model formulation; for the same discrete-continuous problem, some formulations may be tractable while others may not (Grossmann and Trespalacios, 2013).

We claim that most MILP/MINLP formulations arise from attempts to model intuitively disjunctive (logical this-OR-that) problems. Generalized Disjunctive Programming (GDP), proposed by Raman and Grossmann (1994), provides a high-level modeling construct for expressing this application logic intuitively in the form of disjunctions and then applying systematic solution strategies. For example, unit existence in a process network is a disjunction: a candidate unit may exist—and its performance constraints enforced—or it may be absent—and its extensive quantities set to zero (Chen and Grossmann, 2017). Grossmann and Trespalacios (2013) gives the general form of GDP and MILP/MINLP models. Once a GDP model is posed, we can apply reformulations to convert the model structure to a MILP/MINLP and use available mixed-integer solvers, or we can solve the GDP directly using logic-based algorithms. The flexibility of being able to describe a problem using a single GDP model, and then to attempt different solution strategies is invaluable in addressing difficult PSE optimization challenges.

GDP modeling and optimization capability exists in both the open source Pyomo (Hart et al., 2017) and commercial GAMS (GAMS, 2017) algebraic modeling languages (AMLs),

making it accessible to a wide audience of modelers. Both AMLs support the formulation of disjunctions and standard conversions to MILP/MINLP: the big-M (BM) reformulation described in Raman and Grossmann (1994) and the hull reformulation (HR) described in Furman et al. (2016). However, the flexibility to develop and utilize more advanced GDP solution techniques may be necessary to tackle challenging problems.

A key challenge in reformulating GDP models is the generation of MILPs/MINLPs with tight continuous relaxations without growing the problem to an intractable size (Ruiz and Grossmann, 2017). BM leads to smaller formulations, but HR yields tighter relaxations. There has therefore been interest in advanced reformulations to better address this trade-off. Trespalacios and Grossmann (2016) introduced a method for strengthening the BM formulation using cutting planes derived from the HR, leading to a hybrid reformulation tighter than the BM, but smaller than the HR. Another recent development has been the introduction of basic steps for GDP by Ruiz and Grossmann (2012) building on the disjunctive programming concept introduced by Balas (1985). By performing basic steps, the GDP formulation is progressively transformed into disjunctive normal form, tightening the HR of the GDP, and consequently its continuous relaxation. For difficult problems, these advanced reformulations can make a difference in tractability; however, automatic generation of basic steps does not yet exist in a modeling environment, limiting their application.

In some cases, particularly for nonlinear GDP models with higher complexity in the continuous space, it is useful to solve the GDP using direct decomposition algorithms such as logic-based outer approximation (LOA) (Türkay and Grossmann, 1996). However, the LOGMIP 2.0 solver (Vecchietti and Grossmann, 1997) in GAMS 23.7+ no longer supports LOA. A capability gap thus exists for a flexible LOA solver that modelers can tune to fit their needs and adapt to incorporate domain-specific knowledge.

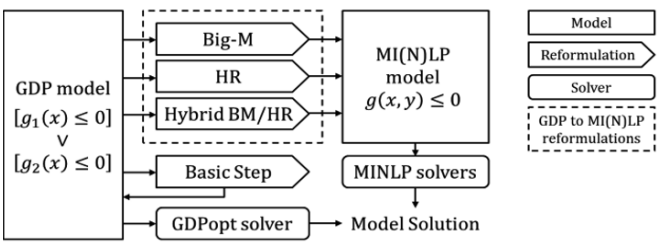


Figure 1. GDP model solution approaches with Pyomo.GDP

In this paper, we present new implementations in Pyomo.GDP enabling use of advanced reformulations such as hybrid BM/HR cutting planes and basic steps, and a flexible LOA implementation in the new direct GDP solver GDPopt. Figure 1 shows how these new capabilities fit in for the solution of a GDP model. In section 2, we discuss the advanced reformulations. In section 3, we present details on GDPopt. We demonstrate hybrid cuts and GDPopt with a case study in section 4, and we conclude in section 5.

2. Advanced GDP reformulations to MILP/MINLP

When reformulating a GDP model to MILP/MINLP, the main challenge is achieving the right compromise between problem size and tightness of the continuous relaxation. We introduce two new automatic tools that modelers can use to find a good trade-off: hybrid big-M cutting planes and GDP basic steps.

2.1. Hybrid BM/HR cutting plane algorithm

The hybrid BM/HR reformulation is based on the idea that cutting planes can be derived from the HR to strengthen the BM relaxation without explicit addition of many new variables and constraints (Trespalacios and Grossmann, 2016). We refer the reader to Section 6 and Figure 13 of Vecchietti et al. (2003) for theoretical background on the BM/HR cutting plane algorithm implemented in Pyomo.GDP.

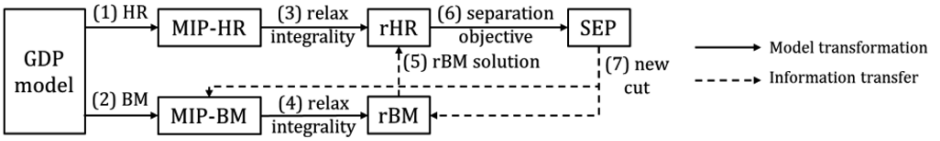


Figure 2. Hybrid BM/HR cutting plane algorithm

The Pyomo.GDP cutting plane implementation relies on pre-existing automatic reformulations for BM and HR. The algorithm is outlined in Figure 2. Note that cuts are generated via a separation problem between the rBM solution and the rHR feasible region. Given a model `m`, the following command will generate a MILP/MINLP representation:

```
TransformationFactory('gdp.cuttingplane').apply_to(m)
```

Following reformulation, the MILP/MINLP may be solved using any standard solver.

2.2. GDP Basic Steps

Basic steps in GDP bring a formulation incrementally from conjunctive normal form to disjunctive normal form by intersecting disjuncts with constraints or other disjuncts. Ruiz and Grossmann (2012) show that by doing so, the HR of the problem is tightened at the expense of a growth in the number of disjuncts. The reformulation that applies a basic step between disjunctions "d" and "e" is implemented with the following command:

```
TransformationFactory('gdp.basic_step').apply_to(m, targets=[m.d, m.e])
```

The result is a reformulation from a GDP model to another GDP model with a basic step applied. The tightened GDP could then be further reformulated to MILP/MINLP or used in the context of the hybrid BM/HR cutting plane algorithm.

3. GDPopt: a flexible logic-based nonlinear GDP solver for Pyomo

We introduce GDPopt, a logic-based solver with a LOA implementation built on top of the Pyomo.GDP framework. GDPopt offers an alternative to reformulation of nonlinear GDPs into MINLPs. For PSE applications with complex nonlinear functions in the continuous space, such as those resulting from mixing, reaction, and equilibrium relations in process networks, LOA allows solution of reduced space NLP sub-problems—including only constraints in selected disjuncts—rather than the full-space representations encountered when solving MINLPs, thereby improving robustness.

Figure 3 illustrates the LOA algorithm. From the original nonlinear GDP model, (1) an initialization procedure is used to generate a linearization, yielding a linear GDP model. (2) An automatic reformulation is then applied to the linear GDP model to yield the MILP master problem. The solution to the MILP master problem provides a lower bound on the overall (minimization) problem as well as (3) a proposed realization of the discrete variables. By (4) fixing the disjunctions and Boolean variables of the nonlinear GDP

model based on the optimal MILP master problem solution, we obtain a reduced space NLP model omitting the constraints in inactive disjuncts. The optimal NLP solution gives an upper bound on the overall (minimization) problem as well as the optimal continuous variable values for (5) generating an outer approximation (OA) cut. GDPopt uses OA/ER cuts described in Viswanathan and Grossmann (1990). The algorithm then loops back through steps 2-6 until a convergence criterion is met. GDPopt also supports the hybrid mixed-integer/GDP models described in Vecchietti and Grossmann (1997).

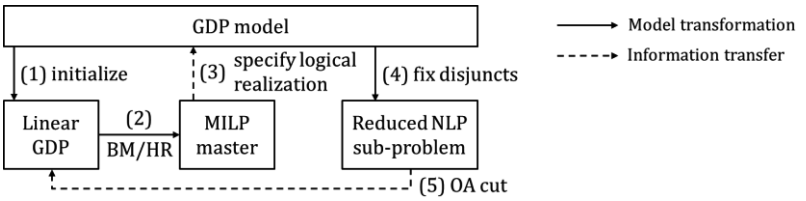


Figure 3. Logic-based outer approximation flow diagram

Two major challenges in applying LOA are the initial linearization of the nonlinear GDP and the solution of the NLP sub-problems. Türkay and Grossmann (1996) introduce a set-covering algorithm to ensure that linearizations are generated for each disjunct. We propose a modification whereby the set cover only needs to include disjuncts that contain nonlinear constraints. Taking advantage of Pyomo model characterization capabilities, GDPopt supports this new set-covering scheme, reducing the number of initialization sub-problems needed for LOA. Furthermore, because GDPopt is written in Python, it offers advanced modelers great flexibility to shape the algorithm according to their needs. GDPopt supports the use of custom initialization schemes for LOA as well as the capability to define callbacks before or after every LOA step outlined above. These callbacks have full access to the underlying Pyomo models used by LOA, and they can even be used to implement custom initialization schemes for the NLP sub-problems, providing warmstart information to the NLP solver that can be important for convergence (Biegler, 2010).

Usage of GDPopt is similar to the invocation of any other Pyomo solver:

```
SolverFactory('gdpopt').solve(m)
```

4. Case study

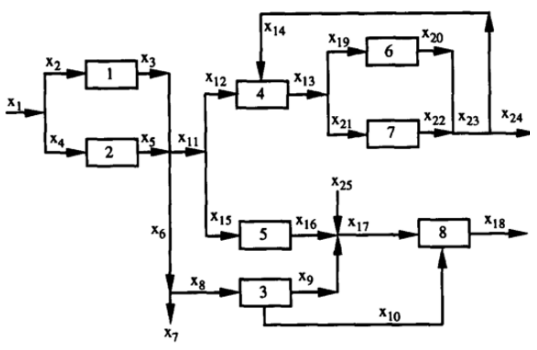


Figure 4. Eight process problem superstructure

We demonstrate the application of advanced reformulations and GDPopt using a disjunctive formulation of a process design problem from Duran and Grossmann (1986). Figure 4 shows the superstructure for the selection among eight process units.

The goal is to minimize annual cost of operation for a process with eight candidate units by determining the optimal flowsheet structure and flowrates. The mathematical formulation for this problem can be found in Appendix A of Türkay and Grossmann

(1996). For brevity, we illustrate below a selection of the model-building and solution steps for the case study in Pyomo.GDP. The complete code is publicly available as a test problem in the Pyomo Github repository.

All models begin with declaration of an empty model object:

```
m = ConcreteModel(name='Eight Process Problem')
```

We can then define model variables, constraints, and disjunctions, such as the one between selection of unit 1 versus unit 2, as shown in Equation 2:

$$\begin{bmatrix} Y_1 \\ \exp(x_3) - 1 = x_2 \\ x_4 = x_5 = 0 \end{bmatrix} \bigvee \begin{bmatrix} Y_2 \\ \exp\left(\frac{x_5}{1.2}\right) - 1 = x_4 \\ x_2 = x_3 = 0 \end{bmatrix} \tag{2}$$

```
m.use_unit1 = Disjunct(); m.use_unit2 = Disjunct()
m.use_unit1.inout1 = Constraint(expr=exp(m.flow[3])-1 == m.flow[2])
m.use_unit1.no_unit2_flow1 = Constraint(expr=m.flow[4] == 0)
m.use_unit1.no_unit2_flow2 = Constraint(expr=m.flow[5] == 0)
m.use_unit2.inout2 = Constraint(expr=exp(m.flow[5]/1.2)-1 == m.flow[4])
m.use_unit2.no_unit1_flow1 = Constraint(expr=m.flow[2] == 0)
m.use_unit2.no_unit1_flow2 = Constraint(expr=m.flow[3] == 0)
m.use1or2 = Disjunction(expr=[m.use_uni1, m.use_uni2])
```

Table 1. MINLP statistics following BM, HR, and Hybrid reformulations

	NLP Relaxation	# vars	# binary	# constraints
HR	67.4	84	12	196
BM	-435.4	37	12	83
Hybrid BM/HR	-140.29	37	12	84

Once defined, the GDP model can be transformed or solved. Table 1 shows the problem size and root node NLP solution for the standard reformulations, demonstrating the trade-off between relaxation quality and problem size. All three examples solve to the optimal solution of 68.0 within 1 CPUs using the DICOPT solver. This example demonstrates the impact of the hybrid BM/HR cutting planes on the relaxation, as the addition of a single constraint improves the BM relaxation by 50%.

Table 2. Iteration results for GDPopt

Iteration	Z_{NLP} (UB)	Z_L (LB)
Set cover #1	68.0	
Set cover #2	100.3	
1		68.0

Another option is to solve the GDP model directly using GDPopt. Table 2 displays the GDPopt iteration results. Compared to Table 5 in Türkay and Grossmann (1996), only two set covering sub-problems are necessary, reducing the number of NLPs that must be solved for initialization. Subsequently, the algorithm converges after the first master iteration.

5. Conclusions

In this work, we introduce new solution capabilities for the Pyomo.GDP ecosystem, including the hybrid BM/HR cutting planes reformulation, GDP basic steps, and the GDPopt logic-based solver. These additions highlight how the Pyomo algebraic modeling language provides a first-class framework for GDP modeling and optimization of discrete-continuous problems in PSE. By modeling in Pyomo.GDP, users can directly express problem logic and leverage a variety of automated solution techniques. Furthermore, as an open tool, advanced modelers are able to modify the reformulations and the GDPopt solution algorithm to incorporate their domain-specific knowledge.

Acknowledgements: Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

References

E. Balas, 1985. Disjunctive Programming and a Hierarchy of Relaxations for Discrete Optimization Problems. SIAM J. Algebr. Discret. Methods 6, 466–486.

L.T. Biegler, 2010. Nonlinear Programming. Society for Industrial and Applied Mathematics.

Q. Chen, I.E. Grossmann, 2017. Recent Developments and Challenges in Optimization-Based Process Synthesis. Annu. Rev. Chem. Biomol. Eng. 8, 249–283.

M.A. Duran, I.E. Grossmann, 1986. An outer-approximation algorithm for a class of mixed-integer nonlinear programs. Math. Program. 36, 307.

K.C. Furman, N. Sawaya, I.E. Grossmann, 2016. A computationally useful algebraic representation of nonlinear disjunctive convex sets using the perspective function. Optimization Online.

GAMS Development Corporation, 2017. General Algebraic Modeling System (GAMS) Release 24.9.2.

I.E. Grossmann, F. Trespalacios, 2013. Systematic modeling of discrete-continuous optimization models through generalized disjunctive programming. AIChE J. 59, 3276–3295.

W.E. Hart, C.D. Laird, J.-P. Watson, D.L. Woodruff, G.A. Hackebeil, B.L. Nicholson, J.D. Siirola, 2017. Pyomo — Optimization Modeling in Python. 2nd ed. Springer.

R. Raman, I.E. Grossmann, 1994. Modelling and computational techniques for logic based integer programming. Comput. Chem. Eng. 18, 563–578.

J.P. Ruiz, I.E. Grossmann, 2017. Global optimization of non-convex generalized disjunctive programs: a review on reformulations and relaxation techniques. J. Glob. Optim. 67, 43–58.

J.P. Ruiz, I.E. Grossmann, 2012. A hierarchy of relaxations for nonlinear convex generalized disjunctive programming. Eur. J. Oper. Res. 218, 38–47.

F. Trespalacios, I.E. Grossmann, 2016. Cutting Plane Algorithm for Convex Generalized Disjunctive Programs. INFORMS J. Comput. 28, 209–222.

M. Türkay, I.E. Grossmann, 1996. Logic-based MINLP algorithms for the optimal synthesis of process networks. Comput. Chem. Eng. 20, 959–978.

A. Vecchietti, I.E. Grossmann, 1997. LOGMIP: a disjunctive 0-1 nonlinear optimizer for process systems models, in: PSE '97-ESCAPE 7. Pergamon Press Ltd, pp. S427–S432.

A. Vecchietti, S. Lee, I.E. Grossmann, 2003. Modeling of Discrete/Continuous Optimization Problems: Characterizaton and Formulations of Disjunctions and their Relaxations. Comput. Chem. Eng. 27, 433–448.

J. Viswanathan, I.E. Grossmann, 1990. A combined penalty function and outer-approximation method for MINLP optimization. Comput. Chem. Eng. 14, 769–782.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50144-0

Mixed-Integer Nonlinear Decomposition Toolbox for Pyomo (MindtPy)

David E. Bernal*, Qi Chen, Felicity Gong, and Ignacio E. Grossmann

Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh 15213, PA, USA
bernalde@cmu.edu

Abstract

In this paper, we present a new software framework developed in Pyomo, MindtPy (Mixed-Integer Nonlinear Decomposition Toolbox for Pyomo), which implements several decomposition methods for solving Mixed-Integer Nonlinear Programs (MINLP). These methods decompose the MINLP into a Mixed-integer Linear Program (MILP) and a Nonlinear Program (NLP). The methods implemented in this toolbox are designed to find the global optimal solution of convex MINLP problems. This toolbox allows the user to specify certain algorithmic options for these methods, which were not available in a single solver. Using Process Systems Engineering appications; the computational implementation and performance are discussed.

Keywords: Mixed-integer nonlinear programming, decomposition methods, Pyomo.

1. Introduction

Optimization problems in Process Systems Engineering (PSE) can be expressed with algebraic equations and decision variables, and be solved via mathematical programming. When these problems include nonlinear equations in the objective and/or constraints, and both continuous and discrete variables, they become Mixed-Integer Nonlinear Programs (MINLP).

The nature of the functions involved in optimization in PSE motivated the development of solution methods for both generalized and specialized MINLP models. Convex MINLP models are of special interest since their continuous relaxation gives rise to convex problems with a unique optimum value.

The general form of a MINLP problem is shown in Eq. (1).

$$\begin{aligned} Z = \min_{w,y} \; & f(w,v,y) \\ \text{s.t.}: \; & g(w,v,y) \leq 0 \\ & A_1 w + A_2 v + By \leq b \\ v \in V \subseteq \mathbb{R}^{n_v}, & w \in W \subseteq \mathbb{R}^{n_w}, y \in Y \subseteq \mathbb{Z}^{n_y} \end{aligned} \tag{1}$$

In Eq. (1) we partition the continuous variables as $[w\ v]^T = x \in X = W \times V$ with w being those continuous variables involved in only linear terms in the constraints, and v the variables involved in nonlinear terms.

MINLP problems like the one shown in Eq. (1) can be solved with Branch and Bound (B&B) methods and decomposition methods. In B&B, a systematic enumeration of the

possible combinations of discrete variables is performed, followed by the solution of simpler continuous subproblems obtained by fixing the discrete variables.

Decomposition methods iterate between a relaxation of the original MINLP, using linear approximations of the original feasible region resulting in a Mixed-Integer Linear Programming (MILP), and the solution of a restricted subproblem where the original problem is projected in a certain discrete combination in the form of a Nonlinear Programming (NLP) problem. The relaxation provides a lower bound over the objective function, while the restriction, when feasible, provides an upper bound. The decomposition methods are designed to provide a monotonic improvement on the lower bound while providing new integer combinations to the subproblem to update the upper bound. If the bounds meet, the convex MINLP problem is solved to optimality.

In convex MINLP problems, decomposition methods such as Generalized Benders Decomposition (GBD) (Geoffrion, 1972), Outer-Approximation (OA) (Duran and Grossmann, 1986), and Partial Surrogate Cuts (PSC) (Quesada and Grossmann, 1992) are proven to converge to the global optimum. Another method is the Extended Cutting Plane (ECP) method (Westerlund and Pettersson, 1995), which avoids solving the nonlinear subproblem, which is also proven to converge to the global optimum.

The motivation of this work is that the importance and difficulty of solving MINLP problems require the development of new methods and software. The decomposition methods rely on the solution of the building blocks of MINLP, MILP and NLP; which has improved considerably in recent years (Bonami et al., 2008).

Methods such as OA or ECP have been implemented as part of commercial MINLP solvers as DICOPT, α-ECP, and BONMIN, among others (Bussieck and Vigerske, 2011). In this manuscript, we present an open-source implementation of the OA, ECP, GBD, and PSC methods inside the Python Optimization Modeling Objects (Pyomo) library (Hart et al., 2017) in a single algorithmic framework called MindtPy. MindtPy exploits several advantages of Pyomo, e.g. optimization object-oriented programming, access to commercial NLP and MILP solvers, and third-party Python packages apart from Pyomo.

The main advantage that MindtPy presents in comparison with other MINLP solvers is its flexibility, by offering the possibility to use methods usually available in separate solvers, of which not all are available for Pyomo. This flexibility is similar to the one offered by other open-source solvers like BONMIN, LaGO, and COUENNE, where the user can easily modify the code to specialize it to a certain type of MINLP. Other solvers are non-open-source for all users, limiting the reach that open-source software like Pyomo can provide. Unlike COUENNE, MindtPy is only able to find the global optimum to convex MINLP problems, but by avoiding the solution of a spatial B&B it may find solutions to these problems efficiently.

2. Decomposition methods

A common classification of MINLP regards its continuous relaxation, where if all the constraints and the objective are convex functions, then the MINLP is denoted as a convex MINLP (Lee and Leyffer, 2012) although MINLPs themselves are nonconvex.

The decomposition methods assume that there is a subset of the problem variables that when temporarily fixed, make the remaining optimization problem considerably more tractable. This is the case with MINLP problems, where the complicating variables are

the integer variables, which when fixed, give rise to an NLP problem. The OA, GBD, and PSC methods solve an MILP problem, which provides a lower bound and an integer combination, which when fixed results in an NLP.

The master problem of the GBD method at the iteration k is presented in Eq. (2).

$$\begin{aligned} & Z_{LB,GBD}^{k} = \min_{\eta,y} \eta \\ \text{s.t.:} \quad & \eta \geq f(v^k, w^k, y) + (\mu^k)^T g(v^k, w^k, y) \;\; \forall k \in K_{feas} \\ & (\mu^k)^T g(v^k, w^k, y) \leq 0 \;\; \forall k \in K_{infeas} \\ & A_1 w^k + A_2 v^k + By \leq b \\ & \eta \in \mathbb{R}^1, y \in Y \subseteq \mathbb{Z}^{n_y} \end{aligned} \tag{2}$$

Where μ^k are the Lagrange multipliers of the nonlinear constraints, and K_{feas} and K_{infeas} are the set of iterations where the NLP subproblems were feasible and infeasible, respectively. This problem adds one cut to the master problem per iteration.

In the OA method, the nonlinear constraints are replaced by linear constraints given by the 1st order Taylor approximations. Therefore, we obtain the master problem in Eq. (3) which is a valid relaxation of the MINLP.

$$\begin{aligned} & Z_{LB,OA}^{k} = \min_{\eta,x,y} \eta \\ \text{s.t.:} \quad & \eta \geq f(v^k, w^k, y^k) + \nabla f(v^k, w^k, y^k)^T [v - v^k \quad w - w^k \quad y - y^k]^T \;\; \forall k \in K \\ & g(v^k, w^k, y^k) + \nabla g(v^k, w^k, y^k)^T [v - v^k \quad w - w^k \quad y - y^k]^T \leq 0 \;\; \forall k \in K \\ & A_1 w + A_2 v + By \leq b \\ & \eta \in \mathbb{R}^1, v \in V \subseteq \mathbb{R}^{n_v}, w \in W \subseteq \mathbb{R}^{n_w}, y \in Y \subseteq \mathbb{Z}^{n_y} \end{aligned} \tag{3}$$

This method adds as many cuts as constraints plus the objective cut at each iteration. It can still converge to the optimal solution in finite iterations if only the cuts corresponding to the active constraints for each problem are added, reducing the master problem size.

The PSC method addresses the tradeoff between the master problem size and the strength of the derived cuts. This method can be seen as a combination of the OA and the GBD methods, where the cuts are derived from the gradient-based linearizations of OA, and the KKT conditions as in GBD. The PSC master problem is shown in Eq. (4).

$$\begin{aligned} & Z_{LB,PSC}^{k} = \min_{\eta,v,y} \eta \\ \text{s.t.:} \quad & \eta \geq f(v^k, w, y) + \begin{bmatrix} \lambda^k \\ -\mu^k \end{bmatrix}^T \begin{bmatrix} g(v^k, w, y) & 0 \\ 0 & A_2 \end{bmatrix} \begin{bmatrix} 1 \\ v - v^k \end{bmatrix} \;\; \forall k \in K_{feas} \\ & \begin{bmatrix} \lambda^k \\ -\mu^k \end{bmatrix}^T \begin{bmatrix} g(v^k, w, y) & 0 \\ 0 & A_2 \end{bmatrix} \begin{bmatrix} 1 \\ v - v^k \end{bmatrix} \leq 0 \;\; \forall k \in K_{infeas} \\ & A_1 w^k + A_2 v + By \leq b \\ & \eta \in \mathbb{R}^1, v \in V \subseteq \mathbb{R}^{n_v}, w \in W \subseteq \mathbb{R}^{n_w}, y \in Y \subseteq \mathbb{Z}^{n_y} \end{aligned} \tag{4}$$

An interesting result is that the cuts in the GBD master problem are surrogates of those derived from the PSC methods, which are surrogates of the cuts in OA (Quesada and Grossmann, 1992). This results in the relation between the solutions of each master problem in Eq. (5).

$$Z^k_{LB,GBD} \leq Z^k_{LB,PSC} \leq Z^k_{LB,OA} \tag{5}$$

After solving each master problem, a candidate discrete combination y^{k+1} is found and fixed to solve an NLP subproblem. The NLP solution $(v^{k+1}, w^{k+1}, y^{k+1})$ is used to generate the next cuts in the master problem, and if feasible, it provides an upper bound.

As in OA, the ECP method also relies on the 1st order Taylor approximation cuts to generate a master problem, but instead of solving an NLP subproblem to obtain a new linearization point, it uses the previous master MILP solution. At each iteration, all the nonlinear constraints violated by a ε-tolerance are linearized and added to the master problem, converging to the optimal solution within that ε-tolerance.

Given the tradeoff between the size of the master problem and the strength of the lower bound predicted by it, none of the methods dominates the others and the performance depends on each problem. Another tradeoff is between solving or not solving the NLP subproblems, given that they can be expensive to solve but provide a better linearization point and an upper bound to the problem.

3. Implementation

The decomposition methods for convex MINLP can be seen as meta-algorithms, where the solution of the optimization problem relies on solutions of other subproblems provided by other solvers, namely MILP and NLP.

Given a convex MINLP written in Pyomo, our toolbox solves it using any of the previously discussed decomposition methods. The object-oriented structure of the optimization models in Pyomo allows us to generate a `Block` component for our purposes. Inside the original model, the block `MindtPy_linear_cuts` stores all the data used such as the cuts in the master problem, or intermediate solutions while solving the model.

The first step after receiving the model is to perform an initialization for the decomposition methods. We require an initial integer combination to solve the nonlinear program, or a starting solution to that problem to populate the dual and gradient information required for the initial master problem. Given this, we implement several initialization procedures: 1) solve a continuous relaxation of the MINLP, providing a valid lower bound and a linearization point to generate the master problem; 2) use the user provided initial point as a fixed discrete combination and use this to solve the master problem; 3) maximize the sum of the discrete variables subject to the linear constraints in an assignment problem to obtain this initial discrete combination.

The gradients of the nonlinear constraints are calculated using the `differentiate` function in Pyomo, which performs an exact differentiation of the equations defining the corresponding derivatives. Given that these equations are encoded in expression trees, the differentiation is performed efficiently.

Other options, such as the inclusion of disjunctive cuts generated to ignore the previous solutions, known as integer cuts; the selection of different subsolvers for the MILP and NLP problems; and the convergence tolerances are available for the user to modify. Other algorithmic options include the use of different primal heuristics for finding feasible solutions, namely the Center-cut algorithm and the Feasibility pump, and the

use of regularization for the decomposition methods. The implementation can be found in the public repository of the project (Bernal and Chen, 2017).

4. Numerical example

To illustrate the use of MindtPy with different methods, we solve the problem in Eq. (6). Defining the problem as an optimization model inside Pyomo called `SimpleMINLP`, the command to solve this problem for example with OA and solving the relaxed NLP as an initial strategy we execute within Python, importing Pyomo:

```
opt=SolverFactory('MindtPy')
opt.solve(SimpleMINLP, strategy = 'OA', initi_strategy = 'rNLP')
```

$$
\begin{aligned}
Z = \min_{x,y}\ & y_1 + 1.5y_2 + 0.5y_3 + x1^2 + x2^2 \\
& \text{s.t.}: (x_1 - 2)^2 - 2y_1 \leq 0 \\
& \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ -1 & 0 \\ 0 & 1 \\ -1 & -1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \leq \begin{bmatrix} 0 \\ 4 \\ -1 \\ 0 \\ -1 \\ -1 \end{bmatrix} \\
& 0 \leq x \leq 4, x \in \mathbb{R}^2, y \in \{0,1\}^3
\end{aligned} \tag{6}
$$

We solve the problem using the four different methods and an initial point $x^0 = 0$, $y^0 = 1$. The results for each iteration are presented in Table 1. The strength of the lower bound results in a convergence in fewer iterations for OA and PSC, while the ECP method takes more iterations since it does not solve the subproblem, although every iteration is computationally cheaper since it only requires the solution of an MILP.

Table 1. Numerical example: Upper and lower bounds against iteration for each method.

Method	OA		PSC		GBD		ECP
Iteration	LB	UB	LB	UB	LB	UB	LB
1	1	11	1	11	-24	11	0.5
2	1.5	5	1.5	5	-23.5	11	0.5
3	3.5	3.5	3.5	3.5	-3.5	5	1
4					-2.5	5	2.02
5					3.5	3.5	3.5

5. Computational implementation and performance

In this section, we solved several convex MINLP with PSE applications from the MINLPLib2 (Vigerske, 2014). We used Gurobi7.5 and IPOPT3.12 as subsolvers. The tests were performed on a desktop running in Ubuntu16.04, with an Intel Core 2 Duo processor, a CPU of 3.4 GHz and 16 GB of RAM using Pyomo 5.2. We solve every instance with the four methods using the rNLP as initial strategy. The results are presented in Table 2. There we can see that for all the examples, OA managed to converge in fewer or equal iterations than PSC, which also converged in fewer iterations than GBD as mentioned in Eq. (5). We can see that ECP also converged in more

iterations but the computational time per iteration was smaller, showing the trade-offs in performance.

Table 2. Solution details for different test cases using the 4 methods solved to $\varepsilon = 10^{-5}$. Termination without proof of optimality in less than 1000 iterations are marked with *

Method	OA		PSC		GBD		ECP	
Instance	Iters.	Time (s)	Iters.	Time (s)	Iters.	Time (s)	Iters.	Time (s)
flay03m	9	0.875	9	0.880	385	57.393	46	3.521
batchdes	2	0.247	4	0.402	45	16.703	4	0.265
ex4	3	1.152	369	147.713	430	182.987	8	1.526
synthes3	7	0.833	12	1.360	86	10.380	19	1.369
enpro48pb	3	1.064	3	1.002	*	*	6	1.239

6. Conclusions

An open source toolbox for MINLP solutions based on decomposition methods implemented in Pyomo was presented in this paper. This toolbox allows the users to modify several algorithmic options for these decomposition methods, allowing him/her to test easily different solution approaches to MINLP problems. MindtPy also provides algorithm designers with a platform to easily test and prototype ideas with access to all the capabilities of Pyomo, such as access to specialized subsolvers. We successfully implemented and tested using PSE related MINLP problems showing the flexibility of the toolbox.

References

Bernal, D.E., Chen, Q., 2017. MindtPy [WWW Document]. URL https://github.com/bernalde/pyomo/tree/mindtpy

Bonami, P., Biegler, L.T., Conn, A.R., Cornuéjols, G., Grossmann, I.E., Laird, C.D., Lee, J., Lodi, A., Margot, F., Sawaya, N., Wächter, A., 2008. An algorithmic framework for convex mixed integer nonlinear programs. Discret. Optim. 5, 186–204. https://doi.org/10.1016/j.disopt.2006.10.011

Bussieck, M.R., Vigerske, S., 2011. MINLP Solver Software, in: Wiley Encyclopedia of Operations Research and Management Science. https://doi.org/10.1002/9780470400531.eorms0527

Duran, M.A., Grossmann, I.E., 1986. An outer-approximation algorithm for a class of mixed-integer nonlinear programs. Math. Program. 36, 307–339. https://doi.org/10.1007/BF02592064

Geoffrion, A.M., 1972. Generalized Benders decomposition. J. Optim. Theory Appl. 10, 237–260. https://doi.org/10.1007/BF00934810

Hart, W.E., Laird, C.D., Watson, J.-P., Woodruff, D.L., Hackebeil, G.A., Nicholson, B.L., Siirola, J.D., 2017. Pyomo — Optimization Modeling in Python, Springer Optimization and Its Applications. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-58821-6

Lee, J., Leyffer, S., 2012. Mixed integer nonlinear programming. Springer.

Quesada, I., Grossmann, I.E., 1992. An LP/NLP based branch and bound algorithm for convex MINLP optimization problems. Comput. Chem. Eng. 16, 937–947. https://doi.org/10.1016/0098-1354(92)80028-8

Vigerske, S., 2014. Towards MINLPLib 2.0 Model instance collections.

Westerlund, T., Pettersson, F., 1995. An extended cutting plane method for solving convex MINLP problems. Comput. Chem. Eng. 19, 131–136. https://doi.org/10.1016/0098-1354(95)87027-X

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50145-2

Simultaneous Parameter Estimation in Reactive-Solvent-Based Processes

John C. Eslick[a*], Paul T. Akula[b], Debangsu Bhattacharyya[b*], David C. Miller[a]

[a]*National Energy Technology Laboratory, 626 Cochrans Mill Rd., Pittsburgh, PA 15236, USA*
[b]*Department of Chemical and Biomedical Engineering, West Virginia University, Morgantown, WV 26506, USA*
John.Eslick@netl.doe.gov

Abstract

Accurate estimation of model parameters is difficult for reactive solvent-based processes due to coupled effects of mass transfer, heat transfer, and reaction kinetics. While commercial process simulators provide capabilities for parameter estimation, it is generally difficult to simultaneously regress parameters of different submodels (e.g. parameters of physical properties and mass transfer models). Morgan et al. (2018) and Chinen et al. (2018) have demonstrated the value of the simultaneous parameter estimation approach in developing highly predictive models. The approach used by Chinen et al. (2018) embedded a simulation of a CO_2 capture process in an external derivative free optimization (DFO) framework. While this approach works well in many cases, it is computationally inefficient, difficult to set up, and limited by the amount of data that can be used. This paper presents an extension to that work where the process model and a generalized parameter estimation tool are developed in the Institute for Design of Advanced Energy Systems (IDAES) process systems engineering (PSE) framework. The IDAES framework allows an entire process model to be implemented in an equation-oriented framework with access to state-of-the-art optimization methods. Improved computational efficiency and ease of implementation are demonstrated.

Keywords: Simultaneous parameter estimation, optimization, CO_2 capture, IDAES

1. Introduction

Accurate estimation of model parameters is important in the development of predictive process models; however, this is difficult for reactive solvent-based processes due to the coupled effects of mass transfer, heat transfer, and reaction kinetics. Experimental data spanning different scales and operating regimes are required to obtain a process model that is predictive for different scales and widely varying operating conditions. Simultaneous parameter estimation for solvent based CO_2 capture systems has been described and a methodology has been presented using chemical process simulators embedded in a DFO framework (Chinen et al. 2018). The focus of this paper is to extend that work by implementing a generic framework that will support the approach within the IDAES PSE framework, which provides a modular process modeling framework on top of the Pyomo algebraic modeling language (AML) (Hart et al., 2011; Hart et al., 2017).

The novelty of the approach is that it facilitates large-scale parameter estimation for hundreds of parameters at multiple scales embedded in complex process models and submodels in a way that is not generally possible with commercial process simulators.

Depending on the complexity and robustness of the process models, over 100 parameters, thousands of experimental data sets, and several hundred thousand equality and inequality constraints can be handled through decomposition methods. The upper limit of this methodology has yet to be identified.

The IDAES framework provides basic modular flowsheeting features and a model library allowing optimization problems to be implemented much more quickly than in a general AML. The benefit of this approach enables the process model and all submodels to be implemented in a common framework where all variables and parameters are easily accessible. In addition, full derivative information is available enabling the use of state-of-the-art optimization solvers and decomposition methods through Pyomo. These features allow the problem to be implemented in a more straightforward way and solved with significantly reduced computation times and superior optimality conditions. The approach presented here can readily solve problems comprising multiple models and submodels while utilizing data from multiple sources and scales.

2. Simultaneous Parameter Estimation

Simultaneous parameter estimation is a technique of regressing parameters of a process model and various interrelated submodels at the same time. This approach allows data from multiple scales and system configurations to be used simultaneously to improve the predictiveness of a model. Features of each coupled sub-model are better identified over a wider domain by this approach. The simultaneous parameter estimation problem formulation used in this work is given by Eq. 1.

$$\min \sum_{i \in R} \sum_{j \in S_i} \frac{\sum_{k \in T_i} \left(\hat{y}_{i,j,k} - y_{i,j,k} \right)^2}{\sigma_{i,j}^2} \tag{1}$$

$$f_{i,k}\left(\theta, x_{i,k}, \hat{y}_{i,k}, z_{i,k}\right) = 0, \quad \forall\, i \in R, k \in T_i$$

Where: R = Data sets (e.g., packed column data and wetted wall column data)
S_i = Set of output measurements for data set i
T_i = Set of experiments in data set i
$\hat{y}_{i,j,k}$ = Predicted output for measurement j for experiment k in set i
$y_{i,j,k}$ = Measured output for measurement j, for experiment k in set i
$f_{i,k}$ = Model equations for experiment k in set i
$x_{i,k}$ = Measured inputs for experiment k in set i
$z_{i,k}$ = Other model variables for experiment k in set i
θ = Model Parameters (same across all experiments)
$\sigma_{i,j}^2$= Variance of measurement j in data set i (for scaling)

Challenges with the simultaneous parameter estimation approach arise due to the size and complex nonlinear nature of the problem. The number of equations to be solved becomes large when considering large data sets, due to the optimization problem containing equations for a full process model for each experiment. This requires the models to be robust over a wide range of conditions.

To overcome these challenges, the IDAES framework was used for this work. The PySP module in Pyomo provides several decomposition techniques for stochastic programming (Watson et al., 2012). The parameter estimation problem is equivalent to a 2-stage stochastic programming problem with no recourse. In the 2-stage stochastic programming problem, a set of scenarios are generated with different parameter values

based on parameter uncertainty distributions, and the weighted sum of scenario objectives is optimized. First-stage variables must be the same for each scenario, while second-stage variables can be adjusted in each scenario to provide recourse against uncertainty. This problem can be decomposed by solving each scenario separately using one of several algorithms to iteratively converge to a solution. In the case of parameter estimation, each experiment corresponds to a scenario, and the inputs to the experiment correspond to the parameter values for a scenario. The parameters being estimated correspond to first-stage variables, and there are no second-stage (recourse) variables.

3. Case study: Amine solvent-based CO_2 absorption

CO_2 capture using aqueous amines is a well-established technology, and a significant amount of bench- and pilot-scale data is available for monoethanolamine (MEA), making it an ideal test case for a parameter estimation framework. This work is applicable to other solvent system as well as other types of process models. Data from two systems are used in this study (Figure 1): (a) a pilot-scale packed column with intercooling (data from the National Carbon Capture Center) and (b) a bench-scale isothermal wetted wall column (WWC) system (Dugas, 2009).

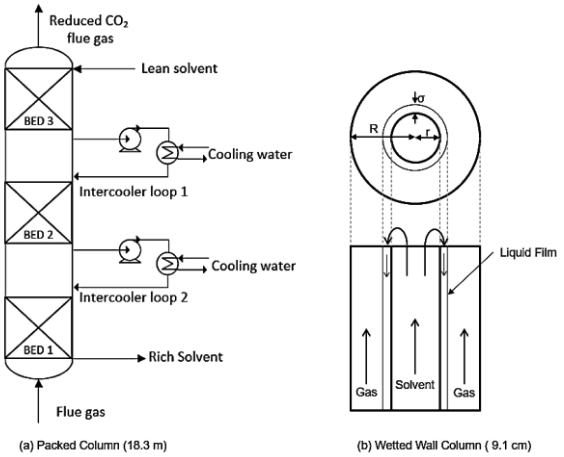


Figure 1: Packed Column and Wetted Wall Column

A 1D rate-based packed absorber column model is implemented. This model is similar to the model of Morgan et al. (2018) with three main differences: (1) an enhancement factor approach is used to characterize the mass transfer resistances through the liquid and gas films as opposed to a rigorous two-film model, (2) the NRTL model is used for calculating the activity coefficients instead of the electrolyte-NRTL (eNRTL) model, and (3) a simplified heat of absorption model is used. A simpler model is used in this work compared to Morgan et al. (2018), and work is ongoing to implement the fully detailed model. The WWC model is developed by considering a column equivalent to the annular space, with appropriate models for the interfacial area and liquid holdup.

Parameters are estimated for the reaction rate constants (A and E in Eq. 2), the vapor phase mass transfer coefficient (C_v, a, and b in Eq. 3), the liquid mass transfer coefficient (C_l in Eq. 4), and the interfacial area (β_1, β_2, and β_3 in Eq. 5) models. The effective interfacial area correlation is applied to the packed-bed model only.

$$k_i = A_i \exp\left(\frac{E_i}{T}\right), \qquad i\epsilon\{\mathrm{H_2O}, \mathrm{MEA}\} \tag{2}$$

$$k_{v,i} = C_v D_{v,i} \left(\frac{a}{\varepsilon_v d_h}\right)^{1/2} \left(\frac{u_v \rho_v}{a \mu_v}\right)^{a} \left(\frac{\mu_v}{\rho_v D_{v,i}}\right)^{b} \tag{3}$$

$$k_{l,CO_2} = 12^{1/6} C_l \left(\frac{u_l}{\varepsilon_l}\right)^{1/2} \left(\frac{D_{l,CO_2}}{d_h}\right)^{1/2} \tag{4}$$

$$\left(\frac{a_e}{\beta_1}\right)^{\beta_2} = {u_l}^{\beta_3} \left(\frac{\rho_l}{\sigma_l}\right) \tag{5}$$

The objective function is given by Eq. 1. The WWC data set is comprised of 35 experiments where CO_2 flux between the vapor and liquid phase is measured. The packed column data set has 17 points where CO_2 capture fraction is measured. The problem is solved using the IPOPT NLP solver (Biegler and Zavala, 2009).

4. Results

The optimization results are presented in Table 1. The initial parameter values were taken from Aboudheir et al. (2003) (Eq. 2) and Morgan et al. (2015) (Eqns. 3-5). The WWC model R^2 was 81.3 % for the initial and 98.3 % for the optimal parameter values. The packed column R^2 was 75.2 % for the initial and 83.3 % for the optimal parameter values. The root mean squared error is 6.4×10^{-5} mol/s/m^2 for the WWC and 1.7 percentage points of CO_2 capture for the packed column models. Some parameters were found to be on their bounds; however, widening the bounds did not yield significant improvement to the fit. Capabilities for parametric uncertainty analysis is currently being developed and will be utilized in future work.

Table 1: Fitted parameters

Model	**Kinetics**				**Mass Transfer**				**Interfacial Area**		
Parameter	A_{H2O}	E_{H2O}	A_{MEA}	E_{MEA}	C_v	a_v	b_v	C_l	β_1	β_2	β_3
Init. value	4.55	3,287	4,610	4,412	0.357	0.750	0.333	0.500	157.0	8.612	1.333
Opt. value*	4.00	4,100	4,800	4,000	0.500	0.495	0.250	0.504	15.0	7.235	-0.50

Figure 2 shows parity plots for the flux in the WWC case and CO_2 capture fraction in the packed column case. The residuals for the packed bed case appear to show a trend suggesting that there are some phenomena not accounted for in the model.

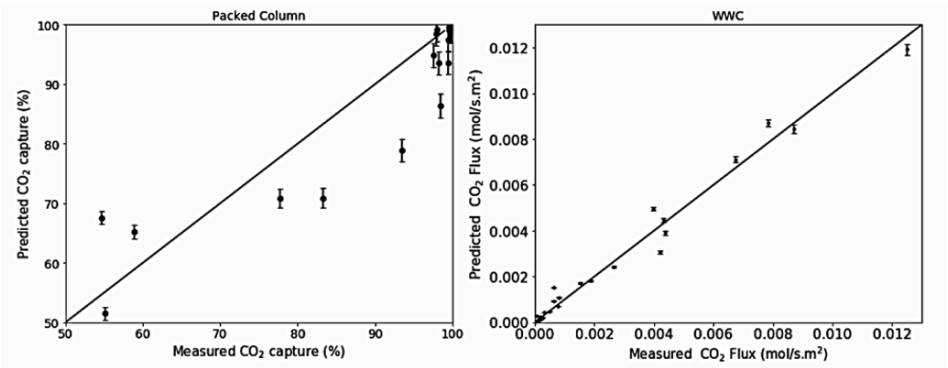


Figure 2: Parity Plots (a) Packed Column, (b) WWC

The packed column temperature profiles in Figure 3 give some insight into this discrepancy among the model used in this work, the Morgan et al. (2018) model, and the plant data. The error bars represent the measurement accuracy based on the type of thermocouples used in the pilot plant. While at low-loading, the temperature profiles agree well, there is a higher discrepancy at high loading. These deviations are within the tolerances given the approximations in the vapor-liquid equilibrium (VLE), mass transfer, and enthalpy models. The e-NRTL model is expected to improve precision by improving the VLE predictions for CO_2 high loading conditions compared with the NRTL model used in this work (Luo et al., 2015; Morgan et al., 2017).

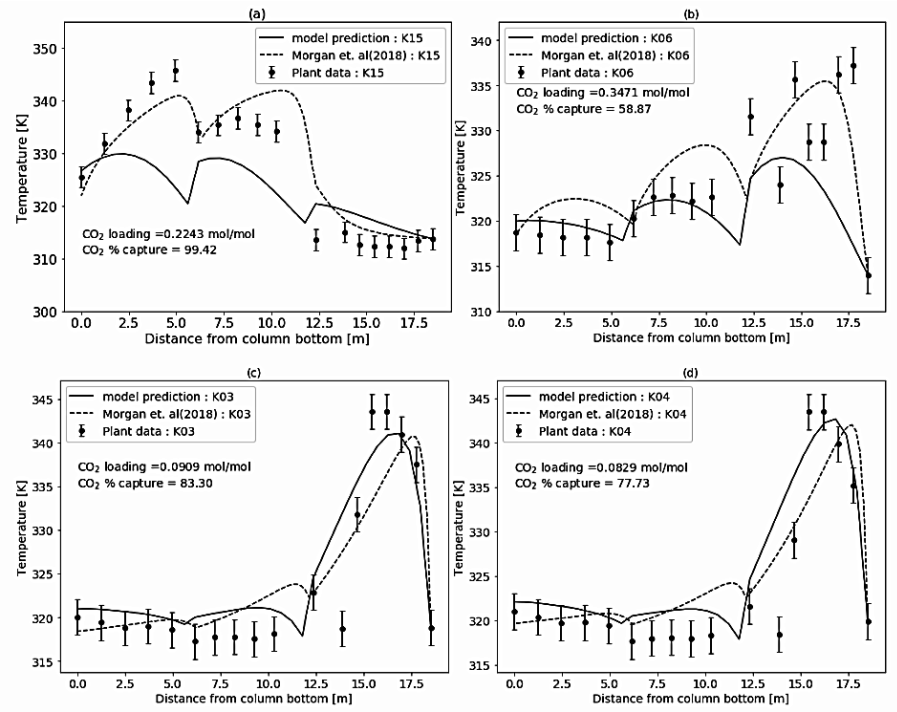


Figure 3: Packed Colum Temperature Profiles, this work, National Carbon Capture Center Data, and Morgan et al. (2018)

The parameter estimation problem in this work consisted of 71,768 constraints, many of which are highly non-linear. The computation time required to solve the optimization problem was 206 seconds on an Intel Xeon E3-1505M CPU, which is a substantial improvement compared with the previous parameter estimation framework (Chinen et al. 2018), which took about 36 hours to run on a similar machine.

5. Conclusions

A new framework for simultaneous parameter estimation has been demonstrated using a system of industrial significance. The novelty of this system is that it facilitates simultaneous parameter estimation for parameters in multiple submodels embedded in complex process models, which is not generally possible with commercial process simulators. The IDAES framework provides basic modular flowsheeting features and a model library allowing optimization problems to be implemented much more quickly than in a general AML. The result showed a significant improvement in computational

efficiency compared with previous methods presented by Chinen et al. (2018), which used a process simulator embedded in a derivative free optimization framework. It is expected that this will lead to the ability to solve larger, multiscale parameter estimation problems with large amounts of data. Future work will provide enhancements to the parameter estimation framework to enable uncertainty both in parameter estimates and measurements to be considered.

Acknowledgment and Disclaimer

This research was funded by the Lawrence Berkeley National Laboratory as part of the U.S. DOE's Institute for the Design of Advanced Energy Systems (IDAES). This paper was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

References

A. Aboudheir, P. Tontiwachwuthikul, A. Chakma. R. Idem, 2003, Kinetics of the reactive absorption of carbon dioxide in high CO2 loaded, concentrated aqueous monoethanolamine solutions, Chemical Engineering Science, 58, 5195-5210.

L. T. Biegler and V. M. Zavala ,2009, Large-scale nonlinear programming using IPOPT: An integrating framework for enterprise-wide dynamic optimization, Computers & Chemical Engineering, 33, 3, 575-582.

A. S. Chinen, J. C. Morgan, B. Omell, D. Bhattacharyya, C. Tong, D. C. Miller, 2018, Development of a gold-standard model for solvent-based CO2 capture. Part 1: hydraulic and mass transfer models and their uncertainty quantification, Under internal review.

R. E. Dugas, 2009, Carbon dioxide absorption, desorption, and diffusion in aqueous piperazine and monoethanolamine, Doctoral dissertation, University of Texas at Austin.

W. E. Hart, J.-P. Watson and D. L. Woodruff , 2011, Pyomo: modeling and solving mathematical programs in Python, Mathematical Programming Computation, 3, 3, 219-260.

W. E. Hart, C.D. Laird, J.-P. Watson, D.L. Woodruff, G.A. Hackebeil, B.L. Nicholson, and J.D. Siirola, 2017, Pyomo – Optimization Modeling in Python, Second Edition.

X. A. Luo, Hartono, S. Hussain, H. F. Svendsen , 2015, Mass transfer and kinetics of carbon dioxide absorption into loaded aqueous monoethanolamine solutions, Chemical Engineering Science, 123, 57-69.

J. C. Morgan, A. S. Chinen, B. Omell, D. Bhattacharyya, C. Tong, D. C. Miller, B. Buschle, M. Lucquiaud , 2018, Development of a gold-standard model for solvent-based CO2 capture. Part 2: steady-state validation and uncertainty quantification with pilot plant data, Under internal review.

J. C. Morgan, B. Omell, D. Bhattacharyya, C. Tong, D. C. Miller, 2017, Thermodynamic modeling and uncertainty quantification of CO2-loaded aqueous mea solutions, Chemical Engineering Science, 168, 309-324.

J. C. Morgan, D. Bhattacharyya, C. Tong, D. C. Miller, 2015, Uncertainty quantification of property models: Methodology and its application to CO2-loaded aqueous MEA solutions. AIChE Journal 61, 6, 1822-1839.

J.-P. Watson, D. L. Woodruff, W. E. Hart, 2012, PySP: modeling and solving stochastic programs in Python, Mathematical Programming Computation, 4, 2, 109-149.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50146-4

Novel MILP-based iteration approach for the equation-oriented simulation and optimization of distillation column

Kunru Yang, Ming Pan, Bingjian Zhang, Qinglin Chen*

School of Chemical Engineering and Technology / Guangdong Engineering Center for Petrochemical Energy Conservation, *Sun Yat-Sen University*, *Guangzhou 510275, China*
chqlin@mail.sysu.edu.cn

Abstract

The material balance, equilibrium, summation and heat balance (MESH) equation describing a distillation column is highly nonlinear, making it very difficult to optimize. A new solution strategy is proposed based on the iterative mixed integer linear programming (MILP) method introduced in the earlier work (Pan et al., 2013). Rigorous phase equilibrium is reformulated by using high dimensional model representation (HDMR), which retains high accuracy (less than 1% of relative deviation) compared with the original thermodynamic models provided by Aspen Plus. Then, the MILP-based iteration approach is utilized to solve the resulting nonlinear programming problems. To validate the proposed approach, a benzene-toluene-o-xylene (BTX) distillation column model is carried out and compared with the results obtained from Aspen Plus. More importantly, the optimization of capital and operating expenses of the system can be achieved based on the iterative MILP-based method which yields an objective of 11.08 with elapsed time of 400 seconds.

Keywords: Simulation, Optimization, Mixed Integer Linear Programming (MILP)–based iteration, Distillation Column

1. Introduction

MESH equation is a system of highly coupled nonlinear equations which is usually solved using Newton or quasi-Newton-type methods. As they are locally convergent, initial guesses that are close enough to the solution are required. Recently, the pseudo-transient continuation (PTC) method was proposed by (Pattison and Baldea, 2014). Through constructing pseudo-transient processes and converting algebraic equations into differential algebraic equations (DAE), the convergence basin of the original model was largely extended.

Traditionally, the design of distillation column is firstly estimated by short-cut method based on simplifying assumptions such as constant molar overflow and constant relative volatility, followed by verification using rigorous tray-by-tray methods. As a result, an optimal synthesis requires a lot of simulations and sensitivity analysis. Another simulation based approach is to combine process simulators with genetic algorithm (GA) reported by (Bravo et al., 2010). The search for optimal synthesis is formulated as a multi-objective problem considering both capital and operating cost.

Besides simulation based approaches, some equation-oriented methods based on mathematical programming were also successfully implemented. (Viswanathan and

Grossmann, 1990) formulated the optimization of the distillation column described by MESH equations as a MINLP problem, which was solved using the outer approximation algorithm. Later, in order to handle the non-existent tray, (Yeomans and Grossmann, 2000) developed generalized disjunctive programming (GDP) model. However, (Dowling and Biegler, 2014) proposed bypass efficiency (ε^j) method, which replaces the stage number by using the summation of continuous variables ε^j. Recently, (Pattison et al., 2016) presented a reformulation of equilibrium tray model as a pseudo-transient DAE system and integrated with previously developed time-relaxed-based optimization algorithm.

The rest of this paper is structured as follows. The reformulation of the thermodynamic model using HDMR method is first introduced in Section 2. In Section 3, the distillation column super structure is illustrated, followed by a detailed MINLP problem statement. Further, an iterative MILP optimization procedure is addressed in Section 4. Finally, Section 5 presents a case study to demonstrate the validity and efficiency of the proposed approach.

2. Thermodynamic model reformulation using HDMR

In MESH equations, the rigorous thermodynamic model describing the phase equilibrium and enthalpy are the most complicated because of the high nonlinearity. Generally, the k value is a function of temperature, pressure, liquid composition and vapor composition, while the enthalpy is a function of temperature and liquid composition or vapor composition. The relations are expressed as follows:

$$k_i = f(T,\ P,\ x_i,\ y_i), \quad H^L = f(T,\ x_i), \quad H^V = f(T,\ y_i) \tag{1-3}$$

To simplify the model while maintain high accuracy, the HDMR method is utilized. In HDMR, the output variable is represented as the summation of a series of functions of input variables. The relation is as follows:

$$y = f_0 + \sum_{i=1}^{N} f_i(x_i) + \sum_{i=1}^{N}\sum_{j=i+1}^{N} f_{ij}(x_i, x_j) + \cdots + f_{12\ldots N}(x_1, x_2, \cdots x_N) \tag{4}$$

N is the number of input variables. Subscript i and j denotes specific variable x. f_0 is the average value of all $f(x)$.

In most of cases, functions containing more than two correlated variables can be omitted. Thus, the simplified form is as follows:

$$y = C + \sum_{i=1}^{N}\sum_{k=1}^{K} A_{i,k} \times x_i^k + \sum_{i=1}^{N}\sum_{j=i+1}^{N}\sum_{k=1}^{K}\sum_{n=1}^{K} B_{i,j,k,n} \times x_i^k \times x_j^n$$
$$i, j \in \{1, 2, 3, \cdots, N\} \qquad k, n \in \{1, 2, 3, \cdots, K\} \tag{5}$$

C is a constant. $A_{i,k}$ and $B_{i,j,k,n}$ are the first and second order coefficients. k and n denote the power of the variables and K is the maximum power.

The parameters of the above model are obtained by feeding the data set to a linear programming model in GAMS, which is formulated to minimized the total deviation. Let y_d denote the calculated values and y_d^* be the real values. D represents the number of the data sets. The objective function is shown as follows:

$$\min \sum_{d}^{D} \left| y_d - y_d^* \right|, \quad d \in \{1, 2, 3, \cdots, D\} \tag{6}$$

Besides, the maximum allowable deviation is added as a constraint:

$$\left| y_d - y_d^* \right| \le \delta, \quad d \in \{1, 2, 3, \cdots, D\} \tag{7}$$

3. Distillation column superstructure and MINLP model

Figure 1 shows the column representation, which is similar to the approach used by (Yeomans and Grossmann, 2000). Three binary variables yf , yr and $mref$ are introduced to determine the feed tray, reflux tray and whether the tray is existent. The trays that are above the reflux tray can be seen as non-existent.

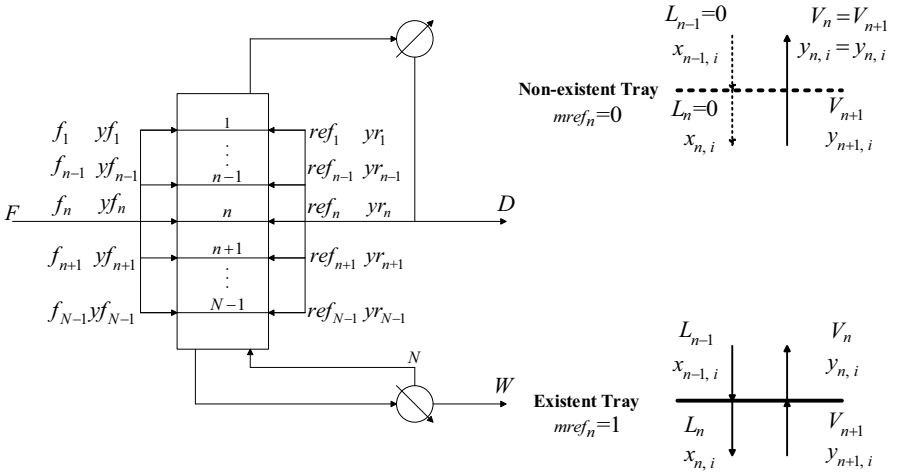


Figure 1: Column superstructure representation

Consider the following set definitions for the model: S is the set of stages n in the column. I represents the set of components i . B is the subset of trays $n \in S$ corresponding to the bottom stage and the reboiler. T is the subset of trays $n \in S$ corresponding to the first tray. As the trays of the column ($n = 1,\ 2,\ \cdots,\ N$) from top to bottom, the top stage is tray 1 and the reboiler is number N .

The objective of the problem is to minimize the capital and operational expenses represented by the summation of the number of stages NT and reflux ratio rr (Viswanathan and Grossmann, 1993), as shown in Eq.(7).

The feed condition, distillate rate and reflux ratio are specified. Eqs.(8)-(14) are representative of mass balance, phase equilibrium, summation and heat balance. Eq.(15) is the mass balance of feed and products. Eqs.(16)-(18) are used to calculate the reflux rate and reflux temperature. Eq.(19) represents the constraint that the reflux tray must be above the feed tray. Eq.(20) is combined with Eq.(21) to determine whether tray n is existent. Eqs.(22)-(25) ensured that the liquid flowrate is set to 0 while the vapor output flowrate and composition are equal to that of input. Eqs.(26)-(31) are the constraints that feed and reflux enter one and only one tray. The modelling equations are as follows:

$$\min \ z = rr + NT \tag{7}$$

$$f_n x_i^{feed} + ref_n x_i^{reflux} + L_{n-1} x_{n-1,\,i} + V_{n+1} y_{n+1,\,i} - L_n x_{n,\,i} - V_n y_{n,\,i} = 0,\ n \in S\ \ i \in I \tag{8}$$

$$k_{n,\,i} = f(T_n,\ P_n,\ x_{n,\,i},\ y_{n,\,i})^{HDMR},\ y_{n,\,i} - k_{n,\,i} x_{n,\,i} = 0,\ n \in S\ \ i \in I \tag{9-10}$$

$$\sum_i x_{n,\,i} = 1,\ \sum_i y_{n,\,i} = 1,\ n \in S\ \ i \in I \tag{11}$$

$$H^{Feed} f_n + H^{Reflux} ref_n + H_{n-1}^{L} L_{n-1} + H_{n+1}^{V} V_{n+1} - H_n^{L} L_n - H_n^{V} V_n = 0,\ \ n \in S\ \ i \in I \tag{12}$$

$$H_n^{L} = f(T_n,\ x_{n,\,i})^{HDMR},\ H_n^{V} = f(T_n,\ y_{n,\,i})^{HDMR},\ \ n \in S\ \ i \in I \tag{13-14}$$

$$Fx_i^{F} - (Dx_i^{D} + Wx_i^{W}) = 0,\ \ i \in I;\qquad T = f(P_{condenser},\ x_{1,\,i})^{HDMR},\ \ i \in I \tag{15-16}$$

$$V_1 = reflux + D,\ reflux = rr \cdot D;\quad \sum_n n \cdot yr_n - \sum_n n \cdot yf_n \geq 1,\ \ n \in S \tag{17-19}$$

$$mref_n = yr_n,\ \ n \in Top;\qquad mref_n = mref_{n-1} + yr_n,\ \ n \in S \setminus Top \tag{20-21}$$

$$L_n \leq M \cdot yr_n \tag{22}$$

$$V_{n+1} - M \cdot mref_n \leq V_n \leq V_{n+1} + M \cdot mref_n,\ \ n \in S \tag{23}$$

$$y_{n+1,\,i} - M \cdot mref_n \leq y_{n,i} \leq y_{n+1,\,i} + M \cdot mref_n,\ \ n \in S \tag{24}$$

$$T_{n+1} - M \cdot mref_n \leq T_n \leq T_{n+1} + M \cdot mref_n,\ \ n \in S \tag{25}$$

$$\sum_n yr_n = 1,\ \sum_n ref_n = Reflux,\ ref_n \leq M \cdot yr_n,\ \ n \in S \tag{26-28}$$

$$\sum_n yf_n = 1,\ \sum_n f_n = F,\ f_n \leq M \cdot yf_n,\ n \in S \tag{29-31}$$

4. MILP model and iteration algorithm

In the above model, Eqs.(8)-(10), Eq.(12), Eqs.(13)-(14) and Eq.(16) are nonlinear, among which, Eqs.(8)-(10), Eqs.(13)-(14) and Eq.(16) are converted to linear equations using first order Taylor series expansion, while the enthalpy H_n^{L} and H_n^{V} are updated as parameters after solving MILP in each iteration, and thus Eq.(12) is converted to a linear equation. Due to the strong coupled nature of MESH equations, some equations will not hold during iterations. Thus Eqs.(7)-(12), Eq.(16) and Eq.(17) are relaxed by adding slack variables. For non-existent tray, the equilibrium and the summation equations cannot be satisfied together and another couple of slack variables are added to Eq.(9) and take on non-zero value for non-existent tray.

As described in the earlier work (Pan et al., 2013), the objective is formulated to minimize the sum of the infeasibility (the slack variables) in the above equality constraints and the absolute differences between initial and updated variables, subject to a certain capital and operating expenses z'. VS denotes the sum of slack variables and ΔV denotes the absolute difference between initial and updated values of the variables. The equations are shown as follows:

$$\min\ VS + \Delta V,\quad z \leq z' \tag{32-33}$$

The iteration algorithm was successfully implemented in maximizing the retrofit profit of HEN, which was given in details in (Pan et al., 2018). The procedure consists of two

loops: in the first loop, a feasible solution is found under a certain capital and operating expenses z'; in the second iteration loop, an initial z' is first estimated and a feasible solution is obtained within a certain upper bound of z', which will decrease gradually until the optimal value is found.

5. Case study

The base case is presented in Figure 2. The Wilson model is selected for the calculation of liquid activity coefficients. In the system of low pressure, the fugacity coefficient is regarded as 1. As the column pressure is also fixed at 1 bar with no pressure drop, the k value is a function of temperature and liquid compositions. The data sets for HDMR are combinations of different temperatures and compositions that are chosen uniformly within a certain range, which is estimated in advance. In this case, the temperature is ranging from 90 ℃ to 150 ℃ and the molar fractions is set between 0 and 1. With the intervals of temperature (2 ℃) and molar fractions (0.02), over 40,000 groups of data are generated using Aspen Properties Excel Add-in. The HDMR parameters are obtained by feeding the data sets to the proposed LP model (Eqs. (5)-(7)) in GAMS. In this case, the relative deviation is less than 1 %.

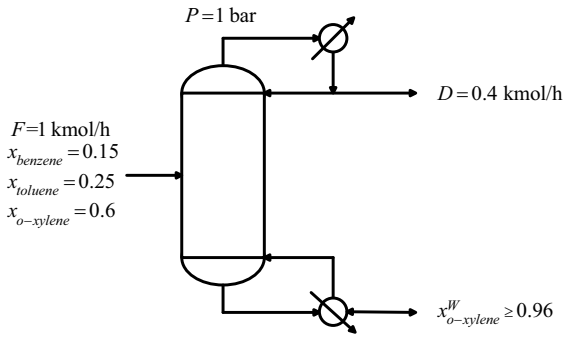


Figure 2: Column Specifications

The aforementioned approaches in Sections 3 and 4 were performed in GMAS 24.4 on an Intel Core i7 processor running at 4.2 GHz. The solvers Antigone, DICOPT and CPLEX are utilized to solve the MINLP and MILP problems.

Table 1: Comparison of Results and Model Statistics

Method	Results	CPU Time(s)	Continuous Variables	Binary Variables	Equations
MILP Iteration	11.08	400	3930	90	3778
MINLP-Antigone	11.08	3600	1639	86	1506
MINLP-DICOPT	11.285	100	1639	86	1506

As shown in Table 1, with the MILP approach, a relatively better objective can be achieved when compared to that obtained by MINLP-DICOPT although the global optimality is not guaranteed. And the time cost is more reasonable than that of MINLP-Antigone approach. At the optimum, the stage number is 8, the feed is at the fourth stage from top and the reflux ratio is 3.08. Finally, the temperature, pressure, flowrate and heat load profiles are compared with Aspen Plus simulation results, and the deviation of major variables are less than 1%, as shown in Table 2. It also can be seen that, the error of toluene (MILP) in bottom product is 11.1 %, which does not affect the accuracy of the MILP results as the key composition in the bottom product is o-xylene with 0.5 % of error compared with the Aspen Plus simulation.

Table 2: Comparison of simulation results of Aspen Plus and MILP approach

	Variables	Values		
Input	Feed composition (benzene, toluene, o-xylene)	0.15/ 0.25/ 0.6		
	Total Stage Number/ Reflux Ratio	8/ 3.08		
	Distillate Rate (kmol/h)	0.4		
	Feed Rate(kmol/h)/ Temperature(℃)	1/ 90		
	Colum Pressure(bar)	1		
Output	Bottom Product Compositions (benzene, toluene, o-xylene)	Aspen Plus: 0.000/ 0.045/ 0.955	MILP approach: 0.000/ 0.040/ 0.960	Error (%): 0/ 11.1/ 0.5
	Top/Bottom Temperature(℃)	96.5/ 141.5	96.4/ 143.1	0.1/ 0.98
	Condenser/Reboiler Duty(kW)	15.3/ 17.4	15.3/ 17.5	0.0/ 0.5

6. Conclusions

The simulation and optimization of distillation column is a challenging task due to the highly coupled nature and high nonlinearity of MESH equations concerning the thermodynamic model. In this work, the equations describing the phase equilibrium and enthalpy model are reformulated using HDMR method and further the nonlinear equations are linearized using first order Taylor series expansion. A MILP-based iterative method has been applied for the equation-oriented simulation and optimization. The case study of BTX separation column has proven that the HDMR reformulation remain high accuracy of the rigorous thermodynamic model, and the proposed approach is more efficient than the original MINLP model.

Acknowledgement

Financial supports from the National Natural Science Foundation of China (No. U1462113) and the Science and Technology Planning Project of Guangzhou city (No. 201704030136) are gratefully acknowledged.

References

Bravo-Bravo, C., Segovia-Hernández, J. G., Gutiérrez-Antonio, C., Durán, A. L., Bonilla-Petriciolet, A., & Briones-Ramírez, A. (2010). Extractive dividing wall column: design and optimization. Industrial & Engineering Chemistry Research, 49(8), 3672-3688.

Dowling, A. W., Biegler, L. T. (2014). Rigorous optimization-based synthesis of distillation cascades without integer variables. Computer Aided Chemical Engineering, 33, 55-60.

Pan, M., Bulatov, I., Smith, R. (2013). New MILP-based iterative approach for retrofitting heat exchanger networks with conventional network structure modifications. Chemical Engineering Science, 104, 498-524.

Pan, M., Bulatov, I., Smith, R. (2018). Heat Transfer Intensification for Retrofitting Heat Exchanger Networks with Considering Exchanger Detailed Performances. AIChE Journal.

Pattison, R. C., Baldea, M. (2015). Equation‐oriented flowsheet simulation and optimization using pseudo‐transient models. Aiche Journal, 60(12), 4104-4123.

Pattison, R. C., Gupta, A. M., & Baldea, M. (2016). Equation‐oriented optimization of process flowsheets with dividing‐wall columns. Aiche Journal, 62(3), 704-716.

Viswanathan, J., Grossmann, I. E. (1993). An alternate MINLP model for finding the number of trays required for a specified separation objective. Computers & chemical engineering, 17(9), 949-955.

Viswanathan, J., Grossmann, I. E. (1990). A combined penalty function and outer approximation method for MINLP optimization. Computers & Chemical Engineering, 14(7), 769-782.

Yeomans, H., & Grossmann, I. E. (2000). Optimal design of complex distillation columns using rigorous tray-by-tray disjunctive programming models. Industrial & engineering chemistry research, 39(11), 4326-4335.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50147-6

Fully Automated Approach for Bio-crude Mixture Modelling Based on GC-MS and Elemental Analyses

Boris Brigljević[a], Petar Žuvela[a,#], J. Jay Liu[a,*], Hee-Chul Woo[a], Jae Hyung Choi[b]

[a]*Department of Chemical Engineering, Pukyong National University, 365 Sinseon-ro, 48547, Busan, Korea*
[b]*The Institute of Cleaner Production Technology, Pukyong National University, 45 Yongso-ro, 48513, Busan, Korea*
[#]*Present address: Department of Chemistry, National University of Singapore, 3 Science Drive 3, 117543, Singapore*
jayliu@pknu.ac.kr

Abstract

The prominence of biofuel research is growing, as do the global energy policies on renewable energy technologies. Accurate process design and simulation is required when evaluating technological and market capabilities of large scale, novel, fuel production processes. Thermochemical decomposition employed in various biofuel production routes (pyrolysis, liquefaction, etc.) yields complex liquid mixtures (bio-crudes) containing numerous compounds. Process simulation of such processes must accurately represent the physical, thermodynamic and chemical properties of bio-crudes, while reducing complexity to a point where it can be handled by a process simulator in a time effective manner. In this work, a software employing automated modelling of bio-crudes based on raw experimental data, has been developed. The program output is a ready-to-use reduced mixture, including all product phases. The mixture is in mass balance with the Proximate and Ultimate analyses of the feedstock biomass material. As there are many approaches to bio-crude modelling, the novelty of this method lies in the combination of minimization of the number of components needed and the minimization of the level of artificiality introduced in the system. Automation of the method allowed for fast reduction and optimization of seven experimental data sets which were subsequently validated by process simulation.

Keywords: Biocrude modelling, Biofuel, Thermochemical decomposition, Process simulation.

1. Introduction

While fossil fuels still, without precedent, remain the life stream of modern societies, their finite nature is becoming imminent. Coupled with the large greenhouse gas surplus their consumption is producing, it is evident that a new energy carrier will soon have to replace them. These concerns gave rise to utilization of biomass, and in recent years especially, aquatic biomass for biofuel production (i.e., microalgae and macroalgae)[1,2]. Other than the obvious advantage that aquatic biomass does not require land for cultivation and that the available area is far larger, marine biomass can absorb carbon dioxide more effectively (6–8 % photosynthetic efficiency) than any terrestrial biomass (1.8-2.2 % photosynthetic efficiency)[3]. Thermochemical conversion processes of biomass offer the most direct conversion, when compared to biochemical. In them, product phase distribution

and composition are open to modifications with varying process parameters such as temperature, pressure and heating rate[4–6]. Accurate process simulation of thermochemical decomposition of biomass, such as pyrolysis is challenging due to the complex nature of the generated products. Pyrolysis liquid products or bio-crudes are complex organic mixtures containing up to hundreds of compounds[7,8], mostly oxygenated hydrocarbons, organic acids, and other organics[7] with varying molecular masses and boiling points. Process simulators, such as Aspen Plus (Aspen Tech, Cambridge, MA, USA), which was used in this work, are equipped with databases containing compounds with well-defined properties[9]. Often, large number of experimentally identified compounds cannot be found in the databases or if they can be found, their properties cannot be well defined. Another common issue in simulation of bio-crudes is defining a mixture with a minimum number of compounds, which, in all respects (chemical composition, density, heating value, etc.), accurately represents the complex bio-crude mixtures[10–12]. The reduced mixture which is representative of the liquid product, together with products from other phases, must be in atomic mass balance with the feedstock. One approach to these challenges is modelling the pyrolysis kinetics using key biopolymer components[12,13]. Although this approach possesses predictive capabilities, it lacks in-built minimizing strategy of components and does not fully employ experimental data of the product characterization. Another approach is the one of Jones et al.[10], and Bonalumi[14] where a set of pseudo-components is defined from the available components in the database of the process simulator, manually mixing-and-matching to tune the properties of the reduced mixture to the original bio-crude. This could save the user time in constructing the reduced mixture, but results in considerable reduction of accuracy. Similar approach was employed by Cruz et al.[15] where real components and pseudo-components were combined, with the ultimate goal of reproducing the true boiling point (TBP) curve of the bio-crude. Although some real components were used, the total number of components in the reduced mixture was quite large (40 components). Finally, Carrasco et al.[16] manually picked components, which are known to be present in both the bio-crude and the process simulator database. This approach is perhaps the most accurate of and closest to the true bio-crude mixture. However, it does not utilize raw characterization data, and requires a great deal of tedious manual reduction. It also requires extensive a priori knowledge about bio-crude composition making its applicability and accuracy questionable. All the above-mentioned approaches introduce a degree of artificiality and bias to the process simulation.

In this work, we aimed to develop a widely applicable, user-friendly and fully automated reduction method, which directly utilizes raw experimental data. Components which comprise the reduced mixture for bio-crude simulation, are derived from GC-MS data. They are thereby unbiased, and accurately represent the original mixture over key characterization parameters while their number is minimized. As the time required for the manual execution of this method and similar methods is considerable (up to two hours per dataset on average), a software has been developed to automate the procedure. The outputs of the software, in a spreadsheet form, are the reduced mixture representing the liquid pyrolysis product, which is a ready-to-use process simulator input, in which the atomic composition of the biomaterial (feed) is in optimal mass balance with the complete set of products of thermochemical decomposition. This alleviates the need for the time consuming manual reactor balancing.

2. Materials and methods

2.1. Datasets and experimental conditions

The method was tested on seven datasets (Table 1) originating from published works by Choi et al[17–19]. In them, brown macroalgae *Saccharina japonica* was subjected to pyrolysis for investigation in biofuel application. Pyrolysis was performed in fixed- and fluidized-bed reactors under various process conditions (temperature, heating rate, fluidizing velocity, etc.). Proximate and ultimate elemental analyses of the solid feedstock as well as GC-MS analyses of the liquid products were performed for characterization. The higher heating values (HHV) of the feedstock and bio-crudes were calculated using models by Channiwala et al.[20] and Dulong-Bertolet[21], respectively.

2.2. Key concepts of the reduction method

Raw experimental data obtained using GC-MS represent the core of the developed reduction method. GC-MS analysis provided: compound IUPAC names, retention times (min), normalized peak areas (%), and chemical group. BPs and CAS numbers were taken from external literature sources (Chemical Abstracts Service, Columbus, OH, USA). Objective of the reduction method is maximum reduction of the number of components from the original (complex) mixture, while minimizing the difference of key parameters that are points of comparison. Key parameters chosen to compare characteristics and control the level of reduction between original and reduced mixtures were: (i) relative ratios of components in specific chemical groups, (ii) weighted averages of chemical group molecular masses (MM), (iii) weighted averages of chemical group boiling points (BP), and (iv) total atomic amounts of all chemical groups within a single phase. Total amount of atoms (C, H, N, O; kmol) from the entire mixture was calculated from the sum of relative amount contributions from every component in the mixture. Subsequently, components from the same chemical group are identified and classified into subgroups using a threshold retention time value for their discrimination. The threshold retention times for each compound group were optimized using a Genetic Algorithm[22] for a minimum number of compounds within a mixture, with maximum absolute difference in weighted average MM and BP difference as a constraint function. Maximum BP absolute difference of 30°C was taken as a limit beyond which components in the same chemical group are viewed as not similar. Finally, components viewed as structurally similar were reduced. Components with the largest normalized peak area were taken as a representative for a subgroup while others are eliminated. The values of normalized peak areas from eliminated components are then added to the normalized peak area of the remaining component, thus increasing it. Applying this to all chemical groups within the mixture, the total number of components is reduced while the parameters do not considerably change, when compared to raw data (Figure 1).

Table 1. Experimental datasets and their conditions

#	Pyrolysis temperature / reactor type / pre-treatment
1	350 °C / fluidized bed / water-washed
2	375 °C / fluidized bed / water-washed
3	400 °C / fluidized bed / water-washed
4	425 °C / fluidized bed / water-washed
5	450 °C / fluidized bed / water-washed
6	375 °C / fluidized bed / no pre-treatment
7	450 / fixed bed / acid-washed

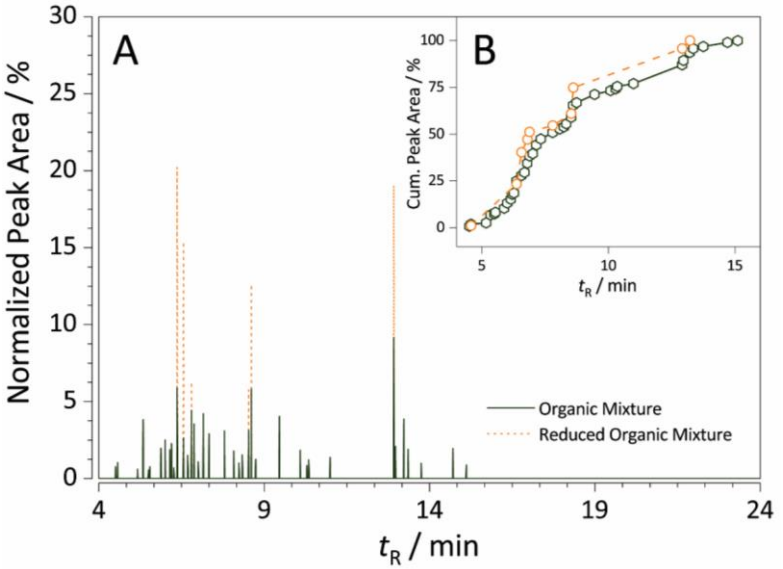


Figure 1. Exemplary depiction of reduction method on the organic phase of dataset 3: A) a normalized GC-MS chromatogram, B) comparison of trends of cumulative peak area over retention time. Experimental organic phase is denoted in full line, while reduced organic phase is denoted in dashed line.

2.3. Automation, performance evaluation and validation of the reduction method

The reduction method and reactor balance calculations were automated in Matlab 2017a (Mathworks, Sherborn, MA, USA). To evaluate the performance of the reduction method, an index, the reduction factor was introduced. It is defined as the ratio of the total number of components in the organic and aqueous phase identified by GC-MS and the total number of components in the reduced mixture. Results of the program (reduced mixtures and reactor balances) were validated by simulation in Aspen Plus 10 (Aspen Tech, Cambridge, MA, USA) at a scale of 100 tonnes per hour of dry seaweed. In order to test the statistical significance of the differences between the experimental and simulated data, two sample *t*-test and one-way analysis of variance (ANOVA) with the Tukey test[23] were used as means of further validation in OriginPro2017. All the statistical hypotheses were tested at a significance level of 95 % ($\alpha = 0.05$).

3. Results and discussion

Table 2. Reduction method performance evaluation

Dataset	1	2	3	4	5	6	7
	Number of compounds						
Experimental*	41	48	64	63	58	55	120
Reduced mixture	15	18	15	16	15	17	22
Reduction factor	2.73	2.67	4.27	3.94	3.87	3.24	5.45
Runtime / min	2.81	3.09	3.5	3.45	3.3	3.19	5.11

3.1. Reduction evaluation parameters

The developed software has reduced seven mixtures by a factor no less than 2.6, while the maximum reduction factor was 5.5 (average 3.7, Table 2). Wall-time was measured, and the lowest wall-time was 2.8, while the highest was 5.1 min. The total number of components is reduced while their normalized peak areas are increased in accordance with the reduction method (Figure 1A). Figure 1B depicts the comparison of trends of cumulative peak area in respect to increasing retention time. The trends are in an excellent agreement and are significantly similar ($t = 0.222$; $p = 0.825 > t$ assuming equal variances; $t = 0.232$; $p = 0.820 > t$ not assuming equal variances).

3.2. Comparison of validation parameters

Reduction method was validated on a simulation of a biofuel production process based on pyrolysis of macroalgae *Saccharrina japonica*. For that purpose, the following validation parameters were used: organic phase mass yield (MY), water content (WC), higher heating value (HHV), and chemical energy recovery (ER). From Figure 2A. It can be observed that the results of the simulation, using the reduced mixture, adhere closely to the experimental values with RMSE values of 2.66 %, 2.32 %, 3.24 MJ kg-1 and 0.65 % for ER, MY, HHV, and WC. Besides that, all the differences between mean values of simulated and experimental validation parameters were found to be insignificant ($p = 0.993 > \alpha$, $p = 0.999 > \alpha$, $p = 0.982 > \alpha$, and $p = 1.000 > \alpha$ for ER, MY, HHV, and WC, respectively).

Figure 2B shows various data points with calculated ER for experimental and simulated organic phases in respect to temperature. Arching lines with represent data sets 1-5 (Table 1) from left to right respectively. Two data points below represent data sets 6-7. Simulated values of ER closely follow experimental ER trend for each data set (representing different reaction conditions). Experimental and simulated water contents are depicted in Figure 2C in respect to simulated density.

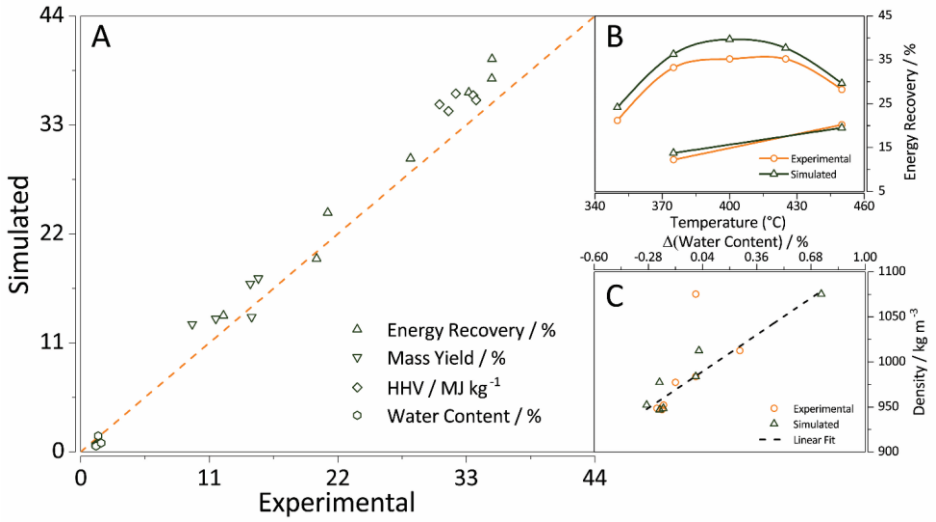


Figure 2. Validation of the automated reduction method. Dependence of A) experimental and simulated validation parameters, B) simulated and experimental energy recovery in respect to pyrolysis temperature, and C) difference in simulated and experimental water content.

4. Conclusions

In conclusion, this is the first time that an automated modelling method for bio-crudes has been developed. It has been tested on seven data sets requiring only 3.5 min wall-time per data set, thereby demonstrating its wide applicability and fast execution. The developed method was validated through process simulation and it has been shown that it produces reduced mixtures, which accurately represent bio-crudes based on statistical inference of parameters commonly valued in biofuel process simulation. Finally, the ease-of-use, wide applicability, execution speed, and accuracy make this methodology a powerful tool in routine biofuel process simulation and design.

References

[1] A. Cole, Y. Dinburg, B.S. Haynes, Y. He, M. Herskowitz, C. Jazrawi, M. Landau, X. Liang, M. Magnusson, T. Maschmeyer, A.F. Masters, N. Meiri, N. Neveux, R. de Nys, N. Paul, M. Rabaev, R. Vidruk-Nehemya, A.K.L. Yuen, 2016, From macroalgae to liquid fuel via waste-water remediation, hydrothermal upgrading, carbon dioxide hydrogenation and hydrotreating, Energy & Environmental Science, 9, 1828–1840.
[2] L.M.L. Laurens, J. Markham, D.W. Templeton, E.D. Christensen, S. Van Wychen, E.W. Vadelius, M. Chen-Glasser, T. Dong, R. Davis, P.T. Pienkos, 2017, Development of algae biorefinery concepts for biofuels and bioproducts; a perspective on process-compatible products and their impact on cost-reduction, Energy and Environmental Science, 10, 1716–1738.
[3] K. Gao, K.R. McKinley, 1994, Use of macroalgae for marine biomass production and CO2 remediation: a review, Journal of Applied Phycology, 6, 45–60.
[4] J.S. Rowbotham, P.W. Dyer, H.C. Greenwell, M.K. Theodorou, 2012, Thermochemical processing of macroalgae: a late bloomer in the development of third-generation biofuels?, Biofuels, 3, 441–461.
[5] H.B. Goyal, D. Seal, R.C. Saxena, 2008, Bio-fuels from thermochemical conversion of renewable resources: A review, Renewable and Sustainable Energy Reviews, 12, 504–517.
[6] J.H. Choi, S.S. Kim, H.V. Ly, J. Kim, H.C. Woo, 2017, Thermogravimetric characteristics and pyrolysis kinetics of high-density-aquacultured Saccharina Japonica: Effects of water-washing, Fuel, 193, 159.
[7] A. Sharma, V. Pareek, D. Zhang, 2015, Biomass pyrolysis—A review of modelling, process parameters and catalytic studies, Renewable and Sustainable Energy Reviews, 50, 1081–1096.
[8] J.H. Choi, 2015, 2015, Pyrolysis and Catalytic Upgrading of Microalgal Biomass for Liquid Biofuel Production, Pukyong National University.
[9] I. Aspen Technology, 2000, Aspen Plus ® User Guide.
[10] S. Jones, Y. Zhu, D. Anderson, R.T. Hallen, D.C. Elliott, 2014, Process Design and Economics for the Conversion of Algal Biomass to Hydrocarbons: Whole Algae Hydrothermal Liquefaction and Upgrading.
[11] M.M. Wright, J. a. Satrio, R.C. Brown, D.E. Daugaard, D.D. Hsu, 2010, Techno-economic analysis of biomass fast pyrolysis to transportation fuels, National Renewable Energy Laboratory, 89, S2–S10.
[12] M. Shemfe, S. Gu, B. Fidalgo, 2017, Techno-economic analysis of biofuel production via bio-oil zeolite upgrading: An evaluation of two catalyst regeneration systems, Biomass and Bioenergy, 98, 182–193.
[13] J.F. Peters, S.W. Banks, A. V. Bridgwater, J. Dufour, 2017, A kinetic reaction model for biomass pyrolysis processes in Aspen Plus, Applied Energy, 188, 595–603.
[14] D. Bonalumi, 2016, Preliminary Study of Pyrolysis and Gasification of Biomass and Thermosetting Resins for Energy Production, Energy Procedia, 101, 432–439.
[15] P.L. Cruz, E. Montero, J. Dufour, 2017, Modelling of co-processing of HDO-oil with VGO in a FCC unit, Fuel, 196, 362–370.
[16] J.L. Carrasco, S. Gunukula, A.A. Boateng, C.A. Mullen, W.J. DeSisto, M.C. Wheeler, 2017, Pyrolysis of forest residues: An approach to techno-economics for bio-fuel production, Fuel, 193, 477–484.
[17] J.H. Choi, S.S. Kim, H.V. Ly, J. Kim, H.C. Woo, 2017, Effects of water-washing Saccharina japonica on fast pyrolysis in a bubbling fluidized-bed reactor, Biomass and Bioenergy, 98, 112–123.
[18] H.V. Ly, S.S. Kim, J.H. Choi, H.C. Woo, J. Kim, 2016, Fast pyrolysis of Saccharina japonica alga in a fixed-bed reactor for bio-oil production, Energy Conversion and Management, 122, 526–534.
[19] H.V. Ly, S.S. Kim, H.C. Woo, J.H. Choi, D.J. Suh, J. Kim, 2015, Fast pyrolysis of macroalga Saccharina japonica in a bubbling fluidized-bed reactor for bio-oil production, Energy, 93, 1436–1446.
[20] S.A. Channiwala, P.P. Parikh, 2002, A unified correlation for estimating HHV of solid, liquid and gaseous fuels, Fuel, 81, 1051–1063.
[21] A. Demirbas, D. Gullu, A. Çaglar, F. Akdeniz, 1997, Estimation of Calorific Values of Fuels from Lignocellulosics, Energy Sources, 19, 765–770.
[22] P. Žuvela, J.J. Liu, 2016, On feature selection for supervised learning problems involving high-dimensional analytical information, RSC Adv., 6, 82801–82809.
[23] J.W. Tukey, 1949, Comparing Individual Means in the Analysis of Variance, Biometrics, 5, 99.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50148-8

Applicability domains of a minimal-calibration model for effective online monitoring of pure components' concentrations in the pharmaceutical continuous manufacturing processes

Shojiro Shibayama[a], Kimito Funatsu[a,*]

[a]*The University of Tokyo, School of Engineering, Bunkyo-ku, Tokyo, 113-0068, Japan*
funatsu@chemsys.t.u-tokyo.ac.jp

Abstract

Continuous manufacturing is an emerging topic in the pharmaceutical industry. To realize continuous manufacturing, quality assurance during processes is essential. To assure the product quality, calibration models are constructed, usually using data that are collected in advance. However, the cost for collection of calibration data is not so small, because model construction is conducted frequently to sustain the model's accuracy. Iterative optimization technology (IOT) is a key technology involving infrared spectroscopy and predicting the concentrations of the pure components in the mixtures. IOT can be a powerful and useful tool, because the concentrations of the pure components in the mixtures, especially active pharmaceutical ingredient, are monitored through the pharmaceutical processes. However, IOT aims, in its original paper, to be applied for the process development step, not for the production step and its accuracy is not reliable enough for monitoring of the manufacturing process. In this work, we propose a method to evaluate quantitatively prediction accuracy for IOT prediction before acquiring the observed mole fractions. Assurance of the prediction accuracy is important as well as construction of accurate models in the actual operation. We estimate the prediction errors of the concentrations using the mixture infrared spectra simulated using Monte Carlo method.

Keywords: Process analytical technology; Applicability domains; Minimal-calibration method; Infrared spectroscopy

1. Introduction

The pharmaceutical industry conducts much research on the continuous processes and online monitoring. The main issue of online monitoring is to assure the product quality at every step in the manufacturing processes. The sensors and quantitative methods for online monitoring are called process analytical technology (PAT). Softsensor is a virtual sensor that enables to monitor variables hard to be observed, such as particle size or the compositions. When predicting active pharmaceutical ingredient (API) content, calibration models that input infrared spectra are employed. However, acquisition of calibration samples usually increases costs of the ingredients and labor for measurements. For instance, API content is measured by high-performance liquid chromatography, which takes hours.

To overcome this problem, iterative optimization technology (IOT), which is based on Beer's law, has been proposed (Muteki *et al.*, 2013). IOT has been shown to predict

concentrations of the pure components in the mixtures from only the pure components' spectra and the mixture spectra with no calibration sample. A wavelength selection method for IOT (WLSEA) (Shibayama, Kaneko and Funatsu, 2016) realizes to predict concentrations stably by IOT. WLSEA requires at least one set of the mixture spectrum and the known composition for the mixture. Therefore, WLSEA-IOT, is a minimal-calibration method in this paper.

Meanwhile, reliability of prediction results can be assessed for predictive models by defining applicability domains (AD) of the models. If the calibration models are constructed, AD is defined as distance from the training samples (Kaneko and Funatsu, 2014). When applying IOT, there is few calibration samples, while the physical model, Beer's law, exists. Therefore, we employed a method like Monte Carlo method, incorporated into Beer's law, for defining AD (Shibayama, Kaneko and Funatsu, 2016). However, it has not been investigated deeply.

The objective of this research is to verify the performance of AD for IOT. We verify whether AD, computed by Monte Carlo method, functions well or not. We apply AD to the binary mixtures that consist of water and acetic acid.

2. Method

2.1. Iterative optimization technology

Iterative optimization technology (IOT) (Muteki *et al.*, 2013) is a calibration-free method based on Beer's law that expresses the linearity between concentrations and absorbances of the pure components in the mixture. Beer's law is incorporated into optimization algorithm in the IOT framework. IOT is carried out by solving the following minimization problem:

$$\min_{r_1,r_2,\ldots,r_c} \sum_{j=1}^{v} \left| x_{\mathrm{mix},j} - \sum_{i=1}^{c} r_i x_{\mathrm{pure},i,j} \right|^2, \quad \text{such that} \quad 0 \le r_i \le 1, \sum_{i=1}^{c} r_i = 1, \tag{1}$$

where i denotes the index of the pure component, c denotes the number of the pure components, j denotes the index of wavelength, v denotes the number of wavelengths, r_i denotes the mole fraction of i, $x_{\mathrm{mix},j}$ denotes the observed absorbance of the mixture at wavelength j, and $x_{\mathrm{pure},i,j}$ denotes the absorbance of i at j.

2.2. Wavelength selection based on excess absorption

Wavelength selection based on excess absorption (WLSEA) is introduced by the authors (Shibayama, Kaneko and Funatsu, 2016). WLSEA eliminates wavelengths where large noises or large spectral changes because of molecular interactions are observed. IOT provides good performance in prediction of mole fractions if the linear relationship between the mole fractions and the spectra of the pure components. To improve prediction accuracy of IOT, WLSEA eliminates wavelengths where the linear relationship does not held.

The prediction method incorporating WLSEA into IOT is called WLSEA-IOT, which is a minimal-calibration method because WLSEA uses at least a set of the mole fractions of the pure components and the mixture spectrum. WLSEA-IOT shows good performance for simple mixtures.

2.3. Estimation of prediction errors of mole fractions

Let the absorbance of the simulated mixture spectrum at wavelength j be $x_{\mathrm{mix,sim},j}$ and the vector of the simulated mixture spectrum be $\boldsymbol{x}_{\mathrm{mix,sim}}=(x_{\mathrm{mix,sim},1},\ x_{\mathrm{mix,sim},2},\ \ldots,x_{\mathrm{mix,sim},v})$. Those are provided through Monte Carlo simulation as follows:

$$\boldsymbol{x}_{\mathrm{mix,sim}} = \sum_{i=1}^{c} r_i \boldsymbol{x}_{\mathrm{pure},i} + k\boldsymbol{e}, \quad \text{such that} \quad e_j \sim N\left(0, s_j^2\right),\ k = 1,2,3,\ldots, \tag{2}$$

where e_j denotes the measurement noise at wavelength j, and s_j denotes the standard deviation, k denotes a factor to scale the measurement noise across wavelengths. k assumes values from 1 to 15 in this paper. The measurement noise e_j follows a normal distribution with mean zero and standard deviation s_j. s_j^2 in Eq.(2) is calculated from the training samples whose mole fractions are known.

Substitute r_i in Eq. (2) for those of training samples. After simulating the observed mixture spectra, WLSEA-IOT is applied for those mixture spectra. The predicted mole fractions have the possible prediction errors of the pure components under the possible measurement noise in $\boldsymbol{x}_{\mathrm{mix,sim}}$.

2.4. Indicator for the assessment of the prediction errors

We employ the prediction errors for the mole fractions and the mixture spectra that are both observed and simulated. The prediction errors of the mole fractions are called mean absolute error (MAE). The indicator to estimate the prediction errors is called root mean square error (RMSE). RMSE is employed for the errors of the mixture spectra because it is the objective function in Eq.(1), which is related to the prediction errors of the mole fractions.

Both indicators are defined as follows:

$$MAE = \frac{1}{c}\sum_{i=1}^{c}\left|r_{obs,i} - r_{pred,i}\right|, \quad RMSE = \sqrt{\frac{1}{v}\sum_{j=1}^{v}\left|x_{mix,j} - x_{mix,pred,j}\right|^2}, \tag{3}$$

where $r_{\mathrm{pred},i}$ and $x_{\mathrm{mix,pred},j}$ denote respectively the predicted mole fractions of the pure component i and the predicted mixture spectrum at the wavelength j, based on Beer's law.

Table 1. Hyper-parameters for SG filter

Polynomial order	Differentiation order	Window size (half)
1	0	20

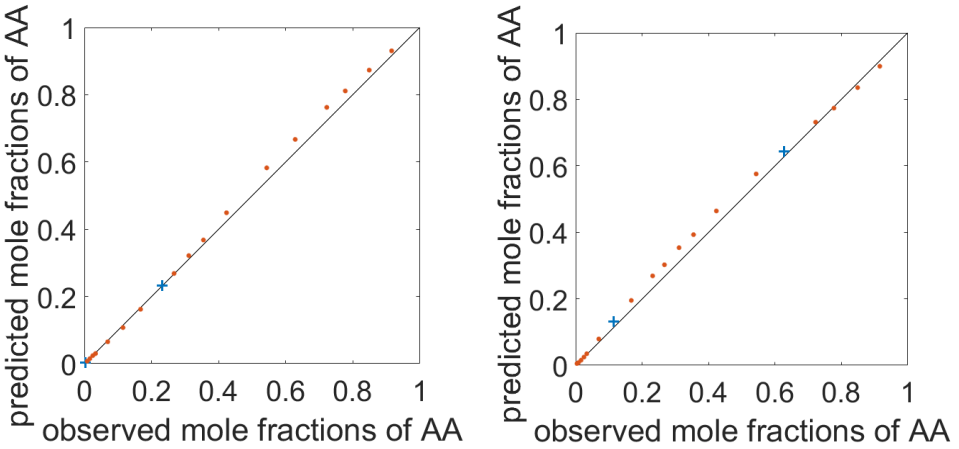
3. Case study

We used the binary liquid mixtures that consist of water (H_2O) and acetic acid (AA). The number of mixtures is nineteen. Our ultimate goal is to predict concentrations of API in the actual process, assuring the model's accuracy. In this study, we started with the simple mixtures to verify that the concept of AD works properly. The spectra of the mixtures showed the strong nonlinear relationship between the absorbances and the mole fractions caused by spectral changes because of water's molecular interactions.

We applied Savitzky-Golay filter (Savitzky and Golay, 1964) (SG filter) with the parameters shown in Table 1 to reduce the effect of noise and baseline shift. The hyper-parameters for SG filter was determined as the prediction errors of the mole fractions of AA was minimized using all samples for WLSEA-IOT training.

Prediction using WLSEA-IOT was carried out with the two sets of training samples whose mole fractions of AA were the 1st and 9th, and the 7th and 15th smallest. The first set of calibration samples would never be selected in the real analysis, because the concentrations of the pure components in the mixtures were not equally divided; however, we employed this set to attempt to verify the performance of the prediction-error estimation method. In the actual case, the calibration samples to be taken were determined by design of experiments.

The prediction results of mole fractions are shown in Figure 1. The blue crosses and orange dots indicate the prediction results for training and test samples. In the first result, in Figure 1(a), the prediction results for the samples between the two calibration samples show good accordance with the observed values. Figure 1(b) shows that the prediction results for the samples whose mole fractions of AA were both from zero to the 7th smallest value, and from the 15th smallest value to one were good. Those figures imply that WLSEA-IOT prediction provided good results for the mixtures in which one component's mole fraction was near one. This implication can be interpreted as follows:



(a) Using the training samples whose mole fractions of AA are the 1st and 9th smallest.

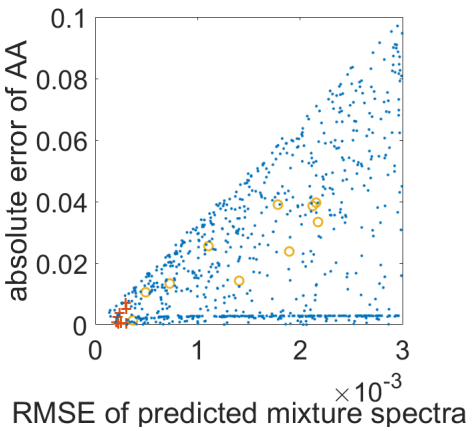
(b) Using the training samples whose mole fractions of AA are the 7th and 15th smallest.

Figure 1. Prediction results of mole fractions for H_2O-AA using WLSEA-IOT. Blue crosses and orange dots denote the calculation results for training samples and the test samples for test samples.

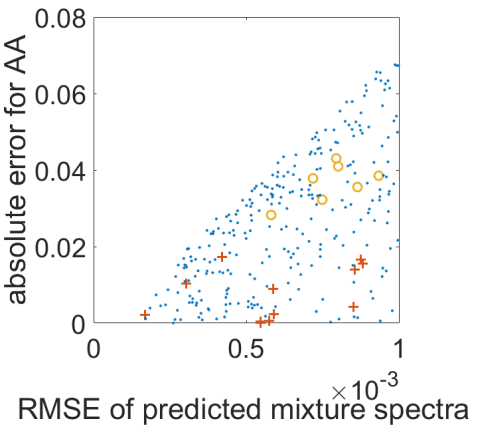
the mixtures in which the mole fraction of a pure component was nearly one accord with Beer's law, because the spectral changes caused by molecular interactions were weak in those mixtures. Thus, we call the samples whose mole fractions varied between those of the training samples and the pure components interpolated samples in the following part.

Next, the relationship between MAE for the mole fractions of AA and RMSE for the mixture spectra is shown in Figure 2. In the figures, the blue dots, orange crosses, and yellow circles denote the results for simulated samples, those for the interpolated samples, and those for the not interpolated samples. The orange crosses and the yellow circles correspond to the results for the observed mixtures; therefore, it can be said that those markers are desirably put in the areas that the blue dots indicate. The scatters of the blue dots in the two figures form the triangles in the different range, because the simulated samples were based on the training samples' mole fraction. In addition, the interpolated samples are predicted well, while not interpolated (or extrapolated) samples were not. The results still follows the assumption that the larger RMSE of the mixture spectra, the smaller the prediction errors are.

As Figure 2 indicates, it can be said that AD worked well as we intended. The prediction errors for the observed mixtures cannot be measured in the actual situation; however, the simulated samples on Figure 2 illustrates how IOT provides the prediction errors for the real mixtures without observed mole fractions. We know the possibly maximum errors of the mole fractions in the IOT prediction from AD. Therefore, we can ensure whether IOT prediction is properly carried out by referring to AD.



(a) When using the mixtures whose the mole fractions of AA are 1st and 9th smallest.



(b) When using the mixtures whose the mole fractions of AA are 7th and 15th smallest.

Figure 2. Estimation of prediction errors of mole fractions from prediction errors of mixture spectra. Markers denote results for a simulated (blue dot), interpolated observed samples (orange cross), and extrapolated observed samples (yellow circle) respectively.

4. Conclusion

We proposed a method to estimate possible prediction errors of mole fractions using a calibration-free method, IOT. The performance of AD for IOT was verified with the binary mixtures. We concluded that AD can estimate the prediction errors of the mole fractions in comparison with the prediction errors for Monte Carlo simulation samples.

However, AD that was applied in this paper did not consider the differences between the absorbances of the pure components at different wavelengths. This would cause the variances in the prediction errors of the mixture spectra. Therefore, we are going to propose a scaling strategy for IOT and corresponding AD in the future.

Acknowledgment

The authors acknowledge the support of the Core Research for Evolutionary Science and Technology (CREST) project "Development of a knowledge-generating platform driven by big data in drug discovery through production processes", under grant number JPMJCR1311, of the Japan Science and Technology Agency. The authors wish to thank Dr. K. Muteki at Pfizer Inc. and Dr. T. Goto at Kwansei University for providing the actual mixture data.

References

Kaneko, H. and Funatsu, K. (2014) 'Applicability Domain Based on Ensemble Learning in Classification and Regression Analyses.', Journal of chemical information and modeling, 54, pp. 2469–2482.

Muteki, K., Blackwood, D. O., Maranzano, B., Zhou, Y., Liu, Y. A., Leeman, K. R. and Reid, G. L. (2013) 'Mixture component prediction using iterative optimization technology (Calibration-Free/Minimum Approach)', Industrial and Engineering Chemistry Research, 52(35), pp. 12258–12268.

Savitzky, A. and Golay, M. J. E. (1964) 'Smoothing and Differentiation of Data by Simplified Least Squares Procedures', Analytical Chemistry, 36(8), pp. 1627–1639.

Shibayama, S., Kaneko, H. and Funatsu, K. (2016) 'Iterative optimization technology combined with wavelength selection based on excess absorption for a process analytical technology calibration-minimum approach', Chemometrics and Intelligent Laboratory Systems. Elsevier B.V., 156, pp. 137–147.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50149-X

Constraints for Set-Membership Parameter Estimation

Falco Jung[a], Adel Mhamdi[a] and Alexander Mitsos[a*]

[a]*RWTH Aachen University, AVT.SVT, Aachen, Germany*
amitsos@alum.mit.edu

Abstract

This contribution analyzes different formulations for the constraints in set-membership parameter estimation. For model discrimination and experimental design, parameter estimation and identifiability play a key role. In particular, it is desirable to determine the feasible parameter set, i.e., parameters for which the model outputs remain within defined bounds of available measurements, which are constraints to an optimization problem. The bounds may be defined by an absolute or relative error of single measurement points. As the errors are often unknown a-priori, we present a methodology to estimate them from available experimental data. Alternatively, the bounds can be defined by a statistical analysis of the model deviation from all available measurements. Both methods are applied in a case study of a microgel polymerization reaction involving spectroscopy measurement data. The approximations of the feasible region are compared to confidence ellipsoid predicted with an established method for identifiability based on linearization of the model.

Keywords: parameter, parameter estimation, identifiability, set inversion

1. Introduction

Mathematical process models are used in many tasks of engineering, ranging from model-based experimental design to model-based control. The estimation of unknown model parameters is usually an important step in model development and its results impact the validity and predictivity of the model. Hence, a thorough analysis of the estimation results is important. One state of the art approach for nonlinear models is to determine an ellipsoidal approximation of the parameter confidence region based on linearization (Franceschini and Macchietto, 2008). This method is based on the assumption of a normally distributed measurement error. Should this assumption be violated, the result may become inaccurate (Gottu Mukkula and Paulen, 2017).

Set-membership parameter estimation (Milanese and Vicino, 1991) accounts for the problem of deficient knowledge about the measurement distribution by assuming that the measurement error is random and bounded (Gottu Mukkula and Paulen, 2017). The method aims to determine the set of model parameters for which the respective model outputs remain within defined bounds of the experimental data. In simulation studies, it has been applied in parameter estimation and identifiability (Paulen et al., 2016), and model-based experimental design settings (Gottu Mukkula and Paulen, 2017). In this contribution, we aim to provide a method for the determination of the bounding measurement error based on available experimental data. Additionally, the method is applied to a case study involving real experimental data to show its applicability in practice.

The process under investigation is an isothermal homo-polymerization of Vinylcaprolactam based microgel in a laboratory-scale reactor of 300 ml (Janssen et al., 2017). One measured quantity is the monomer weight fraction in the batch reactor. It is measured via online Raman spectroscopy (Meyer-Kirschner et al., 2016).

In the following, we briefly outline the methodology of set-membership parameter estimation and suggest approximations of the bounding error based on the available experimental data. Additionally, we show that set-membership parameter estimation can be applied with the assumption of normally distributed errors by modification of respective constraints. Then, we consider the case study of the dynamic model in polymerization. Finally, we apply the methodology to estimate the unknown model parameters and compare the respective results to an ellipsoidal approximation of the feasible parameter set based on linearization of the model.

2. Set-Membership Parameter Estimation

The parameters $\boldsymbol{p} \in \mathbb{R}^{n_\mathrm{p}}$ of a given model are mapped to the model outputs $\boldsymbol{y}(t;\boldsymbol{p}) \in \mathbb{R}^{n_\mathrm{y}}$ by $\boldsymbol{y}(t;\boldsymbol{p}) = \boldsymbol{M}(t,\boldsymbol{p})$ with $\boldsymbol{M}\colon \mathbb{R} \times \mathbb{R}^{n_\mathrm{p}} \to \mathbb{R}^{n_\mathrm{y}}$. In the case study presented in Section 4, the model is a system of ordinary differential equations and one algebraic output equation. In set-membership parameter estimation, we aim at determining the feasible parameter set P_e for which model predictions are within defined bounds of the available measurements (Gottu Mukkula and Paulen, 2017).

$$P_\mathrm{e} \coloneqq \{ \quad \boldsymbol{p} \in P\colon \tag{1a}$$

$$B\big(y_\mathrm{i}(t_\mathrm{j};\boldsymbol{p}), \tilde{y}_\mathrm{i}(t_\mathrm{j}), \varepsilon_{\mathrm{i,j}}\big) \leq 0, \tag{1b}$$

$$\boldsymbol{y}(t;\boldsymbol{p}) = \boldsymbol{M}(t,\boldsymbol{p}) \ \}, \tag{1c}$$

where P is an a-priori estimate of the parameter space, $\tilde{y}_\mathrm{i}$ are unperturbed measurements $\mathrm{i} = 1, \ldots, n_\mathrm{y}$ at times $t_\mathrm{j} \in [t_0, t_\mathrm{end}]$ with $\mathrm{j} = 1, \ldots, N$ and B is a function depending on the assumption about the distribution of the measurement error. The errors $\varepsilon_{\mathrm{i,j}}$ are a measure for the acceptable deviation between model outputs and measurements. Herein, only for illustrative purposes, we discretize the a-priori estimate of the parameter space with an equidistant grid. At each grid point, we determine the respective model outputs and check whether their deviation from unperturbed measurements is within the defined bounds. If the deviation is within the bounds, we classify the point as feasible and collect it. The collection of points approximates the feasible region. This brute-force approximation, used herein for illustration purposes, is obviously only applicable in low-dimensional problems.

3. Constraints for Feasible Parameter Set

The bounding deviation of model outputs and experimental data depends on the assumptions made about the measurement error. Here, we consider bounded and normally distributed measurement errors.

3.1. Bounded Measurement Error

For the assumption of bounded measurement errors, a parameter vector $\boldsymbol{p}$ is considered infeasible, if the absolute deviation between any model output $y_\mathrm{i}(t_\mathrm{j};\boldsymbol{p})$ from the respective unperturbed measurement $\tilde{y}_\mathrm{i}(t_\mathrm{j})$ at any time t_j is greater than defined bounds $\varepsilon_{\mathrm{i,j}}$. Hence, the respective constraint formulation is

$$\left|y_{\mathrm{i}}(t_{\mathrm{j}};\boldsymbol{p})-\tilde{y}_{\mathrm{i}}(t_{\mathrm{j}})\right|-\varepsilon_{\mathrm{i,j}}\leq 0, \qquad \forall\,\mathrm{i},\forall\,\mathrm{j}, \tag{2}$$

Hence, the total number of constraints is equal to the number of data points. If a-priori knowledge is available, the bounding error can be chosen based on it. This may be information about the precision of a certain sensor provided by the manufacturers or a defined performance requirement (Hast et al., 2015). However, when dealing with new sensors and unknown systems, the bounding values may be unknown. Here, we suggest to estimate its value based on the variance of the available experimental data. If the measurement error is assumed to be uniformly distributed with zero-mean and an upper and lower bounding error of equal magnitude and invariant with time, the variance can be expressed as

$$\sigma_{\mathrm{i}}=\int_{-\varepsilon_{\mathrm{i}}}^{+\varepsilon_{\mathrm{i}}}\frac{1}{2\varepsilon_{\mathrm{i}}}\zeta^2 d\zeta)=\frac{1}{3}\varepsilon_{\mathrm{i}}^2. \tag{3}$$

By rearrangement we find that for the stated assumptions the bounding error can be estimated based on the variance of the measurement data as $\varepsilon_{\mathrm{i}}=\sqrt{3}\sigma_{\mathrm{i}}$, which we insert in Eq.(2) to define the constraints of the feasible parameter set. We assume that the variance of the available data sample is close to the variance of the population.

3.2. Normally Distributed Measurement Error

If the measurement errors are described by a normal distribution with zero mean and variance σ_{i}^2, a parameter vector $\boldsymbol{p}$ is considered infeasible, if the weighted sum of deviations between model outputs $y_{\mathrm{i}}(t_{\mathrm{j}};\boldsymbol{p})$ and unperturbed measurements $\tilde{y}_{\mathrm{i}}(t_{\mathrm{j}})$ at all measurement time points N is statistically significant (Franceschini and Macchietto, 2008). To analyze the statistical significance, we determine the sum of the deviations of all data points, rather than inspecting the residuals separately. Hence, the respective constraints are

$$\sum_{\mathrm{j}=1}^{N}\left[\left(y_{\mathrm{i}}(t_{\mathrm{j}};\boldsymbol{p})-\tilde{y}_{\mathrm{i}}(t_{\mathrm{j}})\right)^{T}\left(y_{\mathrm{i}}(t_{\mathrm{j}};\boldsymbol{p})-\tilde{y}_{\mathrm{i}}(t_{\mathrm{j}})\right)\right]-\hat{\varepsilon}_{\mathrm{i}}^2\leq 0, \qquad \forall\,\mathrm{i},\forall\,\mathrm{j}, \tag{4}$$

where ε_{i} is the statistically acceptable deviation. It can be determined from the Fisher distribution F_{dist} as (Franceschini and Macchietto, 2008)

$$\hat{\varepsilon}_{\mathrm{i}}^2=\sigma_{\mathrm{i}}^2 n_{\mathrm{p}}F_{\mathrm{dist}}(n_{\mathrm{p}},N-n_{\mathrm{p}},\alpha), \tag{5}$$

where α is the desired confidence level, usually chosen to be 95 % or 99 %. The number of constraints reduces to the number of sensors. A linearization of the respective model in Eq.(4) yields an ellipsoidal approximation of the feasible parameter space (Franceschini and Macchietto, 2008).

4. Case Study: Microgel Polymerization

The radical precipitation polymerization of N-Vinylcaprolactam in solution with water has received considerable attention recently as the product is attractive for a wide range of application including novel biomedical products (Plamper and Richtering, 2017). A number of authors have investigated radical polymerization and respective modeling approaches are well established (Kiparissides, 1996). Here, we make the assumption of a chain initiation reaction at quasi-steady state and a negligible chain transfer reaction.

Additionally, we combine termination by disproportionation and combination to one effective term. The polymerization is carried out in a batch reactor under isothermal conditions at 70 °C. The major reactants are the initiator, the monomer N-Vinylcaprolactam and the active radical polymer chains. Assuming a constant reaction volume, the molar concentration balances of the respective reactants are

$$\dot{c}_{\mathrm{I}}(t;\,\boldsymbol{p}) = -k_{\mathrm{d}} c_{\mathrm{I}}(t;\,\boldsymbol{p}), \qquad c_I(0) = 1.2\ \mathrm{mol\ m^{-3}} \tag{6a}$$

$$\dot{c}_{\mathrm{M}}(t;\,\boldsymbol{p}) = -k_{\mathrm{p}} c_{\mathrm{R}}(t;\,\boldsymbol{p}) c_{\mathrm{M}}(t;\,\boldsymbol{p}), \qquad c_M(0) = 103\ \mathrm{mol\ m^{-3}} \tag{6b}$$

$$\dot{c}_{\mathrm{R}}(t;\,\boldsymbol{p}) = +2 f k_{\mathrm{d}} c_{\mathrm{I}}(t;\,\boldsymbol{p}) - k_{\mathrm{t}} c_{\mathrm{R}}^2(t;\,\boldsymbol{p}), \qquad c_R(0) = 0\ \mathrm{mol\ m^{-3}} \tag{6c}$$

$$w_{\mathrm{M}}(t;\,\boldsymbol{p}) = c_{\mathrm{M}}(t;\,\boldsymbol{p}) M_{\mathrm{M}}/\rho, \tag{6d}$$

where c_{k} is the concentration of species k. The index I represents the initiator, M the monomer, and R the active radical chains. w_{M} is the weight fraction of monomer. The known parameters are the decomposition rate of initiator k_{d}, $1.2 \times 10^{-4}\ \mathrm{s^{-1}}$, the molecular weight of the monomer M_{M}, $1.39 \times 10^{2}\ \mathrm{g\ mol^{-1}}$, and the density of the solution ρ, $0.98 \times 10^{3}\ \mathrm{kg\ m^{-3}}$. The propagation rate constant k_{p}, $1\ \mathrm{m^3 mol^{-1} s^{-1}}$, was determined from quantum chemical calculations (Kröger et al., 2017). The initiator efficiency $f \in [0.1,1]$ and rate constant of termination $k_{\mathrm{t}} \in [0.01,100]\ \mathrm{m^3 mol^{-1} s^{-1}}$ are unknown. The available data for their estimation are monomer weight fraction measurements taken in three repetitions of the same experimental setup using Raman spectroscopy (Janssen et al., 2017). The parameter vector $\boldsymbol{p}_{\mathrm{e}}$ is determined by minimizing the residual of model outputs and experimental data using the enhanced scatter search metaheuristic implemented in the toolbox MEIGO (Egea et al., 2014) in Matlab®. During the estimation the objective function is evaluated approximately 5000 times, while the final parameter values are obtained after only approximately 250 function evaluations. The optimization is constrained by the upper and lower bounds of the parameter values as defined above. Hence, the determined parameter values are assumed to minimize the residual. However, global optimality cannot be guaranteed. It is assumed that model outputs are close to the unperturbed measurements, i.e. $w_{\mathrm{M}}(t_{\mathrm{j}};\,\boldsymbol{p}_{\mathrm{e}}) = \tilde{w}_{\mathrm{M}}(t_{\mathrm{j}})$.

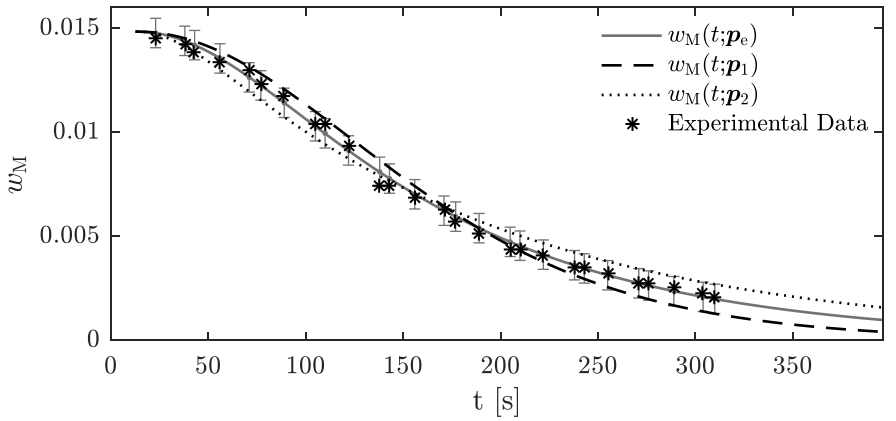


Figure 1: Agreement of model predictions and experimental data with error bounds.

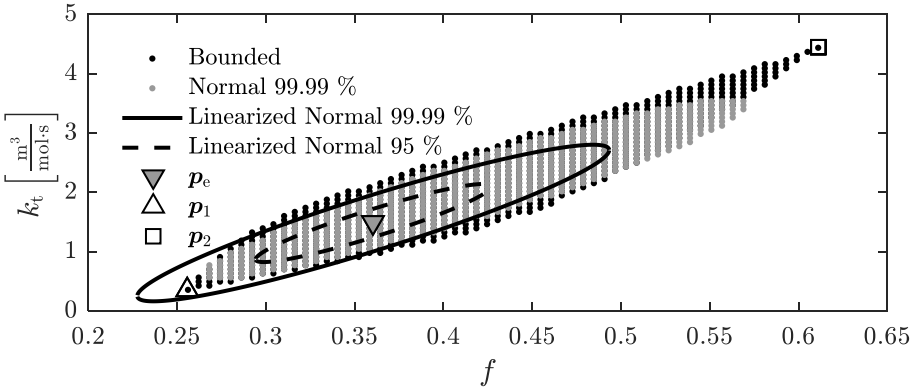


Figure 2: Prediction of the feasible parameter set with $\boldsymbol{p}_e$ based on parameter estimation $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ and chosen at the edges of the feasible space.

Figure 1 presents the agreement between the experimental data and model predictions for the weight fraction of monomer. The whiskers indicate the error bounds determined based on the methodology presented in Section 2. If the method for the estimation of the bounding error is well suited for this case, we expect (1) that the experimental data is within the determined error bounds and (2) that some measurements are close to the determined bounds. This ensures that error bounds are neither too tight nor too lax, which would cause a too optimistic or too pessimistic approximation of the feasible parameter set respectively. The figure shows that both criteria are fulfilled in the presented case study. All measurements are within the respective bounds, but some are close to them, see measurement at 100 and 140 seconds. Hence, this indicates that the methodology for the determination of the error bounds presented in Section 2 is well suited in this case. Figure 2 presents the approximation of the feasible parameter set for the assumption of a bounded and normally distributed error. Additionally, the figure presents ellipsoidal approximations determined via linearization of the model with the assumption of a normally distributed error (Franceschini and Macchietto, 2008). Respective confidence levels are shown in the legend. The values for the acceptable deviation between measurements and model outputs are $\varepsilon_{w_M} = 7 \times 10^{-4}$, $\hat{\varepsilon}^2_{w_M}(95\,\%) = 0.003$ and $\hat{\varepsilon}^2_{w_M}(99.99\,\%) = 0.01$. The nonlinearities in the model cause a shift between the linearized and non-linearized approximations of the feasible sets based on a normally distributed error. While the differences do not affect conclusions in this study critically, it might be significant in other applications. The approximations of the feasible sets for a bounded and normally distributed error cover a comparable range of parameter values for the unusually high confidence level of 99.99 %. We chose this unusually high value to illustrate at which confidence levels the two approximations become comparable. However, for confidence levels at the commonly selected 95 % the feasible set includes considerably fewer parameter values. To visualize the influence of changes in the parameter values on the predicted outputs, Figure 1 presents the model predictions for the parameter vectors marked in Figure 2.

5. Conclusion

We suggest a method to determine the bounding measurement error based on experimental data and present an application to a case study in polymerization. The results shows that the suggested methodology for the determination of the bounding

measurement error yields reasonable values, as discussed in Section 4. However, while a good agreement with experimental data can be achieved, the unknown parameter values can vary considerably without violating the constraints on the model outputs for the assumption of a bounded measurement error. For the assumption of a normally distributed measurement error similar results are achieved for a confidence level of 99.99 %. However, with the commonly chosen confidence level of 95 % the parameters can be determined within a smaller range. Drawing a conclusion on parameter identifiability from the size of the feasible parameter set is difficult, as the range of feasible parameter values, as expected, depends on the chosen confidence level and the assumption about the distribution of the measurement error.

6. Acknowledgements

We thank the Deutsche Forschungsgemeinschaft (DFG) in the Collaborative Research Center SFB 985 Functional Microgels and Microgel Systems for financial support. We thank Franca Janssen, Preet Joy and Johannes Faust for valuable discussions and Julian Meyer-Kirschner, Jörn Viell, Michael Kather and Andrij Pich for experimental data.

References

Egea, J. A., Henriques, D., Cokelaer, T., Villaverde, A. F., MacNamara, A., Danciu, D.-P., Banga, J. R., Saez-Rodriguez, J., 2014. MEIGO: An open-source software suite based on metaheuristics for global optimization in systems biology and bioinformatics. BMC Bioinformatics 15, 136.

Franceschini, G., Macchietto, S., 2008. Model-based design of experiments for parameter precision: State of the art. Chemical Engineering Science 63 (19), 4846–4872.

Gottu Mukkula, A. R., Paulen, R., 2017. Model-based design of optimal experiments for nonlinear systems in the context of guaranteed parameter estimation. Computers & Chemical Engineering 99, 198–213.

Hast, D., Findeisen, R., Streif, S., 2015. Detection and isolation of parametric faults in hydraulic pumps using a set-based approach and quantitative–qualitative fault specifications. Control Engineering Practice 40, 61–70.

Janssen, F. A. L., Kather, M., Kröger, L. C., Mhamdi, A., Leonhard, K., Pich, A., Mitsos, A., 2017. Synthesis of poly(n-vinylcaprolactam)-based microgels by precipitation polymerization: Process modeling and experimental validation. Industrial & Engineering Chemistry Research 56 (49), 14545-14556.

Kiparissides, C., 1996. Polymerization reactor modeling: A review of recent developments and future directions. Chemical Engineering Science 51 (10), 1637–1659.

Kröger, L. C., Kopp, W. A., Leonhard, K., 2017. Prediction of chain propagation rate constants of polymerization reactions in aqueous NIPAM/BIS and VCL/BIS systems. The Journal of Physical Chemistry. B 121 (13), 2887–2895.

Meyer-Kirschner, J., Kather, M., Pich, A., Engel, D., Marquardt, W., Viell, J., Mitsos, A., 2016. In-line monitoring of monomer and polymer content during microgel synthesis using precipitation polymerization via raman spectroscopy and indirect hard modeling. Applied Spectroscopy 70 (3), 416–426.

Milanese, M., Vicino, A., 1991. Optimal estimation theory for dynamic systems with set membership uncertainty. Automatica 27 (6), 997–1009.

Paulen, R., Villanueva, M. E., Chachuat, B., 2016. Guaranteed parameter estimation of non-linear dynamic systems using high-order bounding techniques with domain and cpu-time reduction strategies. IMA Journal of Mathematical Control and Information 33 (3), 563–587.

Plamper, F. A., Richtering, W., 2017. Functional microgels and microgel systems. Accounts of Chemical Research 50 (2), 131–140.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50150-6

Identification of Reaction Systems using Spectroscopic Measurements and Micro-reactors

Manokaran Veeramani[a], Sridharakumar Narasimhan[a], Nirav Bhatt[a*]

[a]*Systems & Control Group; Department of Chemical Engineering; IIT Madras, Chennai, India*
niravbhatt@iitm.ac.in

Abstract

Micro-reactor technology has found increasing usage in the synthesis of several organic compounds and pharmaceutical chemicals in the last decade. The main advantages of micro-reactors are good mixing characteristics, less hold-up volume of reagents and better control as compared to traditional batch reactors. Micro-reactors have recently been used in the laboratory to study the underlying reactions and develop a kinetic models. Further, it is possible to monitor reactions using in-situ spectroscopy in micro-reactors, and collect spectral data at different residence time. In this work, we propose a calibration-free method to analyse spectral data obtained from micro-reactors to identify reaction kinetics. Then, the proposed method can be applied to spectral data directly to identify the parameters of the proposed reaction kinetic model. The method is illustrated using a numerical case study of base catalyzed Knoevenagel condensation reaction system.

Keywords: Reaction kinetics, Model identification, Micro-reactors, Kinetic data, Spectral data.

1. Introduction

Traditional methods for studying the kinetics of reaction systems involves generation of kinetic data, i.e., measurement of concentration of species as a function of time in a batch reactor. The reactions carried out in batch reactor are limited by heat and mass transfer and consumes large amount of reactants. Recently, micro-reactors have been used for kinetic studies as it possess considerable advantages over batch reactors such as better mixing, reduced chemical consumption, better handling of toxic reagents etc [Moore et al (2016)]. Spectroscopy as a non-destructive method of analysis is used extensively in the study of reaction systems. Inline analytical instruments with micro-reactors enable real time monitoring and optimization of reactions.

Mozharov et al (2011) proposed a method for studying the reaction kinetics using flow manipulation technique where a system initially flowing at a steady flow rate is changed to higher flow rate by a step input. Inline Raman spectrometry was used for studying Knoevenagel condensation reaction with steady state measurements and dynamic measurements using flow manipulations. Moore et al (2014) discussed the use of online IR analysis in micro-reactors with flow manipulations for faster generation of kinetic data for the Paal-Knorr reaction. Moore et al (2016) studied the kinetics of aminocarbonylation of aryl halides using inline infrared spectroscopy. In the literature, typically, a calibration model is developed to predict concentrations from the spectral measurements. Then, the predicted concentrations are used to estimate the parameters of kinetic model. Since we have to build a new calibration model for each reaction system, more reagents are used in this approach. Further, building calibration is a time-

consuming exercise. Hence, there is a need for a calibration-free method for identifying reaction systems. In this work, the main objective is to develop a calibration-free methods for identifying reaction systems from spectral data on similar line of the one developed in Armhein (1998).

The paper is organized as follows. Section 2 discusses the model equations for micro-reactors. Then, we propose a calibration-free method to estimate parameters of kinetic models from spectral data collected from micro-reactors. In Section 3, Knoevenagel condensation reaction system is used to demonstrate the proposed method. The results are discussed in Section 4. Section 5 concludes the paper and discusses future extensions. (Note: All vector and matrices are denoted in boldfaced lower and upper case, respectively.)

2. Methodology

In this section, we will discuss the modelling of micro-reactors, and calibration-free method for identification of reaction systems.

2.1 Model equation for Micro-reactors

We consider a homogeneous reaction system with S species and R independent reactions (see Amrhein, 1998 for independent reactions definition). Micro-reactors with low dispersion can be modelled as plug flow reactor (PFR) as mentioned in Nagy et al (2012). For the channel dimension and experimental conditions listed in Section 3 of the paper, the conditions for plug flow behaviour are met, and hence flow behaviour is modelled as plug flow. Further, it is assumed that the reactor is at steady state. Then, the material balance equations for S species in micro-reactors are written as:

$$\frac{d\boldsymbol{f}}{dV_r} = \boldsymbol{N}^{\boldsymbol{T}}\boldsymbol{r}(\boldsymbol{c},\boldsymbol{\theta}),\ \ \boldsymbol{f}(\boldsymbol{0}) = \boldsymbol{f}_{\boldsymbol{0}} \tag{1}$$

where $\boldsymbol{f}$ is an S-dimensional vector of molar flow rates of the species, $\boldsymbol{N}$ is the $R\times S$ stoichiometric matrix, $\boldsymbol{r}$ is the R-dimensional independent reaction rates and V_r is the control volume. $\boldsymbol{f}$ can be expressed in terms of concentrations as follows: $\boldsymbol{f} = q\boldsymbol{c}$, where q is volumetric flow rate. Under the assumption of constant fluid velocity u, and cross section area, Eq. (1) is written in terms of residence time τ and concentrations as follows:

$$\frac{d\boldsymbol{c}}{d\tau} = \boldsymbol{N}^{\boldsymbol{T}}\boldsymbol{r}(\boldsymbol{c},\boldsymbol{\theta}),\ \ \boldsymbol{c}(\boldsymbol{0}) = \boldsymbol{c}_{\boldsymbol{0}} \tag{2}$$

where $\boldsymbol{c}$ is the S-dimensional concentration vector, $\boldsymbol{f_0}$ and $\boldsymbol{c_0}$ are the initial conditions.

2.2 Model for Spectral data

Concentrations of species in micro-reactors are measured using spectroscopy in two manners (i) Spectral data are measured at the end of the micro-reactor at different residence times (τ) at steady-state condition, and (ii) Measurement of spectral data at different locations of micro-reactors after steady-state is established. In both these methods, the spectral data are obtained as a function of residence time. Since the spectral data indirectly measure concentrations, it can be seen that concentrations are measured as different residence times. The spectral measurement model relating

absorbance and concentration according to Beer Lambert's law for a unit path length is written as

$$\boldsymbol{a}^T(i) = \boldsymbol{c}^T(i)\boldsymbol{E} \tag{3}$$

where $\boldsymbol{a}$ is the M-dimensional vector of absorbance's for ith residence time for M wavelengths, $\boldsymbol{E}$ is the $S\times M$-dimensional matrix of pure component spectra for S species, and $\boldsymbol{c}$ is the S dimensional vector corresponding to the ith residence time. The absorbance are measured for L different residence times, Then, Eq. (3) can be written as

$$\boldsymbol{A} = \boldsymbol{CE} \tag{4}$$

where $\boldsymbol{A}$ is the $L\times M$ spectral data matrix, and $\boldsymbol{C}$ is the $L\times S$ concentration matrix at different residence times. For predicting the concentration of the species, a calibration model is required under the assumption that number of absorbing species is equal to S. For reactive systems, the rank of spectral data matrix $\boldsymbol{A}$ can be less than S. This problem is labelled as rank deficiency problem. Hence, the standard calibration methods cannot be applied to the data. The rank deficiency problem of spectral data matrix has been studied by Armhein (1998) and two methods to rectify this issue were proposed. Next, we will formulate a parameter estimation problem to estimate parameters from the spectral data $\boldsymbol{A}$ without using calibration model. Note that $\boldsymbol{C}$ is function of reaction rates and kinetic parameters.

2.3 Calibration-free Approach: Parameter estimation from Spectral data

It is assumed that the minimum number of absorbing species is equal to R. Further, it is assumed that the stoichiometric matrix and the rate expressions are also available. Then, the objective is to estimate the unknown parameters $\boldsymbol{\theta}$ of kinetic model. A constrained objective function J_A can be formulated to minimize the measured spectral data at different residence times with the one obtained by the simulating as follows.

$$\begin{aligned} &\min_{\theta} J_A, \quad J_A = \|\boldsymbol{A} - \boldsymbol{A}(\tau,\boldsymbol{\theta})\| = \|[I_L - \boldsymbol{C}(\tau,\boldsymbol{\theta})\boldsymbol{C}^{+}(\tau,\boldsymbol{\theta})]\boldsymbol{A}\|_F^2 \\ &\text{s.t.} \quad \frac{d\boldsymbol{C}(\boldsymbol{\theta})}{d\tau} = \boldsymbol{N}^T\boldsymbol{r}(\boldsymbol{C},\boldsymbol{\theta}) \\ &\boldsymbol{\theta} \in [\boldsymbol{\theta}^U, \boldsymbol{\theta}^L], \boldsymbol{\theta} \geq 0 \end{aligned} \tag{5}$$

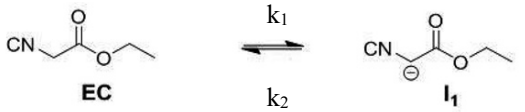
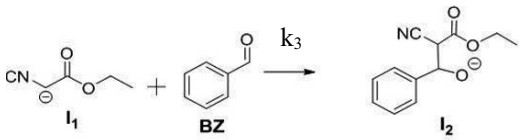
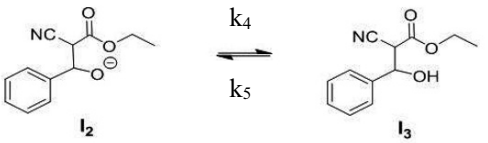
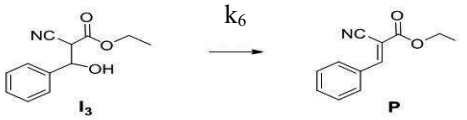
where $\boldsymbol{A(\theta)}$, and $\boldsymbol{C(\theta)}$ are the $L\times M$ and $L\times S$-dimensional matrices computed from the dynamic models with the parameter $\boldsymbol{\theta}$. In Eq. (5), the objective function is the square of the Frobenius norm of the errors between experimental and simulated spectral matrices. Since the pure component spectra are not available, the second expression is minimized to estimate parameters $\boldsymbol{\theta}$. The second expression can be interpreted as projection errors of $\boldsymbol{A}$ on the column space of $\mathbf{C}(\boldsymbol{\theta})$. These projection errors are minimized to find the best estimates of the unknown parameters $\boldsymbol{\theta}$ of the kinetic model. The non-negativity constraints on kinetic parameters with $\boldsymbol{\theta}^U$ and $\boldsymbol{\theta}^L$ as upper and lower limits of parameters are imposed to find the best estimates.

3. Case study: Knoevenagel Condensation Reactions

In this section, the proposed method in Section 2 is illustrated via a simulation study of base catalyzed Knoevenagel condensation reaction in a micro-reactor [Mozharov et al (2011)]. The reactions take place in solution phase where the solution 1 is composed of Benzaldehyde (BZ) and catalyst 1, 4 Diazobicyclo [2,2,2] octane (DABCO) dissolved in methanol and the solution 2 is composed of Ethyl Cyanoacetate (EC) dissolved in

methanol. The reactants are assumed to be pumped into the micro-reactor using an appropriate pump. An Inline Raman spectrometer probe is assumed to be placed at the end of the micro-reactor for measuring spectral data for recording the spectral data continuously with acquisition and delay time of 2 min. The experiment is simulated for different flow rates of reactants (0.2 − 20 μL/min). The simulated steady-state measurements of spectral data are taken after a time period of 1.5 times the residence time for each flow rate. The number of independent reactions in the system is 4, i.e., $R = 4$) and number of species involved in the reaction are equal to 6, i.e., $S = 6$. The reaction scheme and the proposed kinetic model are given in Table 1.

Table 1: Reaction Mechanism and kinetic models

S.No.	Reaction	Kinetic Model
R1		$r_1 = k_1 c_{EC} - k_2 c_{I_1}$
R2		$r_2 = k_3 c_{I_1} c_{BZ}$
R3		$r_3 = k_4 c_{I_2} - k_5 c_{I_3}$
R4		$r_4 = k_6 c_{I_3}$

In the proposed kinetic model, species I_1, I_2 and I_3 are the intermediates of reactions and the number of absorbing species is equal to 4. Under the given experimental conditions catalyst concentration remains constant [Mozharov et al (2011)]. So the catalyst concentration is lumped with kinetic parameters in the rate expressions. The reaction rate vector $\boldsymbol{r}$ is given as: $\boldsymbol{r} = [r_1\ r_2\ r_3\ r_4]^T$ and the stoichiometric matrix **N** is

$$\boldsymbol{N} = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix} \tag{6}$$

The intensity (I_υ) of Raman spectra in terms of concentration of species is given by [J.J. Laserna (1996)]:

$$I_\upsilon = cI_0 V K_\upsilon \tag{7}$$

where I_υ is the Raman intensity of band υ, I_0 is the intensity of the exciting radiation, V is the volume of sample illuminated by the source and viewed by the spectrometer, c is the sample concentration, and K_υ is a constant characteristic for each band. The parameters V, I_0 and K_υ are clubbed together to give the parameter ϕ_υ similar to E in Eq. (3) and Eq. (7) is rewritten as

$$\boldsymbol{I_\upsilon = c\phi_\upsilon} \tag{8}$$

Eq. 8 is the similar to the Eq. 3 with $\boldsymbol{I_\upsilon = a}$ and $\boldsymbol{\phi_\upsilon = E}$. Hence, the method proposed in Section 2 can be applied to the Raman spectra data. The spectral data at different residence times with L = 20 and M = 30 for a selected choice of residence times are shown in Fig. 1. For simulation purpose, the Intensity of Raman spectra $\boldsymbol{I_\upsilon}$ matrix is corrupted with additive zero mean Gaussian noise and the standard deviation σ of each observation is taken as 10 % of maximum intensity from each observation.

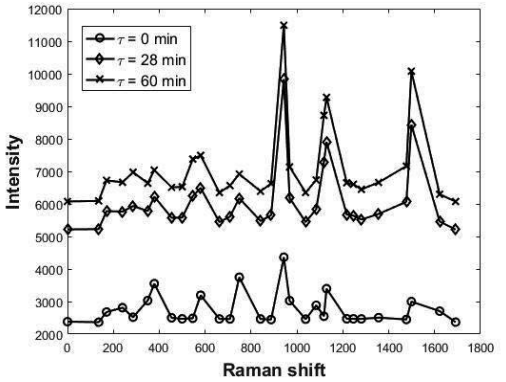


Figure 1: Raman Spectra for reaction mixtures as a function of τ

4. Results and Discussions

The unknown parameters in the model $\boldsymbol{\theta} = [k_1, k_2, k_3, k_4, k_5, k_6]^T$ are estimated from noisy data as mentioned in Section 3 and reported in Table 2. Table 2 also provides the true values and the initial guess values used in the simulation along with the estimated parameter values. The 95% confidence Intervals of parameters are estimated using Bootstrap method and shown in Table 2.

Table 2: True, initial and estimated values of six parameters along with 95 % confidence intervals for the simulation studies.

Parameter	True value	Initial Value	Estimated value	Confidence Interval
k_1 (min^{-1})	0.090	0.050	0.088	[0.081, 0.095]
k_2 (min^{-1})	0.389	0.100	0.385	[0.311, 0.458]
k_3 ($dm^3\ mol^{-1}\ min^{-1}$)	0.308	0.100	0.302	[0.248, 0.356]
k_4 (min^{-1})	15.35	10	15.35	[15.275, 15.426]
k_5 (min^{-1})	512.0	500	512.0	[511.98, 512.02]
k_6 (min^{-1})	294.9	200	294.9	[294.88, 294.91]

The concentration profiles of the species are simulated using the estimated parameters shown in Fig. 2.

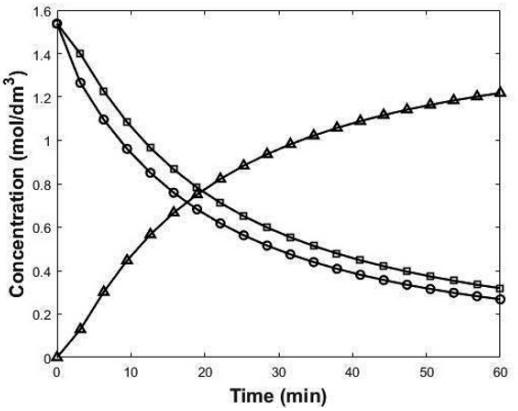


Figure 2: Concentration profiles of the species using the true and estimated values (o – C_{EC}, □ – C_{BZ}, Δ - C_P, solid lines represents the concentrations obtained by the estimated values.)

5. Conclusions

In this paper, it has been shown that spectral data can be collected using two manners in micro-reactors for kinetic studies. A calibration-free approach to identify the kinetic parameters from spectral data obtained in micro-reactors has been proposed. The proposed approach does not require time-consuming calibration and hence, reduces usage of reagents for kinetic modelling using micro-reactors. The proposed approach has been illustrated using a simulated Knoevenagel condensation reaction in micro-reactor. The results show that the proposed approach provides good estimates of kinetic parameters with tight confidence intervals. In future, we will extend the approach to estimate the diffusivities and kinetic parameters from spectral data by relaxing plug flow behaviour assumption.

References

M. Amrhein, 1998, Reaction and Flow Variants/Invariants for the Analysis of Chemical Reaction Data, PhD Thesis, EPFL.

J. J. Laserna,1996, Modern Techniques in Raman Spectroscopy, John Wiley Sons :New York.

J. S. Moore, K. F. Jensen,"Batch kinetics in flow: Online IR analysis and continuous control", 2014, Angew. Chem, 126, 480-483.

J. S. Moore, C. D. Smith, K F. Jensen, 2016, Kinetic analysis and automated online screening of aminocarbonylation of aryl halides in flow, React. Chem. & Eng., 1, 272-279.

S. Mozharov, A. Nordon, D. Littlejohn, C. Wiles, P. Watts, P. Dallin, J M. Girkin, 2011, Improved Method for Kinetic Studies in Microreactors Using Flow Manipulation and Noninvasive Raman Spectrometry, J. Am. Chem. Soc., 133, 3601-3608.

K.D.Nagy, Bo Shen, T.F Jamison, K.F Jensen, 2012, Mixing and Dispersion in Small- Scale Flow Systems, Org. Process Res. Dev, 16, 976-981.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50151-8

A Flexible Framework and Model Library for Process Simulation, Optimization and Control

Andrew Lee[a*], Jaffer H. Ghouse[b], Qi Chen[c], John C. Eslick[b], John D. Siirola[d], Ignacio E. Grossman[c], David C. Miller[b]

[a] *KeyLogic Systems, National Energy Technology Laboratory, Pittsburgh, PA, USA*
[b] *National Energy Technology Laboratory, Pittsburgh, PA, USA*
[c] *Department of Chemical Eng., Carnegie Mellon University, Pittsburgh, PA, USA*
[d] *Sandia National Laboratories, Albuquerque, NM, USA*
andrew.lee@netl.doe.gov

Abstract

A new framework for optimizing process flowsheets has been developed, which enables a greater degree of flexibility and automation to facilitate the development of hierarchical models suitable for rigorous optimization and control problems under both steady-state and dynamic operating conditions. A library of models for common process unit operations has been developed along with the framework to allow rapid development of flowsheets, with a focus on adaptability and extensibility to allow users to extend and apply these models to novel processes and configurations. This paper presents the development of the framework and model library which support conceptual design, process design/integration, and dynamic optimization/control.

Keywords: Process modeling, simulation, optimization, control, Pyomo.

1. Introduction

Advances in computational power and numerical optimization routines have enabled the possibility of applying rigorous simulation and optimization techniques to large scale problems such as those associated with the design, optimization and control of integrated chemical processes and energy systems. To facilitate the development of process models, a number of simulation packages, such as Aspen Plus®, gPROMS, ProSim, PRO/II®, have been developed which provide robust and easy to use tools, including libraries of common process unit operations, thermo-physical property models, and efficient algorithms to solve large, sparse systems of nonlinear differential algebraic equations. These tools allow engineers to solve very large process flowsheets under both steady-state and dynamic conditions. However, most process simulation packages focus primarily on solving well defined simulation problems, and often have only limited capabilities for advanced optimization, such as conceptual design or optimization under uncertainty. In contrast, there exist several platforms and modeling languages designed specifically for solving large scale optimization problems; however, these platforms lack the specialized infrastructure and model libraries necessary for easily simulating chemical processes and energy systems. In these cases, models must be laboriously assembled for specific systems, often requiring specialized initialization procedures. Developing and applying these models to chemical processes also requires expert knowledge in modeling and optimization not available to the typical process engineer. Thus, there exists a gap in the capabilities of existing software for optimizing process flowsheets.

The U.S. Department of Energy's Institute for the Design of Advanced Energy Systems (IDAES) is addressing this need by developing a next-generation process systems engineering framework (Miller at al., 2018) that is built for optimization from the ground up, enabling the use of modern optimization solvers with a framework for advanced process modeling. The IDAES framework utilizes the Pyomo (Hart et al., 2017) algebraic modeling language (AML), an open source framework for formulating large scale optimization problems based on the Python programming language, which can interface with a wide range of optimization solvers. As part of the IDAES framework, a library of models has been developed, along with a standard modular framework for linking the models in a flowsheet and solving the resulting problem. The goal of these tools and libraries is to reduce the amount of effort required to develop a process model and facilitate rapid development and screening of process configurations.

2. Framework Development

Despite the wide variety of chemical engineering process unit models, they share a number of common features, allowing for a standardized core modeling structure. The process for developing a unit model can be summarized as follows:

1. Specify the operating mode (dynamic or steady state).
2. Define calculations for the thermophysical, transport and reaction kinetic properties of the material(s) of interest.
3. Specify material and energy inlets and outlets, along with the information (state variables) transferred in each (e.g., total flowrate, pressure, temperature, and mole fractions).
4. Define material, energy, and pressure balances.
5. Define additional constraints describing the behavior of the unit (e.g., pressure-flow relations, or heat transfer correlations).

Of these elements, the material, energy and pressure balances are central to the formulation of the model, with the other elements serving mainly to inform aspects of the balance equations. This would suggest that it is possible to write the balance equations in a generic form, and to substitute expressions or variables for the generic terms as needed. Terms which are not needed in a specific model can be excluded by substituting a value of 0 into the balance equations, and the exact form of the expressions can be adapted to the problem at hand to improve tractability. Doing this allows for a completely generic formulation of the balance equations that can be reused between unit models, whilst allowing for formulations which can adapt to the differing needs of each problem. Further, the generic form of the balance equations is inherently unitless, which means that by changing the units of measurement used in the individual terms, the units of the entire model can be changed.

2.1. Material, Energy and Pressure Balances

Mass, energy and pressure balances may contain many terms depending on the operation type including: generation of species via chemical reactions, pressure differentials and heat, work and mass transfer. A framework for modeling these systems needs to accommodate all of the possible terms. At a more fundamental level, there are many ways that material and energy balances can be expressed – whether by mass or moles, by element or component, by component flows or total flow and composition, etc. The preferred formulation of the balance equations is situation dependent, and often one form is more convenient or provides for a more tractable problem formulation.

Despite the variations that may occur, the general form of the balance equations is constant, consisting of terms that may or may not be present in each specific case. These can be summarized as follows:

1. inlet and outlet terms,
2. accumulation terms, which are needed only in models which consider dynamic behavior,
3. generation of species due to kinetic and equilibrium based reactions (including phase equilibria), which are primarily functions of the material state, and,
4. transfer terms which are dependent on the behavior of the unit as well as the material state.

For unit models where consideration of spatial variations are not required (0-dimensional models), this results in the following general forms for the material, energy and pressure balance equations (Eqs. (1)–(3));

$$\frac{\partial M_{p,j}}{\partial t} = F_{p,j,in} - F_{p,j,out} + R_{p,j,kin} + R_{p,j,eq} + K_{p,j} \tag{1}$$

$$\frac{\partial E}{\partial t} = \sum_{p} H_{p,in} - \sum_{p} H_{p,in} + W + Q \tag{2}$$

$$0 = P_{in} - P_{out} + \Delta P \tag{3}$$

Here $\frac{\partial M_{p,j}}{\partial t}$ is the accumulation of species j in phase p within the unit, $F_{p,j,in}$ and $F_{p,j,out}$ are the flows of species j in phase p in and out, $R_{p,j,kin}$ and $R_{p,j,eq}$ are the generation of species j in phase p by kinetic and equilibrium controlled reactions respectively. Similarly $\frac{\partial E}{\partial t}$ is the accumulation of energy in the unit, $H_{p,in}$ and $H_{p,out}$ are the flows in and out of energy in phase p, and P_{in} and P_{out} are the pressure of the material at the inlet and outlet respectively. $K_{p,j}$, W, Q and ΔP represent the mass, work, heat and pressure transfer terms within the unit respectively.

The traditional approach to expressing these equations within an AML would be to either write custom equations for each model which contain only the necessary terms, or to write the balance equations using generic terms and to then provide additional constraints to describe each term. The first alternative results in a more compact problem with fewer variables and constraints, whilst the second alternative allows for greater flexibility and the reuse of code for the balance equations. By exploiting the object oriented nature of Python, it is possible to write the balance equations dynamically, using procedural code to automatically substitute the terms of the balance equations as they are constructed. This can be used to completely automate the construction of the balance equations, using a small set of instructions to inform the code which terms should be constructed for each unit model.

2.2. Property Calculations

Most available software packages and correlations use intensive forms for the property calculations, as these are dependent only on the state variables of the system, and not the flowrates of material. However, there are situations where using extensive forms yields a more tractable formulation for the overall problem (e.g. adjusting when and where bilinear terms appear in the problem structure). Thus, the framework needs to allow for both extensive and intensive formulations in order to provide the flexibility

necessary to achieve the most tractable problem structure. Further, in many cases the developer of a process model is not the developer of thermophysical property models, and may not know the details of the property model formulation. Thus, there needs to be a way for the property calculations to integrate with the unit models without the need for the modeler to be aware of the full details of the property models they are using.

In the IDAES modeling framework, the property module defines the extensive terms in the balance so property models are specific to the application whilst the balance equations are generic. Further, the property calculations become the only part of the model structure where units of measurement are inherently present, which means that the units of measurement for the entire model stem from the property calculations. However, this requires that the units be internally consistent and that property models contain information on the extensive flows of material, which has traditionally been kept separate from the otherwise intensive property calculations.

2.3. Inlets and Outlets

The final consideration when developing a model for a unit operation is what information needs to be passed between units within a flowsheet – i.e. state information for the inlet and outlet streams. The necessary state information and the specific form it takes is dependent on the specific problem at hand. Considering that any two unit models connected together in a flowsheet are by necessity using the same material, and in most cases the same set of property calculations, it follows that the best form for the definition of the inlet and outlet streams of each unit should be the same. Rather than defining a single standard to which inlets and outlets must conform, the IDAES framework allows each property module to define the best set of state variables to pass between units. In cases where the state variables required by two connected units do not match (e.g. a change in property modules between units), a translator block can be used. Connections between two units within a flowsheet are then implemented as equality constraints between the state variables in each connector.

2.4. Models with Spatial Variations

In cases where spatial variations cannot be neglected (e.g. plug flow reactors), it is necessary to include spatial domains and partial differential equations describing these variations within the balance equations. The framework discussed above can be readily extended to include these derivative terms, and Pyomo includes support for partial differential equations through the Pyomo DAE toolbox (Nicholson et al., 2017), which allows for automatic discretization of domains and numerical approximation of the derivatives for an arbitrary number of domains. All that is necessary to extend the framework is to replace the extensive material, energy and pressure terms in the balance equations with differential forms and to use the inlet (and/or outlet) flows as boundary conditions for the differential equations. An example for a 1-dimensional system is shown below, where x is the spatial domain:

$$\frac{\partial M_j}{\partial t} = -\frac{\partial F_j}{\partial x} + R_{j,kin} + R_{j,eq} + K_j \tag{4}$$

$$\frac{\partial E}{\partial t} = -\frac{\partial H}{\partial x} + W + Q \tag{5}$$

$$0 = -\frac{\partial P}{\partial x} + \Delta P \tag{6}$$

2.5. Model Library

Using this framework, a library of models for the most common process unit operations has been developed as part of the IDAES PSE Framework. Models have been developed for operations such as mixers and splitters, compressors, equilibrium separators, and ideal reactors, which support both steady-state and dynamic optimization as well as heterogeneous phases. All of these models have been developed to provide a basic representation for these types of operations that can be easily extended and adapted by users to suit their specific applications. The models in the library have been developed to be the simplest representations of these pieces of equipment possible, with the intent that users will build upon these and add additional correlations as needed to obtain the necessary level of detail and rigor for their particular application.

3. Framework Example

In order to demonstrate the application of the framework, consider the workflow for generating a steady-state model for an isentropic compressor unit. When creating the model, the user must specify a property package containing the necessary property calculations and a set of instructions on how to build the balance equations. For a compressor, it is clear that the pressure drop and work transfer terms will be required, whilst reactions, heat and mass transfer are generally neglected. Each unit model contains a default set of instructions for the terms most commonly used; however, the option is provided for the user to override these if required.

From there, the automatic framework first retrieves the time domain for the problem from the flowsheet. Based on the instructions provided, the framework automatically generates the mass, energy and pressure balances. Taking the mass balances as an example, the framework first eliminates all unneeded terms by substituting a value of zero (such as the accumulation and reaction terms). It then checks with the provided property package to determine the best form to use for the extensive flow terms in the equations. For this example, assuming that the property package uses total flowrate, component mole fractions, temperature and pressure as state variables, the expression $F_p y_{pj}$ would be returned, where F_p is the total flowrate of phase p and $y_{p,j}$ is the mole fraction of component j in phase p. These are then substituted into the mass balance equation, and an equation is automatically written for each component and phase, as shown in Eq. (7).

$$0 = F_{p,in} y_{p,j,in} - F_{p,out} y_{p,j,out} + 0 + 0 + 0 \tag{7}$$

Similarly, the pressure and energy balances are written as shown in Eqs. (8) and (9), where h_p is the specific enthalpy for phase p, as calculated by the property package.

$$0 = \sum_p \left(F_{p,in} h_{p,in}\right) - \sum_p \left(F_{p,out} h_{p,out}\right) + W + 0 \tag{8}$$

$$0 = P_{in} - P_{out} + \Delta P \tag{9}$$

The framework then checks the property package again to determine the state variables to include in the inlet and outlet from the unit (in this case F_p, $y_{p,j}$, P and temperature). The user may then provide the constraints necessary to calculate the remaining terms in the balance equation, notably the isentropic constraints relating ΔP and W. As the entire problem structure is open and accessible to the user, this is as straightforward as adding additional Pyomo constraints to the problem. Users may also easily deactivate existing constraints within the framework if they wish to replace these with more rigorous calculations.

4. Conclusions

A flexible and hierarchical framework for developing and optimizing process flowsheet has been developed which automates much of the development of the mass, energy and pressure balances. The framework has been designed from the ground up to be suitable for advanced optimization and control applications, under both dynamic and steady-state conditions. The framework was developed based on the Pyomo AML and leverages the capabilities of modern optimization solvers by using an equation oriented approach with access to complete first and second derivatives. A library of unit models has been developed as part of the IDAES PSE Framework which includes models for many of the common unit operations used in chemical processes. The models are adaptable and extensible to allow users to customize the models to suit the needs of their specific problem.

 KeyLogic Systems, Inc.'s contributions to this work were funded by the National Energy Technology Laboratory under the Mission Execution and Strategic Analysis contract (DE-FE0025912) for support services. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This material is based upon work supported by the U.S. Department of Energy, Office of Science, under Contract No. DEAC02-05CH11231.

References

W.E. Hart, C.D. Laird, J.-P. Watson, D.L. Woodruff, G.A. Hackebeil, B.L. Nicholson, and J.D. Siirola, 2017, Pyomo – Optimization Modeling in Python. Second Edition. Vol. 67. Springer.

W.E. Hart, J.-P. Watson, and D.L. Woodruff, 2011, Pyomo: modeling and solving mathematical programs in Python, Mathematical Programming Computation 3, 3, 219-260.

D.C. Miller, J.D. Sirrola, D. Agarwal, A.P. Burgard, A. Lee, J.C. Eslick, B. Nicholson, C. Laird, L.T. Biegler, D. Bhattacharyya, N.V. Sahinidis, I.E. Grossmann, C.E. Gounaris, and D. Gunter, 2018, Next Generation Multi-Scale Process Systems Engineering Framework, PSE 2018, San Diego.

B.L. Nicholson, J.D. Siirola, J.-P. Watson, V.M. Zavala, L.T. Biegler. 2017, pyomo.dae: A Modeling and Automatic Discretization Framework for Optimization with Differential and Algebraic Equations, Accepted to Math Programming Computation.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50152-X

Towards a Generic Algorithm for Identifying High-Quality Decompositions of Optimization Problems

Andrew Allman[a], Wentao Tang[a], Prodromos Daoutidis[a,*]

[a]*Department of Chemical Engineering and Materials Science, University of Minnesota, 421 Washington Ave. SE, Minneapolis, MN, USA 55455*
daout001@umn.edu

Abstract

Optimization is a ubiquitous tool in process systems engineering which is being used on models of ever-increasing size and complexity. Decomposition solution approaches can be powerful tools for solving difficult optimization problems but are dependent on finding a partition of variables and constraints amenable to the solution approach. In this paper, we propose an automated, generic method for decomposing optimization problems using community detection. This method generates subproblems which are tightly interacting amongst themselves but weakly interacting with other subproblems, thus minimizing the amount of coordination required in the solution approach. We demonstrate the ability of our method to find solutions faster than solving the original problem in many cases.

Keywords: optimization, decomposition, community detection

1. Introduction

One of the major goals of process systems engineering is the development of computational tools that aid decision making in the chemical process industry. Optimization is a tool that has become pervasive within the field since it provides guarantees that one is making the best decision for a given situation. Optimization is used for a wide range of tasks, from model predictive control to supply chain optimization. Models used for optimization of systems relevant to chemical engineering can be nonconvex, mixed-integer, multi scale, and uncertain. Optimization problems with these characteristics will innately scale poorly with size. Thus, a major and open challenge in the field is finding ways to solve optimization models that are large-scale and contain nonconvex model equations in a reasonable amount of time. There exist many different ways that one can improve computational tractability of a large optimization problem, but one of the most powerful tools is to apply a decomposition. Doing so requires two steps: first, one must decompose the optimization problem, identifying groups of variables and constraints within an optimization problem to act as subproblems which may share some variables and/or constraints. Once this is done, a decomposition solution approach can be applied to solve the smaller subproblems individually and then coordinate them in some way to find a solution to the original optimization problem. In the simplest case, one simply defines a master problem which updates values of shared primal or dual variables for each iteration of solving subproblems. Various advanced decomposition solution approaches exist which can be broken into two major classes. The first is the case where subproblems share complicating variables. In this case, Bender's decomposition is most commonly used. It was first introduced as an approach for partitioning mixed integer linear programs (Benders, 1962) and has since been generalized to a wider range of optimization problems (Geoffrion, 1972). Conversely, subproblems can be coupled by complicating constraints. These types of problems are commonly solved using a Dantzig-Wolfe, or Lagrangean decomposition. The Dantzig-Wolfe decomposition was introduced

first and solves linear programs using an alternating solution strategy to the subproblems (Dantzig and Wolfe, 1960). Lagrangean decomposition was introduced later and makes use of duality to solve a wider range of problems (Guignard and Kim, 1987). Many examples of other hybrid primal-dual and problem specific decompositions also exist, a few of which are discussed in textbooks (Conejo et al., 2006) and review articles (Grossmann, 2012). All of the aforementioned decomposition solution approaches require that one can find a decomposition of the original optimization problem. Typically this is done using knowledge of the physical problem, such as in our previous work which optimized the design of a plant which produced both renewable energy and biofuels (Allman and Daoutidis, 2017). Ideally, it would not be necessary to rely on intuition to generate decompositions, but to date there does not exist a generic method for generating decompositions. In this work, we propose such a generic method using the network theory method of community detection. The remainder of this paper details this new method and is structured as follows: in section 2, the concept of community detection is introduced in more detail. Then, section 3 describes the method of identifying communities within an optimization problem. Section 4 tests the method in multiple case studies using MINLP test problems. Finally, we conclude and propose avenues for future work in section 5.

2. Community Detection

The concept of community detection has emerged in network science as a method for finding groups within complex systems through represented on a graph. In contrast to more traditional decomposition methods which seek a strict block diagonal or block triangular structure, community detection methods find subnetworks with statistically significantly more links between nodes in the same group than nodes in different groups (Girvan and Newman, 2002). Central to community detection is the notion of modularity, a metric that captures this difference:

$$Q = \frac{1}{2m}\sum_{i,j}\left(A_{ij} - \frac{k_i k_j}{2m}\right)\delta(g_i, g_j) \tag{1}$$

Here, Q is the modularity, A_{ij} is the edge weight between nodes i and j, k_i is the total weight of all edges connecting node i with all other nodes, and m is the total weight of all edges in the graph. The Kronecker delta function $\delta(g_i, g_j)$ will evaluate to one if nodes i and j belong to the same group, and zero otherwise. Modularity is a property of how one decides to partition a network: networks that are not partitioned and those that place every node in its own community will both have modularity equal to zero. The goal of community detection, then, is to find communities that maximize modularity. Although modularity maximization is an NP-hard integer program, many efficient algorithms exist to solve it approximately, including spectral clustering (Newman, 2006) and fast unfolding (Blondel et al., 2008). Figure 1 shows a few networks with increasing maximum modularity; note how the community structure becomes increasingly apparent as this value increases. Previous efforts in our group have applied community detection to chemical plant networks by creating an equation graph of the corresponding dynamic model (Moharir et al., 2017). By doing so, communities of state variables, inputs, and outputs can be obtained which are tightly interacting amongst themselves but weakly interacting with other communities. As such, these communities can form the basis of distributed control architectures (Jogwar and Daoutidis,, 2017) which typically perform better than other distributed control architectures that one may obtain from “intuition” (Pourkargar et al., 2017). For a more comprehensive review of the use of community detection in distributed control, we refer the reader to Daoutidis et al., 2017. An alternative method for finding communities for distributed model predictive control is to apply a decomposition on the optimization problem as a whole (Tang et al., 2017).

However, community structure can exist in all optimization problems, thus it makes sense to extend this method to any generic optimization problem, and this work proposes to do so. The advantage of using community detection to find decompositions is that subproblems generated will have statistically minimal interactions, through complicating variables or constraints, and thus require minimal coordination through the decomposition solution method. The proposed method is generic, applicable to any optimization problem or decomposition solution approach, and scalable, using computationally efficient graph theory algorithms.

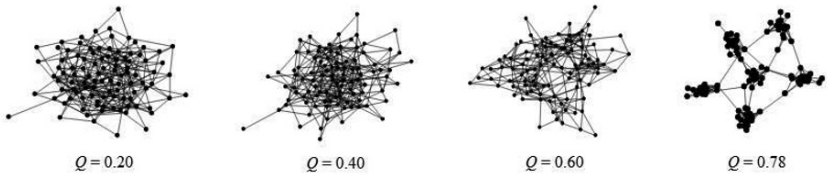


Figure 1. Networks of different maximum modularity.

3. Decomposition Method

To be able to apply community detection algorithms for decomposing an optimization problem, it is first necessary to represent the optimization problem as a graph. To this end, a "variable-constraint bipartite graph" is proposed. To construct such a graph, consider an optimization problem which, without loss of generality, can be written as follows:

$$\begin{aligned} &\text{minimize } \textstyle\sum_{i=1}^{N_v} f_i(x_i) \\ &\text{s.t.} \quad g_j(x) = 0, j = 1, \dots, N_e \\ &\qquad h_k(x) \leq 0, k = 1, \dots, N_i \end{aligned} \tag{2}$$

This formulation only requires that the objective function be a sum of functions of single variables, which can always be satisfied by introducing appropriate auxiliary variables. Then, two sets of nodes are included: a set of N_v variable nodes and a set of $N_e + N_i$ constraint nodes. Edges are then placed between a constraint node and a variable node if the variable appears in the constraint, or, equivalently, if $\frac{\partial g_j}{\partial x_i} \neq 0$ or $\frac{\partial h_k}{\partial x_i} \neq 0$.

Once an equation graph is obtained, community detection can be applied to obtain communities of equations and variables that have maximum modularity. These communities will represent subproblems to be optimized separately by the desired decomposition solution approach. Variables (or constraints) which connect to multiple communities are considered to be shared. This is best illustrated through an example. Consider the following optimization problem:

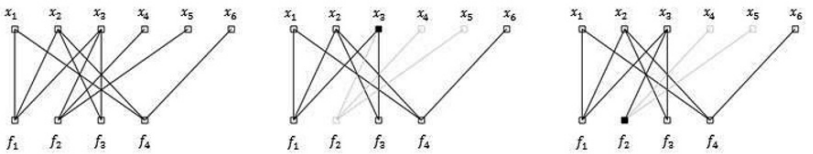


Figure 2.: Graph representation of Eq. (3) (left). Color-coded communities shown with shared variables (center) or constraints (right). Shared nodes are filled in black.

$$\begin{aligned} &\text{minimize } x_1x_2 + 7x_4 + x_5^{0.7} \\ &\text{s.t.} \quad f_1(x) = x_1 + 16x_2 - 9x_3 \leq 0 \\ &\qquad f_2(x) = x_3 \ln x_3 - x_4x_5 = 0 \\ &\qquad f_3(x) = x_2^2 + x_3^2 - 5.8 = 0 \end{aligned} \tag{3}$$

To convert this problem to the form specified by Eq. 2, an auxiliary variable x_6 is introduced and an additional equation $f_4(x) = x_6 - x_1x_2 = 0$ is added. A bipartite graph is constructed with 6 nodes for variables, 4 nodes for equations, and edges added where variables appear within a constraint. The variable constraint graph for this problem is shown in Figure 2. Also shown in Figure 2 are the results of applying community detection to the graph. These communities represent subproblems to be solved; for example, when variables are shared, one subproblem would optimize $x_1, x_2,$ and x_6, a separate subproblem would optimize x_4 and x_5, and a master problem would coordinate these two subproblems by optimizing over x_3.

4. Case Studies

To test the method, problems from the MINLP test library (MINLP World) are used. In particular, feasible problems that did not solve to an optimality gap of 10% or less within an hour using the BARON solver are tested. Some representative results are shown in the following subsections.

4.1. Improving the Upper Bound

"Waterx" is a problem with 68 variables and 52 constraints which, after 1 hour of solving in BARON, had a feasible solution of 984.03 with a lower bound of 240.2, for an optimality gap of 296%. The variable-constraint graph of this problem with communities found in 16.7 s using a fast unfolding algorithm is shown in Figure 3. In this figure, variable nodes are circles, constraint nodes are triangles, shared variables are black asterisks, and nodes in the same community have the same color. This network has a maximum modularity of 0.3937, which gives four separate subproblems. Assigning every constraint to a community gives 18 shared variables, of which 10 can be varied independently while the remaining 8 are coupled to these other variables through equality constraints. To test the performance of the decomposition found by community detection, each community is treated as a subproblem to be repeatedly solved in parallel using BARON. Subproblems are coordinated by a master problem which optimizes shared variables using a derivative-free downhill simplex method. The method is stopped when the objective values of each point on the simplex differ by no more than 10^{-5}.

The results of our method give an upper bound on the solution, and are compared to results from solving the original problem using BARON as shown in Figure 3. In this figure, it is easy to see that after some small initialization time likely due to a poor initial guess, our algorithm is able to find feasible points with objective values less than the best

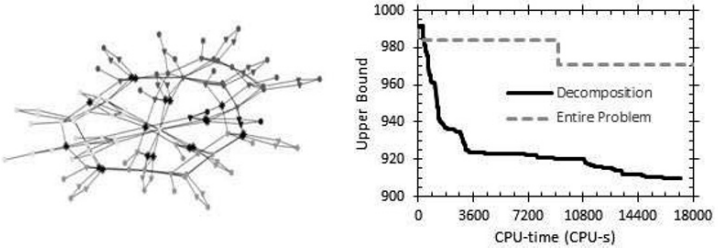


Figure 3. Graph representation of problem "waterx" (left). Performance of decomposition algorithm vs. solving entire problem (right).

value found solving the entire problem with BARON over the same amount of CPU-time. It is important to note that for this problem, the algorithm used does nothing to guarantee that we have found the global optimal, as our algorithm does not find a lower bound to the problem. However, because we were not able to prove global optimality to begin with for this NP-hard problem, we consider the performance of our algorithm satisfactory.

4.2. Improving the Lower Bound

"Feedtray" is a problem with 90 variables and 80 constraints which, after 1 hour of solving in BARON, had a feasible solution of -13.41 with a lower bound of -68.68, for an optimality gap of 81%. The variable-constraint graph of this problem with communities found in 141 s using a fast unfolding algorithm is shown in Figure 4. This network has a maximum modularity of 0.5302, which gives four separate subproblems. Assigning every constraint to a community gives 31 shared variables. However, of the 4 subproblems, only one contains variables which are present in the objective function (the "objective subproblem"); the remaining three subproblems are simply checks of feasibility. It is expected that performing a single feasibility check will be faster than performing three, so the three feasibility checks are combined into a single subproblem. The decompositions are tested using an algorithm which optimizes the objective subproblem using BARON, checks feasibility using the determined values of the shared variables, and generates new cutting plane constraints on the shared variables based on the feasibility check. This process is repeated until a feasible result is obtained.

The results of our method compared to solving the entire problem using BARON are also shown in Figure 4, where the CPU-time neglects time to find communities. Note that unlike the previous case study, intermediate results are not feasible but instead represent a lower bound on the solution. Thus, when the algorithm has converged to a feasible point, this point is a proven global optimal solution. Such a solution can be found in this problem due to its unique decomposition structure of one objective subproblem and one feasibility check subproblem. Figure 4 shows that our algorithm is able to find a better lower bound on the solution than BARON in the same amount of CPU-time and converges on the upper bound determined by BARON. Thus, in this case, the decomposition is used to prove an optimal solution quicker than considering the problem as a whole.

5. Conclusions

In this paper, we have presented a new, generic method for finding decompositions of optimization problems using community detection to find partitions of variables and constraints with high modularity. To implement this method, we developed a novel

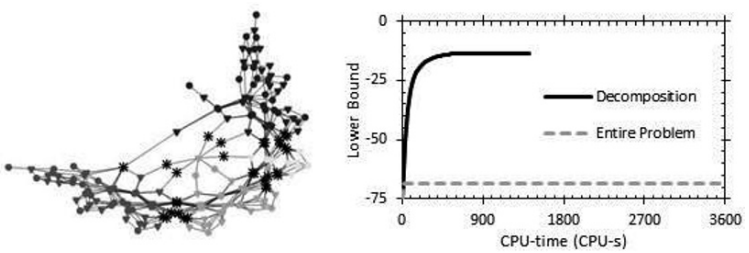


Figure 4. Graph representation of problem "feedtray" (left). Performance of decomposition algorithm vs. solving entire problem (right).

variable-constraint bipartite graph representation of an optimization problem. The algorithm was tested on problems from the MINLP test libraries, and examples were shown where this algorithm can be used to improve both upper and lower bounds of an optimization solution. In the future, we aim to improve our algorithm by considering additional factors which may be important in a decomposition, such as load balancing and constraint convexity.

6. Acknowledgements

Financial support from the National Science Foundation and the University of Minnesota Initiative for Renewable Energy and the Environment (IREE) is gratefully acknowledged.

References

Allman, A., Daoutidis, P., 2017, "Optimal design of synergistic distributed renewable fuel and power systems." *Renew. Energy*, 100, pp. 78-89.

Benders, J.F., 1962, "Partitioning procedures for solving mixed-variables programming problems." *Numer. Math.,* 4, pp. 238-252.

Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E., 2008, "Fast unfolding of communities in large networks." *J. Stat. Mech. Theor. Exp.*, P10008.

Conejo, A.J., Castillo, E., Minguez, R., Garcia-Bertrand, R., 2006, *Decomposition Techniques in Mathematical Programming: Engineering and Science Applications.* Springer Science and Business Media.

Dantzig, G.B., Wolfe, P., 1960, "Decompostion principle for linear programs." *Oper. Res.,* 8, 101.

Daoutidis, P., Tang, W., Jogwar, S.S., 2017 "Decomposing complex plants for distributed control: Perspectives from network theory." *Comput. Chem. Eng.* in press. DOI: https://doi.org/10.1016/j.compchemeng.2017.10.015

Geoffrion, A.M., 1972, "Generalized Benders decomposition." *J. Optimiz. Theory App.,* 10, 237.

Girvan, M., Newman, M.E.J., 2002, "Community structure in social and biological networks." *P. Natl. Acad. Sci. USA,* 99, pp. 7821-7826.

Grossman, I.E., 2012, "Advances in mathematical programming models for enterprise-wide optimization." *Comput. Chem. Eng.,* 47, pp. 2-18.

Guignard, M., Kim, S., 1987, "Lagrangean decomposition: a model yielding stronger lagrangean bounds." *Math. Program.,* 39, pp. 215-228.

Jogwar, S.S., Daoutidis, P., 2017, "Community-based synthesis of distributed control architectures for integrated process networks." *Chem. Eng. Sci.,* 172, pp. 434-443.

MINLP World. *MINLP Library.* <www.gamsworld.org/minlp/minlplib.htm>

Moharir, M., Kang, L., Daoutidis, P., Almansoori, A., 2017, "Graph representation and decomposition of ODE/hyperbolic PDE systems." *Comput. Chem. Eng.,* 106, pp. 532-543.

Newman, M.E.J., 2008, "Modularity and community structure in networks." *P. Natl. Acad. Sci. USA,* 103, pp. 8577-8582.

Pourkargar, D.B., Almansoori, A., Daoutidis, P., 2017, "Impact of decomposition on distributed model predictive control." *Ind. Eng. Chem. Res.,* 56, pp. 9606-9616.

Tang, W., Allman, A., Pourkargar, D.B., Daoutidis, P., 2017, "Optimal decomposition for distributed optimization in nonlinear model predictive control through community detection." *Comput. Chem. Eng.,* submitted.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50153-1

Application of strategy switching mechanism with improved strategy for heat exchanger network design

Takeru Kono[a], Chika Kawahara[a], Naoki Kimura[a*], Yoshifumi Tsuge[a]

[a]*Department of Chemical Engineering, Faculty of Engineering, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395 Japan*
nkimura@chem-eng.kyushu-u.ac.jp

Abstract

Heat exchanger network (HEN) is one of the major technique of energy saving. We have proposed a HEN design system using pinch technology and multiagent. In this system, the maximum heat recovery is derived from pinch technology. And then, an initial HEN can be obtained according to the AreaTargeting method in order to meet the maximum heat recovery. The initial HEN has complexity and a number of heat exchangers due to the AreaTargeting method. Therefore it is necessary to simplify the HEN without deterioration of the energy saving efficiency. In our system, simplification strategy is defined to prescribe the priority order of procedure and/or the part of HEN. The number and variety of resulted HENs increase along with the number of strategies. It is useful if a large number of HENs with wide variety is obtained because the preferable HEN must be different due to the process objectives or conditions. We implemented several HEN design agents (HENDAs) with different simplification strategies.

In this study, we introduced several improvements in simplification strategy of HENDAs using strategy switching mechanism. Several HENDAs have their own strategy, and then they exchange the HENs on the way—that is “strategy switching” from the standpoint of a HEN. We investigated the effect of improvements in the simplification strategies and strategy switching mechanism.

Keywords: Heat exchanger networks, pinch technology, multiagent framework, simplification, strategy switching.

1. Introduction

The utility consumption in a chemical process can be reduced by heat exchange between the process fluids, which leads to reduction of utility cost and energy consumption. However, it takes too much time and load to manually design the optimal Heat Exchanger Network (HEN). Hence, various methods have been proposed. In our laboratory, we have developed the HEN design support system using Multiagent System. We set a different "strategy" to each Heat Exchanger Network Design Agent (HENDA). And HENDAs simplify HENs according to their own strategy. It brought us various HENs. Using the new mechanism "Strategy Switching Mechanism" in this system, we succeeded in increase of the variation of HENs for a small-scale chemical processes which have two hot streams and two cold streams. In this study, we verified whether Strategy Switching Mechanism is applicable to the larger scale processes.

2. Method

We aim to build a system that can propose alternative plans that meet the users' requirement in various bases of evaluation. The outline of our framework is shown in Figure 1. First, ProjectManager receives PFD from user engineer. Next, the Targeting Agent obtains the maximum amount of the heat recovery by pinch analysis, and designs the initial HEN whose utility usage is minimized by AreaTargeting method. Each HENDA respectively designs a lot of HENs based on the initial HEN. EquipmentCostManager narrows down the designed plans based on the total cost. Furthermore, ProjectManager narrows down the plans based on the maintainability from the plans that EquipmentCostManager narrowed down. Finally, the several selected plans are proposed to user.

2.1. Design of initial heat exchanger network

In this section, we explain about a method to design energy-saving HEN. We decide to use pinch analysis to obtain the maximum amount of the heat recovery. We explain the method using a simple process (4SP1) shown in Table 1. The composite curves are made from the heat demands and the minimum approaching temperature difference of the target process shown in Figure 2. This figure illustrates that the shaded area is the heat recovery area by 5480 kW, the minimum usage of the cold utility is 120 kW and the minimum usage of the hot utility is 960 kW. The place is called a "pinch point" where the heat exchange with the minimum approaching temperature difference is carried out. Unless there is heat exchange crossing the pinch point, the utility usage is kept to the minimum. In order to achieve the HEN with the minimum utility usage, the AreaTargeting method is adopted to design the initial HEN shown in Figure 3. In this initial HEN, the utility usage and the total heat transfer area is the minimum. However, because the number of heat exchangers and splits are very large, the equipment cost is high and the maintainability is low. Therefore, it is necessary to simplify the structures of the initial HEN.

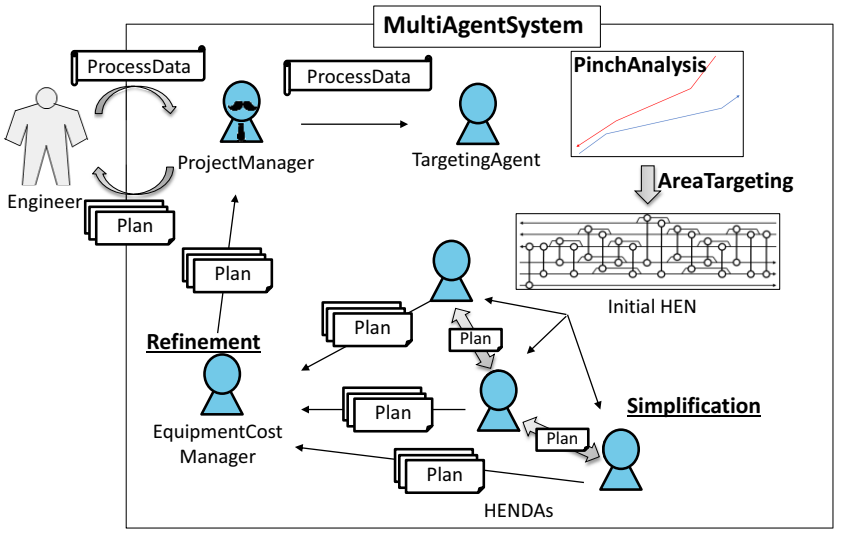


Figure 1: The outline of our framework.

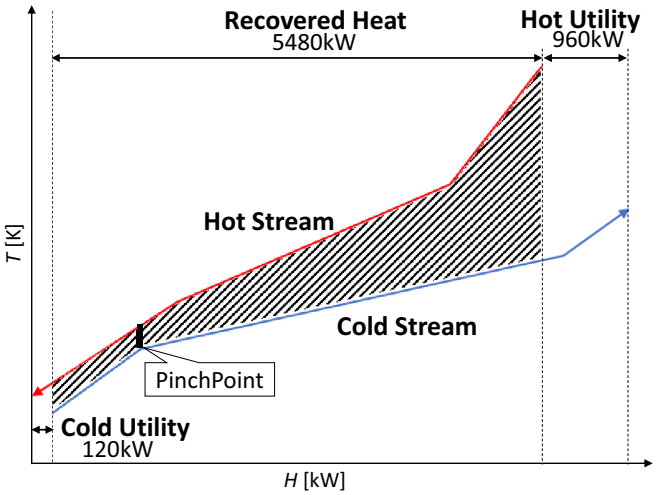


Figure 2: Composite curves of 4SP1.

Table 1: Heat demands of 4SP1.

Stream	T_{in}[K]	T_{out}[K]	CP[kW/K]	ΔH[kW]
Hot1	453	353	20	2000
Hot2	403	313	40	3600
Cold1	333	373	80	3200
Cold2	303	393	36	3240
Steam	378	378	-	
Water	293	298	-	

$\Delta Tmin$ is 10[K].

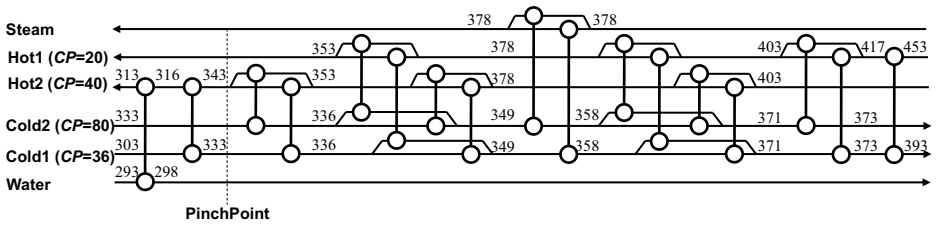


Figure 3: The initial HEN of 4SP1.

2.2. Simplification of HEN

In order to simplify the structures of the initial HEN, several simplification procedures are prepared as shown in Figure 4. We can obtain new alternative plans by every single procedure. In the initial HEN, a lot of partial structures indicated in Figure 4 are found. However the structure of HEN after the application of a procedure differs depends on the order of simplification procedure. Then, we prepared several simplification "strategy" which contains the preference orders of the procedures and the segments to be simplified. For example, we prepared strategies "OutLTDL", "InHTDWSL",

"InHSLTDL". These strategies are intended to design wide variety of HENs. In detail, the strategy "OutLTDL" (Out-Low-Temperature Delete-Loop) focuses on segments in ascending order of exit temperature and preferentially carries out Delete Loop. The details of other strategies are described in our previous study (Kou 2010). We assigned each strategy to each HENDA who is in charge of simplification. HENDAs simplify based on the strategy. After every procedure, it is regarded as the HENs are one of the alternative plans. It means that a HENDA with a single strategy can make a lot of HENs according to the progress of the simplification procedures. However, if the individual HENDA simplifies without "strategy switching," the optimal solutions may not be included, because there are limitations on the number of designed HENs. Furthermore, as shown in Figure 5, each HENDA can hand over partially simplified HEN to another HENDA to continue simplification based on another HENDA's strategy. It is possible to obtain the larger number of HENs. We call this mechanism "strategy switching". In Figure 5, the expression "HEN C2-A3" means that the HEN is simplified twice with strategy C and subsequently simplified once with strategy A.

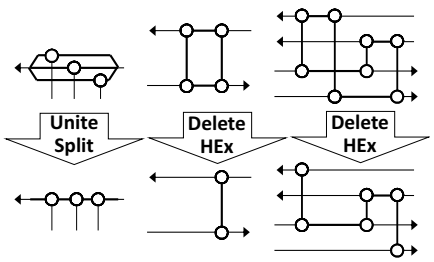


Figure 4: The simplification procedures.

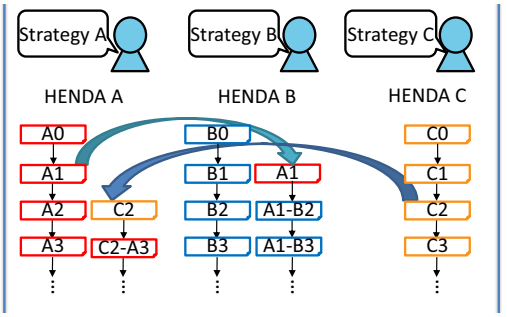


Figure 5: Outline of strategy switching.

3. Simulation

3.1. Application of Strategy Switching Method

In our previous study (Kimura 2016), we verified that the Strategy Switching Mechanism is effective in a small scale processes (4SP1). In this study, we carried out the simulations of several processes (4SP2, 9SP, 10SP). In this paper, we show the result of 4SP2, which is the smallest among the target processes due to the space limitations. The heat demands of 4SP2 are shown in Table 2. In 4SP2, whereas we could only obtain 225 kinds of HENs without switching, we could obtain 498 kinds of HENs with switching. Figure 6 shows the simplest HEN in the simulation with switching. This HEN has the 10 heat exchangers and the 4 splits. However, it is possible to reduce the number of heat exchangers to 6 and splits to 0 theoretically.

3.2. Strategy Improvement

In accordance with the inadequate result mentioned in section 3.1, we tried to improve the strategy.

There is a constraint that the temperature difference of all the heat exchangers must be larger than ΔT_{min}. If a HEN does not satisfy the constraint (that is '*inconsistent HEN*'), the HENDA discarded the HEN. However, we found that there is a possibility that HENDA can obtain consistent HENs by simplifying the inconsistent HEN subsequently. Then, we improved the strategies to make HENDA allowable to

temporarily design an inconsistent HEN. We had expected that consistent HENs are designed from an inconsistent HEN after introducing this mechanism, however, we could not obtain.
In most cases, the inconsistent heat exchangers in the inconsistent HENs are near the pinch point. The heat exchanges near the pinch point must satisfy the following equation:

$$(CP \text{ of stream entering the pinch point}) \leq (\text{the sum of } CP \text{ of stream exiting from the pinch point})$$

Therefore, we improve the strategy to meet this constraint. As a result, HENDAs could design simpler HENs.

Table 2: Heat demands of 4SP2.

Stream	T_{in}[K]	T_{out}[K]	CP[kW/K]	ΔH[kW]
Hot1	443	333	30	3300
Hot2	423	303	15	1800
Cold1	293	408	20	2300
Cold2	353	413	40	2400
Steam	450	450	-	
Water	353	413	-	

$\Delta Tmin$ is 5.6[K].

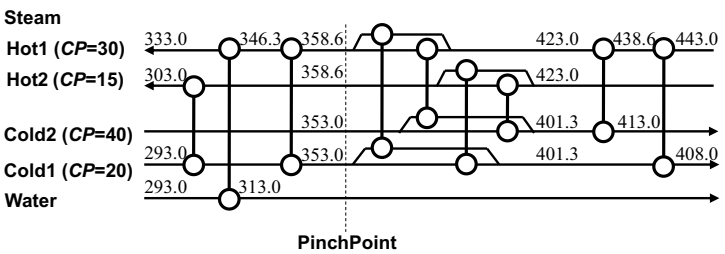


Figure 6: The simplest HEN of 4SP2 before improvement of strategies.

3.3. Result

In the simulation with the improved strategy, whereas we could only obtain 920 kinds of HENs without switching, we could obtain 2,232 kinds of HENs with switching. We show the number of kinds of HEN obtained in each simulation in Table 3, and the simplest HEN in the simulation with switching in Figure 7.
In this study, we compared the results on the total cost [$/year] (the equipment cost plus the utility cost), the numbers of the heat exchangers and splits in the most inexpensive HENs in Table 4. In table 4, the results are by (I) Yerramsetty (2008), (II) Pariyani (2006) and (III) our improved method with strategy switching. Despite implement of the improvement mentioned in section 3.2, the HEN whose cost is lower than the results in other methods could not be obtained.

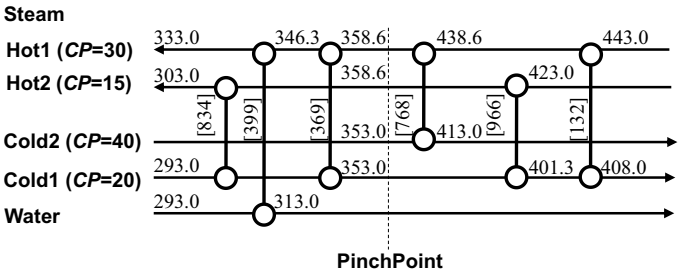


Figure 7: The simplest HEN of 4SP2 after improvement of strategies.

Table 3: The comparison result of the numbers of the HEN in 4SP2.

Supplement		No. of unique HEN
Before improvement	w/o switching	225
	switching	498
After improvement	w/o switching	902
	switching	2,232

Table 4: The comparison result of the costs of the HEN in 4SP2.

Method	Total Cost [$/year]			No. of	
		Equipment	Utility	Heat Exchanger	Split
(I) Differential Evolution (Yerramsetty)	84,222	76,222	8,000	6	1
(II) Randomized Algorithm (Pariyani)	85,307	77,307	8,000	6	1
(III) This study	84,648	76,648	8,000	6	0

4. Conclusion

Strategy switching and simplification of HENs could be proceeded by the improvements of the simplification algorithm. It brought that a large number of HENs with a wide variety can be made for some chemical processes. In our future work, it is necessary to implement a decision-making mechanism to evaluate the huge number of alternatives from various aspects.

References

N. Kimura, 2016, A Multiagent Approach to Heat Exchanger Network Design using Strategy Switching Mechanism, in Proc. of The 7th International Symposium on Design, Operation and Control of Chemical Processes (PSE Asia 2016), Tokyo, July 2016

T. Kou, 2010, Strategy Evaluation using Edit Distance for Heat Exchanger Network Design, in Proc. of The 5th International Symposium on Design, Operation and Control of Chemical Processes (PSE Asia 2010), P126, 482–490, Singapore, July 2010

K.M. Yerramsetty, 2008, Synthesis of cost-optimal heat exchanger networks using differential evolution, Computers & Chemical Engineering, 32, 1861–1876

A. Pariyani, 2006, Design of heat exchanger networks using randomized algorithm, Computers & Chemical Engineering, 30, 1046–1053

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50154-3

Systematic Comparison of Aggregation Methods for Input Data Time Series Aggregation of Energy Systems Optimization Problems

Holger Teichgraeber[a*], Adam R. Brandt[a]

[a]*Department of Energy Resources Engineering, 367 Panama St., Stanford, CA, USA*

hteich@stanford.edu

Abstract

We present a systematic comparison of how aggregation methods represent input data in the operational part of complex energy systems optimization problems. We compare both traditionally used methods such as k-means, k-medoids, and hierarchical clustering, and shape-based clustering methods such as dynamic time warping and k-shape in the domain of the objective function of sample operational optimization problems. Centroid-based approaches show improved performance as the number of clusters increases, whereas medoid-based approaches do not show a clear pattern as to when they best represent the objective function of the optimization problem of the full representation. Furthermore, clustering based on shape-based algorithms can improve performance compared to the traditionally used algorithms.

Keywords: Time-series aggregation, Clustering, Energy systems optimization

1. Introduction

The optimization of energy systems and of design problems of energy-intensive processes is computationally expensive and is often computationally intractable for the full set of input data (e.g. electricity prices, heating and cooling demand, renewable generation).

Clustering for time series aggregation has been used recently to address this challenge. It can reduce computational times for optimization problems by 1-2 orders of magnitude. For example, an energy system can be optimized for a small set of 5-10 representative days instead of a whole model year. Various studies have used different aggregation methods. Brodrick et al. (2015) and Teichgraeber et al. (2017) use k-means clustering for carbon capture applications, Bahl et al. (2017) use k-means clustering for decentralized energy supply systems, and Kotzur et al. (2018) use k-means, k-medoids, and hierarchical with the medoid for various energy supply systems.

All the above mentioned studies jointly investigate the representation in the operational and design domain of these 2-stage problems. In this study, we analyze only the representation of input data in the operational part of the problem in order to decompose the representation of the clustering algorithm by removing the design decisions, which are influenced by extreme values. We compare the previously mentioned aggregation methods as well as clustering methods that are shape-based and have not yet been applied to the optimization of energy systems: k-shape clustering and dynamic time warping barycenter averaging (DBA) clustering.

We formulate two simple operational problems that can be solved for the full set of electricity price time series data, so that the objective function error due to clustering can be exactly computed. These problems capture characteristics of real world energy-intensive industrial applications such as storage, load shifting, and on-site power generation. We then compare the clustering methods in the domain of the objective function of these optimization problems. This paper proceeds as follows: First, we introduce the aggregation methods that we compare, and we then introduce the sample optimization problems and the data. After that, we compare the performance of the aggregation methods on the optimization problems and draw conclusions.

2. Aggregation Methods

In this section, we introduce the aggregation methods that are compared in this paper.

We aggregate a time-series consisting of N daily price vectors $\mathbf{p_1}\ldots\mathbf{p_N}$ into K clusters. We introduce the following sets for notation: N = {1,...,N} is the set of daily indices, T={1,...,T} is the set of hourly time-step indices within each day, K={1,...,K} is the set of cluster indices, and C_k with k∈K are the disjoint sets of daily price vectors belonging to the respective clusters k.

Aggregation typically consists of clustering the days into K clusters and then choosing a representation of the cluster. The cluster center **c** is often, but not always, chosen as the representation.

We compare partitional clustering algorithms (k-means, k-medoids, dynamic time warping, k-shape), and hierarchical clustering algorithms (hierarchical clustering with centroid, hierarchical clustering with medoid).

2.1. Partitional clustering algorithms

All partitional clustering algorithms compared in this work minimize the within-cluster sum of squared distances and result in the assignment of daily price vectors $\boldsymbol{p}_i$ to the sets $C_1^* \ldots C_K^*$:

$$C_1^* \ldots C_K^* = \underset{C_1 \ldots C_K}{\operatorname{argmin}} \sum_{k \in K} \sum_{\boldsymbol{p}_i \in C_k} \operatorname{dist}(\boldsymbol{p}_i, \boldsymbol{c}_k)^2 \tag{1}$$

These algorithms differ in their choice of distance measure and cluster center but can all be solved following the general structure of Lloyd's algorithm (Lloyd, 1982): At the beginning, the cluster representations are initialized randomly. Then, the algorithm iteratively performs two steps until it converges or reaches a maximum number of iterations: In the first step (assignment), each day $\mathbf{p_i}$ is assigned to the closest cluster center, and in the second step (refinement), the cluster centers are updated in order to reflect the changes in cluster assignments.

Cluster centers can be centroid-based or medoid-based. The algorithm yields locally converged solutions based on the random cluster initializations. In this study, we run all

clustering algorithms until convergence and do not consider any non-converged solutions. In the following, we present the partitional clustering algorithms and their respective distance measure and cluster center:

K-means clustering: K-means clustering (Lloyd, 1982) minimizes the Euclidean distance. It uses the centroid for cluster representation, which can be computed with the arithmetic mean.

K-medoids clustering: Similarly to k-means clustering, k-medoids clustering uses the Euclidean distance as distance measure. However, it represents the cluster by its medoid. The medoid is an actual day $\mathbf{p}_i$. Besides being solved with the outlined partitional algorithm, the k-medoids clustering problem can be formulated as a Binary Integer Program (BIP) (Kotzur et al., 2018). The BIP formulation can be solved to global optimality for the number of days considered in this study.

Dynamic Time Warping based clustering: Dynamic Time Warping (DTW) (Petitjean et al., 2011) takes into consideration that the shape of two time-series can be similar but shifted because it is able to align different hours of the day. The optimal non-linear alignment between two days can be calculated using dynamic programming. The algorithm can be constrained to a so-called "warping window", where only hours i and j, $|i - j| \leq b$, can be aligned. DTW allows us to calculate the assignment step in the partitional clustering algorithm. For the refinement step, we use DTW barycenter averaging (DBA), which is an iterative algorithm, to calculate the cluster centroid.

K-shape clustering: K-shape clustering has recently been proposed as an alternative to DTW as a shape-based clustering method (Paparrizos and Gravano, 2015). It uses a shape-based distance SBD. SBD finds the optimal shift w of two time-series vectors **u** and **v** such that the cross correlation between these vectors is maximized. SBD can be calculated efficiently using a Fast Fourier Transform algorithm. Each cluster is represented by its centroid, which can be computed analytically.

2.2. Hierarchical clustering algorithm

Hierarchical clustering starts with k=N clusters and proceed by merging the two closest days into one cluster, obtaining k=N-1 clusters. The process of merging two clusters to obtain k-1 clusters is repeated until we reach the desired number of clusters K. We use the Euclidean distance to find which clusters to merge, using the algorithm proposed by Ward (1963). The final cluster assignments are then represented by either the centroid or the medoid. Hierarchical clustering is deterministic, which means it is reproducible. However, it is also greedy, which means that it yields local solutions.

3. Optimization Problem and Data

We formulate an electricity storage problem and a gas turbine power generation problem. These problems are simple operational problems formulated as linear programs that can be solved directly for 365 days. This gives us the true value of the objective function, to which the value of the objective function with clustered input data is compared. In combined design and operations problems, the optimal decision variables are of primary interest. However, in this study, we consider the operational part of the optimization problem. Because the design decision variables of a potential design and operations problem are strongly influenced by the operational profitability (our objective function), we use the objective function as the measure to compare the performance of the clustering algorithms.

The problem is formulated in a way that maximizes operating profit for one day in hourly time-steps Δt and is carried out for each clustered day. The yearly profit is then calculated as a weighted addition of the individual clustered days' profits.

The electricity storage problem is formulated in Eq.(2):

$$\begin{aligned}
&\max_{\bar{\bar{E}},\bar{E},S} \sum_{k\in K} N_k \sum_{t\in T} (\bar{\bar{E}}_{k,t} - \bar{E}_{k,t}) p_{k,t} \\
&\text{s.t.} \\
&0 \le \bar{\bar{E}}_{k,t} \le P_{max}\Delta t && \forall\, k \in K,\, t \in T \\
&0 \le \bar{E}_{k,t} \le P_{max}\Delta t && \forall\, k \in K,\, t \in T \\
&0 \le S_{k,t} \le E_{max} && \forall\, k \in K,\, t \in T \\
&S_{k,t+1} = S_{k,t} + \eta_{in}\bar{E}_{k,t} - \frac{\bar{\bar{E}}_{k,t}}{\eta_{out}} && \forall\, k \in K,\, t \in T \\
&S_{k,1} = S_{k,T+1} \wedge S_{1,1} = S_{k,1} && \forall k \in K
\end{aligned} \tag{2}$$

Here, $\bar{\bar{E}}_{k,t}$ and $\bar{E}_{k,t}$ are hourly energy that flows out of or into the battery from the market, $p_{k,t}$ is the hourly electricity price, and $N_k = |C_k|$ is the number of days in cluster k. P_{max} is the maximum power, and E_{max} is the maximum storage capacity of the battery. $S_{k,t}$ is the amount of energy stored in the battery, and η_{in} and η_{out} are the charge and discharge efficiencies. We assume maximum power of 100 MW and maximum storage capacity of 400 MWh, with charge and discharge efficiencies at a typical value of 95%.

The gas turbine dispatch problem is formulated in Eq.(3):

$$\begin{aligned}
&\max_{\bar{\bar{E}}} \sum_{k\in K} N_k \sum_{t\in T} \bar{\bar{E}}_{k,t} p_{k,t} - \frac{\bar{\bar{E}}_{k,t}}{\eta_{gt}} p_{gas} \\
&\text{s.t.} \\
&0 \le \bar{\bar{E}}_{k,t} \le P_{max}\Delta t && \forall\, k \in K,\, t \in T
\end{aligned} \tag{3}$$

Here, $\bar{\bar{E}}_{k,t}$ is the hourly electric energy generated from the gas turbine, and p_{gas} is the gas price. We assume a gas price of 6.8 \$/GJ, and assume a combined cycle efficiency η_{gt} of 60% as can be found in modern combined cycle gas turbines.

The different clustering approaches are evaluated on hourly electricity price data from the day-ahead market in Germany from 2015. The mean price is 31.6 EUR/MWh, and the standard deviation is 12.7 EUR/MWh. A visual example of cluster outcomes of k-means on the data can be found in Teichgraeber et al. (2017).

We z-score normalize the data. Z-score normalization is commonly applied in statistics before clustering algorithms are applied, it shifts the mean to zero and the standard deviation to one. We use element-wise z-score normalization, which normalizes each hour of each daily sample by the same mean and standard deviation. Shape-based methods often use sequence-wise z-score normalization, which we suspect to work well in this case as well and will be part of future work.

4. Results

These optimization problems are operational, and the results presented here thus show the representation of the operational problem. We compare the performance of the aggregation method based on the objective function value, and we calculate the objective function value of the respective optimization problem for the full representation (365 days of electricity prices) for comparison. Note that unlike for our simplified problem, it is often computationally intractable for a complex problem involving design and operations to optimize the full representation.

Figure 1 shows the objective function value normalized by the objective function value of the full representation for the battery problem (Figures 1a and 1b) and the gas turbine problem (Figures 1c and 1d). First, we investigate conventionally used clustering methods (Figures 1a and 1c). We observe that k-means clustering and hierarchical clustering with centroid perform similarly on both problems. They generally increase in objective function value with increasing number of clusters (except for two notable decreases from 4 to 5 and 6 to 7 clusters of k-means in the gas turbine problem). Compared to the centroid-based clustering algorithms, the medoid-based algorithms perform less predictably. As the number of clusters increases, the objective function either increases toward or decreases away from the full representation, with no clear pattern emerging. If the full representation is unknown, choosing the number of clusters used in a specific case study is challenging.

The performance of the shape-based clustering methods DTW and k-shape with element-wise normalization is shown in Figures 1b and 1d. We compare DTW with a Sakoe-Chiba warping window of 0 to 3 and k-shape clustering. DTW with a warping window of 0 yields the same results as k-means, which confirms our implementation of the algorithm. A warping window of 1 increases the objective function for both optimization problems. As the warping window increases further, the objective function increases on average, but it also increases toward and decreases away from the full representation with less of a pattern emerging. As we increase the warping window size, the clustering algorithm is able to compare hours further and further away, which at some higher warping window size may lead to clustering peaks and valleys in price that may not belong together physically. Increasing the size of the warping window to a number too large can more generally be seen as overfitting.

K-shape on the other hand overestimates the objective function value for the storage optimization problem, and fluctuates as the number of clusters increases.

5. Conclusions

It is often computationally intractable to solve design and operations optimization problems of energy systems for the full set of input data. In this study, we present a comparison of conventional and shape-based clustering algorithms, and how they affect the operational representation in terms of the objective function value of the full representation. We find that centroid-based clustering algorithms behave more predictably with increasing number of clusters than medoid-based clustering algorithms, and we also find that shape based algorithms can improve performance over k-means. Future work will include comparing shape-based algorithms using sequence-wise instead of element-wise normalization.

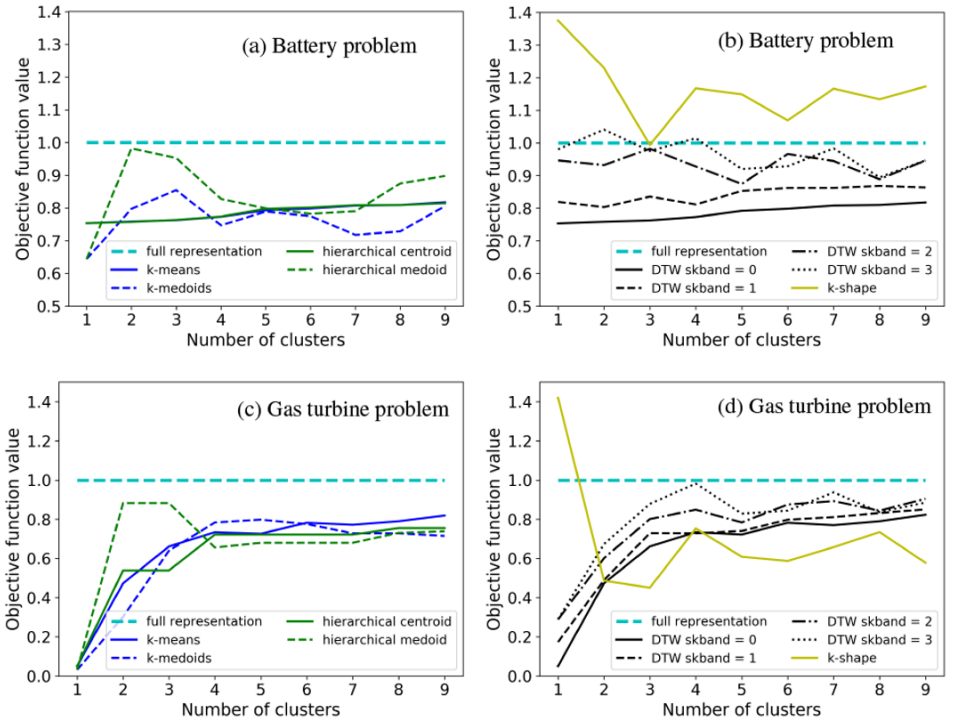


Figure 1: Comparison of normalized objective function values of different clustering methods: (a) conventional clustering algorithms on the storage optimization problem, (b) shape-based clustering algorithms on the storage optimization problem, (c) conventional clustering algorithms on the gas turbine optimization problem, and (d) shape-based clustering algorithms on the gas turbine optimization problem. Price data from Germany in 2015.

References

Bahl, B., Kümpel, A., Seele, H., Lampe, M., Bardow, A., 2017. Time-series aggregation for synthesis problems by bounding error in the objective function. Energy 135, 900–912.

Brodrick, P.G., Kang, C.A., Brandt, A.R., Durlofsky, L.J., 2015. Optimization of CCS-Enabled Coal-Gas-Solar Power Generation. Energy 79, 149-162.

Kotzur, L., Markewitz, P., Robinius, M., Stolten, D., 2018. Impact of different time series aggregation methods on optimal energy system design. Renewable Energy 117, 474-487

Lloyd, S.P., 1982. Least Squares Quantization in PCM. IEEE Trans. Inf. Theory 28, 129–137.

Paparrizos, J., Gravano, L., 2015. k-Shape: Efficient and Accurate Clustering of Time Series. Acm Sigmod 1855–1870.

Petitjean, F., Ketterlin, A., Gancarski, P., 2011. A global averaging method for dynamic time warping, with applications to clustering. Pattern Recognit. 44, 678–693.

Teichgraeber, H., Brodrick, P.G., Brandt, A.R., 2017. Optimal design and operations of a flexible oxyfuel natural gas plant. Energy 141, 506–518.

Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. 58, 236–244.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50155-5

Comprehensive Investment Framework for Optimal Exploitation of Liquid-Rich Unconventional Reservoirs and Natural Gas Distribution

Jorge Chebeir[a]*, Hope Asala[b], Vidhyadhar Manee[a], Ipsita Gupta[b], Jose Romagnoli[a]

[a] *Louisiana State University, Cain Department of Chemical Engineering, 3307 Patrick F. Taylor Hall, Baton Rouge, LA, 70803, USA*
[b] *Louisiana State University, Craft and Hawkins Department of Petroleum Engineering, 3207 Patrick F. Taylor Hall, Baton Rouge, LA, 70803, USA*
jchebe1@lsu.edu

Abstract

A techno-economic framework is proposed for the optimal exploitation of unconventional shale reservoirs and the delivery of shale gas products within the Marcellus region. A 3-D compositional reservoir model is constructed to simulate realistic drilling, completion, stimulation (DCS) strategies and pad production processes. A feed-forward neural network (FNN) is implemented to understand re-frac well candidature so as to aid development of alternative development strategies throughout a project planning horizon. In addition, optimal well configurations and DCS strategies for well-pad development are developed in line with best petroleum engineering practices. Simulation results are incorporated in a long-term strategic planning model used for determining the optimal shale gas supply chain network and its operations. For this, a mixed-integer non-linear programing (MINLP) formulation is developed to maximize net present value (NPV). Natural gas demand is predicted by running long short-term memory (LSTM) neural networks based on the historical data of critical predictors. Results of this framework depict the influence of reservoir engr., completion engr., and machine learning in predicting project profitability and establishing optimal strategies for shale gas production and distribution.

Keywords: 3-D compositional reservoir model; MINLP model; Feed-forward Neural Network; LSTM neural network.

1. Introduction

The United States has an enormous potential in the energy sector given the abundant presence of natural gas in shale formations. These formations are mainly characterized by low permeability (from 10^{-6} - 10^{-2} md), low porosity, low strength resistance and water sensitivity. In the last decade, the rapid development of technologies including extended-reach directional drilling, hydraulic fracturing, and biodegradable diverter technology have enabled more economic exploitation of shale reservoirs (EIA, 2013).
The development of shale gas resources represents a multidimensional challenge due to the complexities involved in its exploitation. The quality and quantity of shale resources inferred in resource assessments and the exact location of the zones with the highest

productivity ('sweet spots') or richer content in liquids of higher economic value remain uncertain until actual exploratory drilling occurs. Geologists and engineers have to determine pay zones with superior reservoir and geomechanical characteristics to ensure that drilling and completion operations occur within these high-quality intervals. In this context, 3D compositional models that capture the reservoir characterization, the physics of shale gas adsorption and incorporate realistic development strategies can be used as a support tool to analyse multi-shale pad operations. The access to necessary infrastructure is also critical to ensure the optimal allocation of subsequent shale gas products. Once a pad is constructed and wells are producing, the next step is delivery to final consumers. Gathering pipelines, compression stations, water management infrastructure and other support facilities are constructed to ensure the movement of natural gas to in-state distribution pipelines. In the presence of liquid-rich gas, processing is required for separating heavier and more valuable hydrocarbons. These factors have a critical influence on gas developments and encourage gas producers to re-think the short and long-term strategies implemented.

In recent years, large-scale production of shale gas has accompanied the plunge in natural gas prices. Although this slump of gas prices has made the U.S. attractive for developing other industries, it also poses financial difficulties to gas operators and asks significant questions about the future of shale gas production. This factor accompanied by the customary sharp decline of producing shale gas wells has encouraged gas producers to invest more in liquid rich shale plays. For shale oil and gas production, larger expenditures are required for infrastructure, and the economic feasibility of such projects is still a concern in the presence of fluctuating demand. Therefore, a comprehensive approach that combines formation evaluation, reservoir simulation and long-term strategic planning would suffice for predicting the success of these ventures.

2. Problem statement

A techno-economic approach is proposed for the development of shale gas assets. Fig 1 illustrates the workflow for the optimization framework.

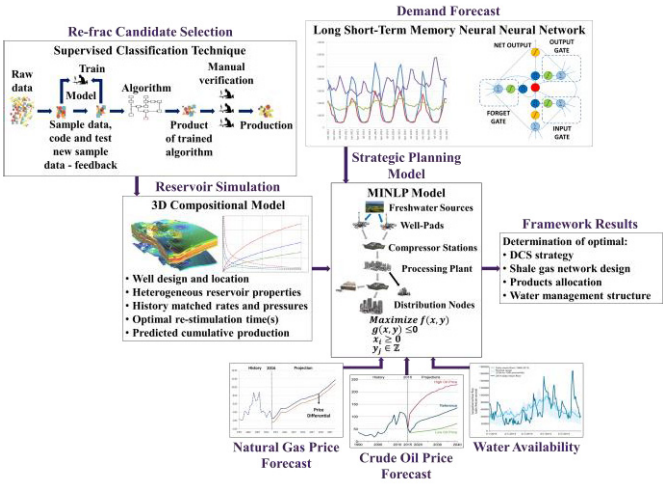


Figure 1: Integrated optimization framework for economic development of shale gas.

In this approach, a supervised classification technique is implemented to determine acceptable candidates for re-stimulation (Hope *et al*., 2017). An FNN with an architecture

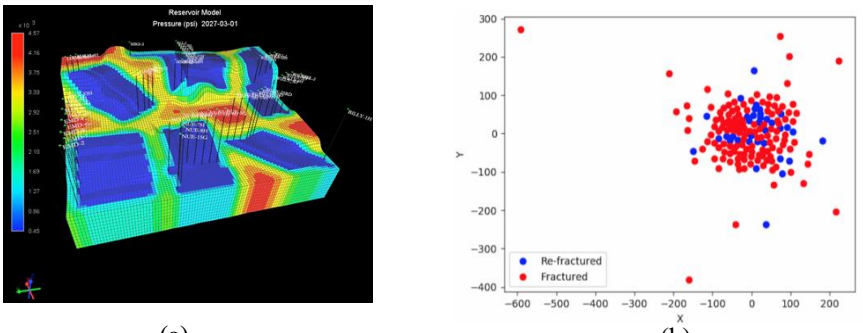
of 23-128-2 was used to determine re-frac candidates from 250 wells. This approach coupled with petroleum engineering practices are used to define potential DCS strategies for the reservoir. Considering geological properties of the formation, these strategies are implemented in a reservoir simulator. After running the simulations, shale gas production, re-fracturing times, and wastewater rates among other parameters are determined for each strategy and utilized as input data to the strategic planning model. This model is described by an MINLP formulation which determines the optimal DCS strategy, infrastructure, operation, and product deliverability, while maximizing enterprise profitability. Natural gas demand trends during the planning horizon are determined by an LSTM neural network and included in the strategic planning model. Other parameters including natural gas and crude oil prices as well as water availability are also inputs to the MINLP model.

3. Reservoir Model, Development Strategy and Candidate Selection

A reservoir model was built for the case study in question in order to simulate alternative field development strategies that will optimize gas recovery and project NPV. The reservoir model was developed using CMG's Builder and run using an adaptive implicit solver (GEM). A single component Langmuir isotherm was used. The near wellbore region of all 54 planned wells employed 7x7x1 local grid refinement (LGR). This aided in capturing important transient effects close to the wellbore and fractures as well as manage the computational effort required to run the compositional model. Fig. 2a illustrates a depletion map for strategy II after 10 years of development.

Three main development strategies (incorporating drilling, stimulation, and re-stimulation) were determined as optimal alternatives for optimizing recovery from the shale reservoir case study. The developed strategies (I, II and III) are the product of reservoir engineering/economic optimization models, current practices in the area, and results of a feedforward neural network algorithm developed for re-frac candidate selection. Details of this approach were discussed by (Hope *et al*., 2017). Six well pads (with variable number of wells) are planned for development having 4000ft - 5000ft laterals, five possible stimulation stage designs and three major well orientations.

A feedforward single layer NN algorithm was developed using well stimulation, completion, reservoir, operational and production data for existing wells in the area of interest. Fig. 2b illustrates mapping of these well datasets. The results of the t-SNE and NN gave insights to the ratio of wells that should be re-fractured among planned wells.



(a) (b)

Figure 2: (a) Illustration of reservoir model showing depletion map at the end of planning horizon, according to strategy II. (b) t-SNE visualization of well data set.

4. Strategic planning model

4.1. MINLP model

For the 10-year planning horizon considered, an MINLP model was developed for shale gas production, transportation and water management, for a venture located in the liquid-rich region of the Marcellus play. The shale gas production in each pad is :

$$SP_{i,t} = \sum_{st \epsilon ST} Y_{st} \cdot spr_{i,st,t} = \sum_{j \epsilon J} FJ_{i,j,t} \quad \forall\, i \epsilon I, t \epsilon T \tag{1}$$

where Y_{st} is a binary variable that equals 1 when a DCS strategy is selected. $spr_{i,st,t}$ is the pad production for each strategy. $SP_{i,t}$ is total shale gas production at each pad. $FJ_{i,j,t}$ represents the shale gas transported from well pads to compression stations.
The variability of total dissolved solids (TDS) in water is considered in this model. This has bilinear terms and introduces non-linearity in our formulation (Yang *et al.*, 2016).
In each compression station, it is possible to establish component balances as follows:

$$FPM_{j,p,t} = mcj_{j,t} \cdot FP_{j,p,t} \quad \forall\, j \epsilon J, p \epsilon P, t \epsilon T \tag{2}$$

where $FPM_{j,p,t}$ is methane transported from compression stations to processing plants. $mcj_{j,t}$ stands for methane composition at each compression station. $FP_{j,p,t}$ is total gas transported from compression station to processing plant. The second term of the equation also includes a bilinear term. This expression can be extended to the other gas constituents. In the case of methane, the flow balance in the processing plat is given by:

$$\sum_{j \epsilon J} FPM_{j,p,t} = \sum_{m \epsilon M} FMM_{p,m,t} + \sum_{u \epsilon U} FMU_{p,u,t} \quad \forall\, p \epsilon P, t \epsilon T \tag{3}$$

where $FMM_{p,m,t}$ is methane transported from processing plant to distribution node. $FMU_{p,u,t}$ is methane transported from processing plant to underground reservoir.
The supply of methane has to satisfy a demand constraint as follows:

$$\sum_{p \epsilon P} FMM_{p,m,t} + \sum_{u \epsilon U} FMUM_{u,m,t} \leq dm_{m,t} \quad \forall\, p \epsilon P, t \epsilon T \tag{4}$$

where $FMUM_{u,m,t}$ is the methane transported from underground reservoirs to distribution nodes. $dm_{m,t}$ stands for gas demand at each distribution node. Similar equations (without storage) can be defined for other shale gas constituents.
The objective function is the maximization of the net present value (NPV), given by:

$$Max\ NPV = \sum_{t \epsilon T} \frac{CFlow_t - CapEx_t}{(1+dr)^{t-1}} - CapExI \tag{5}$$

where NPV is determined as cash flow minus shale wells' investment and capital investment required for gas transportation and processing.

4.2. LSTM neural network for natural gas demand prediction

Time series forecasting, at its heart, is the search for an optimal function that maps historical predictors to forecasted data. Since forecasted data is not available a priori, the algorithm is trained by dividing the historical data into past and future data. Neural network models like LSTM have been shown to outperform traditional time series methods on temporal processing tasks. LSTM's replace the neurons in traditional neural networks with cells that have the ability to retain useful information and overwrite

extraneous data. In this paper, the model is built by stacking an LSTM, sigmoid and linear layer one over the other to maximally increase the accuracy. The number of nodes/layer, epochs and batch size are design parameters that were optimized using exhaustive enumeration. The input to the model comprises of predictors like the price of natural gas, crude oil price, regional population, regional temperature and past natural gas demand. Results of implementing LSTM can be observed in Fig. 3.

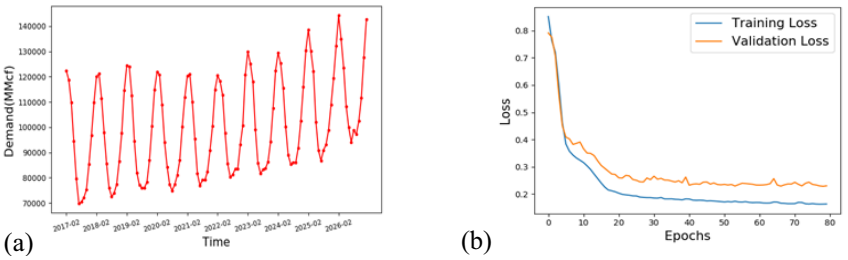


Figure 3. (a) 10-year natural gas demand forecast. (b) Training and validation loss.

5. Results and discussion

The compositional reservoir model was developed with CMG's Builder using an Intel(R) Xeon(R) CPU @ 3.10GHz, 64-bit operating system. Simulation took 10 h of CPU time. Maximum material balance error reported was 2.8113e-5 %. The MINLP model was implemented in an Intel(R) Core(TM) CPU @ 4GHz, 64-bit operating system. The global optimization solver utilized in GAMS was BARON (Tawarmalani and Sahinidis, 2005) with sub-solvers: CONOPT, MINOS and SNOPT. Convergence was achieved in 1.5 h of CPU time with an optimality gap of 2.91%. Python platform was utilized for developing the FNN and LSTM algorithms. The optimal infrastructure is given by Fig. 4.

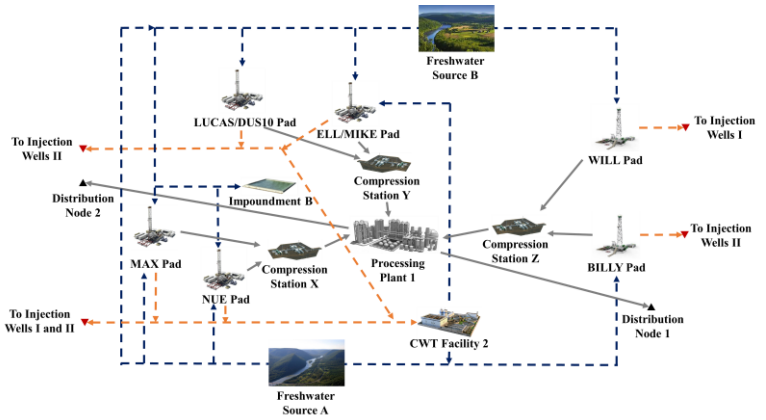


Figure 4: Optimal infrastructure for shale gas production and distribution.

Development strategy II was selected during optimization as it maximizes exploitation of the shale gas reservoir and demand for natural gas, with minimum investment costs. 18 wells were planned for re-fracturing. Fig. 4 shows two distribution nodes are supplied without gas storage. This is necessary to handle when production exceeds demand throughout the planning horizon. Economic results reveal an NPV of 289.83 MMUS$ for

the 54 wells distributed in 6 production pads in the field. The investment required for drilling, stimulation, and completion operations as well as multiple re-stimulation campaigns summed up to 156.626 MMUS$. The costs incurred in drilling and completion far outweighed re-stimulation costs incurred over the planning horizon. This justifies the advantage of re-fracturing selected wells rather than drilling infill wells. Infrastructure costs including pipeline gathering and processing plant summed up to 242.32 MMUS$ (Fig. 5a). Operating and DCS costs are illustrated in Fig. 5b.

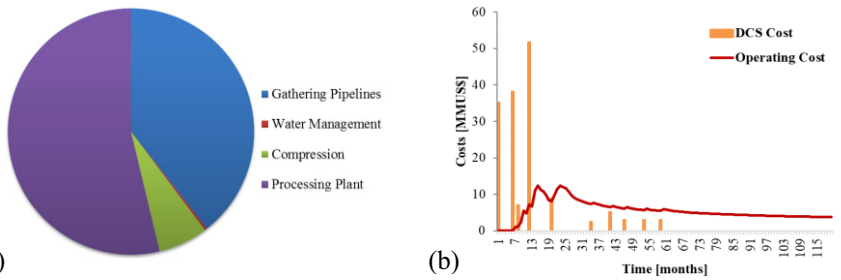


Figure 5: (a) Infrastructure expenditure. (b) Operating and DCS costs.

6. Conclusions

This work demonstrates the importance of an integrated approach for solving the techno-economic problem of shale gas exploitation and supply chain optimization. The simulation of unconventional reservoirs offers a more realistic methodology to define optimal field development strategies which incorporate re-fracturing. Background reservoir engineering optimization should be incorporated in shale gas project planning to support well placement and efficient recovery of natural gas.

Data analytic techniques such as the t-SNE can be analysed with feed-forward algorithms to delineate potentially successful re-frac candidates. This eventually provides a decision support tool for implementing the best initial stimulation and re-stimulation strategies. True NPV optimization was achieved after LSTM demand trends were predicted. Understanding future demand (distribution node dynamics) gave insights to the specific development strategy needed for shale gas exploitation. Future work integrates Machine learning predictions of underground storage, re-frac candidate selection, external/internal demand, and water availability with MINLP optimization.

References

U.S. Energy Information Administration (EIA), 2013, International Energy Outlook 2013 with projections to 2040, US Department of Energy, Washington, DC.

H. Asala, J. Chebeir, W. Zhu, A. Dahi Taleghani, J. Romagnoli, 2017, A Machine Learning Approach to Optimize Shale Gas Supply Chain Networks, Paper SPE-187361-MS was prepared for presentation at the SPE Annual Technical Conference and Exhibition, 9-11 October, San Antonio, Texas, USA. https://doi.org/10.2118/187361-MS

M. Tawarmalani, N. Sahinidis, 2005, A polyhedral branch-and-cut approach to global optimization, Math. Prog., 103(2), 225-249. https://doi.org/10.1007/s10107-005-0581-8

L. Yang, I. Grossmann, M. Mauter, R. Dilmore, 2015, Investment optimization model for freshwater acquisition and wastewater handling in shale gas production, AIChE J., 61, 1770–1782. http://onlinelibrary.wiley.com/doi/10.1002/aic.14804/abstract

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50156-7

Bioenergy investments in sugarcane mills: an approach combining portfolio theory with neural networks

Victoria M. Mutran*, Celma O. Ribeiro, Claudio A. O. Nascimento, Erik E. Rêgo

Universidade de São Paulo, Av. Prof. Almeida Prado, Trav. 2, n. 83, São Paulo, 05508-900, Brazil
victoriamutran@usp.br

Abstract

Investments to increase bioenergy generation have been discussed in the sugarcane industry, as public policies opened a broad range of opportunities for bioelectricity in Brazil. The investment decisions are usually based on market conditions, while those projects may implicate relevant changes in product mix decisions and the process conditions in mills. Thus, this study aims to address the nonlinearities associated to process modeling by an optimization framework combining portfolio theory with artificial neural networks, a widely used approach to approximate functions. The results indicate that changes in process conditions impact decision-making for investments. In conclusion, this study evidences the importance of considering process conditions for optimal economic decisions.

Keywords: sugarcane, portfolio optimization, neural networks.

1. Introduction

In industrial processes, the use of residues as a source of revenue has been a practice that enhances the development of strategies to improve energy recovery from waste. In fact, in some sectors, bioelectricity cogeneration already holds a relevant share of the business revenue. This is the case of the Brazilian sugarcane industry, for which there is a large-scale generation of biomass. The bagasse, one of the main residues of the process, represents approximately 27 % of all sugarcane processed. Thus, it is a challenge to find a proper destination for it, which consequently motivates studies for economically feasible applications.

Cogeneration plants were firstly implemented in sugarcane mills as an environmental-friendly destination for the bagasse, applied at the time solely for internal demands of electricity in the process, due to a lack of market for commercialization. Later on, as a consequence of the implementation of public policies to increase renewable energy in Brazil, bioelectricity from bagasse has undergone an expressive growth, now holding a share of approximately 9% of the national matrix of electricity generation.

As a consequence of the increasing opportunities in the energy sector, producers currently face difficulties for decisions with regards of energy recovery projects. Their willingness for investments is usually based on their expectation of profitability, which in turn is highly dependent on market conditions and prices. In the scenario of the Brazilian electricity market, there are two environments of commercialization: the free environment and the regulated environment of commercialization. The first one consists

of bilateral short to mid-term contracts, for which prices are based on the spot market, subject to price volatilities. In the regulated environment, on the other hand, the contracts are usually 20 to 25-years-long, for which there are no variations in electricity prices. Hence, in addition to the investment decisions, sugarcane producers shall also choose a commercialization option for the additional electricity based on their risk-taking profiles.

Price risks have always been an inherent challenge in the sugarcane business, since sugar and ethanol prices have been historically volatile. This way, annual production decisions are usually based on the producer's awareness of price risks. Modern Portfolio Theory, firstly introduced by Markowitz (1952), has been vastly applied as a strategy for such types of risks mitigation. The author provided a method - through an optimization model - to identify how good a given portfolio is, based on the risk and the expected return of assets. Thereupon, the development of risk management theory led to a significant improvement in studies of risk measures. A more recent measure proposed by Rockafellar and Uryasev (2000), the Conditional Value at Risk (CVaR), has been applied in a variety of problems in literature and may be appropriate to deal with this kind of investment decisions, since it is single-tailed. However, the CVaR function is not tractable, resulting in a complex optimization problem to solve and indicating the need for approximation and smoothing techniques (Barrosa et al., 2016).

Moreover, the traditional portfolio optimization approach focuses on economic aspects. Nevertheless, in additional to the high investment costs, the implementation of new bioenergy generation projects implicates extensive process design problems, since they yield changes in process conditions and, consequently, in the availability of surplus electricity. While the process models are nonlinear, and also may considerably change optimal economic decisions, surrogate models emerge as an alternative to make intelligent approximations for complex optimization problems.

Strategies for optimization in process system engineering have been a widely discussed topic in literature, indicating the use of surrogates for modeling as a promising field. In chemical engineering, artificial neural networks (ANN) have been successfully applied to approximate process functions and thus simplify optimization models. Thus, this study aims to propose an optimization framework to solve such types of portfolio decision problems, combining portfolio theory and ANNs. This way, it may contribute to literature in optimization models to support investment decisions on new energy recovery processes, as well as the use of neural networks as surrogate models to portfolio optimization.

This paper is organized as follows: the proposed optimization framework is described in section 2, followed by a case study with an application, in section 3, where results are presented and discussed. Finally, conclusions of this work are in section 4.

2. The proposed framework

In finance, a broadly used measure for risks in a portfolio is the Value at Risk, which states the greatest loss that might occur at a confidence level β, denoted by α. Although popular among portfolio managers, the VaR presents some drawbacks that hinder its use in some applications. Thereupon, the Conditional Value at Risk (CVaR) was proposed by Rockafellar and Uryasev (2000), being a risk measure that accounts for the expected shortfalls at a limited percentile β, as presented in Eq. (1).

$$\mathrm{CVaR} = \mathrm{F}_{\beta}(\mathrm{x},\alpha) = \alpha + \frac{1}{(1-\beta)} \int_{\mathrm{y}\in\mathbb{R}^{\mathrm{m}}} [\mathrm{f}(\mathrm{x},\mathrm{y}) - \alpha]^{+}\mathrm{p}(\mathrm{y})\mathrm{dy} \quad (1)$$

where p(y) is the density function of a random vector *y* that satisfies the cumulative distribution function for the loss associated with the decision vector *x*. Given the relevant volatility in prices for products in the sugarcane sector, the decision-maker has to face a trade-off between expected profitability and risks of significant falls in prices. This way, CVaR may be an adequate measure to deal with such types of product mix and investment decisions in agro-industrial plants.

The portfolio optimization model proposed for this decision problem is presented below. Let $x \in \mathbb{R}^{\mathrm{N}}$ be a decision vector of quantities of each product in a sugarcane mill (sugar, ethanol, free-market electricity, regular-market electricity) and p is the vector of prices for each product in the portfolio. Further, c is the vector of production costs, $\$_{\mathrm{j}}$ is the investment cost of option $j \in \mathrm{J}$, the set of investment options assessed, and $u_j \in \{0,1\}$ is the investment decision. Thus,

$$\begin{aligned} &\min_{x\epsilon X} \mathrm{CVaR}(x,\mathrm{u}) \\ &\text{s. t.} \\ &\textstyle\sum_{i=1}^{N}(\mathrm{E}(\mathrm{p}_i) - c_i)x_i \; - \sum_{j=1}^{J} \$_j u_j \geq r_0\,, \quad x \epsilon \mathbb{R}_{+}^{N}, \;\; u \in \{0,1\} \end{aligned} \quad (2)$$

where X is the set of operational constraints, and r_0 is the lower bound for profit determined by the decision-maker. To build the efficient frontier of risk and expected profit, one should solve the optimization problem for a variety of levels of r_0. Clearly from Eq.(1), the CVaR is not mathematically tractable. Its nonlinearity turns the formulation presented in Eq. (2) into a difficult optimization problem. Thereby, Rockafellar and Uryasev (2000) presented an approximation approach, through the use of historical series or Monte Carlo simulation. However, this method is extremely sensitive to the quantity of simulated scenarios (Barrosa et al, 2016).

In addition, the operational constraints in the model are built from the modeling of processes in the sugarcane mill. Those models also present several nonlinearities that shall be incorporated into the portfolio optimization problem, causing a great increase in the model's complexity. In order to deal with this challenge, it is proposed the use of ANNs, since they have been successfully applied in a variety of chemical engineering problems, including process modeling and optimization. Villarrubia et al (2018) discuss the use of ANNs to approximate the objective function in optimization problems, proposing a framework for which the output function from the ANN is used as an approximation of a nonlinear objective function. Similarly, for the bioenergy investment decisions in the Brazilian sugarcane mills, the portfolio optimization model may be approximated through an ANN. The input variables in the neural network are the amount of sugar (1) and ethanol (2) produced in each product mix, the binary investment decisions (3), and the amount of electricity commercialized to the free (4) and to the regulated market (5). The availability of surplus electricity, which may be destined to either one of the two markets, depends on the process conditions resultant from the first three inputs variables. The simulated process data later used to train and test the ANN is contemplating those inherent variations in process conditions.

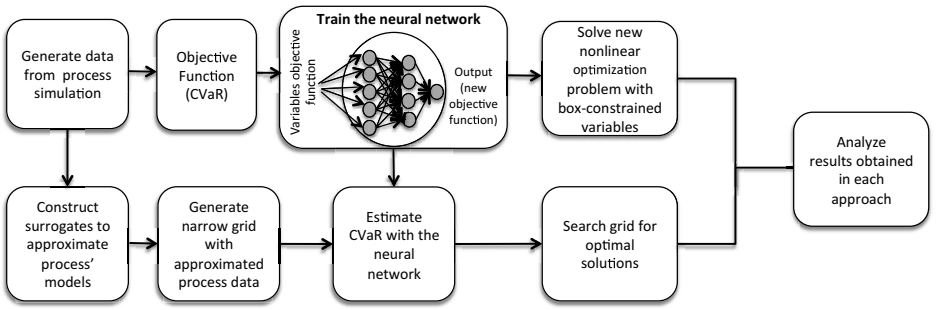


Figure 1 – Optimization framework (Source: Authors)

The optimization framework proposed is presented in Figure 1, summarizing the two possible approaches to use ANN to approximate the solution of this complex problem. The simulated data for decision variables and the CVaR calculated for each product mix, based on the price distributions of all products in the portfolio, are used to fit the neural network, from which the approximated new objective function shall be extracted. The output function obtained from the ANN may be used to find optimal portfolios through grid search or as a new objective function for the optimization problem. In the latter application, the output function is also nonlinear, however, its complexity is reduced, allowing the use of nonlinear optimization software to solve it, while respecting the original box-constraints for the decision variables.

3. Application

In order to illustrate the optimization framework presented in this paper, a simplified case study was performed. Process conditions were simulated from a phenomenological model of a sugarcane mill with processing capacity of 3,000,000 t of sugarcane per year. The original process scenario consists of cogeneration with a Rankine cycle, and the investment decision is to increase electricity generation through the implementation of a backpressure turbine. Opportunity costs associated with other investment options were not considered in the problem. The main production decision is the share of sugarcane's juice that will be diverted to the production of sugar, which implicates changes in ethanol production, as well as in the internal consumption of steam and electricity, changing the amount of surplus electricity available for commercialization.

3.1. Model Calibration

The scenarios were constructed for the two configurations of bioenergy generation, being treated as a binary decision in the optimization problem and implicating investment costs if chosen. For each configuration, the share of juice for sugar was varied from 0 to 100 %, providing annual productions of sugar (t/y), ethanol (m^3/y) and the surplus electricity available for commercialization (MWh/y). Further, scenarios for shares of energy destined to the regulated or the free market were generated randomly, respecting the constraint that all electricity available should be sold. This way, the domain of the variables was created with 440 scenarios generated for combinations of bioenergy configurations (2), shares of juice used for sugar production (11), and shares of energy commercialized in the regulated market (20).

The purpose of this surrogate is to approximate the risk (CVaR) as a function of the decision variables, which in this case are annual production of sugar and ethanol, annual

amount of electricity sold to the free market and the regulated market, and the binary decision on bioenergy investments, accounting also for the important variation in prices associated with variables. Thus, the CVaR was calculated based on a set of historical prices containing 100 monthly observations, from 2005 until 2015. The production costs were assumed to be constant and were based on literature (Grisi et al., 2012), as well as the investment cost, for which the equivalent annual cost (EAC) was obtained from Dantas et al. (2013). All monetary values are in 2010 US dollars. The results obtained are presented in the following subsection.

3.2. Results and Discussion

The neural network that presented the best fit with the training set has one hidden layer with four neurons. Although differentiable, the ANN output function is nonlinear. Thus, its numerical solution is still difficult to determine. The Augmented Lagrangian Method was chosen among the nonlinear optimization packages in R to solve this problem, subject to box-constraints for decision variables, aiming to maintain feasibility of process conditions in the final solution. The optimization problem was solved also for a variety of levels of profits, in order to build the efficient frontier risk-profit.

The use of the ANN as a new objective function for the optimization problem presented some disadvantages, as the algorithm showed a relevant sensitivity to the starting points for search, in some cases resulting in the inability to find solutions within the constrained space. In regards of the results obtained, the portfolios built presented a tendency of increasing sugar production as expected values of profits were incremented, and it also showed a preference for the regulated market. This latter appears to be a manner to reduce CVaR, as this asset is risk-free. However, the choice was done at the expense of feasibility for process conditions, since there was a clear tendency to increase ethanol participation to maintain the profit levels, extrapolating the relation sugar-ethanol production. Thus, to obtain more accurate results, it is necessary to incorporate the nonlinear relations between process variables into the optimization problem, which may potentially increase its complexity.

In order to analyze the potential impacts of those changes in process conditions on the risk-profit results, a second approach of searching within a narrow grid built from the ANN was implemented, which in turn presented more robust results (Figure 2). For the portfolio solutions, quantities of sugar were also increased as expected profit grows. However, instead of increasing ethanol production to unfeasible levels, the portfolios showed an increment in the electricity commercialized to the free market, which increases profits while raises the overall risks. Consequently, levels of risks were higher in the grid search solutions. The changes in the risk levels in relation to the first approach evidence the need of considering process models to obtain more accurate results and assist investor's decisions. Further, in Figure 2, the comparison of frontiers shows that the investment option in bioenergy generation causes a gain in efficiency in the frontier of risk-profit. It is relevant to highlight that, although the second approach was more robust in finding feasible solutions, the grid search method is computationally demanding and may be not viable for large-dimension problems. Thus, an alternative to improve the nonlinear optimization approach for future studies is to combine other surrogate models to perform feasibility analysis within the search-space, as proposed by Ibrahim et al (2017).

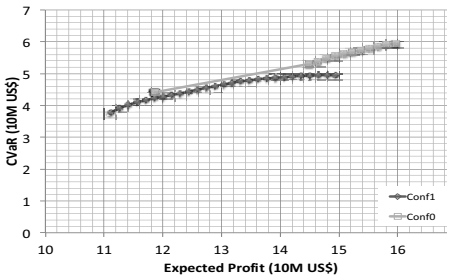


Figure 2 – Efficient Frontiers for Grid Search method (Source: Authors)

4. Conclusions

This paper aimed to discuss the use of artificial neural networks to assist product mix and investment decisions in Brazilian sugarcane mills. The proposed model combined Portfolio Optimization with surrogate models, through the use of ANNs, in order to assess the impact of bioenergy investments in the financial and process conditions. Two approaches were applied showing that nonlinear optimization of the ANN output function was not sufficiently robust to find feasible solutions with regards of process conditions. On the other hand, the grid-search within the results of the ANN presented important insights into the problem, evidencing that the investment in projects to increase energy recovery from waste may present a gain in efficiency in the trade-off risk-profit of bio-refineries. For future studies, a hybrid approach will be developed in order to incorporate feasibility analysis into the optimization framework.

Acknowledgments

This work was supported by FAPESP (São Paulo Research Foundation) under FAPESP-NERC grant 2015/50684-9 and FAPESP grant 2017/17530-3.

References

M.R. Barrosa, A.V. Salles, C.O. Ribeiro, 2016, Portfolio optimization through Kriging methods, Applied Economics, 48, 4894-4905.

G.A. Dantas, L.F.L. Legey, A. Mazzone, 2013, Energy from sugarcane bagasse in Brazil: An assessment of the productivity and cost of different technological routes, Renewable and Sustainable Energy Reviews, 21, 356-364.

E.F. Grisi, J.M. Yusta, R. Dufo-López, 2012, Opportunity costs for bioelectricity sales in Brazilian sucro-energetic industries, Applied Energy, 92, 860-867.

D. Ibrahim, M. Jobson, J. Li, G. Guillén-Gosálbez, 2017, Surrogate Models combined with a Support Vector Machine for the Optimized Design of a Crude Oil Distillation Unit using Genetic Algorithms, Comp. Aided Chem. Eng., 40, 481-486.

H. Markowitz, 1952, Portfolio selection. The Journal of Finance, 7, 77-91.

R.T. Rockafellar, S. Uryasev, 2000. Optimization of conditional value-at-risk. Journal of Risk, 2, 21-42.

G. Villarrubia, J.F. Paz, P. Chamoso, F. De la Pietra, 2018, Artificial neural networks used in optimization problems. Neurocomputing, 272, 10-16.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50157-9

A Projection-based, Data-Driven Method for High-Dimensional Black-Box Optimization

Ishan Bajaj, M.M. Faruque Hasan*

Artie McFerrin Department of Chemical Engineering, Texas A&M University, 3122 TAMU, College Station, TX 77843, USA.
hasan@tamu.edu

Abstract

A derivative-free optimization method based on projection of samples onto a univariate space is proposed to solve multi-dimensional box-constrained black-box problems. This space is defined using sum of the decision variables. When the objective function values are projected onto this space, a multivalued function is generated. It is shown that a univariate function exists on this map that contains the global minima of the original function. A two-step strategy is proposed to identify the function. Sensitivity theorem is employed to predict the values of the function and a trust-region based algorithm is used to correct the predicted point. An algorithm is proposed that identifies the univariate function and then optimizes it such that it corresponds to the actual optima. The proposed algorithm is applied to a test suite of 393 test problems comprising of sets of convex nonsmooth and nonconvex smooth black-box problems and compared with other competitive model-based solvers.

Keywords: Derivative-free Optimization, Univariate Projection, Black-box Optimization, Simulation-based Optimization

1. Introduction

In this paper, we consider the following box-constrained black-box problem:

$$\min\ f(x) \qquad (1)$$

$$s.t.\ x \in [x^L, x^U]$$

where $x \in R^n, f(x) : R^n \to R,$ is a continuous black-box function, whose values are obtained by performing computer simulations. For a variety of reasons, the derivatives are unavailable or unreliable. There are numerous applications of such problems in engineering, physics, medicine, business, and operations research. The methods to solve these problems are referred as derivative-free optimization (DFO). DFO techniques have been applied to many chemical engineering problems including the recent works on process and flowsheet optimization (e.g., Caballero and Grossmann, 2008; Boukouvala and Ierapetritou, 2013). The growing number of potential applications has led to development of several algorithms for solving box-constrained or unconstrained black-box problems. These algorithms can be classified into two major categories, namely direct and model-based (Rios and Sahindis (2013)). Direct algorithms use the function values directly to make progress, and model-based algorithms develop inexpensive surrogate models using simulation data to determine the search direction. Many of the algorithms are currently also available as software packages including ORBIT (Wild et al., 2008), SNOBFIT (Huyer and Neumaier, 2008), BOBYQA (Powell,

2009), IMFIL (Gilmore and Kelley, 1995). Rios and Sahindis (2013) extensively compared the performance of 27 solvers on a set of 502 problems.

In this paper, we propose an algorithm based on projection of samples onto a univariate space defined by a linear combination of the decision variables. This results in a point-to-set map/multivalued function since multiple function values exist corresponding to a single variable value. We identify a univariate function representing the lower envelope of this projection such that its minima is also the minima of the original *n*-dimensional function. We identify the points on this function by applying combination of sensitivity theorem (prediction step) and a trust-region method (correction step). An algorithm is also developed that optimizes the lower envelope such that it converges to the actual optima. We compare the proposed algorithm with BOBYQA, ORBIT, SNOBFIT, IMFIL on 161 convex nonsmooth and 232 nonconvex smooth problems.

2. Main Idea

If an auxiliary variable is added to (1), it is equivalent to projecting the original function in the *t* space. This results in the following optimization problem:

$$\begin{aligned} &\min_{x,t} f(x) \\ &s.t. \frac{\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i^L}{\sum_{i=1}^{n} x_i^U - \sum_{i=1}^{n} x_i^L} = t \\ &\quad x \in [x^L, x^U], t \in [0,1] \end{aligned} \tag{2}$$

While, there are many choices of defining auxiliary variable, the one shown above is a simple choice. The auxiliary variable added is a mathematical artifact and is independent of the objective function. Assuming certain regularity conditions, it can be proved using KKT conditions that the optima of (1) and (2) are the same. Furthermore, we can write problem defined in (2) as follows:

$$\begin{aligned} &\min_{t} G(t) \\ &s.t.\ G(t) = \min_{x} f(x) \\ &\qquad s.t.\ \frac{\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i^L}{\sum_{i=1}^{n} x_i^U - \sum_{i=1}^{n} x_i^L} = t \\ &\qquad\quad x \in [x^L, x^U], t \in [0,1] \end{aligned} \tag{3}$$

Expressing problem as shown in (3) allows us to decompose the problem in two steps. The lower problem defines the problem of identifying the samples on the lower envelope, while the upper problem involves optimizing the lower envelope. Solving the lower problem given in (4) is a linearly constrained black-box problem and needs to be solved for different parameter values, t such as t^p to obtain $G(t^p)$, where t^p is the p^{th} element in the set, $Z_{k'} := \{t^1, \dots, t^Q\}$:

$$\begin{aligned} G(t) = \ &\min_{x} f(x) \\ &s.t.\ g_i^1 := x_i^L - x_i \le 0 && \forall i \in \{1, \dots, n\} \\ &\quad g_i^2 := x_i - x_i^U \le 0 && \forall i \in \{1, \dots, n\} \\ &\quad h_1 := \frac{\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i^L}{\sum_{i=1}^{n} x_i^U - \sum_{i=1}^{n} x_i^L} = t \end{aligned} \tag{4}$$

The conservation of the minima on projection onto a univariate space is illustrated using a 2-dimensional Branin function. Figure 1(a) give the surface plot of the function. It can be observed that the problem has 3 minima, all of which are also global minima. The global minimum value of the function is 0.397887 and is achieved at $\boldsymbol{x}^* = (-\boldsymbol{\pi}, \mathbf{12.275})$, $(\boldsymbol{\pi}, \mathbf{2.275})$, and (9.424, 2.475). Figure 1(b) illustrates the point-to-set map obtained by projecting the samples on the *t* space. It also shows the lower envelope obtained as a result of minimizing $\boldsymbol{f}(\mathbf{x})$ at all *t* values. Figure 1(c) shows that the three minima are conserved after projection.

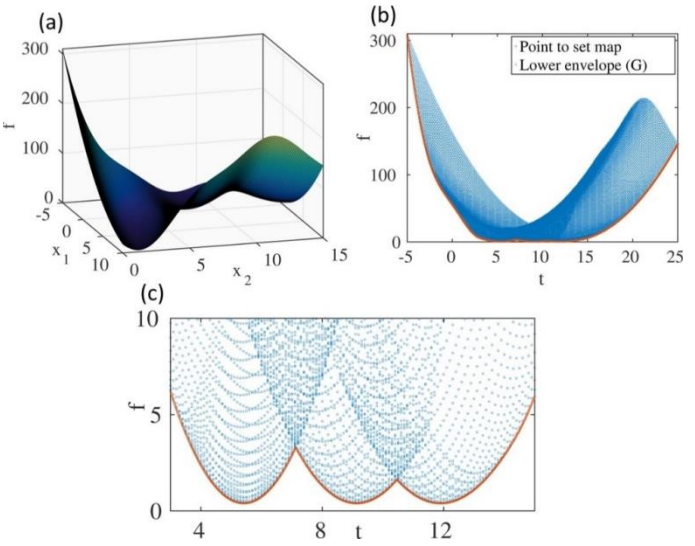


Figure 1. (a) *Branin Function*, (b) *Point-to-set map*, (c) *Lower envelope illustrating the minima.*

3. Properties of *G*(*t*)

It can be observed that (4) is a parametric problem in *t* and therefore, the results about convexity, concavity, continuity, differentiability can be borrowed from sensitivity literature (Fiacco, 1984) to illucidate key properties of $G(\mathrm{t})$.

Property 1. If $f(\mathrm{x})$ is continuous, then $G(\mathrm{t})$ is continuous.

Property 2. Assuming KKT conditions, LICQ hold at the optimal point of (4) at parameter value, t^p ($\bar{x}^p$), then *G*(*t*) is differentiable at t^p.

Property 3. If $f(\mathrm{x})$ is convex (concave), then $G(\mathrm{t})$ is convex (concave).

Property 4. Assume that KKT conditions, SOSC, LICQ and SCS hold true at $\bar{x}^p$ with associated multipliers $\bar{v}$ and $\bar{\lambda}$, then in the neighborhood of t^p, the optimal function *G*(*t*) is twice continuously differentiable.

Property 5. $G(t^L) = f(x^L)$, $G(t^U) = f(x^U)$.

3.1. Prediction of G(t)

Solving (4) from an arbitrary point is challenging. However, a good initial guess can greatly expedite the convergence. If the optima at t^p is known, we desire to obtain the optima at the next parameter value t^{p+1}. Fiacco (1984) proposed a technique to estimate the sensitivity of a local solution of a nonlinear program to small perturbation in the problem parameters. A linear approximation of Fiacco's sensitivity theorem is given as follows:

$$\begin{bmatrix} \Delta\hat{x}_p \\ \Delta\hat{v}_p^1 \\ \Delta\hat{v}_p^2 \\ \Delta\hat{\lambda}_p \end{bmatrix} = -\left(\widehat{M}\left(y(t^p)\right)\right)^{-1} N(t^p)(t^{p+1} - t^p) \tag{5}$$

where $\Delta\hat{x}_p = \hat{x}(t^{p+1}) - \bar{x}^p$, $\Delta\hat{v}_p^1 = \hat{v}^1(t^{p+1}) - \bar{v}^1(t^p)$, $\Delta\hat{v}_p^2 = \hat{v}^2(t^{p+1}) - \bar{v}^2(t^p)$, $\Delta\hat{\lambda}_p = \hat{\lambda}(t^{p+1}) - \bar{\lambda}(t^p)$, and let $\Delta\hat{y}_p = \left[\Delta\hat{x}_p, \Delta\hat{v}_p^1, \Delta\hat{v}_p^2, \Delta\hat{\lambda}_p\right]$.

For definition of $\widehat{M}$ and N, refer Fiacco (1984). $\widehat{M}$ requires the Hessian of the Lagrangian function (L). Since the optimization problem is black-box, the Hessian of the surrogate model is used to approximate the Hessian of the original Lagrangian function. Let the corresponding matrix be defined as $\widetilde{M}$ and the solution given by Fiacco's sensitivity theorem using $\widetilde{M}$ be denoted by $\Delta\tilde{y}_p = \left[\Delta\tilde{x}_p, \Delta\tilde{v}_p^1, \Delta\tilde{v}_p^2, \Delta\tilde{\lambda}_p\right]$. For convenience in notation, $\widehat{M}(y(t^p))$ and $\widetilde{M}(y(t^p))$ will be represented as $\widehat{M}^p$ and $\widetilde{M}^p$ respectively. We derive the bounds on the difference of the solution given by using exact Hessian and the solution obtained if the Hessian of the surrogate model is used.

Theorem 1. Assuming that the surrogate model is fully quadratic, the assumptions of Theorem 1 are satisfied at t^p, then

$$\left\|\Delta\tilde{y}_p\right\| - \left\|\Delta\hat{y}_p\right\| \leq \left((\kappa_{\widetilde{M}^p} - 1) + \frac{\kappa_{\widetilde{M}^p}\kappa_H\Delta}{\left\|\widetilde{M}^p\right\|} \right) \alpha \frac{\kappa_{\widehat{M}^p}\|N\|}{\|\widehat{M}^p\|} \tag{6}$$

Proof. Rewriting first-order sensitivity approximation compactly when the Hessian is known and taking the norm: $\Delta\hat{y}_p = -\left(\widehat{M}^p\right)^{-1} N^p\alpha \;\Rightarrow \|\Delta\hat{y}_p\| \leq \alpha\|\left(\widehat{M}^p\right)^{-1}\|\|N^p\|$. We also know that $\widetilde{M}^p\Delta\tilde{y}_p = \widehat{M}^p\Delta\hat{y}_p$. Taking norm, we get $\frac{\|\Delta\tilde{y}_p\|}{\|\Delta\hat{y}_p\|} \leq \|\left(\widetilde{M}^p\right)^{-1}\|\,\|\widehat{M}^p\|$. We can also show that $\frac{\|\Delta\tilde{y}_p\| - \|\Delta\hat{y}_p\|}{\|\Delta\hat{y}_p\|} \leq \frac{\|\kappa_{\widetilde{M}^p}\widehat{M}^p - \widetilde{M}^p\|}{\|\widetilde{M}^p\|}$, where $\kappa_{\widetilde{M}^p}$ refers to the condition number of $\widetilde{M}^p$. Now, aassuming Frobenius norm is used, the following can be verified:

$$\left\|\kappa_{\widetilde{M}^p}\widehat{M}^p - \widetilde{M}^p\right\|_F = (\kappa_{\widetilde{M}^p} - 1)\left\|\widehat{M}^p\right\|_F + \|\kappa_{\widetilde{M}^p}\nabla^2 L - \nabla^2 L^r\|_F - (\kappa_{\widetilde{M}^p} - 1)\|\nabla^2 L^r\|_F \tag{7}$$

$$\frac{\left\|\Delta\tilde{y}_p\right\| - \left\|\Delta\hat{y}_p\right\|}{\|\Delta\hat{y}_p\|} \leq (\kappa_{\widetilde{M}^p} - 1) + \kappa_{\widetilde{M}^p}\frac{\|\nabla^2 L - \nabla^2 L^r\|}{\|\widetilde{M}^p\|} \tag{8}$$

Assume that the model is fully quadratic, $\|\nabla^2 L - \nabla^2 L^r\| \leq \kappa_H\Delta$ and using Eq. (8), we finally get the bounds given in Eq. (6).

4. Algorithm

The overall algorithm is given in Figure 2. We know the optimal point at $t = t^L$ and therefore, the algorithm is started by setting t^l to t^L and a positive starting step length, α is selected by dividing the t-space into N_p parts. Basic sensitivity theorem (details provided in Section 3.1) acts as a prediction step by estimating the optima at t^{p+1} given the optima at t^p is known. Due to approximations involved, the predicted point, $\tilde{x}^{p+1}$ may not always correspond to the exact optima of (4) and therefore, the predicted points need to be corrected. This mechanism is referred as the correcting algorithm. We adapt the trust-region model-based method of Powell (2009) to handle linearly constrained

problems and use it as the correcting algorithm. The details of the algorithm are not provided here to stay within the page limit. We chose to use trust-region based method since we expect the prediction step to give a good initial guess and trust-region methods are known to converge rapidly if a good initial point is provided. Once an optimal point, $\bar{x}^{p+1}$ is obtained, the objective function value at $\bar{x}^{p+1}$ is compared to that at $\bar{x}^{p}$. If the objective function value at $\bar{x}^{p+1}$ is less than that at $\bar{x}^{p}$, a new step length α is chosen based on the number of evaluations taken by the correcting algorithm to solve (4) at parameter value t^{p+1}. If the number of function evaluations corresponding to the parameter value t^{p+1} is less than that taken at t^{p}, the step length is increased. If the objective function at t^{p+1} is more than that at t^{p}, this implies that $G(\mathrm{t})$ has started to increase and that optima (at least local) lies within $[t^{l}, t^{p+1}]$. A univariate surrogate model to approximate $G(t)$ is then constructed using samples in the interpolation set, Z_{kl}. $G(t)$ is then minimized, and the bounds on t are updated. This procedure is repeated until the bounds on t are smaller than a pre-specified tolerance.

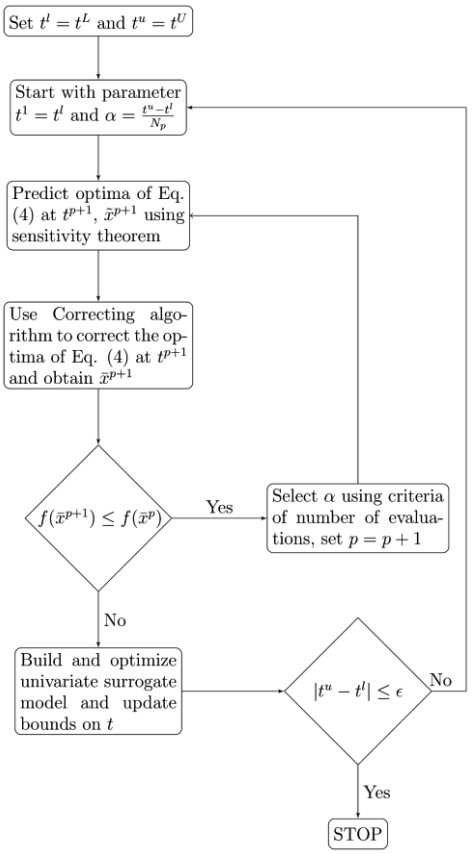


Figure 2. *Optimization framework.*

5. Numerical Results

The algorithm is applied to a test suite of 393 test problems from (Rios and Sahinidis (2013)) consisting of 161 convex nonsmooth and 232 nonconvex smooth problems. The performance of the proposed algorithm is compared to that of BOBYQA, SNOBFIT, IMFIL and ORBIT. All the solvers are allowed to perform 100,000 function evaluations. Each evaluation requires the same computation time for different solvers. Default parameters are used for all the solvers. A solver is said to have solved a problem if a point x^s is obtained such that $f(x^s) \leq \max(1.01 f(x^*), f(x^*) + 0.01)$, where x^* is the minima. The performance is first compared on a set of 232 nonconvex

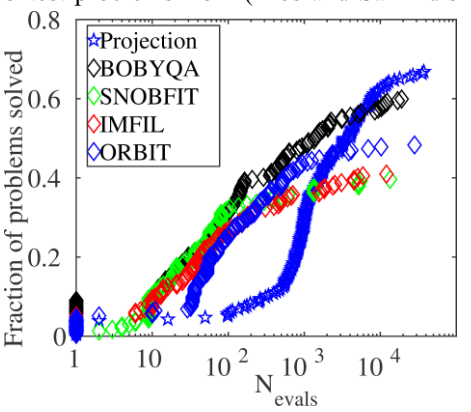


Figure 3. *Data profile illustrating the fraction of problems solved with number of evaluations for 232 nonconvex smooth problems.*

smooth problems. Figure 3 illustrates that the projection based approach is able to obtain solution within 1% of the global minima for 155 of the 232 problems, BOBYQA solves a total of 155 problems. ORBIT, SNOBFIT and IMFIL solves 112, 81 and 81 problems, respectively to global optimality. This illustrates that the proposed approach has a better mechanism to explore the global space more effectively. Figure 4 provides the data profile of different model-based solvers on a test suite of 161 convex nonsmooth problems. The proposed algorithm solves 49 of the 161 problems while ORBIT, BOBYQA, IMFIL and SNOBFIT solves 17, 15, 7 and 2 problems respectively.

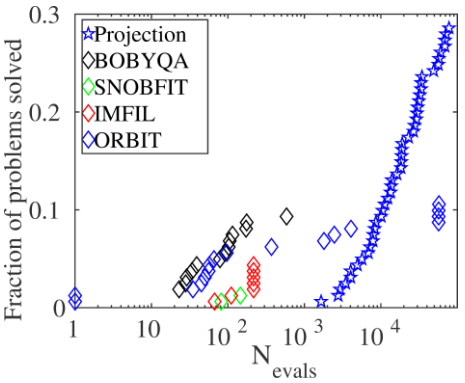


Figure 4. *Data profile illustrating the fraction of problems solved with number of evaluations for 161 convex nonsmooth problems.*

6. Conclusions

We proposed a method for optimizing box constrained black-box problems. The method is based on the concept of projecting multidimensional variables to a univariate space defined by linear combination of the original decision variables. A special function is identified in this space optimizing which is equivalent to optimizing the original problem. An algorithm is developed which identifies the points on the one-dimensional function and optimizes it. A two-step strategy is employed that identifies the function. The first step involves a prediction step based on sensitivity theorem while the second step is a correction step based on trust-region method. It is shown that the difference of the solution given by the sensitivity theorem with the hessian of the surrogate model and the exact Hessian is bounded. The performance of the algorithm is illustrated on 393 problems and compared with BOBYQA, SNOBFIT, ORBIT and IMFIL.

References

A. V. Fiacco, 1984, Introduction to sensitivity and stability analysis in nonlinear programming. Academic Press, New York.

P. Gilmore, and C.T. Kelley, 1995, An implicit filtering algorithm for optimization of functions with many local minima, SIAM Journal on Optimization, 5(2):269-285.

J.A. Caballero, and I.E. Grossmann, 2008, An algorithm for the use of surrogate models in modular flowsheet optimization, AIChE Journal, 54(10):2633-2650.

W. Huyer, and A. Neumaier, 2008, SNOBFIT-Stable Noisy Optimization by Branch and fit, ACM Transactions on Mathematical Software (TOMS), 35(2):9.

M.J.D. Powell, 2009, The BOBYQA algorithm for bound constrained optimization without derivatives. Cambridge NA Report NA2009/06, University of Cambridge, Cambridge.

F. Boukouvala, and M.G. Ierapetritou, 2013, Surrogate-based optimization of expensive flowsheet modeling for continuous pharmaceutical manufacturing, Journal of Pharmaceutical Innovation, 8(2):131-145.

L.M. Rios, and N.V. Sahinidis, 2013, Derivative-free optimization: a review of algorithms and comparison of software implementations, Journal of Global Optimization, 56(3):1247-1293.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50158-0

Efficient Surrogate Model Development: Impact of Sample Size and Underlying Model Dimensions

Sarah E. Davis, Selen Cremaschi, Mario R. Eden*

Department of Chemical Engineering, Auburn University, Auburn, AL 36849, USA
edenmar@auburn.edu

Abstract

This study compares eight surrogate-model construction approaches using computational experiments. The construction approaches considered include: Artificial Neural Networks (ANNs), Automated Learning of Algebraic Models using Optimization (ALAMO), Radial Basis Networks (RBNs), Extreme Learning Machines (ELMs), Gaussian Progress Regression (GPR), Random Forests (RFs), Support Vector Regression (SVR), and Multivariate Adaptive Regression Splines. Each approach is used to construct surrogate models for predicting the outputs of thirty-four test functions. The data used to construct surrogate models were generated using Latin Hypercube Sampling (LHS), Halton and Sobol sampling methods. We studied the impact of sampling method, sample size, and challenge function characteristics on the accuracy of surrogate model predictions. Our results revealed that for smaller sample sizes, on average, surrogate models trained using the data points generated by Sobol sequences provided the best estimation for all surrogate model construction approaches. As the sample size increased, the impact of sampling methods diminished. The comparison of surrogate-modeling construction approaches at large sample sizes revealed that surrogate models trained using ANN, ALAMO, and ELM yielded smaller root mean squared error and higher adjusted R-squared values than the models trained using the rest of the approaches.

Keywords: Surrogate Model, Process Design, Optimization

1. Introduction

Chemical engineering design problems have evolved in recent years from large scale bulk processes to targeted small batch approaches. As a result, computer aided design is reaching beyond the traditional solutions to classical chemical engineering problems. Take for example, the removal of CO_2 from flue gas. The computer aided molecular design framework, CAMD, can be used to identify the ideal ionic liquid for CO_2 removal (Davis et al., 2016). One method that can be used to reduce the computational burden is surrogate modeling. Surrogate models seek to understand input and output relationships that are either unknown or complex via reasonably accurate simpler functions. Many surrogate models have been defined and studied for different applications. Our previous study compared different surrogate modeling approaches in terms of their accuracies using root mean square error (RMSE) and maximum absolute error (MaAE) as the performance metrics. We evaluated the impact of function shape and number of inputs on the accuracies of surrogate model predictions with a sample size of 1,000 data points (Davis et al., 2017). When the results were considered in terms of RMSE, the models developed using Artificial Neural Networks (ANNs) consistently provided the lowest values for all numbers of inputs functions, while the models

developed using Automatic Learning of Algebraic Models for Optimization (ALAMO) were comparable to ANN models for functions with four and 10 inputs. When the performance of surrogate model types based on the shape of the challenge functions were considered, ANN models yielded the smallest RMSE for all function shapes. The current work extends these initial comparisons to understand the relationship between the accuracy of the surrogate model and the following characteristics: (1) sampling method; (2) sample size; (3) block-box model shape; and (4) number of inputs. Eight surrogate modeling approaches were used in the analysis. These approaches are: ANN, ALAMO, Extreme Learning Machines (ELM), Support Vector Regression (SVR), Radial Basis Networks (RBN), Gaussian Process Regression (GPR), Random Forests (RF), and Multivariate Adaptive Regression Splines (MARS). A brief overview of these surrogate modeling approaches is provided in Section 2. Training data was generated using thirty-five challenge functions with various surface shapes and number of inputs for each surrogate model. Some brief introductions of the challenge functions are given in Section 3. Each challenge function-surrogate model combination was trained using data generated using three different sampling methods, Latin Hypercube Sampling (LHS), Sobol Sequence and Halton Sequence. The sample size was varied from 50 to 12800, doubling each time. Section 4 provides the procedure for computational experiments. In addition to the surrogate model training and evaluation times, six performance metrics, MaAE, RMSE, Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), R-Squared, and R-Squared Adjusted were calculated for each challenge function and surrogate model combination based on the difference between the true values and the surrogate model predictions using a test set of 100,000 data points generated using the Sobol sampling method. The comparison of surrogate model training and evaluation times, and performance metrics for different sampling methods and surrogate model construction approaches are presented in Section 5. Concluding remarks are given in Section 6.

2. Surrogate Models Compared in This Study

ANNs are inspired by the brain. The neurons and the synapses that relay messages use signal strengths in an adaptive feedback loop. ANN was developed by relating the artificial neurons with their biological counterpart and synapses with weights. Weights and biases associated with artificial neurons are optimized by minimizing an error metric (generally mean-squared error) between ANN predictions and training data (Haykin, 2009). An approach which was developed to reduce complexity while maintaining accuracy is ALAMO. This surrogate model takes advantage of many different types of basis functions including: polynomial, multinomial, exponential, logarithmic, and trigonometric. ALAMO decides which basis functions to incorporate in the surrogate model along with the parameters of the basis functions incorporated using global optimization techniques (Cozad et al., 2014). ELMs is based on the same premise as ANN, but differs in the learning scheme. Instead of using a learning algorithm, ELM randomly assigns the weights. This difference allows the training problem to be solved as a linear equation system which provides a significantly faster solution (Huang et al., 2006). Another surrogate model that seeks to transform the data and relationships into a linear system is SVR. This application of support vector machines fits the data onto a linear plane where the data can be analyzed (Jin et al., 2001). RBN, while similar to ANN, differs due to its reliance on the radial spread of each RBF function in each dimension (Jin et al., 2001). GPR is a surrogate model that places all the data on continuous domain, such as time or space, and then uses a linear

combination of the inputs to generate the outputs (Mirbagheri, 2015). The RF model was developed based on the use of decision tress as a prediction method. Input values are processed through a set of decision trees which in turn use probability to estimate the outputs (Breiman, 2001). MARS is the final surrogate model considered. This method overfits the model using a linear summation of spline functions, then removes the additional terms until the model is optimized (Friedman, 1991).

3. Challenge Functions

Thirty-five challenge functions with defined shapes and numbers of inputs were utilized for analysis from the Virtual Library of Simulation Experiments (Surjanovic, 2013). Simon Fraser University created this resource to provide users with functions to evaluate new simulation and optimization methods. The optimization group provides 47 challenge functions divided by shape including: 16 multi-local minima shaped functions, seven bowl shaped functions, five plate shaped functions, four valley shaped functions, three ridges/drops shaped functions and 12 other differently shaped functions. Thirty-five functions with defined shapes were utilized for the analysis. Additionally, the challenge functions were compared based on the number of inputs.

4. Computational Experiments

The sampling methods that were utilized to generate training data from the challenge functions include LHS, Sobol Sequence and Halton Sequence. LHS is a stratified sampling technique, and it splits the range of each input variable into N intervals of equal probability, where N is the number of sample points. Then, each of the N partitions is sampled once, and randomly combined. Sobol and Halton sequences are both quasi-random low-discrepancy sequences, which seek to distribute samples uniformly across the input space. These methods were selected as they are shown to sample input space uniformly with limited sample sizes for functions up to ten dimensions (Dife and Diwekar, 2016). From each challenge function, input-output pairs were generated using the three sampling methods for nine sample sizes. Each surrogate model was trained using these data sets. Then, the performances of all surrogate models were evaluated using a data set of 100,000 input-output pairs that were generated according to the Sobol sampling technique. The error, the difference between the surrogate model prediction and the actual function output, is calculated for each of these 100,000 input-output pairs. The errors and the number of model parameters (when necessary) are then used to calculate performance metrics, which are MaAE, RMSE, Akaike Information Criteria, Bayesian Information Criteria, R-Squared, and R-Squared Adjusted (Mathworks, 2017) Additionally, the training time was recorded during the phase of the program where the model was trained, and evaluation time was recorded during the phase when the final outputs were generated. All computations were carried out on a HP Spectre X360 X64-based PC with 16 GB RAM using MATLAB 2017b.

5. Results and Discussion

5.1. Results Based on Sampling Method and Size

Figure 1 shows the change in average adjusted R-squared values with sample size. The error bars in Figure 1 correspond to standard deviations. At sample sizes of 1,600 or more, the surrogate models trained using any of the three sampling methods performed similarly. However, for smaller sample sizes, surrogate models trained using data generated by Sobol sequence, on average, outperformed models trained using data

generated by Halton sequence and LHS. Similar trends were observed with Akaike Information Criteria, and Bayesian Information Criteria.

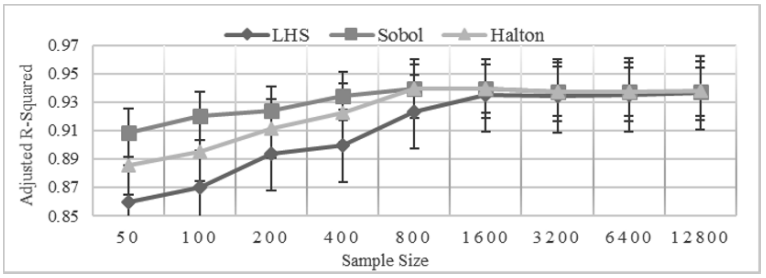


Figure 1: Adjusted R-Squared based on Sample Size

When considering MaAE and RMSE, similar trends were observed, therefore a plot of RMSE is given in Figure 2 to illustrate the trends. As the size of training data points is increased above 400, the RSME of surrogate models trained using any of the sampling methods was about the same. When the sample size was less than 400, all sampling methods error rates increased.

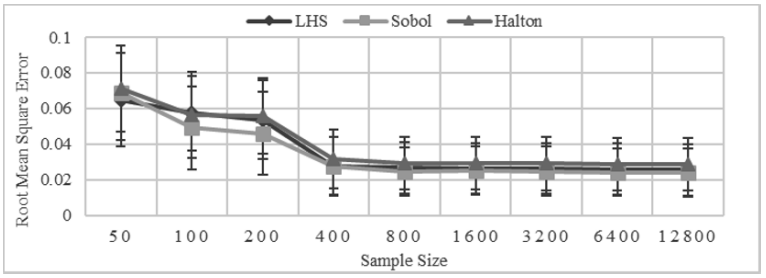


Figure 2: Root Mean Square Error based on Sample Size

Figure 3 shows how the training time changes, as sample size is increased by surrogate model type. In average, surrogate models constructed using ELM were trained the fastest, and ALAMO the slowest.

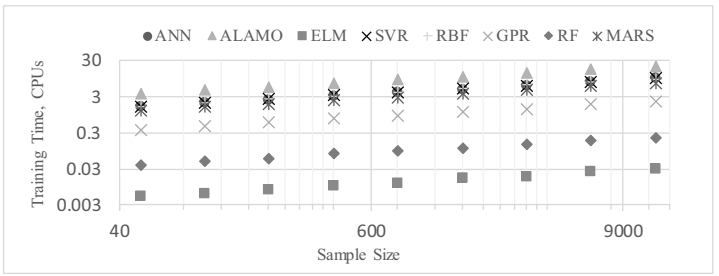


Figure 3: Training times for all Surrogate Models based on Sample Size

All training times increased exponentially with increasing sample sizes. The average surrogate model evaluation times were 0.006953, 0.009465, 0.005953, 0.006648, 0.006656, 0.006652, 0.006642, 0.006652 for surrogate models constructed using ANN, ALAMO, ELM, SVR, RBF, GPR, RF, MARS, respectively. ALAMO was shown to

have the longest evaluation with other models requiring about the same amount of evaluation time.

5.2. Results Based on Challenge Function Shape and Inputs

Two performance measures, MaAE and RMSE, were calculated for each challenge function and surrogate model combination. Both error calculations provide similar inferences, so RMSE was used to illustrate the results. Based on the results previously discussed, the surrogate models were compared using a sample size of 12,800 produced with the Sobol sampling method.

In terms of RMSE (Figure 4) based on the number of input functions for a sample of size of 12,800 using the Sobol Sampling Sequence, the models developed using ANNs consistently provided the lowest values for all functions, where the RMSEs of the models developed using ALAMO were comparable to the RMSEs of ANN models for functions with 2, 3, and 5 inputs. Overall, these comparisons suggest that the models developed using ANNs and ALAMO have comparable accuracy for functions with different dimensions.

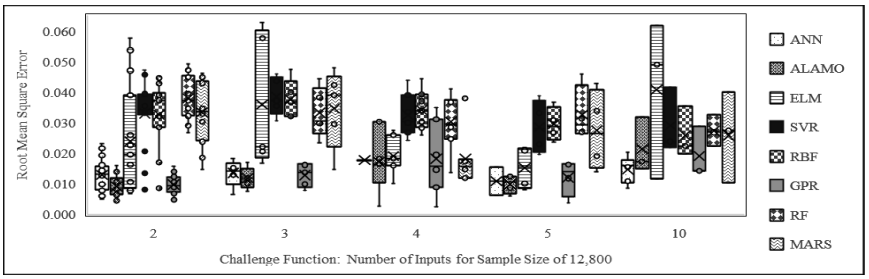


Figure 4: Challenge Function – Number of Inputs for Sample Size of 12,800

When the performance of surrogate model types are compared based on the shape of the challenge functions for a sample of size of 12,800 using the Sobol Sampling Sequence as shown in Figure 5, the error for ALAMO and GPR was almost equal for Multilocal Minima, Plate, Bowl, and Valley Shapes. ANN provided the lowest error for Ridges and Drops, however provided nearly as good results as ALAMO on other shapes.

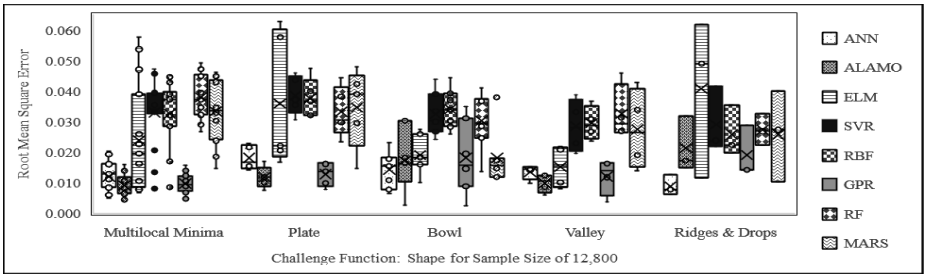


Figure 5: Challenge Function – Shape for Sample Size 12,800

6. Conclusions

This study compared the relationship between the surrogate model and the system variables including: (1) sampling method; (2) sample size; (3) surface shape of the input function; and (4) complexity of the input function. In addition to the training and evaluation time, six performance measures: Maximum Absolute Error (MaAE), Root

Mean Square Error (RMSE), Akaike Information Criteria, Bayesian Information Criteria, R-Squared, and R-Squared Adjusted were analyzed to provide guidance when selecting the optimum surrogate model. All methods and models provided a better fit at larger sample sizes. Sobol provided the best estimation over the full range of sample sizes. Overall, ANN, ALAMO, and ELM models all provided better estimation than the other models tested.

References

Akkoyunlu, A., 2010, "A Neural Network-Based Approach for the Prediction of Urban SO_2 Concentrations in the Istanbul Metropolitan Area." International Journal of Environment and Pollution 40.1, 310-321

Beck, J., Friedrich, D., Brandani, S., Fraga, E.S. 2015, "Multi-Objective Optimzation using Surrogate Models for the Design of VPSA Systems", Computers and Chemical Engineering 82, 318-329

Breiman, L., 2001, "Random Forests", Machine Learning 45.1, 5-32

Cozad, A., Sahinidis, N., Miller, D., 2014, "Learning Surrogate Models for Simulation Based Optimization". AIChE Journal 60.0, 2211-2227

Davis, S.E,, Hada, S., Herring, R.H., Eden, M.R., 2015, "Multivairate Characterization, Modeling, and design of ionic liquid molecules. Computers and Chemical Engineering 81, 310-322.

Davis, S.E., Cremaschi, S., & Eden, M.R. (2017). Efficient Surrogate Model Development: Optimum Model Form Based on Input Function Characteristics. In Computer Aided Chemical Engineering (Vol. 40, pp. 457-462). Elsevier.

Dife, N., Diwekar, U.M., 2016, "Novel Sampling Technique for High Dimensional Stochastic Optimization/Stochastic Programming Proglem", AIChE Annual Meeting, Nov. 13-18, San Francisco, CA.

Friedman, J.H., 1991, "Multivariate Adaptive Regression Splines", Annals of Statistics 19.1, 1-67.

Haykin, S., 2009, "Neural Networks and Learning Machines", Prentice Hall PTR.

Huang, G., Zhu, Y., Siew, C., 2006, "Extreme Learning Machine: Theory and Applications". Neaurocomputing 70.1-3, 489-501

Jin, R., Chen, W., Simpson, T., 2001, "Comparative Studies of Metamodeling Techniques under Multiple Modeling Criteria", Structural and Multidisciplinary Optimization 23.1, 1-13

Mathworks. (2016). Statistic and Machine Learning Toolbox: User's Guide (r2017b). Retrieved October 17, 2017. from www.mathworks.com/help/pdf_doc/gads/gads_tb.pdf

Mirbagheri, S., 2015, "Evaluation and Prediction of Membrane Fouling in a Submerged Membrae Bioreactor with Simulaneous Upward and Downward Aeration using Artifical Neural Network-Genetic Algorithm. Process Safety and Environmental Protection 96, 111-124.

Surjanovic, S., Bingham, D., 2013, "Virtual Library of Simulation Experiments: Test Functions and Datasets. http://www.sfu.ca/~ssurjano.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50159-2

Space mapping based derivative-free optimization framework for supply chain optimization

Atharv Bhosekar, Marianthi Ierapetritou*

Department of Chemical and Biochemical Engineering, Rutgers,The state University of New Jersey, 98 Brett Road, Piscataway 08901, United States,
marianth@soemail.rutgers.edu

Abstract

Derivative-free optimization (DFO) addresses the problem of optimizing over simulations where a closed form of the objective function is not available. Developments in the theory of DFO algorithms have made them useful for many practical applications. However, most of the existing DFO algorithms do not show satisfactory performance on high-dimensional problems. One contributor to this difficulty is the accuracy of the surrogate models used to guide the search for the optimum. Space mapping approach exploits a simplified simulation, which is a physical surrogate of the original problem at hand. As the simplified simulation is computationally efficient, a larger number of samples can be collected for guiding the search. This work aims to assess the potential of space mapping for derivative-free optimization, understand the difficulties and display its performance on a supply chain simulation optimization problem for identifying optimal inventory allocation. Computational results illustrate the potential of this approach.

Keywords: derivative-free optimization, physical surrogate models, multi-fidelity analysis, space mapping, supply chain simulation.

1. Introduction

Simulations are widely used by researchers in various engineering and scientific disciplines due to their ability to capture realistic details of the phenomenon under study. The problem of optimizing over simulations, however, is a non-trivial problem because of high computational requirement and absence of an algebraic model. Derivative-free optimization (DFO) algorithms can be used to address this problem (Conn et al., 2009). There has been a significant research in the area of DFO algorithms in past two decades and these algorithms have achieved significant maturity to be able to solve many problems. A lot of research has been done on strategies to build surrogate models, their accuracy for modelling, and essential considerations for their use in derivative-free optimization (Bhosekar & Ierapetritou, 2018). However, sampling requirement for accurate predictions increases with increase in problem size. Thus, surrogate models fail to provide reliable approximations of the original problem. This poses a difficulty for DFO algorithms (Rios & Sahinidis, 2013).

Another approach for obtaining an approximation of a simulation is to use a simplified but computationally efficient simulation to gather information. These models are referred as physical surrogate models. The idea was introduced and developed under the terminology of space mapping (Bandler et al., 1994) and multi-fidelity analysis (Fernández-Godino et al., 2016). Because of the generality of this idea, it has found

successful applications in various disciplines such as electrical engineering, design simulations, traffic optimization (Osorio & Bierlaire, 2010) and fluid dynamics (Banda, 2011).

Supply chain modelling is one of the areas where simulations have found increasing use. With the help of simulations, one can understand the effects of complex interactions between various supply chain entities on the overall performance of supply chain. Agent based simulations is an efficient way to represent such a complex network of entities (Sahay & Ierapetritou, 2013). Decision making considering such interactions is an interesting but often a non-trivial problem (Sahay & Ierapetritou, 2016).

In this work, an insight into space mapping approaches is provided for their application in optimizing a supply chain simulation. The problem of optimal allocation of inventory for minimizing the total cost of a multi-enterprise supply chain is considered. Simulations are designed to capture the details of the interaction between these enterprises in the form of complex auction mechanisms. Under the influence of complex interactions within a supply chain network, this problem is shown to display a non-convex and non-smooth behavior. Two output mapping strategies are compared for this problem of identifying optimal inventory while minimizing the total cost. Comparisons with existing DFO algorithms show that space mapping shows promising performance. The rest of the article is organized as follows. Section 2 provides an insight into the space mapping methods. In Section 3, specifics of the supply chain simulation under consideration are discussed. Section 4 discusses computational results, potential difficulties and positives of using space mapping. Finally, conclusions and summary are provided in Section 5.

2. Space mapping

Consider the problem of minimizing a simulation output where variable bounds are the only constraints present as shown in the problem (1).

$$\begin{aligned} &\min f(x) \\ &\text{s.t. } \mathrm{LB}_i \le x_i \le \mathrm{UB}_i \; \forall i \in \{1, \dots, n\} \end{aligned} \tag{1}$$

where LB_i and UB_i represent lower and upper bounds on the variable x_i; n is the number of variables; $f: R^n \to R$ is the simulation output which is assumed to be expensive and therefore closed form of f is assumed to be unavailable. Space mapping makes use of an approximation $c: R^n \to R$. This approximation is assumed to be computationally efficient. For the rest of the paper, simulation f is referred to as high-fidelity simulation (HFS) whereas simulation c is low-fidelity simulation (LFS). There are multiple ways in which one can establish a relationship between c and f. One way to do this is by correcting LFS response with the help of additive or multiplicative corrections as given in Equations (2), and (3), respectively.

$$\hat{f}(x) = \rho(x)c(x) \tag{2}$$

$$\hat{f}(x) = c(x) + \delta(x) \tag{3}$$

Corrections $\rho(x)$ and $\delta(x)$ are used to enhance the existing accuracy of the analytical surrogate model with the help of finite HFS evaluations. Another way is to have

comprehensive corrections where multiplicative and additive corrections included together as shown in Eq. (4).

$$\hat{f}(x) = \rho(x)c(x) + \delta(x) \tag{4}$$

In comprehensive corrections, $\rho(x)$ is often considered a constant. Another comprehensive surrogate modelling strategy is co-kriging (Qian et al., 2006). Kriging surrogate can be described as given in Equation (5).

$$\hat{f}(x) = \sum_{i=1}^{m} \beta_j f_j(x) + \epsilon(x) \tag{5}$$

where $f_j(x)$ are m independent basis functions that define the trend of mean prediction at location x; β_j are unknown parameters; $\epsilon(x)$ is a random error at location x that is normally distributed with zero mean. Co-kriging makes use of LFS to build a base Kriging model and relates it to HFS again using a Kriging surrogate as given in Eq. (6).

$$Z_e(x) = \rho Z_c(x) + Z_d(x) \tag{6}$$

where Z_e is the output from HFS and Z_c is the output from LFS; $Z_d(x)$ represents the discrepancy at sample location x. All the above-mentioned mappings consider a relationship between outputs. It is also possible to have a mapping between inputs as given in Eq. (7).

$$x_{LFS} = p(x_{HFS}) \tag{7}$$

where, $p: R^n \rightarrow R^n$ is the input space mapping. This mapping is established iteratively using local secant approximations until mapped HFS and LFS responses align within a certain tolerance limit (Bandler et al., 1995). Since trust-region framework allows any kind of surrogate model to be used, space mapping optimization has been combined with trust-region optimization approach (Osorio & Bierlaire, 2010), (Vicente, 2003). This provides additional properties that help in proving convergence of the optimization approach.

3. Supply chain simulation

In this work, an agent based multi-enterprise supply chain simulation is considered (Sahay & Ierapetritou, 2016). As opposed to centralized supply chains considered in traditional problems, this simulation allows one to handle decentralized decision making by individual entities. Each entity in the supply chain is a separate agent that performs its functions based on a set of policies. These policies determine the behavior of each agent based on its environment. Five different types of agents that are supplier, production site, warehouse, market, and auctioneer are considered. Each entity in this network has its own role. For example, warehouses store products and send shipments to markets according to the demand. Similarly, production sites handle production and ship products to warehouses. Auctioneer agent handles interactions between entities belonging to separate enterprises in the form of auctions. There is a certain cost associated with each agent and the task it performs. Market incurs a backorder cost if the demand is not fulfilled. Warehouses have an inventory holding cost and transport cost to the markets. Production sites have production cost, transport cost as well as holding cost for raw materials and products. Addition of all these costs is the total cost of the supply chain.

There are several differences in the previously proposed simulation (Sahay & Ierapetritou, 2016) and the simulation used in this work. Production sites give preference to the product that is available in the least quantity in the product inventory. Warehouse inventory replenishment is done with continuous review using a policy including reorder level and reorder quantity. The reorder amount of a warehouse depends on the capacity of the warehouse. The demand is considered to be lost with a certain penalty if supply chain fails to satisfy the demand. As production sites and warehouses belong to the same enterprise, production sites aim to satisfy maximum demand from the warehouse. Time delay between an order from warehouse and delivery from production sites is incorporated.

4. Computational results and discussion

4.1. Problem description

A supply chain network consisting of one supplier, three production sites, two warehouses, and two markets is considered in this study. Each market or retailer is considered to belong to a separate enterprise whereas raw material suppliers, production sites, and warehouses are all considered to belong to the same enterprise. The supply chain is considered to produce two products. The problem of optimal allocation of warehouse capacity is considered. Warehouse capacity decides the amount of product inventory maintained at the warehouse. A large amount of inventory may lead to large inventory holding cost. On the other hand, small warehouses make it difficult for the supply chain to satisfy the market demand. This trade-off becomes more complex under the influence of auctions between warehouses and markets. For understanding this, the supply chain simulation is studied over 100 time periods. As the computational cost of the simulation is directly proportional to the number of time periods, a higher number of time periods results in computationally expensive simulations. An LFS for this simulation is obtained by limiting the planning horizon to 10 periods. Demand for each period is approximated as average demand at the corresponding time period if the horizon is divided into 10 subintervals. The aim, in this case, is to minimize the total cost reported by the simulation with respect to warehouse capacity. For the problem under consideration, the total cost can be plotted to study the response. Figure 1 and Figure 2 display the response surface of supply chain total cost with respect to warehouse capacities for both LFS and HFS respectively Close observation of these responses indicates non-smooth behavior.

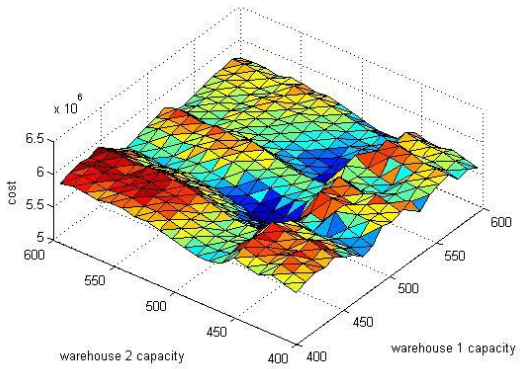


Figure 1:LFS total cost variation with warehouse capacity

4.2. Methods and results

In the computational comparisons, two mappings are considered. First mapping considers a multiplicative correction as given by Equation (2) whereas second mapping considers comprehensive

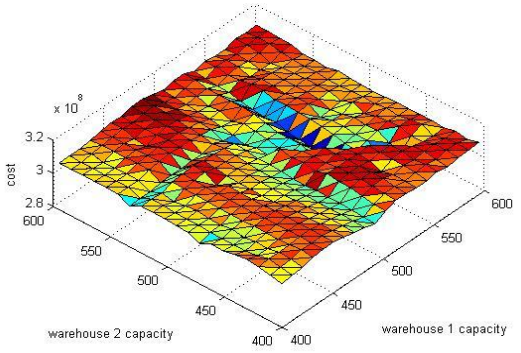


Figure 2:HFS total cost variation with warehouse capacity

correction as given in Equation (4) with a constant $\rho(x)$. These strategies are considered because of their simplicity and because of number of previous applications. Multiplicative correction particularly helps in this case because the total cost of a supply chain can be different by an order of magnitude because the HFS is run for more number of time periods. The ratio and discrepancy of HFS and LFS is approximated using a Kriging model built starting with 50 sample points from the HFS. For optimizing the corrected surrogate model and for optimizing the non-smooth simulation in this work, a discrete gradient method (Bagirov et al., 2008) is used.

Total cost optimization problem is formulated subject to bound constraints on warehouse capacities. Lower and upper bounds are set to be 400 units and 600 units respectively. The best solution found after optimizing the mapping surrogate is 464 units for one warehouse and 466 units for the other warehouse. The same optimum is achieved using both mappings. Total cost at the optimum is \$307110590. Further, HFS is optimized using pattern search and CMA-ES algorithms as these algorithms are not deterministic and make no assumptions on the smoothness of the objective function. These results are tabulated in Table 1.

Table 1: Comparison of space mapping surrogates with existing DFO solvers

Solver name	CMA-ES	Pattern search	Mapping 1	Mapping 2
Total cost	\$306587517	\$304629215	\$307110590	\$307110590

It is clear that pattern search performs better than other algorithms while CMA-ES shows a comparable performance with mapping methodology. Several observations can be made from these results. Relatively worse performance of the mapped surrogates can be attributed to the accuracy of Kriging surrogate used for mapping discrepancy or ratio. As the accuracy of Kriging surrogate depends on the quality of sample set, a model built from a stationary sample set may lead to inaccuracies near optimum. Identical optimal solution achieved indicates that both strategies are equally accurate and lead to a similar response with an identical optimal solution. Finally, as sampling requirements for pattern search and CMA-ES increases steeply with problem size. Therefore, a good simplification with the help of space mapping strategy can be preferred for derivative-free optimization in higher dimensions.

5. Conclusions

In this work, a simplified simulation is used as a surrogate model and its use in optimization of supply chain simulation is studied. Several strategies used by researchers in various disciplines for mapping LFS and HFS are discussed. Two of

these strategies are applied for supply chain simulation optimization. A multi-enterprise supply chain simulation is considered where enterprises interact through multi-attribute auctions. Warehouse capacity allocation problem for minimizing the total cost of the supply chain simulation is considered. This was observed to form a nonconvex as well as non-smooth optimization problem. Results showed that both mapping strategies yielded an identical optimal solution. With the fact that sampling requirements for existing DFO algorithms as well as that for surrogate models grow with problem size, space mapping is a promising approach and needs to be studied for higher dimensional problems.

Acknowledgements

M.G.I. acknowledges financial support from NSF under grant CBET 1159244.

References

Bagirov, A. M., Karasözen, B., & Sezer, M. (2008). Discrete gradient method: Derivative-free method for nonsmooth optimization. *Journal of Optimization Theory and Applications*, *137*(2), 317–334. http://doi.org/10.1007/s10957-007-9335-5

Banda. (2011). Space mapping in gas networks. *Optimal Control Applications and Methods*, *32*(3), 298–312. http://doi.org/10.1002/oca

Bandler, J. W., Biernacki, R. M., Chen, S. H., Grobelny, P. A., & Hemmers, R. H. (1994). Space Mapping Technique for Electromagnetic Optimization. *IEEE Transactions on Microwave Theory and Techniques*, *42*(12), 2536–2544. http://doi.org/10.1109/22.339794

Bandler, J. W., Biernacki, R. M., Chen, S. H., Hemmers, R. H., & Madsen, K. (1995). Electromagnetic Optimization Exploiting Aggressive Space Mapping. *IEEE Transactions on Microwave Theory and Techniques*, *43*(12), 2874–2882.

Bhosekar, A., & Ierapetritou, M. (2018). Advances in surrogate based modeling, feasibility analysis, and optimization: A review. *Computers and Chemical Engineering*, *108*, 250–267. http://doi.org/10.1016/j.compchemeng.2017.09.017

Conn, A., Scheinberg, K., & Vicente, L. (2009). *Introduction to Derivative-Free Optimization*. Society for Industrial and Applied Mathematics. http://doi.org/10.1137/1.9780898718768

Fernández-Godino, M. G., Park, C., Kim, N.-H., & Haftka, R. T. (2016). Review of multi-fidelity models. http://doi.org/10.1016/j.jcp.2015.01.034

Osorio, C., & Bierlaire, M. (2010). A Simulation-Based Optimization Framework for Urban Traffic Control. *Report TRANSP-OR*, *101203*, 1–29. http://doi.org/10.1287/opre.2013.1226

Qian, Z., Seepersad, C. C., Joseph, V. R., Allen, J. K., & Wu, C. F. J. (2006). Building Surrogate Models Based on Detailed and Approximate Simulations. *Journal of Mechanical Design*, *128*(4), 668. http://doi.org/10.1115/1.2179459

Rios, L. M., & Sahinidis, N. V. (2013). Derivative-free optimization: A review of algorithms and comparison of software implementations. *Journal of Global Optimization*, *56*(3), 1247–1293. http://doi.org/10.1007/s10898-012-9951-y

Sahay, N., & Ierapetritou, M. (2013). Supply chain management using an optimization driven simulation approach. *AIChE Journal*, *59*(12), 4612–4626. http://doi.org/10.1002/aic.14226

Sahay, N., & Ierapetritou, M. (2016). Multienterprise supply chain: Simulation and optimization. *AIChE Journal*, *62*(9), 3392–3403. http://doi.org/10.1002/aic.15399

Vicente, I. S. N. (2003). Space Mapping : Models , Sensitivities , and Trust-Regions Methods, (1994), 159–175.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50160-9

A Comparison between NLP and MILP Formulations of Organic Rankine Cycle Systems for Optimization

Vathna Am[a*], Jonathan Currie[b], David I. Wilson[a]

[a]*School of Engineering, Computer & Mathematical Sciences, Auckland University of Technology, Auckland 1010, New Zealand*
[b]*Guidance, Navigation & Control, Rocket Lab Limited, Auckland 2022, New Zealand*
vam@aut.ac.nz

Abstract

In chemical process optimization, often one type of formulation can be more advantageous due to financial constraints, compatibility issues with modelling software, the availability of optimization solvers and/or the importance of establishing global optimality. Historically, the availability and computational efficiency of linear program (LP) solvers, or even mixed integer linear program (MILP) solvers, were significantly better than some nonlinear program (NLP) solvers, which was a key incentive to reformulate nonlinear optimization problems to LPs or MILPs. However, with the current advanced NLP solvers and the recent improvements in optimization modelling platforms and techniques, it might now be redundant for practitioners to reformulate certain real-world NLP problems into integer/linear programming problem. This paper shows that with the right integration between the NLP formulation and the state-of-the-art solvers, it is possible to achieve significantly better optimization performance for real-world organic Rankine cycle systems.

Keywords: Organic Rankine Cycle, Nonlinear Programming, Mixed Integer Linear Programming, Optimization

1. Introduction

Often practitioners invest a considerable amount of time reformulating their optimization problems into LP and MILP models in order to achieve convergence or better optimization performance, as shown in (Borrelli et al., 2006; Yue et al., 2013). However, with the recent advancement in nonlinear optimization algorithms, we have observed significant performance improvements in NLP solvers in handling complex chemical process optimization problems. This is in addition to the improvement in optimization modelling platforms and computational performance of desktop hardware. Consequently, with the current availability of both advanced modelling platforms and optimization nonlinear solvers, it is debatable if there are sufficient compelling advantages for certain real-world chemical process systems to be reformulated into LPs or MILPs.

In this paper, we test this hypothesis on organic Rankine cycle (ORC) systems, with different topologies consisting of real-world thermodynamic processes. The novelty of this research is to illustrate a systematic equation-oriented modelling approach on a spectrum of ORC systems, namely as NLP and MILP formulations, where exact derivatives can be provided. This differs from the common black-box optimization

approach implemented by Clarke & McLeskey (2015) and Ghasemi et al. (2013). In addition, the paper also aims to compare the computational advantages and disadvantages of both formulations and to indicate if it is advisable to formulate linear models for real-world ORC systems. This way, optimization problems that concern with improving the economics or the performance of the ORC systems, such as those carried out by Astolfi et al. (2014) and Budisulistyo & Krumdieck (2015), can be solved efficiently.

2. Optimization Strategy and Modelling Platform Overview

2.1. Nonlinear Programming Strategies

One of the main difficulties in solving NLP problems is dealing with inequality constraints because they can either act as equality constraints or just strictly satisfies the inequality condition at the optimal solution. Most derivative-based NLP solvers, like FMINCON and IPOPT, are Newton-type algorithms and address this issue using the active-set sequential quadratic programming (SQP) strategy or the interior point (IP) method. Unless optimizing convex problems, these solvers can only assure local optimality and are dependent on a good starting point and accurate derivatives. Whereas, spatial branch-and-bound based global NLP solvers, like SCIP and BARON, can guarantee the "true" global solution but are more computationally expensive as the nonlinear problem are successively divided and solved until it converges to the global solution. Modern NLP solvers are now significantly more reliable and are widely adopted due to the development of the quasi-Newton method and globalization strategies, and the more recent filter method (Biegler & Grossmann, 2004), that helped to overcome the difficulties in dealing with indefinite Hessians and poor starting points.

2.2. Mixed Integer Linear Programming Strategies

MILP methods are derived from the branch and bound algorithm, where the MILP problems are "branched" into subproblems, called nodes, and solved as relaxed LP problems at imposed bounds. The MILP problem will continue to recursively split into nodes until the difference between upper bound (integer solution) and lower bound (LP solution) are within a tolerance. Since the 1990s, there have been significant advancements in MILP methods, such as pre-processing techniques and cutting planes, which have greatly improve the performance of MILP solvers (Balas & Fischetti, 2017).

2.3. Modelling Platforms

Solving process optimization problems with fast and reliable optimization solvers is only one aspect of achieving efficient optimization capabilities. The modelling platform also plays a crucial role in optimization. With a better understanding of the advantages and disadvantages of different optimization algorithms, practitioners can now focus more on formulating well-posed models that are tailored to specific solvers and improve the overall performance. Some simulation packages, such as GAMS, AMPL and AIMMS, have direct interfaces to the solvers and can provide exact derivatives using automatic differentiation strategies that can significantly improve the overall optimization performance (Biegler, 2014).

3. Organic Rankine Cycle Models

For this research, four different ORC configurations were modelled, namely the Basaran ORC (using refrigerant R227ea) (Basaran & Ozgener, 2013), DOE Pilot Plant (DiPippo, 2016, p. 550-561), USGeo Plant (DiPippo, 2016, p.562-571) and Magmamax Binary

Power Plant (DiPippo, 2016, p.516-524). For brevity, the construction of both the NLP and MILP formulations will only be briefly described below for the DOE Pilot Plant.

3.1. Nonlinear Programming Formulation

The DOE plant is a dual-pressure ORC system that consists of a high pressure (HP) and a low pressure (LP) ORC utilizing the same geofluid. The NLP model was formulated in equation-oriented mode, which amounts to deriving a set of deterministic algebraic equations that describes the process of the plant and approximating the output characteristic of the unit operations using regression and thermodynamic analysis. The system was assumed a steady state and a steady-flow process; where changes in kinetic and potential energy were neglected, and losses induced by friction were neglected. The geofluid was treated as pure water and the pressure drops across heat exchangers and pipelines were neglected. A similar procedure is discussed by Am et al. (2017) that results in the following algebraic NLP formulation

$$\min \quad f(\boldsymbol{x}), \quad \text{s.t.} \quad \boldsymbol{h}(\boldsymbol{x}) = 0\,, \boldsymbol{g}(\boldsymbol{x}) \;\leq\; 0, \boldsymbol{x} \;\in\; \mathbb{R} \tag{1}$$

where $f(\boldsymbol{x})$ is the objective function, $\boldsymbol{h}(\boldsymbol{x}) = 0$ is a set of equations that represent the system's performance (mass and energy balances, and operational constraints), and $\boldsymbol{g}(\boldsymbol{x}) \leq\; 0$ describes primarily the operational limits or specifications of the DOE plant. As such, the thermodynamic process of the pump and the turbine, namely the input/output power, were regressed with respect to pressure, P, and/or enthalpy, h, using polynomial functions to allow for the calculation of exact derivatives. As a result, the LP turbine output power can be expressed as

$$\dot{W}_T = -0.286P^2 - 5.33 \times 10^{-5}h^2 + 0.00948Ph + 3.5P + 0.0335h - 41.9 \tag{2}$$

In addition, some of the operational constraints/limits were also regressed if they vary with respect to pressure and/or temperature. For instance, the LP turbine inlet enthalpy must be greater or equal to the saturated vapour enthalpy value can be expressed as

$$h_T^{in} \geq\; 0.014P^3 \;+\; -0.697P^2 \;+\; 15.2P \;+\; 545 \tag{3}$$

The geofluid inlet and outlet properties were kept constant as specified in (DiPippo, 2016), as well as the isentropic efficiencies, mass fraction, and the cooling water and pump inlet/outlet temperatures and pressures. Whereas, the HP and LP turbine inlet pressures and temperatures were allowed to vary between 15 to 26.31 bar and 388.71 to 413.71 K, and 6 to 14 bar and 355.37 to 380.37 K, respectively. Since the focus is on the solver performance and optimization formulations, and not the plant performance, variation of ambient conditions and power output were not considered.

3.2. Mixed Integer Linear Programming Formulation

The MILP model was formulated using piecewise linear approximations on the NLP model described above. For all the nonlinear unit operation approximations and operational constraints, shown in (2) and (3), these were replaced with piecewise linear approximations by changing the order of regression curves and specifying fixed breakpoints. For the isentropic turbine work approximations, the triangle method (D'Ambrosio et al., 2010) was used in order to obtain a better piecewise linear fit. Subsequently, all the mass and energy balance bilinear terms, such as $\dot{m}_i h_i$, can be linearly approximated using the method discussed in Section 7.7 of (Bisschop, 2016). If the objective function is a fractional term, the Charnes & Cooper transformation in conjunction with the Glover's linearization scheme (Yue et al., 2013) can be implemented to acquire the following MILP formulation

$$\min \ \mathbf{c}^{\top}\mathbf{x} + \mathbf{d}^{\top}\mathbf{y}, \quad \text{s.t. } \mathbf{Ax} + \mathbf{Ey} = \mathbf{b}, \mathbf{Fx} + \mathbf{Gy} \leq \mathbf{h}, \mathbf{x} \in \mathbb{R}, \mathbf{y} \in \{0,1\} \qquad (4)$$

where $\mathbf{c}^{\top}\mathbf{x} + \mathbf{d}^{\top}\mathbf{y}$ is the objective function, $\mathbf{Ax} + \mathbf{Ey} = \mathbf{b}$ and $\mathbf{Fx} + \mathbf{Gy} \leq \mathbf{h}$ describe the system's performance, the operational limits or specifications of the ORC system, and the piecewise fit constraints.

4. Model Optimization

The ORC models were constructed using MATLAB R2017a and were optimized using a Windows 3.1 GHz Intel Core i5 laptop. In order to efficiently optimize the ORC models, the OPTI Toolbox v2.27 (J. Currie, 2017) was used due to its compatibility with the nonlinear regression and the piecewise linear approximation functions that were developed in MATLAB. The solvers were selected based on their availability for academic use and their compatibility with OPTI and MATLAB. Derivative-free solvers were not considered as they are not suitable for constrained problems (Biegler, 2014) and do not fully benefit from our algebraic model structure. The optimality tolerance was set to 1×10^{-7} for all solvers and the objective is to maximize the following ratio

$$f(\boldsymbol{x}) = \dot{W}_{\text{net}}/(\dot{m}_{\text{cw}} + \textstyle\sum_{i=1}^{n} \dot{m}_{\text{wf}_i}) \qquad (5)$$

where $\dot{W}_{\text{net}}$ is the total turbine work minus total pump work, $\dot{m}_{\text{wf}}$ is the working fluid mass flow and $\dot{m}_{\text{cw}}$ is the cooling water mass flow. The thermodynamic properties of the fluids were obtained from the REFPROP database (Lemmon et al., 2013).

5. Results and Discussion

Table 1 shows both the comparison between the NLP and the MILP formulations, and the performance between the different solvers. As expected, there are considerably more variables and constraints in the MILP formulations due to the piecewise fit approximations. Interestingly, the results showed that the NLP solvers are significantly faster than their MILP counterparts. All the NLP solvers managed to converge to a feasible solution under 13s. Whereas the MILP solvers took between ~1s and ~2200s to converge and, in a few cases, did not find a feasible solution for some of the ORC plants, namely with CBC and GLPK. The fastest MILP solvers were GUROBI and CPLEX, ranging from 0.79s to 106.72s, but they are still, overall, slower than the NLP solvers.

One factor that could have contributed to the poor performance of the MILP formulations is the number of binary variables. There is a possibility that it can run into time limitation with a large number of binary variables, especially if there is a large difference between the feasible solution and the optimal relaxed LP solution (Biegler & Grossmann, 2004). While the number of binary variables can be reduced by decreasing the number of breakpoints of the piecewise approximation, it might not be sensible as it could render the MILP model infeasible or give an inaccurate representation of the ORC system. In addition, unlike NLP solvers, supplying a good initial guess to MILP solvers does not always improve the performance because all the MILP solvers in this research can automatically handle this internally. This was verified by obtaining similar CPU times as shown in Table 1 for most of the solvers when the optimal solution was substituted back to the problem. Consequently, having a large number of variables would require more effort from the solver to finding a feasible starting point and likely leads longer CPU time. For the NLP solvers, they benefited more from the algebraic model structure; therefore, their short computational time can be attributed to the availability of accurate derivatives

and bypassing the use of finite difference. Furthermore, from analysing the Jacobian matrices of the NLP formulations, they are significantly sparse and have the density values of 47.92% (BAS), 17.93% (DOE), 11.87% (USG), and 17.16% (MMM). Therefore, given that the models are algebraic, the solver can exploit this sparsity information to achieve a faster computational time (J. D. Currie, 2014). Moreover, the inequality constraints present a major difficulty for NLP problems due to the complementarity conditions (Biegler & Grossmann, 2004), and they are much harder to solve than equality constraints. Consequently, given that the number of inequality constraints in the NLP models is relatively small, i.e. 3 for BAS, 7 for DOE, 11 for USG, and 9 for MMM, these NLP models might be inexpensive to optimize.

Table 1: Comparison results between the NLP and the MILP formulations, where ORC Plant BAS, DOE, USG, and MMM are Basaran ORC, DOE Pilot Plant, USGeo Plant and Magmamax Binary Power Plant, respectively. The objective function value (top) and the average CPU time (bottom [seconds]) are given for each solver. Please refer to (J. Currie, 2017) for the solvers' references.

	NLP				**MILP**			
ORC Plant	BAS	DOE	USG	MMM	BAS	DOE	USG	MMM
Variables								
Continuous	6	17	25	19	343	862	906	842
Binary	0	0	0	0	161	402	410	394
Constraints								
Linear	0	3	11	4	929	2345	2677	2420
Quadratic	3	7	9	9	0	0	0	0
Nonlinear	4	10	10	9	0	0	0	0
Solver								
BARON[C] (v17.10.13)	-2.71	-5.02	-5.32	-6.55	-2.73	-5.04	-5.32	-6.55
	0.61	0.49	0.55	1.89	8.18	226.67	311.08	193.51
SCIP[NC/C] (v4.0.0)	-2.71	-5.02	-5.32	-6.55	-2.73	-5.04	-5.32	-6.55
	12.77	2.80	1.87	4.16	3.54	24.76	43.31	33.74
FMINCON[C] (v7.6)	-2.71	-5.02	-5.32	-6.55	-	-	-	-
	0.27	0.30	0.42	0.63				
IPOPT[NC] (v3.12.7)	-2.71	-5.02	-5.32	-6.55	-	-	-	-
	0.86	1.15	1.05	0.52				
NLOPT[1,NC] (v2.4.2)	-2.71	-5.02	-5.32	-6.55	-	-	-	-
	0.07	0.56	0.49	0.35				
INTLINPROG[C] (v7.6)	-	-	-	-	-2.73	-5.04	-5.32	-6.55
					3.37	94.48	2265.69	419.38
GUROBI[NC/C] (v7.5.2)	-	-	-	-	-2.73	-5.04	-5.32	-6.55
					0.89	5.00	13.08	8.07
CPLEX[NC/C] (v12.8)	-	-	-	-	-2.73	-5.04	-5.32	-6.55
					0.79	6.22	12.42	106.72
CBC[NC] (v2.9.8)	-	-	-	-	-2.73	-5.04	-5.32	NSF
					6.81	122.19	372.08	3599.63
GLPK[NC] (v4.48)	-	-	-	-	-2.73	-5.04	NSF	NSF
					1.05	37.99	2761.31	3600.00

[1]Algorithm: LD_SLSQP

C: Commercial solver; NC: Non-commercial solver; NSF: no solution found.

6. Conclusions

This paper compared the NLP and the MILP formulations of various real-world ORC systems and the optimization performance of both formulations. The results showed that the optimization performance of the NLP models, and their respective solvers, performed considerably better than the MILP models. This could be attributable to the availability of exact derivatives, especially with the second derivatives, and utilizing state-of-the-art NLP solvers. In addition, the NLP solvers were able to take advantage of the algebraic

model structure and the sparsity information to efficiently solve the optimization problem. Evidently, with the right integration between the solver and the NLP formulation, such as providing analytical derivatives, can lead to better optimization performance than the MILP formulation. Therefore, for chemical processes that are similar to an ORC system, it is conceivably sensible to initially use the NLP formulation if it is applicable, especially if there is a small number of inequality constraints or no highly nonlinear terms. However, since some MILP solvers do not require any initial values, it is possible to take advantage of the MILP formulation to acquire a good initial guess for the NLP model to further improve the performance or when the initial guess is hard to obtain.

References

Am, V., Currie, J., & Wilson, D. I. (2017). A systematic approach to modeling organic Rankine cycle systems for global optimization. In *2017 6th International Symposium on Advanced Control of Industrial Processes (AdCONIP)* (pp. 487–492).

Astolfi, M., Romano, M. C., Bombarda, P., & Macchi, E. (2014). Binary ORC (Organic Rankine Cycles) power plants for the exploitation of medium-low temperature geothermal sources - Part A: Thermodynamic optimization. *Energy*, *66*, 435–446.

Balas, E., & Fischetti, M. (2017). Chapter 5: Integer (Linear) Optimization. In *Advances and Trends in Optimization with Engineering Applications* (pp. 49–63).

Basaran, A., & Ozgener, L. (2013). Investigation of the effect of different refrigerants on performances of binary geothermal power plants. *Energy Conversion and Management*, *76*, 483–498.

Biegler, L. T. (2014). Recent Advances in Chemical Process Optimization. *Chemie Ingenieur Technik*, *86*(7), 943–952.

Biegler, L. T., & Grossmann, I. E. (2004). Retrospective on optimization. *Computers and Chemical Engineering*, *28*(8), 1169–1192.

Bisschop, J. (2016). Integer Linear Programming Tricks. In *AIMMS: Optimization Modeling* (pp. 75–85). AIMMS B.V.

Borrelli, F., Subramanian, D., Raghunathan, A. U., & Biegler, L. T. (2006). MILP and NLP Techniques for centralized trajectory planning of multiple unmanned air vehicles. *2006 American Control Conference*, 5763–5768.

Budisulistyo, D., & Krumdieck, S. (2015). Thermodynamic and economic analysis for the pre-feasibility study of a binary geothermal power plant. *Energy Conversion and Management*, *103*, 639–649.

Clarke, J., & McLeskey, J. T. (2015). Multi-objective particle swarm optimization of binary geothermal power plants. *Applied Energy*, *138*, 302–314.

Currie, J. (2017). OPTI Toolbox. Retrieved October 25, 2017, from https://www.inverseproblem.co.nz/OPTI/

Currie, J. D. (2014). *Practical Applications of Industrial Optimization : From High-Speed Embedded Controllers to Large Discrete Utility Systems* (Doctoral thesis). Auckland University of Technology, Auckland, New Zealand.

D'Ambrosio, C., Lodi, A., & Martello, S. (2010). Piecewise linear approximation of functions of two variables in MILP models. *Operations Research Letters*, *38*(1), 39–46.

DiPippo, R. (2016). *Geothermal Power Plants Principles, Applications, Case Studies and Environmental Impact* (4th ed.). Waltham, MA: Joe Hayton: Elsevier.

Ghasemi, H., Paci, M., Tizzanini, A., & Mitsos, A. (2013). Modeling and optimization of a binary geothermal power plant. *Energy*, *50*(1), 412–428.

Lemmon, E. W., Huber, M. L., & McLinden, M. O. (2013). NIST Standard Reference Database 23: Reference Fluid Thermodynamic and Transport Properties-REFPROP, Version 9.1, National Institute of Standards and Technology.

Yue, D., Guillén-Gosálbez, G., & You, F. (2013). Global optimization of large-scale mixed-integer linear fractional programming problems: A reformulation-linearization method and process scheduling applications. *AIChE Journal*, *59*(11), 4255–4272.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50161-0

Role of analytics within the energy, water and food nexus – An Alfalfa case study

Haile Woldesellasse, Rajesh Govindan, Tareq Al-Ansari*

Division of Sustainable Devlopment, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar
talansari@hbku.edu.qa

Abstract

One of the biggest challenges today, is the need to feed a growing population. Considering climate change and the effect it may have on agriculture systems, the sustainable intensification of food production is a necessity. The objective of the research presented in this paper is to address the issues related to the large-scale stressors that can present a barrier to the current efforts for sustainable intensification of food production in Qatar. It relies on remote sensing imagery as the primary dataset owing to the ability of the satellite sensors to cheaply gather data with good areal coverage and spatial resolution, and with added benefit of repeated surveys from the satellite overpasses over a given area. The steps of the proposed methodology developed in this research include: (a) image processing for training and prediction using neural networks; (b) extraction of atmospheric and surface features that can directly affect greenhouse operations, e.g. those related to land and water resources, and energy requirements; (c) formulation of a mixed integer non-linear programming (MINLP) framework using surrogate modelling for the maximization of the crop productivity objective; (d) development of GIS database augmented with recommender systems. The results obtained so far are promising, with root mean squared percentage error of 11.93 for the crop water demand prediction, and it is envisaged that the proposed development of a smart geospatial framework with predictive capabilities, EWF optimization, and augmented GIS will facilitate for informed policy development aimed at the success of the food security programs in Qatar.

Keywords: EWF nexus, food security, neural network, MINLP, Optimization

1. Introduction

The global population is expected to increase to 9 billion by 2050 (DESA 2010), thus increasing the demand for energy, water and food (EWF) resources to meet the growing demand for products and services. For instance, food production will need to increase by 50 – 100 % (Royal Society, 2009). In addition, uncertainties with respect to climate change and rapid resource utilization necessitates the need to develop integrated resource models. The complex relationship between climate and agriculture systems requires special attention (Grasty 1999). Temperature changes, shifting of rainfall patterns and increase of the global CO_2 levels are the most obvious consequences of climate change which will inevitably impact global food systems. Perhaps, the most vulnerable regions that may suffer as a result of climate change are those characterised by arid and semi-arid climates. In fact, crop yields vary from season to season in these regions which are dry and lower in latitude primarily due to changes in climate which are difficult to predict (Grasty 1999). As such, the importance of integrated modelling and the utilization of a

EWF nexus approach for assessments on the environment and/or resource utilization is especially important for countries such as the State of Qatar. Considering that, the State of Qatar is embarking on large scale food security initiatives, it is important to manage future resource requirements.

(Ali et al. 2017) used feasible generalized least square (FGLS) and heteroskedasticity and autocorrelation (HAC) error techniques to analyse the impacts of the extreme climate phenomena on the agricultural products of several crops, such as rice wheat, sugarcane and maize in Pakistan. The stressors considered include the maximum and minimum temperature, rainfall, humidity and sunshine. The study concluded that wheat production is significantly affected by the maximum temperature, whereas the effect of the minimum temperature is visible in all crops. (Fraiture et al. 2007) suggested that the wheat production in South Asia will decline by about 50% by 2050 which corresponds to 7% of the global crop production. With current trends of global warming, it is suggested that agricultural production in developing countries can decrease by 10% to 25%.

Considering, the integrated manner in which EWF resources are utilized, it is necessary to understand the system wide affect from climate change and how that may contribute to food insecurity. Al-Ansari et al. (2015; 2016; 2017) developed a EWF nexus, a web of interactions that link the EWF sectors, model which was applied to a hypothetical food security scenario in the State of Qatar. The study evaluated through a nexus approach means to which impacts on the environment can be mitigated when increasing domestic food production. However, the effect of climate change is not considered in their study. In a comprehensive study utilizing the EWF nexus for drought prevention, it was found that climate change is the main driver for reduction in agriculture productivity (Zhang et al. 2017). In their study, the authors used an integrated optimized EWF nexus modelling approach as a basis for the development of an effective agricultural drought management system. This was achieved by combining systematically a real-time irrigation management and drought monitoring by using a geospatial and agricultural data obtained from a remotely measuring satellite, meteorological stations. The study found that the crop yield during a drought year can be reduced by 50% when compared to a wet year.

The objective of this study is to develop a model which is capable of predicting crop water demand for an Alfalfa case study in Qatar when challenged with external stressors. The cultivation of Alfalfa is a fundamental component of Qatar's food security aspirations as it progressively expands its domestic livestock systems. A feedforward neural network is used to predict the effect of large scale stressors on the crop water demand. The application of neural network is very promising in a complex system like agriculture, which constitutes of huge amount of data coming from different components. To incorporate a spatial information of the stressors, a satellite data is used and a non-linear optimization technique is applied to allocate the optimum water required by the production fields. To the best knowledge of the authors, this is the first study done for this region that focuses on the application of machine learning methodologies for monitoring the effect of large scale stressors on crop productivity on a national scale.

2. Data Description

For this study, a predictive model for the water demand of alfalfa crop (m^3/ha) within the State of Qatar is developed. The schematic diagram process is shown in figure 1. The model is challenged by four potential large-scale stressors which include; humidity (%), temperature (C), soil moisture (g/cm^3) and evapotranspiration (mm). Except for soil

moisture, all the other variables are obtained from some meteorological stations in Doha. However, the data for soil moisture was obtained from AMSR-E satellite. The spatial resolution of the soil moisture is 25km by 25km. Furthermore, a monthly timescale is adopted for the analysis and the time range covers from 2002 to 2006. The water used for irrigation is assumed to come from a treated sewage effluent (TSE). The TSE plants found in Doha and which are currently in operation, are shown in GIS map below in figure 2. Moreover, the existing Alfalfa fields, potential areas for growing alfalfa crop and preferable transportation means for the distribution of water are also showed in the map. Previously, both ground water and TSE plants were using for irrigation, but groundwater is neglected from consideration because of depleting ground water reserves and increasing salinity of its water.

3. Methodology

The methodology consists of two parts in which the model development procedure is illustrated on figure 1. The first part involves a forecasting model using a Neural Network to make a prediction of crop water demand. A neural network is a computational model works in an analogous way to the way information is processed in a brain by biological neurons. Once the crop water demand is known, then a Non-linear Programming technique is used to make cost-efficient allocations of water resources. The type of neural network considered in the study is a simple feedforward neural network with three layers i.e. the input, hidden and output layer. A linear transfer function is used for the input layer, and a sigmoidal activation function in the hidden and output layer. The data observed is divided into training and validation groups. The model is first trained using training data and then validated using the portion of the remained data. Back-propagation method is used here, which is the most common technique for training ANN model. Once the parameters of the model are optimized and a least error is obtained in the validation processes, then the model is used for forecasting the crop water demand for one year ahead.

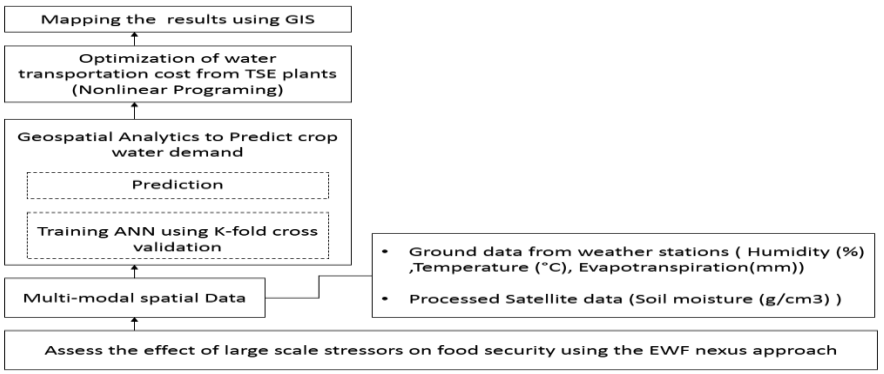


Figure 1. Schematic Diagram of the process

The limitations on the duration of the data available had its own challenge in training the model. To overcome this limitation, a k-fold cross validation was used, where the samples are randomly subdivided into k equal sized parts. The model was trained and analysed using the k-1 subsamples, and a single subsample was left for validation. In this study, k is taken to be 3 due to small data size. The forecasted unit crop water demand (m^3/ha) obtained are used for calculating the amount of water required by the crop production

fields. The areas of the production fields are obtained from the GIS map in figure 2. The irrigation water from the TSE plants are then allocated based on the cost of transportation using Non-Linear Programming. The optimally allocated water demand is finally reported using GIS software. The energy consumption is included in the water utilization and transportation, thereby making the study a holistic approach to the EWF nexus.

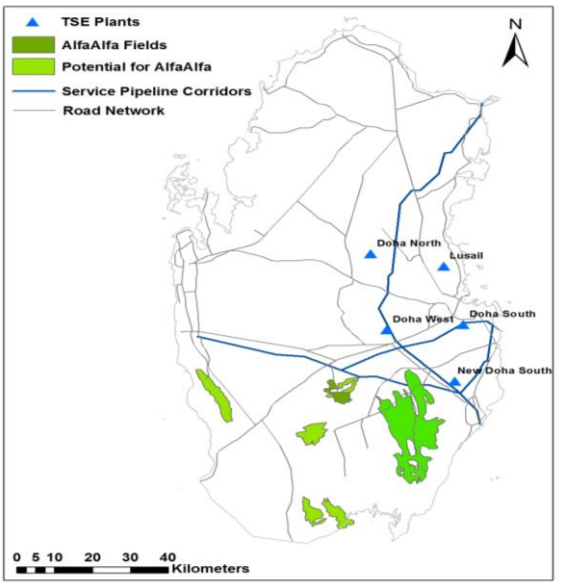


Figure 2. GIS Map of TSE sources and sinks, and transport network options

4. Results and Discussions

In order to demonstrate the objectives of this study two production fields and two nearby TSE plants are considered where the total water demands for field 1 (f1) and field 2 (f2) are 4.25 $x10^6$ m^3 and 5.46$x10^6$ m^3 respectively. The model developed as part of this research is able to capture the variation of the crop water demand during the hot and cold seasons, where the output of the optimized feedforward neural network illustrates agreement with the observed data as illustrated in figure 3(a) below. The accuracy of the output is validated using a correlation coefficient (0.983) and error measurement indices, root mean squared percentage error (11.93). After validation, a scenario is assumed for the one-year forecasting. The temperature observations are multiplied by random numbers generated between 1 and 1.1, by taking the assumption that the temperature will vary up to 10% above the previous observations. Comparison of forecasted year with previous years demonstrates that the crop water demand will increase with the maximum periods from April to August as illustrated in figure 3(b).

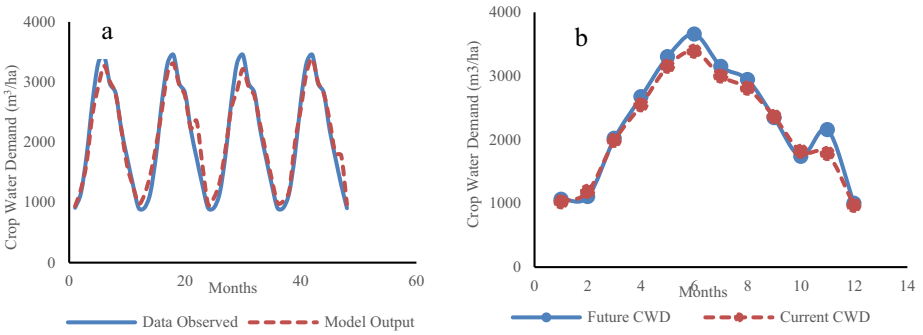


Figure 3: (a) Comparison of Model Output with Observation, (b) Comparison of Future and Current CWD.

The output of the optimization is illustrated in Figure 4. The fraction of water delivered from the TSE plant i to the field j is indicated by qij. It consists of four fractions of water supply distributed from the two sources to f1 and f2, q11, q12, q21 and q22, and optimum location (x,y) for storage water received from TSE plants prior to distribution to f1 and f2.The forecasted unit crop water demand (m^3/ha) is considered uniform throughout all alfalfa fields in Qatar because of the similar regional climatic conditions and the total water demand per production field is calculated using the product of field area and unit crop water demand. The objective of the nonlinear optimization is to minimize the total cost of the network which is a factor of the type of water distribution. The TSE plants which are currently in close proximity to these fields are Doha West and Doha South, with capacities of 7.37×10^6 m^3 and 7.20×10^6 m^3 respectively.

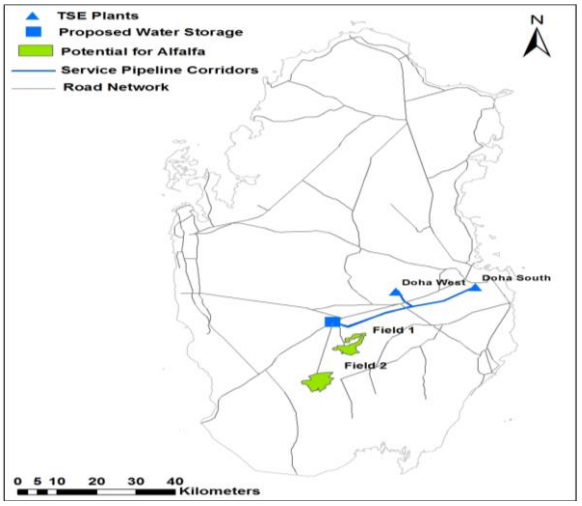


Figure 4. Final output of optimisation.

Water is transported from the TSE plants to the fields through two means: (1) pipeline where the transportation cost is \$0.00061 per m^3.km (Zhou and Tol 2005), (2) and local roads with the price of transportation is assumed to be 10% larger than the cost of pipeline. Based on the analysis, the optimum location of the storage is determined as illustrated on figure 4. Moreover, the results demonstrate that the two TSE plants should distribute 62% of their water capacity to f1, whereas the remaining amount is distributed to f2. The water supplied to f1 is more than the water demand of f1 and thus has an excess of 4.86 $x10^6$ m^3, whereas f2 receives the exact amount of its water demand, therefore remains with no excess water each month. The excess water allocated to f1 can be locally stored and used for next month or through peer to peer water trading system which can be introduced between neighboring fields.

5. Conclusions

In conclusion, ANN is found to be helpful tool in predicting the crop water demand. The non-parametric nature of ANN and its easy adaptation to different data formats give it superiority over traditional modelling tools. In this study, it was found that the impact of drought varies with the growing season of the crop, where the highest impact is observed between Mid- April and Early-June. A number of assumptions is considered in the study, because of coarse resolution of the soil moisture and limited point measurement of temperature available, but the framework is ready enough to work to its capacity when a data with a longer period record and good resolution is obtained. Future work will include: strengthening the model with data to overcome any assumptions considered; model expansion to accommodate other stressors of crop productivity, such as sunshine; and finally, the application of sensitivity analysis to monitor the reaction of the crop productivity for the change of the stressors.

6. References

Al-Ansari, T., Korre, A., Nie, Z. and Shah, N., 2015. Development of a life cycle assessment tool for the assessment of food production systems within the energy, water and food nexus. Sustainable Production and Consumption, 2, pp.52-66.

Al-Ansari, T., Korre, A., Nie, Z., & Shah, N. (2016). Integration of Biomass Gasification and CO_2 Capture in the LCA Model for the Energy, Water and Food Nexus. Computer Aided Chemical Engineering 26th European Symposium on Computer Aided Process Engineering, 2085-2090.

Al-Ansari, T., Korre, A., Nie, Z. and Shah, N.(2017). Integration of Greenhouse Gas Control technologies within the Energy, Water and Food Nexus to enhance the environmental Ali, S., Y. Liu, M. Ishaq, T. Shah, A. Ilyas & I. U. Din (2017) Climate Change and Its Impact on the Yield of Major Food Crops: Evidence from Pakistan. *Foods,* 6**,** 39.

DESA, U. 2010. United Nations, Department of Economic and Social Affairs, Population Division: world urbanization prospects, the 2009 revision: highlights. UN publications, New York, h ttp://esa. un. org/unpd/wup/Documents/WUP2009 Highlights Final. pdf (last access: 23 October 2012).

Fraiture, C. d., V. Smakhtin, D. Bossio, P. McCornick, C. Hoanh, A. Noble, D. Molden, F. Gichuki, M. Giordano & M. Finlayson (2007) Facing climate change by securing water for food, livelihoods and ecosystems.

Grasty, S. (1999) Agriculture and climate change. *Published in TDRI Quarterly Review,* 14**,** 12-16.

Zhang, J., P. E. Campana, T. Yao, Y. Zhang, A. Lundblad, F. Melton & J. Yan (2017) The water-food-energy nexus optimization approach to combat agricultural drought: a case study in the United States. *Applied Energy*.

Zhou, Y. & R. S. Tol (2005) Evaluating the costs of desalination and water transport. *Water resources research,* 41.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50162-2

Integrated techno-economic optimization for the design and operations of energy, water and food nexus systems constrained as non-cooperative games

Sarah Namany, Tareq Al-Ansari*, Rajesh Govindan

Division of Sustainable Devlopment, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar
talansari@hbku.edu.qa

Abstract

The rise in global population and the consumption of energy, water, and food (EWF) resources continues to increase exponentially, placing enormous stresses on the three resource sectors. Therefore, assessing the inter-linkages between EWF resources is a very crucial step to achieve sustainable resource utilization, through increased economic efficiency and enhanced environmental performance. Such integrated assessment is often challenged by the difficulty in presenting a robust formulation which considers the necessary details of embedded systems, the availability of relevant models and data, as well as the representation of their uncertainties, based on which the design and operations of connected supply chains are ideally optimized. The objective of the research presented in this paper is to address the key challenges of the EWF nexus through the development of a techno-economic framework under non-cooperative settings. The methodology developed assumes a bottom-up approach in which the self-interested stakeholders, i.e. the players, are represented based on their own decision variables and objective functions at various levels, i.e. resources and environment systems, engineering systems, and their integration and operations. For this purpose, a combination of optimization models based on linear programming, stochastic programming, and game theoretic approach, represented by a Stackelberg competition, were developed. A case study is set in the State of Qatar with the objective of enhancing food security using hypothetical scenarios. The results obtained demonstrates that interesting interactions between systems, albeit competitive, can potentially result in the achievement of desired objectives under properly regulated markets, and is therefore envisaged as a promising contribution towards sustainable policy development and nexus governance.

Keywords: EWF nexus, food security, stochastic optimization, Stackelberg

1. Introduction

Human activities are driven by a combination of energy, water and food systems. The fact is, water is used in food cultivation and power plants' cooling necessities for electricity generation. Energy is essential for water desalination and treatments as well as for agricultural facilities. In addition, crops produced in agriculture can be converted into energy in the form of biofuels (Garcia and You, 2016). These inherent inter-dependencies between the resources introduces the concept of the energy water food (EWF) nexus. Factors such as exponential population growth, climate change and globalization are considered stressors to EWF systems. As such, to ascertain true resource consumption in

the delivery of products or services, and ensure continuous and sustainable resource utilization, it is necessary to optimize the EWF nexus inter-linkages, enabling the efficient management of trade-offs that will likely exist. Considering environmental issues and the excessive demand for EWF resources, many studies have attempted to utilize a EWF nexus approach to better understand resource utilization and environmental degradation in the delivery of a product or service. For instance, Al Ansari et al. (2015; 2016; 2017) developed a EWF nexus tool that quantifies synergies between EWF systems, in addition to computing environmental impacts in delivering a hypothetical food security scenario. Moreover, the current situation of EWF resources necessitates a holistic analysis that models synergies between the three systems, tackles the inherent competition between their respective supply chains at the decision-making level and optimize the whole performance of the integrated EWF nexus. Yet, this cohesive assessment is usually hindered by modelling complexities, such us multi-scale uncertainties governing the nexus, multi-temporal and spacial structure of supply chains, along with computational limitations associated with optimization (Garcia and You, 2015a).

Process Systems Engineering (PSE) is one of the fields that devoted attention to the modelling and optimization of EWF resources, to enhance their environmental performance and increase their economic efficiency. Karnib (2017) adopted a coupled simulation/optimization framework based on Q-Nexus Model. The latter is based on an input-output theory incorporating water and energy inflows into the production of food, water and energy resources while addressing different interactions amongst the three systems. Garcia and You (2015b) used multi-objective optimization to model water and energy nexus of biofuel products. The study formulated a model that minimizes the water footprint and maximizes economic performance and energy output. Mostly, these studies focused on developing integrated approaches for the evaluation of EWF nexus systems; however, most of them focused mainly on one or two resource systems of the EWF nexus in their optimization models. Thus, investigating the three systems holistically is a very promising research area for PSE community especially with the multiple interactions that exist amongst the EWF subsystems.

The work presented in this research tackles the modelling challenges of the EWF nexus through the implementation of a bottom-up framework aiming to enhance the economic performance of the EWF resource sectors whilst considering the environment. The methodology developed is divided into technology selection and resource allocation. The first problem consists of an integer linear program that determines the best performing combination of technologies under an environmental burden constraint. The second consists of a resource allocation approach which is based on a constraint to resource capacity, natural gas price fluctuations and non-cooperative Stackelberg competition. As part of this study, the developed methodology is applied to the State of Qatar with the target of achieving a hypothetical level of 40% in food self-sufficiency.

2. Technology selection

2.1. Case study

To design resilient and sustainable EWF systems, decision makers, at every level of the nexus, must hedge against negative environmental impacts generated from the interactions between the three sectors. Incidentally, they must do so whilst adhering to economic objectives, including cost minimization and/or profit maximisation. In the context of Qatar and this study, domestic production is assumed to increase to a level which attains a perceived food security self-sufficiency. Al-Ansari et al. (2016;2017) illustrated this target through EWF nexus and a sub-system approach enabling the

evaluation of multiple scenarios. In this paper, a similar design is adopted where the EWF nexus system under consideration is illustrated in Table 1, includes: renewable (i.e. solar energy & biomass) and non-renewable (i.e. natural gas) energy sources representative of the energy sector, a reverse osmosis based seawater desalination system representative of the water sector, and finally the food sector is represented by the production of fertilizers and agricultural activities (i.e. livestock management and application of fertilizers). A carbon capture system is also integrated to the combined cycle driven by both natural gas and gasified biomass, in which the latter is transformed to a negative emission technology (Al-Ansari et al., 2017). The system boundary is limited to a local nexus system where food and energy are produced locally. Thus, import/export relationships are neglected in this model.

Table 1 Technologies by sector

Energy	Water	Food
Combined Cycle Gas Turbine (CCGT)- CCGT with Carbon Capture(CC)-Biomass Integrated Gasification Combined Cycle (BIGCC)/Photovoltaics	Reverse Osmosis (RO)	Agriculture/Fertilizer Production

2.2. Problem formulation

The optimised EWF nexus model proposed in this study is formulated using a linear program where the total cost of the EWF nexus is the objective function to be minimized and the Global Warming Potential (GWP) is considered as the main constraint. This optimization provides an optimum combination of technologies which together will operate at the lowest cost whilst remaining within the assumed value of $5x10^7$ $kgCO_2eq$, being the average of the highest and lowest emissions of the entire nexus. This is represented in equation 1.

$$Min: Total\ Cost = \sum_{i=1}^{n} x_i^e c_i + \sum_{i=1}^{m} x_i^w c_i + \sum_{i=1}^{l} x_i^f c_i \quad (1)$$

The decision variables are x_i^e, x_i^w, x_i^f, the percentages of each technology used in energy, water and food respectively, and c_i (USD)is the capital cost associated with each technology.

Subject to the following set of constraints:

$$\sum_{i=0}^{n} x_i^e = 1 \quad (2)$$

$$\sum_{i=0}^{m} x_i^w = 1 \quad (3)$$

$$\sum_{i=0}^{l} x_i^f = 1 \quad (4)$$

$$\sum_{i=1}^{n} x_i^e t_i + \sum_{i=1}^{m} x_i^w t_i + \sum_{i=1}^{l} x_i^f t_i \leq 5 \times 10^7 \quad (5)$$

Where t_i is the amount of GWP associated with each technology in kg of CO_2eq, while *n* is the number of energy technologies, *m* is the number of water technologies and *l* is the number of food technologies.

As for the capital cost, all technologies are denoted with a value that represents the incurred implementation cost, except for agriculture, it is assumed to be $0 as no leasing costs are associated with agricultural lands in Qatar.

3. Resource allocation

3.1. Baseline scenario

The second step in the bottom-up approach is to find the optimal resource allocation amongst the selected energy and water technologies from the previous step, that will satisfy the additional food self-sufficiency requirement while minimizing the cost (USD).

In this model, it is assumed that only water and energy are the only inflows, since the target is to produce additional food.
The problem is modelled using a LP which is illustrated in equation 6.

$$Total\ Cost = \sum_{i=1}^{m}\sum_{j=1}^{n} s_j^{w-e}\, c_{ij}^{w-e} p_i^{w} + \sum_{i=1}^{m}\sum_{j=1}^{l} s_j^{w-f}\, c_{ij}^{w-f} p_i^{w} + \sum_{i=1}^{n}\sum_{j=1}^{m} s_j^{e-w}\, c_{ij}^{e-w} p_i^{e} + \sum_{i=1}^{n}\sum_{j=1}^{l} s_j^{e-f}\, c_{ij}^{e-f} p_i^{e} \quad (6)$$

The decision variables are $c_{ij}^{w-e}, c_{ij}^{w-f}, c_{ij}^{e-w}, c_{ij}^{e-f}$ denoted as the intersectoral coefficient of allocation (Karnib, 2017). $s_j^{w-e}, s_j^{w-f}, s_j^{e-w}, s_j^{e-f}$ are the inflows from a sector to another. p_i^{w} is the unit price of water and it is equal to US$0.986/m^3 (Al-Karaghouli and Kazmerski, 2012). p_i^{e} is the levelized cost of electricity (LCOE) of energy technologies, which is estimated to be US$ 110/MWh for BECCS and US$ 93/MWh for CCGT (OpenEl, 2015). Subject to the following set of constraints:

$$\sum_{j=1}^{n} c_{ij}^{w-e} = 1 \text{ with } i=1,\ldots,m \quad (7)$$

$$\sum_{j=1}^{l} c_{ij}^{w-f} = 1 \text{ with } i=1,\ldots,m \quad (8)$$

$$\sum_{j=1}^{m} c_{ij}^{e-w} = 1 \text{ with } i=1,\ldots,n \quad (9)$$

$$\sum_{j=1}^{l} c_{ij}^{e-f} = 1 \text{ with } i=1,\ldots,n \quad (10)$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} s_j^{w-e}\, c_{ij}^{w-e} + \sum_{i=1}^{m}\sum_{j=1}^{l} s_j^{w-f}\, c_{ij}^{w-f} \le Maximum\ water\ capacity \quad (11)$$

$$\sum_{i=1}^{n}\sum_{j=1}^{m} s_j^{e-w}\, c_{ij}^{e-w} + \sum_{i=1}^{n}\sum_{j=1}^{l} s_j^{e-f}\, c_{ij}^{e-f} \le Maximum\ energy\ capacity \quad (12)$$

The constraints represented by Eq.(11) and Eq.(12) restrict the additional amount of water and energy used to the capacity of water and energy. Furthermore, the usage of 99% of water in agriculture compared to fertilizers production is added as a constraint to meet the high irrigation requirement (Al-Ansari et al., 2016).

3.2. Stochastic scenario

In the baseline scenario for resource allocation, the levelized cost of electricity for CCGT is assumed to be fixed. In reality, it is affected by the price of natural gas that knows significant fluctuations. To account for these variations and assess their impact on the overall cost, a new stochastic constraint is added to the baseline model, and it is formulated by the following assumed probabilistic requirement:

$$P(y_1 \le y_2) \ge 95\% \quad (13)$$

where y_1 is the levelized cost of electricity following a normal distribution with a \$93/MWh mean and \$5/MWh standard deviation, which is estimated using a Monte Carlo simulation. y_2 is the LCOE of BECCS equal to \$110/MWh (OpenEl, 2015). The problem is solved using chance constraint programming with a confidence level of 95%.

3.3. Stackelberg duopoly scenario

The EWF nexus system presented in this paper is also part of a competitive milieu, where stakeholders have conflicting interests and each one aims to maximize its profit (Yue and You, 2016). For this reason, a decentralized approach based on a non-cooperative Stackelberg game is modelled. The framework focuses on the energy sector only and assumes that CCGT technology option is the leader, whereas BECCS is the follower. The two technologies compete on resources such as water and food, and their common target is to produce an optimal quantity of energy that will lead to maximization of their profits. This objective is formulated as a constraint and added to the baseline resource allocation scenario. Eq.(11) and Eq.(12) are substituted by the following constraints:

$$z_1 \times c^{CCGT-RO} + z_2 \times c^{CCGT-fert} + z_3 \times c^{CCGT-agr} \ge q_l \quad (14)$$

$$z_1 \times c^{BECCS-RO} + z_2 \times c^{BEECS-fert} + z_3 \times c^{BECCS-agr} \ge q_f \quad (15)$$

where z_1, z_2 and z_3 are the total energy requirement for RO, fertilizers production and agriculture respectively. q_l and q_f are the optimal quantities for the leader and the follower, computed based on backward induction method based on Stackelberg Duopoly.

4. Illustrative example and results

By applying the technology selection to the food security scenario, the model yields a configuration which is composed of the CCGT and the BECCS within the energy sector, RO as the only water source, and the production of fertilizers production as the only technology in the food sector. In this case study, the fertilizers are used for the domestically grown crops and not for crops used to raise of livestock which are assumed to be imported. The crop selection profile considered as part of this study is assumed to meet 40 % food self-sufficiency requirements in which the remainder is assumed to be imported (see Table 2). The capital cost of this configuration is 1.88 Billion USD. This value is the most optimal considering the GWP from the proposed technologies. The highest percentage is given to BECCS, since it has a negative CO_2 emission (Al-Ansari et al., 2016; 2017) and a relatively low cost compared to PV. BIGCC, CCGT+CC and PV are not selected because of their enormous relative capital cost. The CCGT has a major representation as it has the lowest capital cost amongst all technologies. The resource allocation part is based on the configuration found from the technology selection optimization. In fact, all energy technologies with 0% participation are eliminated, while the two other sectors are kept the same. The results for the baseline, stochastic and Stackelberg scenarios are summarized in Figure 1.

Table 2 Results for technology selection

Technology	BECCS	CCGT+CC	CCGT	BIGCC	PV	RO	Agriculture	Fertilizers
Capital cost (Million USD)	1,605	1,175	720	1,150	4,180	650.269	0	37.389
GWP (kgCO2eq)	-11,600,00,000	2,530,000	460,000,000	10,000	45,300	414,892,037	1,260,000,000	45,500,000
Percentage	53.72%	0%	46.27%	0%	0%	100%	0%	100%

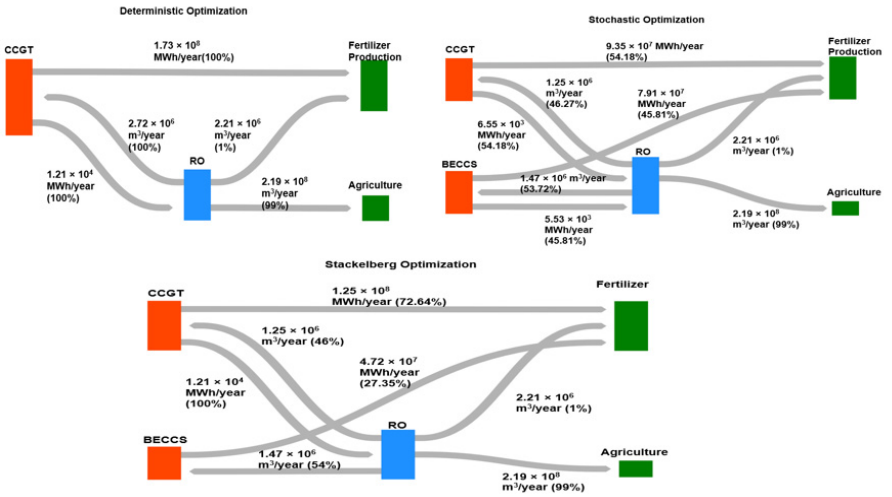


Fig. 1. Summary of the resource allocation results

In comparison to the deterministic optimization, stochastic optimization provides an opportunity for BECCS to contribute to the mix, particularly for both the water and food sectors, as CCGT suffers from an exposure to risk of increasing natural gas prices. Stackelberg duopoly, wherein CCGT is assumed to be the leader and allowed to set the energy prices to reduce its exposure, the contribution of BECCS is reduced in the mix.

5. Conclusions and further improvements

This study presents a techno-economic framework coupled with an optimization methodology for a EWF nexus case study for Qatar. The objective is to achieve a perceived level of food sufficiency at the lowest cost and environmental impact. The bottom-up approach resulted in an optimal configuration of energy technologies. This configuration is then used to come-up with the best resource allocation, considering three scenarios. As a result, the deterministic optimization generated a cost of 1.88 Billion USD with a usage of CCGT only. As for the resource allocation, the cost associated with additional energy and water resources in the baseline scenario 664 Million USD. This amount decreases after adding the natural gas constraint to 227 Million USD. This value decreases further with the game theoretical constraint to 224 Million USD. The extension of this work would include the operating cost of technologies as part of the decision-making process. The Stackelberg competition could be enhanced by considering one of the three sectors of the EWF nexus as the leader whilst the two-remaining act as followers. Moreover, other uncertainties can be considered such as those related to CO_2 prices.

References

Al-Ansari, T., Korre, A., Nie, Z. and Shah, N., (2015). Development of a life cycle assessment tool for the assessment of food production systems within the energy, water and food nexus. Sustainable Production and Consumption, 2, pp.52-66.

Al-Ansari, T., Korre, A., Nie, Z., & Shah, N. (2016). Integration of Biomass Gasification and CO_2 Capture in the LCA Model for the Energy, Water and Food Nexus. Computer Aided Chemical Engineering 26th European Symposium on Computer Aided Process Engineering, 2085-2090.

Al-Ansari, T., Korre, A., Nie, Z. and Shah, N.(2017). Integration of Greenhouse Gas Control technologies within the Energy, Water and Food Nexus to enhance the environmental performance of food production systems. Journal of Cleaner Production, 162, pp. 1592-1606.

Al-Karaghouli, A & Kazmerski, L. (2012). National Renewable EnergyLaboratorGolden, Colorado 80401

Garcia, D. J., & You, F. Q. (2016). The water-energy-food nexus and process systems engineering: A new focus. Computers & Chemical Engineering, 91, 49–67.

Garcia, D. J., & You, F. (2015a). Supply chain design and optimization: Challenges and opportunities. Computers and Chemical Engineering, 81, 153–170.

Garcia, D. J. & You, F. (2015b). Life Cycle Network Modeling Framework and Solution Algorithms for Systems Analysis and Optimization of the Water-Energy Nexus, Processes, 3, 514-539.

Hoff, J., (2011). Understanding the nexus. In: Background Paper for the Bonn 2011.

Karnib, A. (2017) Water-Energy-Food Nexus: A Coupled Simulation and Optimization Framework. Journal of Geoscience and Environment Protection , 5, 84-98.

OpenEI. (2015). Transparent Cost Database. Retrieved November 18, 2017, from https://openei.org/apps/TCDB/

Yue, D., You, F. (2016). Stackelberg-game-based modeling and optimization for supply chain design and operations: A mixed integer bilevel programming framework. Computers and Chemical Engineering.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50163-4

Identification of Design Modifications for Implementing Optimal HEN Cleaning Schedules

Chuei-Tin Chang[a*], Bo-Jyun Liao[a]

[a]*Department of Chemical Engineering, National Cheng Kung University, 1 University Road, Tainan 70101, Taiwan, ROC*
ctchang@mail.ncku.edu.tw

Abstract

Due to fouling in the heat-transfer units, the target temperatures of process streams in a heat exchanger network (HEN) may not always be achievable after a long period of operation. To address this practical issue, the conventional cost-optimal HEN design must be modified to accommodate online cleaning operations via proper allocation of area margins, bypasses, spares and auxiliary heaters/coolers. A solvable MINLP model has been developed in this work to simultaneously generate such design refinements and also the corresponding spare-supported cleaning schedules. The optimization results of one example are presented to demonstrate the feasibility of this approach.

Keywords: heat exchanger network, spare, margin, bypass, cleaning schedule.

1. Introduction

Efficient energy recovery is one of the essential requirements of almost any process design, while a heat exchanger network (HEN) is often utilized to fulfil this need. Since the overall performance of HEN tends to deteriorate over time due to fouling in the embedded heat-transfer units, there are incentives to clean at least some of them when the normal production is still in progress.

The mathematical programming approach was first proposed by Smaïli et al., (1999) to synthesize the optimal cleaning schedules. Since then, this practice has been adopted in numerous realistic applications. For example, Sanaye and Niroomand (2007) produced the optimal HEN cleaning schedule for the urea and ammonia units by minimizing the operating cost. Although these optimal cleaning schedules could indeed be implemented to save energy, the overall heat-exchange rate of a HEN may still gradually decrease over time after cleaning and, furthermore, any defouling operation still calls for temporary removal of online unit(s).

In principle, the aforementioned decrease in total heat recovery level of a HEN may be compensated by (a) introducing a spare unit, (b) adjusting the heat transfer rates of other online units via area margins and bypasses, or (c) raising the hot and cold utility consumption rates in heaters and coolers. Cheng and Chang (2016) developed a MINLP model to create the HEN cleaning schedule and spare replacement schedule simultaneously, while virtually no studies have been performed for systematic implementation of the other options mentioned above. Thus, there are incentives to modify the existing models to optimally allocate spares, margins, bypasses and auxiliary heats/coolers for implementing an effective and energy efficient cleaning schedule.

2. Mathematical representations of network and schedule structures

2.1. HEN networks

It is assumed in this study that, before applying the proposed method, a preliminary HEN design is available. To characterize this given structure, let us first introduce two stream sets, i.e., I and J, to collect and classify the hot and cold process streams respectively. On the other hand, the matches in this given HEN can be characterized mathematically with a function, i.e., the mapping from the set of all heat exchangers (EX) to the set of matches between process steams (MA), i.e., $f: \mathrm{EX} \longrightarrow \mathrm{MA}$. Notice that $f(e) = (i, j)$ is known for every $e \in \mathrm{EX}$. This given structure can be described more clearly by dividing EX into subsets according to hot or cold streams. In particular, $\mathrm{EX}_{i'}$ ($i' \in$ I) and $\mathrm{EX}_{j'}$ ($j' \in$ J) denote respectively the sets of heat exchangers on hot stream i' and cold stream j'. Since the units on each process stream may be placed in series and/or in parallel, they can be further classified into subsets $\mathrm{EX}_{i'}^{m}$ ($i' = 1,2,\cdots, M_{i'}$) and $\mathrm{EX}_{j'}^{n}$ (j$' = 1,2,\cdots N_{j'}$) according to their downstream "mixing points". To clarify this notion, let us consider the bullets (i.e., mixing points) in the fictitious structure in Figure 1. One can identify that

- $M_{i'} = 2$, $\mathrm{EX}_{i'}^{1} = \{1,2\}$ and $\mathrm{EX}_{i'}^{2} = \{3\}$ on hot steam i';
- $N_{j'} = 3$, $\mathrm{EX}_{j'}^{1} = \{4,5\}$, $\mathrm{EX}_{j'}^{2} = \{6,7\}$ and $\mathrm{EX}_{j'}^{3} = \{8\}$ on cold stream j'.

Notice that it is also assumed that a bypass flow may join the output streams from heat exchangers at each mixing point.

2.2. Cleaning and spare replacement schedules

In this work, the time horizon for schedule synthesis (say t_f) is assumed to be given. To simplify calculation, this schedule horizon $[0, t_f]$ is partitioned into n_P different periods and each is further divided into two intervals for performing the cleaning and heat-exchange operations respectively. For computation simplicity, the lengths of aforementioned time periods are fixed, i.e., $t_f = n_P \tau$ and τ is a constant. Also, it is assumed that the durations of all sub-periods required for defouling (f_c) are the same and their values can be determined in advance. Thus, within each partitioned period, four time points can be identified to facilitate accurate presentation of the proposed model, i.e., *bcp* (beginning of cleaning sub-period), *ecp* (end of cleaning sub-period), *bop* (beginning of operation sub-period), and *eop* (end of operation sub-period). Note that time point *ecp* represents the instance just before *bop*. To produce concise model formulation, let us introduce two sets, i.e., P and TP, to characterize the periods and time points respectively.

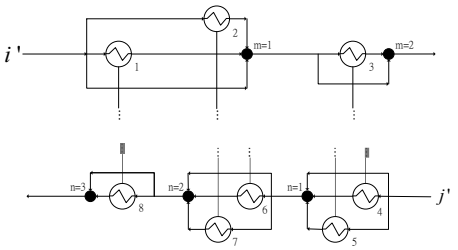


Figure 1. A fictitious stream structure in HEN.

The selections of exchangers to be cleaned in each time period are expressed with binary variables $Y_{e,p} \in \{0,1\}$ (where, $e \in$ EX and $p \in$ P.). Let us also assume that the available spares are given in advance and they are collected in set S. It is clear that the decision to use spare $s \in$ S to replace unit $e \in$ EX in period $p \in$ P can only be made after determining whether this unit should be removed and cleaned in the same period. The corresponding logic constraints can be represented with another set of binary variables $X_{e,p,s} \in \{0,1\}$, i.e., $\left(1 - Y_{e,p}\right) + \sum_{s \in S} X_{e,p,s} \leq 1$

3. Model constraints

3.1. Overall heat-transfer coefficients

As a result of fouling, the overall heat-transfer coefficient of every unit in HEN may decrease with time and this phenomenon can be modelled as follows:

$$U_e(t) = \left[\left(U_e^{cl}\right)^{-1} + r_e(t)\right]^{-1} \tag{1}$$

where, $e \in$ EX; $U_e(t)$ denotes the overall heat-transfer coefficient of exchanger e at time t; U_e^{cl} is the corresponding constant value when the heat-transfer surface is *completely* clean. Notice that the time function $r_e(t)$ is the fouling resistance of unit e at time t, which is expressed with a linear model in this paper, i.e., $r_e(t) = \dot{r}_e t$ and $\dot{r}_e$ is a constant fouling rate. Four scenarios should be considered for determining the overall heat-transfer coefficient of unit e at the aforementioned four time points in period p of a cleaning schedule:

(i) Exchanger e is not cleaned during period p ($p \geq 2$), but in at least one of the prior periods a defouling operation has been performed, i.e., $Y_{e,p} = 0$ and $\prod_{k=1}^{p-1}\left(1 - Y_{e,k}\right) = 0$ for $p = 2,3,\cdots,n_P$.

(ii) Exchanger e is cleaned and replaced with a spare in period p ($p \geq 1$), i.e., $Y_{e,p} = 1$ and $\sum_{s \in S} X_{e,p,s} = 1$ for $p = 1,2,\cdots,n_P$.

(iii) Exchanger e is cleaned but not replaced with a spare in period p ($p \geq 1$), i.e., $Y_{e,p} = 1$ and $\sum_{s \in S} X_{e,p,s} = 0$ for $p = 1,2,\cdots,n_P$.

(iv) Exchanger e has never been cleaned since period 1, i.e., $Y_{e,1} = Y_{e,2} = \cdots = Y_{e,p} = 0$ for $p = 1,2,\cdots,n_P$.

The formulas for computing the overall heat-transfer coefficients in these four scenarios at any time point can be found in Cheng and Chang (2016). It is assumed in this work that the initial and target temperatures of every process stream in a HEN are fixed a priori and these conditions must be kept unchanged throughout the entire horizon. For this operational problem, it is necessary to examine the temperature variations along every process stream in HEN.

3.2. Temperature changes along process streams

Since the temperature of each stream is adjusted with heat exchangers, coolers/heaters and perhaps also mixers according to the structure described in Figure 1, it is important to check if the intermediate temperatures at critical locations satisfy thermodynamic and/or operational constraints. By assuming counter-current flow, let us express the heat transfer rate in unit e at time point tp in period p as $Q_{e,p}^{tp} = \mu_e A_e U_{e,p}^{L,tp} LMTD_{e,p}^{tp}$ and, in this expression, A_e denotes the given heat-transfer area of unit e in the original HEN design; $\mu_e = A_e^M / A_e$ is a dimensionless variable used to represent design margin and A_e^M denotes the enlarged heat-transfer area of unit e; $LMTD_{e,p}^{tp}$ is the log-mean

temperature difference of unit e at time point tp in period p. The corresponding energy balances can be written as

$$Q_{e,p}^{tp} = Fcp_i^H \phi_{e,p}^{H,tp} \left(Ti_{e,p}^{H,tp} - To_{e,p}^{H,tp} \right) = Fcp_j^C \phi_{e,p}^{C,tp} \left(To_{e,p}^{C,tp} - Ti_{e,p}^{C,tp} \right) \tag{2}$$

where, Fcp_i^H and Fcp_j^C respectively denote the heat-capacity flow rates (kW/K) of hot stream i and cold stream j; $Ti_{e,p}^{H,tp}$ and $To_{e,p}^{H,tp}$ respectively denote the inlet and outlet hot stream temperatures of unit e at time point tp; $Ti_{e,p}^{C,tp}$ and $To_{e,p}^{C,tp}$ respectively denote the inlet and outlet cold stream temperatures of unit e at time point tp in period p; $\phi_{e,p}^{H,tp}$ and $\phi_{e,p}^{C,tp}$ denote the flow fractions hot and cold streams, respectively, passing through unit e at time point tp in period p.

Let next consider unit e, its corresponding match (i, j) and its downstream mixing points m and n on hot and cold streams. The mass balances at these mixing points can be written as $\tilde{\phi}_{i,p}^{m,tp} + \sum_{e \in \mathrm{EX}_i^m} \phi_{e,p}^{H,tp} = 1$ and $\tilde{\phi}_{j,p}^{n,tp} + \sum_{e \in \mathrm{EX}_j^n} \phi_{e,p}^{C,tp} = 1$, where $\tilde{\phi}_{i,p}^{m,tp}$ and $\tilde{\phi}_{j,p}^{n,tp}$ denotes the flow fractions of bypasses joining mixing point m on hot stream i and mixing point n on cold stream j at time point tp in period p, respectively. To simplify the final HEN structure, it is assumed that the bypass of each heat exchanger can be placed on its hot or cold stream, but not both.

Notice also that the log-mean temperature difference can also be expressed as a function of inlet and outlet temperatures of the hot and cold streams, i.e.

$$LMTD_{e,p}^{tp} = \frac{\left(Ti_{e,p}^{H,tp} - To_{e,p}^{C,tp} \right) - \left(To_{e,p}^{H,tp} - Ti_{e,p}^{C,tp} \right)}{\ln\left[\left(Ti_{e,p}^{H,tp} - To_{e,p}^{C,tp} \right) \Big/ \left(To_{e,p}^{H,tp} - Ti_{e,p}^{C,tp} \right) \right]} \tag{3}$$

Thus, the outlet temperatures of heat exchanger e can be determined by solving the aforementioned equations simultaneously according to the inlet temperatures. Finally, the mixing point temperatures are calculated in this study with the following formulas:

$$T_{m,i,p}^{H,tp} = \tilde{\phi}_{i,p}^{m,tp} T_{m-1,i,p}^{H,tp} + \sum_{e \in \mathrm{EX}_i^m} \phi_{e,p}^{H,tp} To_{e,p}^{H,tp} \tag{4}$$

$$T_{n,j,p}^{C,tp} = \tilde{\phi}_{j,p}^{n,tp} T_{n-1,j,p}^{C,tp} + \sum_{e \in \mathrm{EX}_j^n} \phi_{e,p}^{C,tp} To_{e,p}^{C,tp} \tag{5}$$

where, $T_{m,i,p}^{H,tp}$ and $T_{n,j,p}^{C,tp}$ respectively denote the temperatures at mixing point m on hot stream i and mixing point n on cold stream j. The last mixing point temperature on a hot stream obviously must be higher than the designated final temperature, while that on a cold stream lower than the given end value. Note that the design targets can always be achieved exactly with cooling and heating utilities.

4. A simple example

The given HEN design in this example can be found in Table 1 (stream data), Table 2 (unit specifications) and Figure 2 (network structure), which is taken from Lavaja and Bagajewicz (2004). For convenience, the overall heat-transfer coefficient of any completely cleaned heat exchanger (U_e^{cl}) is assumed to be constant at $88.1\mathrm{BTU/hr/ft^2/^\circ F}$. Since the linear fouling model is adopted for the present application, the fouling rates ($\dot{r}_e$) of the four units in Table 2 are chosen to be 3.07, 3.27, 3.68 and 3.88 (×

10^7 ft^2°F/BTU), respectively. The unit costs of hot and cold utilities are assumed to be 2.812×10^{-3} and $5{,}624 \times 10^{-4}$ USD/BTU, respectively. The cost of cleaning a heat exchanger is set to be 4000 USD, while that for a spare is 1000 USD. For every unit in HEN, the annualized fixed capital cost is 5500 USD/yr, and the corresponding variable cost coefficient is 15 USD/yr/ft$^{2\alpha}$ ($\alpha = 0.8$). Finally, the time-related model parameters are selected as follows: $t_f = 18$ (mon), $\tau = 1$ (mon) and $f_c = 0.2$ (mon).

By minimizing the total annual cost with only 1 margined unit, one can identify the optimal spare-supported cleaning schedule for the present example. In this schedule, each unit is cleaned in designated periods over the given horizon. Specifically, HE1: 4, 7, 12, 15 (mon); HE2: 9, 12 (mon); HE3: none; HE4: 4, 7, 10, 13, 16 (mon). When HE4 is removed for cleaning, a spare is used to take its place in HEN and the heat-transfer area of this spare unit is 1485 ft^2. Notice also that the optimal design refinements can be found in Figure 3. It can be observed that (1) the margined unit is chosen to be HE3 and its heat-transfer area is enlarged to 1653.9 ft^2 ($\mu_3 = 1.388$), (2) the bypasses are placed at the hot-stream sides of HE2, HE3 and HE4, and (3) one auxiliary heater and three coolers are added and their specifications are given in Figure 3.

It should be noted that, when compared with the scenario in which the given HEN is operated for 18 months without cleaning and without any design refinement, a 27 percent reduction in the total annual cost (TAC) can be achieved with the cleaning schedule suggested by Lavaja and Bagajewicz (2004). On the other hand, a 55 percent saving is realizable by implementing the proposed spare-supported cleaning schedule together with the aforementioned design refinements. This saving can be raised even higher to 84 percent if the number of margined units is increased to 3.

Table 1. Stream data of a given HEN design

Stream	Initial Temp (°F)	Target Temp (°F)	Heat Capacity Flow Rate (BTU/hr/°F)
C1	270	429.67	331862.86
H1	428	374.82	94652.24
H2	513	426.93	51667.7
H3	536	466.01	262274.26
H4	631	536.345	265718.98

Table 2. Unit specifications of a given HEN design.

Unit	Heat-Transfer Area (ft^2)	Cold Inlet/Outlet Temp (°F)	Hot Inlet/Outlet Temp (°F)	Heat Duty (BTU/hr)
HE1	465	270/285.2	428/374.8	5.03×10^6
HE2	287.4	285.2/298.6	513/426.9	4.3×10^6
HE3	1196.1	298.6/353.9	536/466.0	1.84×10^7
HE4	1487.6	353.9/429.7	631/536.3	2.52×10^7

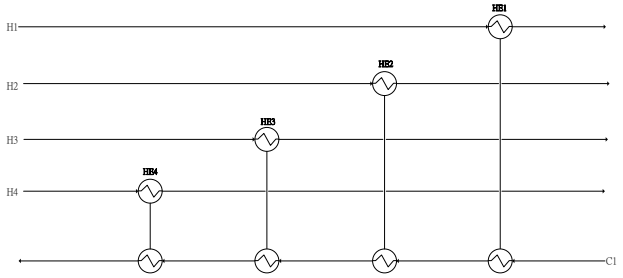


Figure 2. Network structure of a given HEN design

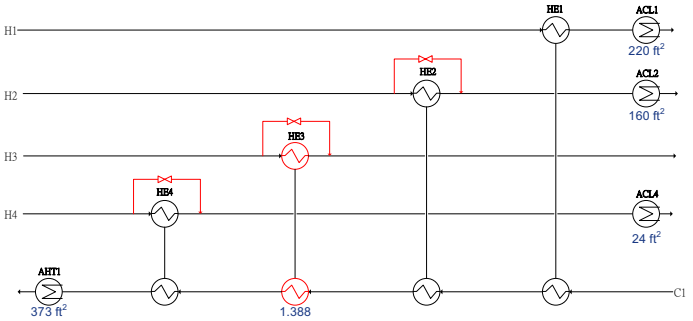


Figure 3. Refined HEN design obtained by introducing area margin to HE3.

5. Conclusions

An improved mathematical programming model has been proposed in this work to synthesize the optimal spare-supported cleaning schedule and the corresponding design refinements for any given HEN design. The effectiveness of this approach has been demonstrated in extensive case studies. It can be observed from the optimization results that spares, margins, bypasses and auxiliary units are viable options for reducing the extra amount of utility consumption due to fouling and/or removing online heat exchangers for cleaning purpose.

References

K. Y. Cheng, and C. T. Chang, 2016. Model Based Approach to Synthesize Spare-Supported Cleaning Schedules for Existing Heat Exchanger Networks, Computers and Chemical Engineering, 93, 413 – 427.

J. H. Lavaja, and M. J. Bagajewicz, 2004, On a New MILP Model for the Planning of Heat-Exchanger Network Cleaning, Industrial & Engineering Chemistry Research, 43(14), 3924 - 3938.

S. Sanaye, and B. Niroomand, 2007, Simulation of Heat Exchanger Network (HEN) and Planning the Optimum Cleaning Schedule, Energy Conversion and Management, 48(5), 1450 - 1461.

F. Smaïli, D. K. Angadi, C. M. Hatch, O. Herbert, V. S. Vassiliadis, and D. I.Wilson, 1999, Optimization of Scheduling of Cleaning in Heat Exchanger Networks Subject to Fouling: Sugar Industry Case Study, Trans IChemE, Part C: Food and Bioproducts Processing , 77(2), 159 - 164.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50164-6

Optimization Model for the Transfer Line Exchanger System

Guang Song[a], Lixin Tang[a,*]

[a] *Institute of Industrial and System Engineering, Northeastern University, Shenyang 110819, PR China*

songguang@ise.neu.edu.cn

Abstract

To avoid the secondary reactions in steam cracking of hydrocarbon feedstocks, which lead to undesirable products such as coke, the high temperature gas leaving from the cracking furnace need be rapidly quenched in Transfer Line Exchanger (TLE). A severe problem in the process is that coke deposits on the walls of the furnace coil and the TLE tubes, which leads to increasing heat transfer resistances and pressure drops. When the coke deposition reaches an allowable maximum, the TLE must be periodically shut down for decoking. This work developed an optimization model for the operation process of TLE system. The proposed model has been considered an integrated model for TLE to optimize the operation process. Firstly, we built a steady-state heat transfer model and a dynamic coke depositing model, respectively. Then, we integrated two models by using quasi-steady assumption, in which the thickness of the coke deposition is considered as a constant in one time interval. According to the integrated model, we chose the COT as the manipulated variable to optimize the operation period between decoking. At last, the reliability and adaptability of the optimization model was validated by one significant case study. The results showed that the optimization model of TLE system could effectively increase the operation period.

Keywords: TLE, coke deposition, optimization model, operation period.

1. Introduction

Steam cracking of hydrocarbon feedstocks includes two parallel reaction pathways, which are the primary reactions and the secondary reactions. To avoid the secondary reactions, which lead to undesirable products such as coke, the high temperature gas leaving from the cracking furnace need be rapidly quenched in Transfer Line Exchanger (TLE). The TLE can not only quench the cracking gas, but also recover heat from the cracking gas to produce high pressure steam. A severe problem in the process is that coke deposits on the walls of the furnace coil and the TLE tubes, which leads to increasing heat transfer resistances and pressure drops. When the coke deposition reaches an allowable maximum, the TLE must be periodically shut down for decoking. In general, there are three coking mechanisms(Bach et al., 1995): 1) the catalytic mechanism; 2) the free-radical mechanism; 3) the physical condensation mechanism. In recent years, researchers not only developed many kinds of coking-inhibitors that can limit the coke deposition by the catalytic mechanism, but also designed many types of TLE tubes that can reduce the coke deposition by the free-radical mechanism(Geem et al., 2009). Therefore, the coke deposition in TLE tubes is mainly generated by the physical condensation(Kopinke, Bach, & Zimmermann, 1993).

Although the coke deposition in the TLE has important influence on the whole process, the research on this field is few and deficient. Manafzadeh et al develop a comprehensive mathematical model for a TLE system, which can simulate the operation process and predict the coke deposition rate(Manafzadeh, Sadrameli, & Towfighi, 2003). Jin et al propose a parametric model based on theoretical analysis, mathematic reduction and parameters estimation, which can predict the TLE outlet temperature(Jin, Li, Du, Wang, & Qian, 2013). However, the above model lacks optimization for the operation process, which only adopts mathematic model to predict the coke deposition and outlet temperature. Thus, a novel model for the TLE system is proposed in this paper to optimize the operation process.

The paper is organized as follows. In section 2, the details about the TLE system are given. In section 3, we introduce the mathematical formulation, which integrates a steady-state heat transfer model and a dynamic coke depositing model by using quasi-steady assumption. In section 4, we illustrate the use of the optimization model through one significant case study.

2. Transfer Line Exchanger System

Transfer Line Exchanger System includes a set of shell-and-tube heat exchangers. As shown in Figure 1, the cracked gas (CG) /steam mixture from the furnace is quenched in tube pass, and in exchange, the saturated boiler feed water recovers heat to generate high pressure steam in shell-side. The TLE inlet conditions are determined by the furnace outlet operation.

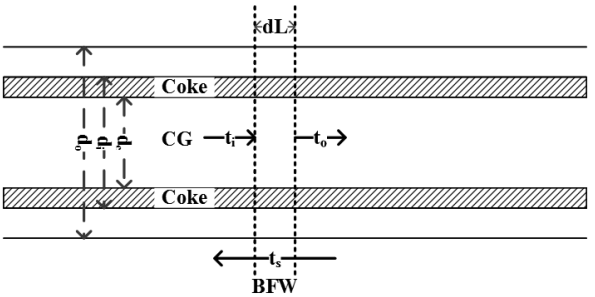


Figure 1. Schematic diagram of TLE tube.

The optimization model is assumed as one dimensional plug flow model which implies the neglect of radial concentration gradients and axial dispersion. The TLE tube is divided into a number of sections, one of which has a length of $\boldsymbol{dL}$. In one section, three assumptions are presented: 1) the thickness of the coke deposition is constant at the same time interval; 2) the temperature of the CG is constant as the inlet temperature; 3) the thermal physical parameters of the CG are constant(Yaws, 1961).

3. Mathematical Model

The optimization model integrates a steady-state heat transfer model and a dynamic coke depositing model by using quasi-steady assumption, in which the thickness of the coke deposition is considered as a constant in one time interval. The mathematical formulation of both models are given below.

3.1. Steady-state Heat Transfer Model.

The steady-state heat transfer model can simulate the temperature distribution in the TLE tubes under the given CG inlet conditions, which include the temperature, the mass flow and the chemical composition.

In one tube section shown in Figure 1, the heat balance is given as equation 1~2, which states that in one tube section the enthalpy change of CG is equal to the heat exchange capacity between CG and BFW.

$$\mathrm{d}Q = m \cdot \mathrm{d}H_i \tag{1}$$

$$\mathrm{d}Q = K \cdot \mathrm{d}A \cdot \Delta t_m \tag{2}$$

Equation 3~4 gives the enthalpy change of CG, in which c_P represents the heat capacity.

$$\mathrm{d}H_i = c_P \cdot \Delta t \tag{3}$$

$$\Delta t = t_i - t_o \tag{4}$$

Equation 5~6 gives the heat transfer area and the logarithmic mean temperature difference.

$$\mathrm{d}A = N \cdot \pi d_o \cdot \mathrm{d}L \tag{5}$$

$$\Delta t_m = \frac{(t_i - t_s) - (t_o - t_s)}{\ln \dfrac{(t_i - t_s)}{(t_o - t_s)}} \tag{6}$$

The overall heat transfer coefficient is given by equation 7, which includes convective heat transfer coefficients inside and outside the tube, the heat conductivity coefficients of the tube wall and coke deposition, and the fouling factors.

$$K = \frac{1}{\dfrac{1}{\alpha_\mathrm{i}} \cdot \dfrac{\mathrm{d_o}}{\mathrm{d_c}} + \dfrac{d_o}{2\lambda_w} \ln \dfrac{\mathrm{d}_o}{\mathrm{d}_i} + \dfrac{d_o}{2\lambda_c} \ln \dfrac{\mathrm{d}_i}{\mathrm{d}_c} + \dfrac{1}{\alpha_\mathrm{o}} + R_o + R_i \dfrac{d_o}{\mathrm{d_c}}} \tag{7}$$

Equation 8 gives the convective heat transfer coefficient inside the tube.

$$\alpha_\mathrm{i} = 0.023 \frac{\lambda_g}{d_c} Re^{0.8} Pr^{1/3} \left(\frac{\mu_g}{\mu_w}\right)^{0.14} \tag{8}$$

Equation 9 gives the convective heat transfer coefficient outside the tube.

$$\begin{aligned} \alpha_o^{4/3} &= h_c^{4/3} + h_r^{4/3} \\ h_c &= 0.62 \left[\frac{gr\rho_v(\rho_l - \rho_v)\lambda_v^3}{\mu_v \, \mathrm{d}_o (t_i - t_s)}\right]^{1/4} \\ h_r &= \frac{\varepsilon\sigma((T_w)^4 - (T_s)^4)}{T_w - T_s} \end{aligned} \tag{9}$$

3.2. Dynamic Coke Depositing Model.

The dynamic coke depositing model can simulate the thickness of coke deposition in the TLE tubes in one operation period under the given temperature distribution.

Equation 10 states that the flow rate of coke deposition is equal to the sum of all the physical condensation components, in which α represents the coke lay-down factor.

$$dm_c = \alpha \cdot dV_c \cdot \sum_i r_i \cdot M_i \tag{10}$$

Table 1 Coking rate and dew points of the physical condensation components

Name	$r_i(\mathrm{mol}/(m^3 \cdot \mathrm{s}))$	Dew Point(℃)	Molar mass(kg/mol)
M-Anthracene	9.624E-04	488	0.1922
Acenaphthalene	2.116E-03	317	0.1521
Anthracene	1.251E-03	381	0.1782
Chrysene	1.677E-03	380	0.2282

Equation 11 gives the rate of the coke thickness.

$$\frac{d\delta_c}{dt} = \frac{dm_c}{\rho_c \cdot dA_C} \tag{11}$$

The tube geometry parameters are given by equation 12.

$$dV_c = \frac{\pi}{4} \cdot {d_c}^2 \cdot dL$$
$$dA_c = \pi \cdot d_c \cdot dL \tag{12}$$

3.3. The integrated model.

We integrated two models by using quasi-steady assumption: 1) one operation period between decoking is divided into a number of time intervals; 2) the thickness of the coke deposition is considered as a constant in one time interval.

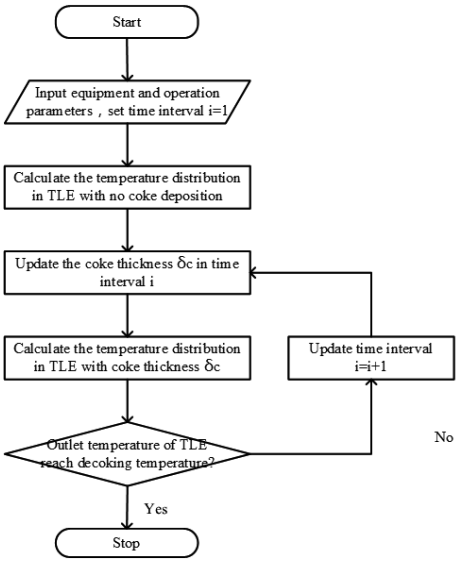


Figure 2. The algorithm diagram of the integrated model.

The algorithm diagram is shown in Figure 2. Firstly, the equipment and operation parameters are input and the time interval i is set to 1. Then, the temperature distribution in TLE with no coke is calculated. According to the temperature distribution, we update the coke thickness δ_c in time interval i. Based on the coke thickness δ_c, the temperature distribution in the time interval $i+1$ is calculated. Then, we check whether the outlet temperature of TLE reaches the decoking temperature. If the answer is "No", the time interval i is updated and the coke thickness δ_c is updated then. If the answer is "Yes", the operation period, coke deposition distribution and the temperature distribution are obtained.

4. Case Study

In this section, a case study of a TLE system is presented to illustrate the use of the proposed optimization model. The TLE configuration and the CG compositions and conditions are given in Table 2.

Table 2 Equipment and operation parameters of a TLE system

Parameters	Values	Component	Mass%
Tube length/m	9.5	H2	0.54
Tube OD/m	0.051	CH4	8.28
Tube ID/m	0.041	C2H4	15.33
Tubes number	50	C2H6	1.90
CG mass flow/(kg/hr)	7500	C3H6	8.07
CG inlet temperature/℃	840~855	C3H8	0.45
CG inlet pressure/MPa	0.072	C4H6	3.01
Steam temperature/℃	325	N-C4H8	0.71
Steam pressure/MPa	13.5	I-C4H8	1.33
Tube inside fouling/(m^2K/W)	0.000688	C5+	10.41
Tube outside fouling/(m^2K/W)	0.000174	H2O	50.00

Based on the data listed in Table 2, the optimal results of the operation period are obtained, which are shown in Figure 3~4. Figure 3 shows that the TLE operation period varies with the CG inlet temperature and the optimal operation period is obtained as 91.03 days with the inlet temperature 843.3℃. Figure 4 gives the TLE temperature distribution of the optimal inlet temperature. At the start of the period, the CG outlet temperature is about 327 ℃ with no coke deposition. As the coke deposition thickness grows, the outlet temperature increases rapidly at first, and increases smoothly then. At the end of the period, the outlet temperature reaches the decoking temperature.

5. Conclusions

This paper proposed an optimization model of the TLE system, which integrated a steady-state heat transfer model and a dynamic coke depositing model. The results of a case study showed that the optimization model could effectively increase the operation period between decoking. In future work, we will add the cracking furnace system into our model, which can optimize the operation period in both systems.

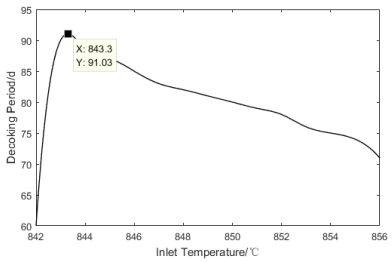


Figure 3. The optimal operation period of the TLE system.

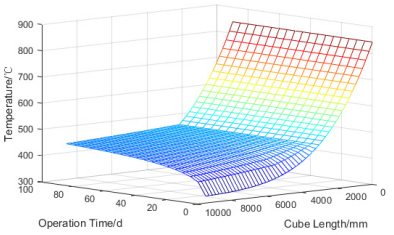


Figure 4. The temperature distribution of the TLE system.

Acknowledgement

This research was supported by the National Key Research and Development Program of China (2016YFB0901900), the Major Program of National Natural Science Foundation of China (71790614), the Fund for Innovative Research Groups of the National Natural Science Foundation of China (71621061), the Major International Joint Research Project of the National Natural Science Foundation of China (71520107004), and the 111 Project (B16009).

References

Bach, G., Zimmermann, G., Kopinke, F. D., Barendregt, S., Oosterkamp, P. V. D., & Woerde, H. (1995). Transfer-Line Heat Exchanger Fouling during Pyrolysis of Hydrocarbons. 1. Deposits from Dry Cracked Gases. *Industrial & Engineering Chemistry Research, 34*(34), 1132-1139.

Geem, K. M. V., Dhuyvetter, I., Prokopiev, S., Reyniers, M. F., Viennet, D., & Marin, G. B. (2009). Coke Formation in the Transfer Line Exchanger during Steam Cracking of Hydrocarbons. *Industrial & Engineering Chemistry Research, 48*(23), 861-868.

Jin, Y., Li, J., Du, W., Wang, Z., & Qian, F. (2013). Outlet Temperature Correlation and Prediction of Transfer Line Exchanger in an Industrial Steam Ethylene Cracking Process. *Chinese Journal of Chemical Engineering, 21*(4), 388-394.

Kopinke, F. D., Bach, G., & Zimmermann, G. (1993). New results about the mechanism of TLE fouling in steam crackers ☆. *Journal of Analytical & Applied Pyrolysis, 27*(27), 45-55.

Manafzadeh, H., Sadrameli, S. M., & Towfighi, J. (2003). Coke deposition by physical condensation of poly-cyclic hydrocarbons in the transfer line exchanger (TLX) of olefin plant. *Applied Thermal Engineering, 23*(11), 1347-1358.

Yaws, C. L. (1961). MATHESON GAS DATA BOOK. *Chem.eng.news*(38), 20.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50165-8

Water networks synthesis for industrial parks respecting to unpredictable scenarios

Linlin Liu[a], Yao Sheng [a], Lei Zhang[a], Jian Du[a*], Qingwei Meng[b]

[a] *Institute of Chemical Process Systems Engineering, School of Chemical Engineering, Dalian University of Technology, Dalian, 116024, China.*
[b] *State Key Laboratory of Fine Chemicals, School of Pharmaceutical Science and Technology, Dalian University of Technology, Dalian 116012, China.*
dujian@dlut.edu.cn

Abstract

Cooperative reuse of water resource across the plants placing in same industrial park would bring significant benefit to each participant plant, the park, and also the realization of resource-saving society. Different from designing water reusing networks merely subjecting to one or several pre-known operation scenarios, this study concentrates on solving the retrofit problem towards the unpredictable scenario which is upcoming yet unrealizable with current water allocation system. A hierarchical procedure is proposed, wherein by introducing series of specific constraints, the mixed integer non-linear programming model formulated for design purpose (Liu et al., 2017) is developed for achieving the optimal retrofit measures respecting to different retrofit specifications of decision-maker requires. At last, an industrial park example containing three retrofit cases at several typical scenarios is illustrated to show the application of the method.

Keywords: industrial park, water network, retrofit, unpredictable scenario

1. Introduction

An industrial park is a cluster of multiple individual plants locating in the same geographical area. This existence creates many extra opportunities for resources saving and sharing beyond plant level. Accordingly, increasing efforts have been contributed to inter-plant water integration due to its remarkable performance on water saving and cleaning for process industries in recent years.

Early work by Olesen and Polley (1996) first addressed water integration across plants based on pinch analysis, where a load table and a capacity diagram were employed. Liao et al. (2007) combined mathematical programming with pinch analysis insights to conduct flexible inter-plant water integration. Then direct and indirect water reusing modes were proposed by Chew et al. (2008) for the same purpose. Lovelady and El-Halwagi (2009) introduced a systematic water integration approach for the design of industrial park based on a source-interception-sink representation, and this work was extended into property-based integration by Rubio-Castro et al. (2013). Chen et al. (2010) developed a novel integration scheme, where water-using units within each plant were interconnected with the use of central and decentralized watermains. Alnouri et al. (2014) included pipe merging purpose into the design of inter-plant water network. Liu et al. (2016) reduced the synthesis complexity by building a plant-based mode for crossing-plant water reusing. In addition to mentioned work, the operation aspect of

inter-plant water integration becomes to gain growing attention in recent years. Faria and Bagajewicz (2011) studied a planning model for time-varying water system, taking the increase in number and mass load of water-using units into account. Burgara-Montero et al. (2013) introduced a multi-period optimization model to involve seasonal variability. Aviso (2014) developed a robust optimization approach for multi-period water integration of direct inter-plant reusing in eco-industrial parks, which satisfied various probable future scenarios. And recently, the same problem was also studied by Liu et al. (2017) by using indirect inter-plant water reusing pattern, accompanying with including in-plant and inter-plant regeneration operations.

Nevertheless, the mentioned study considers a single or several certain operation scenarios only for design purpose. In practice, the variations of many factors like prices of raw material and product, market requirement, government policy etc., will push decision-makers to change production scale so as to gain the maximum profit. In most situations, it is hard to predict these variations in very early advance, for example at the design stage of a process, but requires the plant to timely take retrofit actions on existent water network so as to satisfy the process specifications of coming operation scenario. Concerning this practical necessity, this study undertakes the challenge to handle unpredictable problem of water network retrofit for industrial park. The synthesis method proposed in our former study (Liu et al., 2017)) is extended and developed to catch the optimal network retrofit behaviours against the unanticipated events in future development of the park.

2. Problem statement

The problem addressed in this study is stated as follows: Given is a concerned industrial park with its existing water network, wherein the water reuse and regeneration operations are implemented inside and across individual plants for fresh water saving sake. The network information relating to currently running scenario is known, mainly including the stream connections, flowrates, concentrations and removal ratios of regenerators. However, this management may fail to satisfy a new or unplanned operation scenario so that the appropriate countermeasures are required. For this reason, the purpose of this work is to determine the best fresh water-saving retrofit schemes by re-allocating stream flowrates and re-designing stream connections based on the current network and the retrofit requests that for practical concerns.

3. Method

This retrofit method is developed from the multi-period synthesis model proposed by Liu et al. (2017), which aims at covering all predictable scenarios for network design. The multi-period model is used to build relationships between current network (period 1) and target scenario (period 2). Beyond its design application, we further involve running information of existent network into the model, moreover, new constraints towards the retrofit scope and strategy indicating where and which measures can be taken are also included for optimization. The method can be basically executed in three steps as follows.

Step1. Determine original network. The original network refers to an existent water network in operation. It can be obtained using either single-period water network synthesis method or multi-period synthesis method.

Step2. Determine the scope/part of network where and which retrofit measures are permitted. Production variations commonly occur in one or several plants, rather than the whole industrial park. Therefore, to maintain the stability of other plants, the retrofit measures should be restricted in a certain scope according to process specifications and tolerable fluctuation degree of other plants. With this concern, we can determine which network elements, mainly referring to stream connections and flowrates, are changeable and which are fixed. This information will be used in the mathematical model-based optimization of step 3 for retrofit purpose.

Step3. Optimize retrofit measures. Apply the mathematical model proposed for design purpose to determine the optimal retrofit measures towards a coming park scenario with target of minimum fresh water consumption. The mathematical model used in this study is absorbed from Liu et al., (2017), having a superstructure containing plentiful water allocation schemes formulated. While moreover for retrofit intents, series of specific constraints for stream connections, flowrates and concentrations are also included in the model for retrofit measure optimization according to the scope specified in step 2.

4. Case Study

This study directly takes the active part (in solid lines) of the fifth period in Liu et al. (2017)(Figure 3) as the existent water network, named Case 0, to show the application of the method on solving unpredicted operation problem. The major variation concerned in this work is the significant adjustment of production scale, however other variation scenarios can be also handled in the same way. Assume just the production scale of plant A changes. The six scale adjustment scenarios (process data of original scale in Liu et al., 2017) of +10%, -10%, +20%, -20%, +30%, -30% are investigated and analyzed based on the three retrofit cases set below. To meet these updating water using requirements, new allocations featuring minimum fresh water consumption are desired.

Case 1: *Re-design water network A with holding the rest connections in network, meanwhile adjust all stream flowrates.*

In this case, two parts of retrofit work are simultaneously implemented towards each adjustment scenario: Part 1, re-design the sub-network structure including water stream connections and flowrates for plant A, and Part 2, meanwhile retain stream connections outside plant A in base case Case 0 however assign new flowrates that going through. In such, we extract constraints from Case 0 and act on Case 1, which insure the non-existent connections in the base network section that excludes plant A remain not exist in the new case. Use indicator f_0 (no A) to represent the flowrates of mentioned connection in general, then equation of f_0 (no A) = 0 is included in the model to handle the retrofit problem targeted by Case 1.

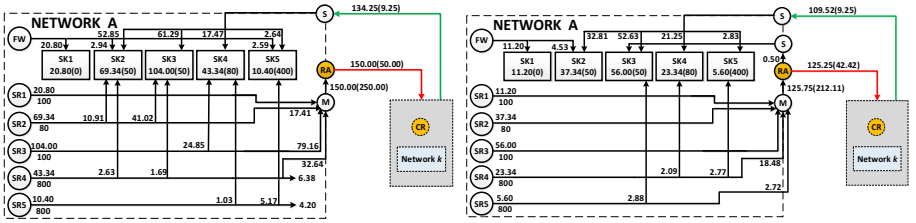
The six scenarios are individually studied, and the new network structures of plant A and new water allocation schemes are achieved accordingly. Optimization results are summarized in Table 1. As indicated, the minimum fresh water consumption alters with production scale. Figure 1 shows the newly designed network of Plant A at production scale of +30% and -30%. With the optimized network retrofit behaviour, the new networks are able to satisfy the new production requirements, meanwhile, conform to the retrofit stipulation of Case 1.

Table 1. Scenario solutions for Case 1

Scale change	0	+10%	-10%	+20%	-20%	+30%	-30%
Fresh water/t·h^{-1}	56	56.02	52.63	59.13	52.81	62.33	51.73

Case 2: *Keep all connections in network unchanged, but adjust stream flowrates, removal ratio and the using of fresh water.*

This case is trying to meet the operation scenario by just altering the water stream flowrates of Case 0, however it may fail because the fixed network structure may cause the flowrate adjustment of no avail. Against this situation, two strategies are proposed to facilitate the retrofit intention.



(a) Increases 30% (+30%) (b) decreases 30% (-30%)

Figure 1. Water network structure of plant A at its production scale changing 30 %

Strategy (1): Enhance removal ratios of regenerators to assist flowrates re-assignment. In this study, the removal ratios are increased in step of 0.01 until to get feasible re-assignment solutions for the new operation scenarios. In such way, Table 2 and 3 have shown the respective removal ratio and fresh water consumption at each scenario. Obviously, not all removal ratios but that of pre-treatment unit in Plant B needs to be intensified. This is because no matter Plant A scale becoming larger or smaller, concentrations of inter-plant streams going forward to centralized regenerator must meet the process constraints, and the flowrates must be sufficient for water sink demands. The base network structure with original removal ratios is not feasible for concerned scenarios, so it has to enhance.

Table 2. Removal ratio of Plant B in Case 2

Scale change	0	+10%	-10%	+20%	-20%	+30%	-30%
RC(B)	0.8	0.81	0.81	0.82	0.82	0.83	0.83

Table 3. Scenario solutions for Case 2

Scale change	0	+10%	-10%	+20%	-20%	+30%	-30%
Fresh water/t·h^{-1}	56	53.6	50.4	55.2	48.8	56.8	47.2

Strategy (2): Re-assign water stream flowrates with involving new fresh water pipelines while fixing network structure and regeneration operations. In such way, Table 4 collects the fresh water quantity added into each scenario and the data show that the more the plant scale is, the more extra fresh water is required for this case. In this study, the newly involved fresh water is added at inlet of centralized regenerator because there are concentration specifications towards streams going into the regenerator, however the fresh water can be also dispersed and directly used into water sinks if the mentioned

concentration limits are ignored. Anyhow, adding fresh water is a choice against the unpredictable scenarios.

Table 4. Added fresh water quantity for each scenario

Scale change	0	+10%	-10%	+20%	-20%	+30%	-30%
Added fresh water / t·h^{-1}	0	1.30	2.40	4.15	4.03	5.80	4.48
Total fresh water /t·h^{-1}	56	54.9	52.8	59.35	52.83	62.6	51.68

Case 3: *Keep all connections in network and operation parameters outside sub-network A unchanged, just adjust the stream flowrates of Plant A and the using of fresh water.*

In this case, the operation adjustment is restricted on the flowrates of connections in plant A and the plant related inter-plant streams, which is to say the elements in other plants including network structure and stream flowrates stay the same with that in base case. Such that according to the inter-plant water allocation diagram presented in Figure 2, the stream flowrates and concentrations should obey the constraints showing on right, meanwhile the inner-plant connections are restricted similarly.

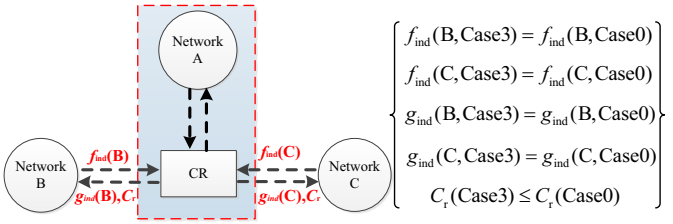
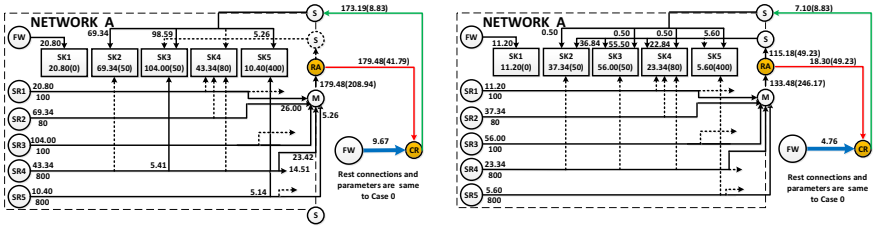


Figure 2. Inter-plant stream allocation diagram

Compared with that of former two cases, the retrofit scope of this case is significantly narrowed, so in order to accommodate the operation scenarios, adding inter-plant fresh water strategy proposed in last case is applied. Through retrofit design, we obtain new allocation solutions for the six unpredictable scenarios. Take +30% and -30% scenarios for example, their required supplementary fresh water quantity are 9.67 t·h^{-1} and 4.76 t·h^{-1}, total fresh water consumption are 66.47 t·h^{-1} and 51.96 t·h^{-1}, respectively. The network information is shown in Figure 3. Anyhow, this case offers the minimal adjustment measures to meet the retrofit purpose.



(a) Increases 30 % (+30 %) (b) decreases 30 % (-30 %)

Figure 3. Water network at production scale of plant A changes 30 %

Observe the three cases some conclusions can be obtained. Firstly, water network structure of a plant changes with its plant scale if the network retrofit is permitted. Secondly, stream flowrates of the same connection in two scale scenarios do not follow the relative ratio of the scales. Thirdly, the allocation details (not only flowrates but also concentrations) of inter-plant streams as well as other in-plant streams are influenced by the variation to a plant if the water using is whole-park considered. Therefore, it is necessary to implement retrofit operation for industrial park in a long-term perspective.

5. Conclusions

This study develops a three-step method for retrofit intentions over water networks of industrial park. The optimal retrofit behaviours are obtained by involving constraints and strategies into the mathematical model having existent network information included.The method is illustrated throguh three retrofit cases with receiving their individual optimal countermeasures against the unpredictable scenario. And the results have shown the ability of this method to provide a variety of feasible retrofit alternatives following decision-makers' intention.

Acknowledgements

The authors gratefully acknowledge the financial support from Natural Science Foundation of China (No. 21406026).

References

S. Y. Alnouri, P. Linke, M. El-Halwagi, 2014. Optimal interplant water networks for industrial zones: Addressing interconnectivity options through pipeline merging. AIChE J. 60(8), 2853-2874.

K. B. Aviso, 2014, Design of robust water exchange networks for eco-industrial symbiosis, Process Safety & Environmental Protection, 92(2):160-170.

O. Burgara-Montero, J. M. Ponce-Ortega, M. Serna-González, M. M. El-Halwagi, 2013, Incorporation of the seasonal variations in the optimal treatment of industrial effluents discharged to watersheds. Industial & Engineering Chemistry Research,52:5145-5160.

C. L. Chen, S.W. Hung, J. Y. Lee., 2010, Design of inter-plant water network with central and decentralized water mains. Computers & Chemical Engineering, 34(9):1522-1531.

I. M. L. Chew, R. Tan, D. K. S. Ng, D. C. Y. Foo, T. Majozi, J. Gouws, 2008, Synthesis of direct and indirect inter plant water network. Industial & Engineering Chemistry Research, 47(23), 9485-9496.

D. C. Faria, M. J. Bagajewicz., 2011, Planning model for the design and/or retrofit of industrial water systems. Industial & Engineering Chemistry Research, 50(7):3788-3797.

Z. W. Liao, J. T. Wu, B. B. Jiang, J. D. Wang, Y. R. Yang, 2007, Design methodology for flexible multiple plant water networks, Industial & Engineering Chemistry Research, 46(14):4954-4963.

L. Liu , J. Wang , H. Song, J. Du, F. Yang, 2016, Water networks synthesis for industrial parks considering inter-plant allocation. Computers & Chemical Engineering, 91,307-317.

L. Liu , J. Wang , H. Song, J. Du, F. Yang, 2017, Multi-period water network management for industrial parksconsidering predictable variations, Computers and Chemical Engineering, 104 :172-184.

E. M. Lovelady, M. M. El-Halwagi, 2009, Design and integration of eco-industrial parks for managing water resources, Environmental Progress & Sustainable Energy, 28(2):265-272.

S. G. Olesen, G. T. Polley, 1996, Dealing with plant geography and piping constrains in water network design, Transactions of the Institude of Chemical Engineers, Part B, 74(4), 273-276.

E. Rubio-Castro, J. M. Ponce-Ortega, M. Serna-Gonzalez, M. M. El-Halwagi, V. Pham, 2013, Global optimization in property-based inter-plant water integration. AIChE J. 59, 813-33.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50166-X

Hybrid Global Optimization Method for the Analysis of Integrated Reaction and Separation Processes

L.-S. Harding[a*], M. Schroeder[a], G. Fieg[a]

[a]*Institute of Process and Plant Engineering, Hamburg University of Technology, Am Schwarzenberg-Campus 4, 21073 Hamburg, Germany*
selin.harding@tuhh.de

Abstract

The reactive distillation and the reactive dividing wall column are examples of integrated reaction and separation processes, which show significant savings in investment as well as operational costs. Due to the high grade of integration the behavior of these processes is strongly nonlinear. This results in several local optima and thus the determination of energy optimal designs is quite challenging. Additionally, it is not trivial to understand and predict how the process performs in detail. Still, a fundamental process understanding is indispensable for a safe and energy efficient design of integrated reaction and separation processes. Therefore, a new hybrid global optimization method is presented, which is based on an evolutionary algorithm. This holistic approach is advantageous for optimization problems where the optimal values of the decision variables depend strongly on a certain decision variable. Dependencies of the process parameters can be deduced easily and a comprehensive process analysis can be carried out.

Keywords: reactive distillation, process integration, energy minimum design, global optimization, evolutionary algorithm

1. Introduction

Process intensification has become an important concept to meet the challenges of an increasing global competition in the chemical industry. A promising possibility of process intensification is the integration of primarily separated process operations into one single shell (Lutze et al. 2010). Examples are the reactive distillation (RD) or the reactive dividing wall column (RDWC) (Figure 1). These apparatuses enable a reaction and a simultaneous separation of the products. In the RD reactive systems with two products can be separated, whereas a separation of reactive systems with three product fractions can be carried out in the RDWC. These fractions can be products of the reaction, non-reacting components or excessive reactants (Müller 2010).

For energy optimal operation the component flows in the column and the resulting realized splits are important parameters. The split in the prefractionator of the RDWC can be characterized by the feedsplit. Latter states the molar fraction of the feed stream that leaves the upper part of the prefractionator (PF). The behavior in the PF of the RDWC equals the one of a reactive distillation, so that the feedsplit is equatable to the distillate/feed-ratio (D/F) in the RD (Schröder et al. 2016; Cici and Fieg 2017).

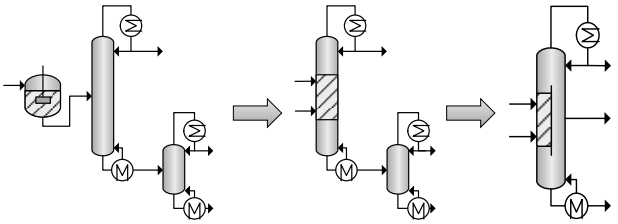


Figure 1: Levels of process integration: reactor and direct distillation column sequence (left), direct reactive distillation sequence (middle) and reactive dividing wall column (right)

Compared to conventional process alternatives integrated reaction and separation processes show significant savings in investment as well as operational costs (Yildirim et al. 2011; Harmsen 2007). These savings increase with the grade of process integration. Though, the complexity of the processes also increases with the grade of process integration. The process behavior of integrated processes is strongly nonlinear. Hence, the design is extremely demanding. A fundamental process understanding as well as reliable design methods are required for a safe and energy efficient design.

Energy minimum designs can be determined by process optimization. The optimization problems for integrated reaction and separation processes belong to the complex type of mixed integer non-linear programming (MINLP). Evolutionary algorithms (EAs) are suitable to solve such problems and are commonly used in science and industry (Chiong 2009). One example of an optimization tool based on EAs is the Advanced Process Optimizer (AdvPO). It is specially tailored for the optimization of chemical processes and is featured with an individual Elite-concept. This consists of an Eliteindex for storing the optimal solutions and an Elitecount for population development. The AdvPO additionally enables the linking of commercial flowsheet simulators and parallel mode calculation (Ernst et al. 2017).

Despite their good performance EAs do not guarantee to find the global optimum. To minimize the risk of convergence to local optima and gain knowledge about the influence of process parameters on the required energy demand at once, a new hybrid global optimization method has been developed. It is based on the algorithm of the AdvPO and is advantageous for optimization problems where the optimal values of the decision variables depend strongly on a certain decision variable. In the following chapter the algorithm will be described with its main characteristics.

2. Methodology

The functional principle of the developed hybrid global optimization method compared to the common method of single objective optimization is shown in Figure 2. In the common optimization method one optimization run yields one process design. Ideally, this result represents the global minimum of the objective. If the influence of a process parameter should be investigated, several optimization runs have to be carried out. Here the parameter of interest is changed manually for each optimization run. In every run the search for the optimum starts at the very beginning. No good genes (values of process parameters) are present from which the individuals (process designs) can profit. This procedure is, especially for fine parameter discretization, computationally intensive and quite time consuming.

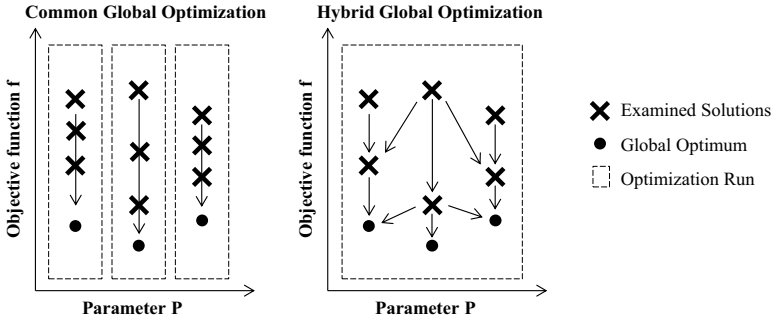


Figure 2: Functional principle of hybrid global optimization method in comparison to common single objective optimization

In comparison the hybrid global optimization method combines the analysis of process parameters and the optimization in one method. The idea behind this innovative concept is that the investigated parameters should only be varied in a practical reasonable interval. The design variables of the resulting optimal designs do not differ in a large range. So, individuals with a specific parameter value can also profit from good genes from individuals with different parameter values. As a result, fast convergence can be achieved for every parameter value.

The algorithm for this new method is based on the EA of the AdvPO, which is described in detail by Ernst et al. (2017). For an equal treatment of all parameter solutions the evaluation and Elite-concept are changed by the implementation of a parameter specific ranking, shown in Figure 3. Initially, the population is sorted by the different parameter values. The individuals of each parameter value are then ranked independently from individuals with different parameter values. For the Eliteindex the ranking is performed for the objective values, whereas for the evaluation the fitness values are ranked. Latter take the objective values as well as the constraints into account. As a result, at least one individual for each parameter value exists with the best rank.

Additionally, a high diversity of the parameter is required in each generation to find the global optimum for each parameter value during the optimization. Therefore, the places in the Elitecount, typically 20 % of the total population, as well as the places for the parents of the next generation are equally distributed to all parameter values. So, for all parameter values the same number of parents is chosen via a tournament selection and the same number of individuals is present in the Elitecount of a generation. Remaining places, which cannot be equally distributed, are randomly assigned to a parameter value. These changes ensure that the diversity of the parameter stays high during the whole optimization without hindering the convergence to optimal process designs. The genetic operations recombination and mutation stay unchanged, so that the genes from one parameter value can profit from good genes of another parameter value.

The progress of the optimization is checked separately for each parameter value. Therefore, the improvement of the objective value over a defined number of generations is analyzed. If no change is measured, the optimization is converged for this parameter value. It is sorted out of the population and the individuals are distributed to the remaining parameter values. The determined optima remains saved in the Eliteindex. The optimization run is terminated as soon as all parameter values are converged.

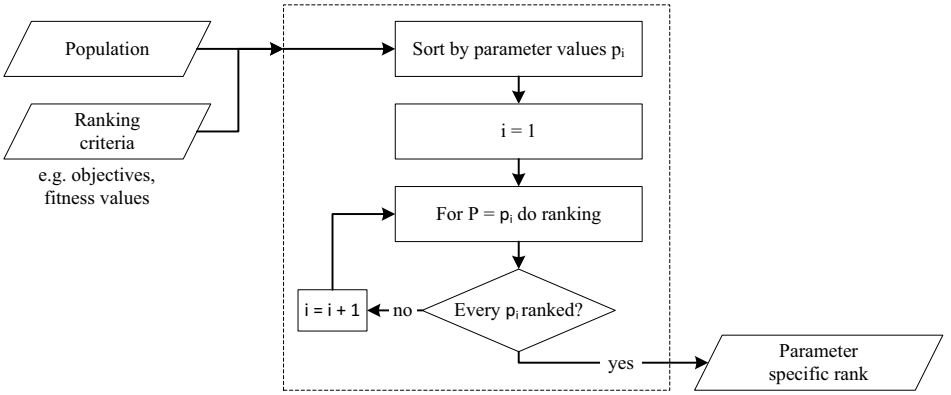


Figure 3: Schematic flow diagram of parameter specific ranking in hybrid global optimization

2.1 Optimization Problem Formulation

The hybrid global optimization method is demonstrated on the example of a reactive distillation. The column model is implemented in Aspen Custom Modeler, which is linked to the optimizer via an DLL-interface. A quinary reactive system given in Eq. (1) is investigated. For a more detailed description of the reaction system the reader is referred to Schröder et al. (2016). The amount of each feed stream into the column is 50 kmol/h with a molar fraction of 10 mol% of inert and a molar fraction of 90 mol% of reactant.

$$A + B + E \rightleftharpoons C + D + E \quad (1)$$

As stated in the introduction, the D/F ratio is an important parameter for energy optimal operation. Thus, the influence of this ratio on the minimum vapor demand of the RD is investigated. So, the objective of the optimization is to minimize the vapor demand of the RD. As a constraint a minimal conversion of 99.5 % of the reactant A is demanded. The decision variables for the optimization with their bounds and step size are presented in Table 1, in which the first decision variable equates the parameter of the D/F ratio.

Table 1: Summary of decision variables with bounds and step size for the optimization of a reactive distillation column

Decision Variable	Lower Bound	Upper Bound	Step Size	Unit
D/F ratio	0.45	0.55	0.01; 0.001	-
Upper feed position	1	1.99	0.01	-
Distance of feed positions	0	20	1	-
Vapor duty	100	400	1	kmol/h
Number of stages top segment	1	50	1	-
Number of reactive stages	60	100	1	-
Number of stages bottom segment	1	50	1	-

3. Results

3.1 Validation

For the validation of the new hybrid global optimization method the obtained results are compared to results of a systematic, complete search of the solution space in Figure 4. As it can be seen for every D/F ratio a minimum vapor demand has been determined by the new method. Nearly all results match the ones from the complete search of the solution space, so that the hybrid global optimization method is validated successfully.

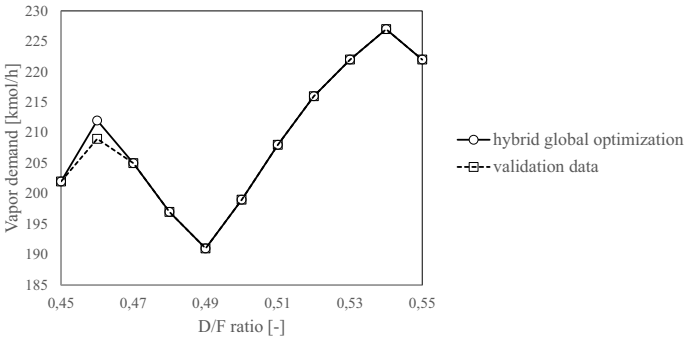


Figure 4: Results of the optimization of the vapor demand of a reactive distillation of the hybrid global optimization method in comparison to validation data

3.2 Analysis of Integrated Reaction and Separation Processes

As a second run the D/F ratio in the region of the global minimum was varied in a smaller step size to get a more detailed distribution. The results are shown in Figure 5. As expected, it can be seen that the vapor demand is highly influenced by the D/F ratio. The correlation of the D/F ratio and the vapor demand shows various saddle points, which hinder the determination of the global optimum with common global optimization methods. The new hybrid global optimization method allows identifying the correlations between the process parameters and thus minimizes the risk of convergence to local optima.

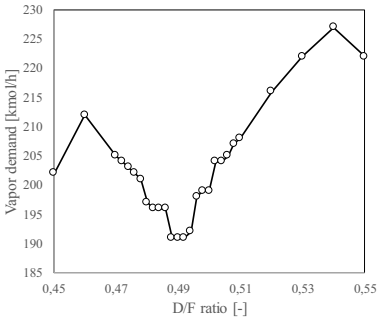


Figure 5: Correlation of minimum vapor demand and D/F ratio in the RD

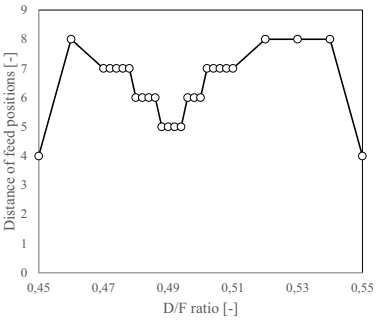


Figure 6: Correlation of distance of feed positions and D/F ratio in the RD

Additionally, further process analysis can be carried out by analyzing the dependencies of other process parameters. In Figure 6 the dependency of the distance of the feed positions and the D/F ratio is presented exemplarily.

4. Conclusions

The proposed hybrid global optimization method is characterized by an EA with a special parameter based ranking and a high diversity of the parameter during the whole optimization. In principle, this method is advantageous for optimization problems where the optimal values of the decision variables depend strongly on a certain decision variable. In these cases, this approach helps to find the optimal solutions more reliably and with less manual effort. An evaluation of this innovative method was carried out on the example of a reactive distillation column. In future work it is planned to investigate this method for further examples. The obtained results showed a good agreement with validation data. Additionally, correlations of the minimum vapor demand, the D/F ratio and the distance of feed positions were identified. So, the hybrid method provides a holistic, time-efficient approach for the design of integrated reaction and separation processes. It enables the determination of the influence of a process parameter and the energy minimum design of apparatuses. Thus, correlations of process parameters can be deduced easily and local optima of the process designs can be identified.

References

R. Chiong, 2009, Nature-inspired Algorithms for Optimisation, Berlin.

L.-S. Cici; G. Fieg, 2017, Influence of non-ideal behavior of the reactive dividing wall column on its energy saving potential, 10th World Congress of Chemical Engineering, Barcelona, Spain.

P. Ernst; K. Zimmermann; G. Fieg, 2017, Multi-objective Optimization-Tool for the Universal Application in Chemical Process Design, Chem. Eng. Technol., 60, 7, p. 1867–1875.

G. J. Harmsen, 2007, Reactive distillation, The front-runner of industrial process intensification, Chemical Engineering and Processing: Process Intensification, 46, 9, p. 774–780.

P. Lutze; R. Gani; J. M. Woodley, 2010, Process intensification, A perspective on process synthesis, Chemical Engineering and Processing: Process Intensification, 49, 6, p. 547–558.

I. Müller, 2010, Theoretische Untersuchung innovativer integrierter Trennverfahren, Munich.

M. Schröder; C. Ehlers; G. Fieg, 2016, A Comprehensive Analysis on the Reactive Dividing-Wall Column, its Minimum Energy Demand, and Energy-Saving Potential, Chem. Eng. Technol., 39, 12, p. 2323–2338.

Ö. Yildirim; A. A. Kiss; E. Y. Kenig, 2011, Dividing wall columns in chemical process industry, A review on current activities, Separation and Purification Technology, 80, 3, p. 403–417.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50167-1

Simultaneous Synthesis of WHEN Based on Superstructure Modelling Considering Thermodynamic and Economic Factors

Yu Zhuang[a], Lei Zhang[a], Linlin Liu[a], Qingwei Meng[b], Jian Du[a*]

[a]*Institute of Process Systems Engineering, School of Chemical Engineering, Dalian University of Technology, Dalian 116012, Liaoning, China*
[b]*State Key Laboratory of Fine Chemicals, Chemical Engineering Department, Dalian University of Technology, Dalian 116024, Liaoning, China*
dujian@dlut.edu.cn

Abstract

To explicitly reveal the interaction mechanism between work and heat integration considering both thermodynamic and economic analysis, an improved superstructure, coupling of work exchange networks (WEN) and heat exchange networks (HEN), is proposed for simultaneous synthesis of work and heat exchange networks (WHEN) in this paper. The formulations based on exergy analysis are developed to determine the cold or hot identity of process streams. Afterwards, economic analysis is performed by formulating a cost-based mixed-integer nonlinear programming (MINLP) model to optimize the sequence of work and heat integration, aiming to minimize the total annual cost (*TAC*). Finally, an example study selected from the open literature is conducted to demonstrate the efficacy of the proposed methodology, where the solutions show a diminution of 40.4% in TAC of HEN prior to WEN, compared with that of WEN prior to HEN.

Keywords: work and heat exchange network; superstructure; exergy analysis; economic analysis; sequence

1. Introduction

Heat-based thermal energy and work-based mechanical energy are two common forms of energy consumed significantly in chemical process industries.(Aida-Rankouhi and Huang, 2017) While heat integration has become a mature technology that is widely used for available thermal energy recovery via heat exchange networks, work integration, however, has not been fully explored. Interestingly, work integration can contribute significantly to mechanical energy recovery through synthesizing work exchange networks (WEN). (Zhuang et al., 2017c)

From the thermodynamics point of view, pressure and temperature are two important thermodynamic parameters for interactive conversion between work and heat. (Fu and Gundersen, 2016) Especially significant at sub-ambient conditions, Wechsung et al. (2011) proposed an optimization formulation allowing for the inter-conversion of work, pressure, and temperature-based exergy to reduce usage of costly cold utility. In addtion, considerable cost savings can be achieved by simultaneous synthesis of work-heat exchanger networks (WHEN) based on the multi-stage superstructure. (Huang and Karimi, 2016) However, the above-mentioned works overemphasize that pressure

manipulation is essential for enhancing heat integration, ignoring the interaction between work integration and heat integration.

In this article, a superstructure for WEN coupled with HEN is proposed to simultaneously synthesize WHEN. Based on exergy analysis, the cold or hot identity of process streams is confirmed prior to WHEN design. Afterwards, economic analysis is performed through a developed cost-based MINLP model to optimize the sequence of work and heat integration, aiming at the minimum *TAC*. Finally, an example study is conducted to demonstrate the efficacy of the proposed methodology.

2. Problem statements

A set of gaseous streams at high pressure (HP) and low pressure (LP) are present in a chemical process, with known flowrates, inlet and outlet pressure, inlet and outlet temperature, heat capacity and heat transfer coefficients etc. The purpose of this paper is to design a network configuration of WEN coupled with HEN to attain the minimized TAC. To simplify the synthesis procedure, the assumptions in our previous work (Zhuang et al., 2017b) are adopted.

3. Model formulation

3.1. Improved superstructure for WEN coupled with HEN

An improved stage-wise superstructure is proposed to achieve simultaneous synthesis of WHEN, consisting of two distinct, but interconnected networks. One (HEN) is for heat integration and the other (WEN) is for work integration, as shown in Figure 1 and Figure 2. From these two figures, it can be found that one concern is the streams should pass through HEN or WEN first, which is optimized in our latter model.

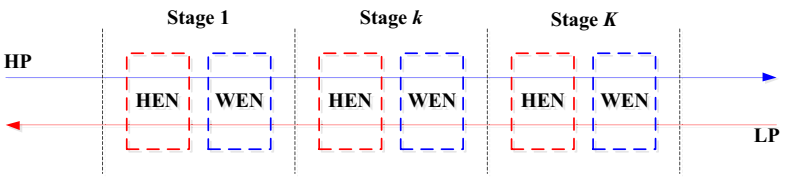


Figure 1 Stage-wise superstructure for synthesizing WHEN

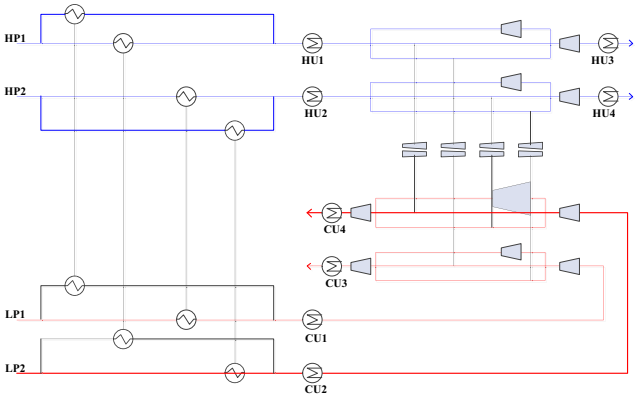


Figure 2 WHEN superstructure in each stage

Another concern is about how to determine the cold or hot identity of HP and LP streams. To settle this problem, an exergy analysis model is established for expanders, compressors and work exchangers, as shown in Figure 3. According to the corresponding exergy balance, the following exergy loss formulations can be obtained.

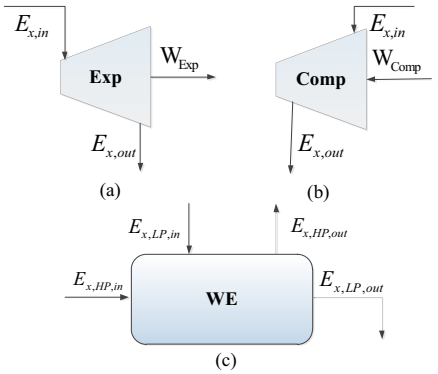


Figure 3 The exergy analysis for expander (a), compressor (b) and work exchanger (c)

$$\Delta E_{X,Exp} = FCpT_0 \ln\{1+\eta\cdot[(\frac{P_{out}}{P_{in}})^{\frac{\gamma-1}{\gamma}}-1]\} + FT_0R\ln(\frac{P_{in}}{P_{out}}) \tag{1}$$

$$\Delta E_{X,Comp} = FCpT_0 \ln\{1+\frac{1}{\eta}\cdot[(\frac{P_{out}}{P_{in}})^{\frac{\gamma-1}{\gamma}}-1]\} + FT_0R\ln(\frac{P_{in}}{P_{out}}) \tag{2}$$

$$\Delta E_{X,WE} = 0 \tag{3}$$

Eqs. (1)-(3) indicate that HEN bears little effect on WEN based on exergy losses analysis. Consequently, since recoverable energy from HP streams increases with inlet temperature while the required energy of LP streams decreases with inlet temperature, thus HP (LP) streams will be treated as cold (hot) streams in HEN.

3.2. Objective function and constraints

In this context, the main objective of the proposed mathematical is to optimize the WHEN synthesis, by minimizing *TAC*, composed of total annual cost of HEN (TAC_{HEN}) and total annual cost of WEN (TAC_{WEN}). Additionally, the TAC for single HEN and WEN both consists of capital investment and operational expenditure, which is based on our previous study presented in Zhuang et al. (2017a) and the famous optimization model for HEN synthesis proposed by Yee and Grossmann (1990).

As for the mathematical formulation, it also refers to the above two open literatures, which includes equipment design equations (such as work exchangers, expanders, compressors, heat exchangers, heaters and coolers), mass and energy balances for heat recovery and pressure manipulation, some logical constraints to ensure feasible solutions and design relationships with physical meaning. The key strategies are to determine the cold or hot identity of streams and optimize the sequence of WEN and HEN synthesis, which facilitates synthesizing WHEN. Since there are more than 40 equality and inequality constraints, thus the detailed representation for each constraint

cannot be shown hereby, due to the limit of six-page manuscript. Despite lack of detailed description, this paper still provides a novel superstructure-based optimization model with thermodynamic and economic factors to simultaneously synthesize WHEN so that the thermal and mechanical energy consumption can be sharply reduced.

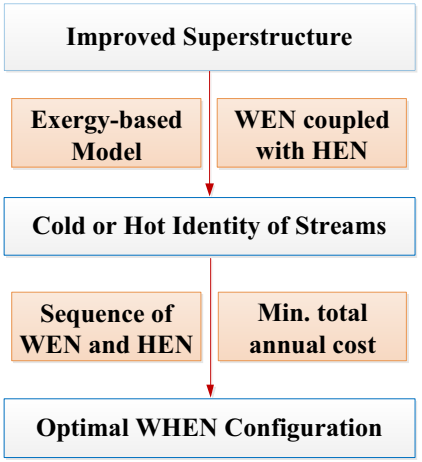


Figure 4 Flow chart of the proposed method.

4. Example study

In this section, the example from Onishi et al. (2014) has two high-pressure streams and two low-pressure streams, where the detailed stream data is shown in Table 1.

Table 1 Process stream data

Stream	*F* (kg/s)	*Cp* (kJ/kg K)	*h* (kW/m^{2}K)	*TIN* (K)	*TOUT* (K)	*PIN* (kPa)	*POUT* (kPa)
HP1	12	2.454	0.1	410	600	900	100
HP2	18	0.982	0.1	355	500	850	150
LP1	15	1.432	0.1	600	350	100	700
LP2	18	2.454	0.1	600	360	100	900

Based on the presented optimization procedure, a two-stage superstructure for WHEN synthesis is introduced to deal with the example. Since the outlet temperature of two high-pressure streams is higher than their respective inlet temperature, each HP stream is treated as cold stream. Similarly, each LP stream is treated as hot stream. Then by targeting the minimized TAC, the formulated MINLP model is solved by BARON solvers in GAMS 24.0.

Interestingly, two cases are studied, one of which considers WEN prior to HEN while the other allows for HEN prior to WEN, represented by Case A and Case B, respectively. In both cases, TAC minimization, composed of capital investment and operational expenditure of both HEN and WEN, is considered to be the objective function. Clearly, the results of these two cases are provided in Table 2 and the corresponding optimal WHEN configurations are shown in Figure 5 and Figure 6.

From Table 2 and Figure 5, 6, it can be seen that our approach adopting Case B yields an optimal configuration with a TAC of 14937 k$/y, a 40.4% decrease to that obtained by Case A. This is because HEN contributes to improving the inlet temperature of HP streams and reducing the inlet temperature of LP streams prior to WEN, which increases the available mechanical energy of HP streams and decreases the demanded energy of LP streams. Thus, the work utility consumption (W_U) with higher cost is considerably decreased despite the increment of heat utility consumption with lower cost.

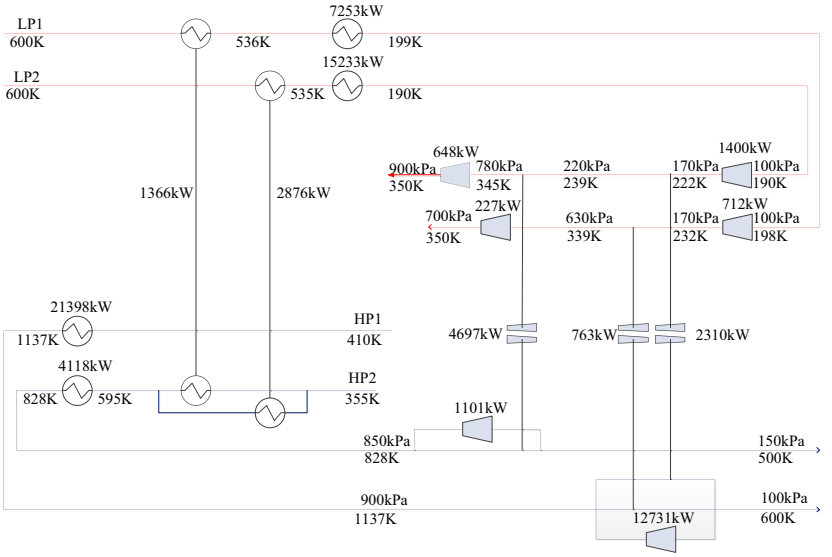


Figure 5 Optimal WHEN configuration for Case A

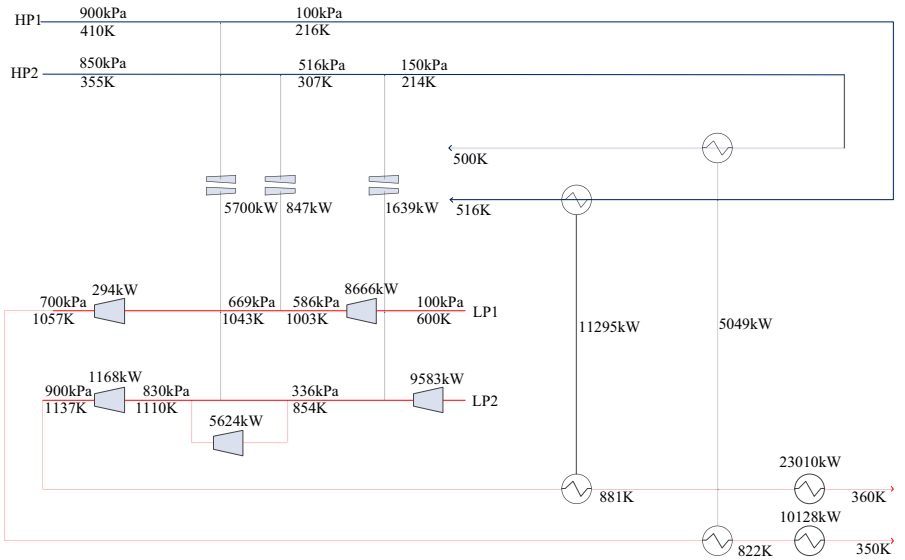


Figure 6 Optimal WHEN configuration for Case B

Table 2 Solution comparison of two different cases

Item	W_U (kW)	Q_U (kW)	TAC_{HEN} (k$/y)	TAC_{WEN} (k$/y)	TAC (k$/y)
Case A	25,335	33,138	279	24,763	25,042
Case B	16,819	48,002	532	14,405	14,937

5. Conclusions

In this paper, a superstructure-based mathematical model for simultaneously synthesizing WHEN is proposed, in order to explicitly reveal the interaction mechanism between work and heat integration considering both thermodynamic and economic analysis. The exergy analysis is conducted to determine the cold or hot identity of process streams. After that, the developed model is formulated as MINLP problem with the objective of minimized TAC. Finally, an example study is carried out to show that our approach yields a HEN-WEN configuration with 40.4% lower TAC than that of WEN-HEN structure.

Acknowledgement

The authors deeply appreciate the financial support from Natural Science Foundation of China (No. 21776035) and (No. 21576036).

References

A. Aida-Rankouhi, Y. Huang, 2010, Prediction of maximum recoverable mechanical energy via work integration: A thermodynamic modeling and analysis difference, AIChE Journal, 63, 4814-4829.

C. Fu, T. Gundersen, 2016, Heat and work integration: Fundamental insights and applications to carbon dioxide capture processes. Energy Conversion and Management, 121, 36-48.

K. Huang, I. Karimi, 2016, Work-heat exchanger network synthesis (WHENS), Energy, 113, 1006-1017.

V. Onishi, M. Ravagnani, J. Caballero, 2014, Simultaneous synthesis of work exchange networks with heat integration, Chemical Engineering Science, 112, 87-107.

A. Wechsung, A. Aspelund, T. Gundersen, PI. Barton, 2011, Synthesis of heat exchanger networks at sub-ambient conditions with compression and expansion of process streams, AIChE Journal, 57, 2090-2108.

T.F. Yee, I.E. Grossmann, 1990, Simultaneous optimization models for heat integration – Heat exchanger network synthesis, Computers & Chemical Engineering, 14, 1165-1184.

Y. Zhuang, L. Liu, J. Du, 2017a, Direct work exchanger network synthesis of isothermal process based on superstructure method, Chemical Engineering Transactions, 61, 133–138.

Y. Zhuang, L. Liu, Q. Liu, J. Du, 2017b, Step-wise synthesis of work exchange networks involving heat integration based on the transshipment model, Chinese Journal of Chemical Engineering, 25, 1052-1060.

Y. Zhuang, L. Liu, L. Zhang, J. Du, 2017c, Direct work exchanger network synthesis of isothermal process based on improved transshipment model, Journal of the Taiwan Institute of Chemical Engineers, 81, 295–304.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50168-3

Optimization of cooling water system with compression refrigeration cycle

Jiaze Ma[a], Yufei Wang[a*], Xiao Feng[b]

[a]*State Key Laboratory of Heavy Oil Processing, China Uni. of Petroleum, Beijing*
[b]*School of Chemical Engineering & Technology, Xi'an Jiaotong University, China*
wangyufei@cup.edu.cn

Abstract

Cooling water system and chilled water system are wildly used in industry to reject waste heat. When cooling water is not capable to cool down the hot stream to target temperature, chilled water is required. So far, cooling water system and chilled water system are designed separately because the two systems use water with different qualities. Desalted water is used in chilled water system but it cannot be used in cooling water system, because water losing is huge in open cooling tower. Meanwhile, fresh water cannot be used in chilled water system because of high concentration of impurities. In this work, closed cooling tower is applied, because there is no lose of recirculating water in closed cooling tower, desalted water can be used in cooling water system. A large amount of water was saved by using closed cooling tower. Cooling water system and chilled water system was integrated into one, cooling duty was redistributed between the two systems. Stage-wise heat exchanger network was employed where water can be reused. For returned hot water, they are firstly sent to closed cooling tower to be cooled down to a certain temperature, and then part of the water is further cooled down by refrigeration cycle. The model is formulated as mixed-integer nonlinear programming (MINLP) problem. The objective is to minimize the cost of integrated cooling and chilled water system, and determine the optimal heat load distribution between closed cooling tower and refrigeration cycle. Results show that optimization model yields significant reduction on total annual cost and energy consumption, in comparison with system where cooling water system and refrigeration cycle are separated.

Keywords: Cooling water, Chilled water, MINLP, Optimization.

1. Introduction

Cooling water system is widely used in industry to reject waste heat. Many works have been done to study the cooling water system and formulate the cost saving and water saving system. Earlier work focus on minimizing water consumption and improving tower efficiency (Kim and Smith. 2001). Later, Ponce-Ortega et al. (2007) used the superstructure approach to optimize cooling water system and formulated cost saving cooling network. It is quite effective to use cooling water as an intermedium to transfer waste heat from industry to environment, owning to the good thermal property of water. However, when the hot streams target temperature is below ambient temperature, the chilled water is needed. Compression refrigeration system produces chilled water that meets the demand of heat exchanging service in industry. Researchers studied the refrigeration system from different perspectives. For example, a recent work conducted

by Yin et al. (2016) formulated cascade vapor compression system that is energy efficiency. Jain et al. (2016) optimized a cascade vapor-absorption refrigeration system and yield a cost saving system with great thermodynamic performance. Both of cooling water and refrigeration system have been studied thoroughly. However, no research has ever integrated two systems into one and explored the interconnections between. This is because different water was used in two systems. Fresh water in cooling water system cannot be used as chilled water. Employing closed cooling tower makes it possible to integrate two systems into one, since the clean recirculating water has no contact with air and the water is qualified to be used as chilled water. Moreover, in cooling water system, because fresh water is replaced by desalted water, the occurrence of scaling in coolers can be prevented and the cost for scaling inhibitor is avoided. In this work, we proposed a method that allows us to optimize two systems simultaneously. Given a set of hot streams, cooling water as well as chilled water is employed to cool down the hot streams to target temperature. For the conventional method, the outlet temperature of cooling tower is specified. The first step is to formulate cooling water network. Since the cooling water cannot cool down some hot streams to target temperature, the chilled water network is formulated. In this study, the proposed method is capable to integrate two systems simultaneously. As shown in figure 1, hot streams were cooled down by cooling and chilled water. Then the cooling water and chilled water mix and enter the closed cooling tower. A partial of outlet water of cooling tower enters heat exchanger network as cooling water, and the rest of water is cooled down by refrigeration cycle as chilled water. The outlet temperature of cooling tower and the flowrate of cooling and chilled water are the optimized variables. Some hot streams are cooled down by cooling and chilled water sequentially, the partition temperature of hot stream is also an optimized variable.

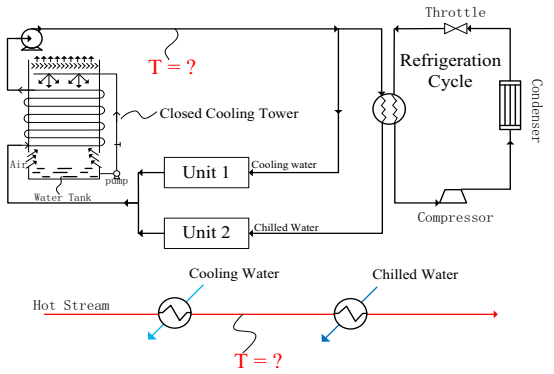


Figure 1. Illustration of two-stage cooling configuration.

2. Model formulation

In this paper, the stage-wise cooler network is employed. The following section describes the equations for the MINLP model. The set HP represents all the hot streams, set CP denotes cold stream which includes cooling and chilled water, set ST represents all the stages in the superstructure. *i*, *j*, *k* represents individual hot, cold streams and stages. In Eq.(1), $th_{i,k}$ is the temperature of hot stream *i* at stage *k*. The hot stream enters at first stage and the temperature of first stage is inlet temperature of hot stream. Target temperature of hot stream is defined as the last stage temperature. In Eq.(2), NOK is the total number of stages, $Thin_i$ and $Thout_i$ are inlet and target temperature of hot stream.

$$th_{i,1} = Thin_i \qquad i \in HP \quad (1)$$

$$th_{i,NOK+1} = Thout_i \qquad i \in HP \quad (2)$$

The cold streams enter at last stage of network and the inlet temperature of cold streams $tc_{j,k}$ equals to the last stage temperature of cold stream i. Similarly, cold streams exit network at first stage and outlet temperature of cold stream i corresponds to the first stage temperature of cold stream i.

$$tc_{j,NOK+1} = Tcin_j \qquad j \in CP \quad (3)$$

$$tc_{j,1} = Tcout_j \qquad j \in CP \quad (4)$$

Eq.(5) and Eq.(6) indicate that the temperature of hot and cold streams decrease constantly through stages of superstructure.

$$th_{i,k} \geq th_{i,k+1} \qquad i \in HP, k \in \mathrm{ST} \quad (5)$$

$$tc_{j,k} \geq tc_{j,k+1} \qquad j \in CP, k \in \mathrm{ST} \quad (6)$$

In Eq.(7), binary variable $Z_{i,j,k}$ is employed to determine the existence of exchanger between hot stream i and cold stream j in stage k. Eh_i is total heat content of hot stream i, $q_{i,j,k}$ is heat load of the exchanger. The heat load of that exchanger is not higher than total heat content of hot stream i.

$$q_{i,j,k} \leq z_{i,j,k} \cdot Eh_i \qquad i \in HP, j \in CP, k \in \mathrm{ST} \quad (7)$$

Energy balance for each match of the superstructure are as Eq.(8) and Eq.(9). Fh_i is heat capacity flowrate of hot stream i, and fc_j is mass flowrate of cold stream j and cp is heat capacity of water.

$$\sum_{i \in HP} q_{i,j,k} = fh_i \cdot \left(th_{i,k} - th_{i,k+1}\right) \qquad i \in HP, j \in CP, k \in \mathrm{ST} \quad (8)$$

$$\sum_{j \in CP} q_{i,j,k} = fc_j \cdot cp \cdot \left(tc_{j,k} - tc_{j,k+1}\right) \qquad i \in HP, j \in CP, k \in \mathrm{ST} \quad (9)$$

In Eq.(10) and Eq.(11), $dt_{i,j,k}$ represent temperature difference for the match between the hot stream i and cold stream j at stage k. $\Omega_{i,j}$ is upper limit for the temperature difference for hot stream i, the value of $\Omega_{i,j}$ is obtained by Eq.(12).

$$dt_{i,j,k} \leq th_{i,k} - tc_{j,k} + \left(1 - z_{i,j,k}\right) \cdot \Omega_{i,j} \qquad i \in HP, j \in CP, k \in \mathrm{ST} \quad (10)$$

$$dt_{i,j,k+1} \leq th_{i,k+1} - tc_{j,k+1} + \left(1 - z_{i,j,k}\right) \cdot \Omega_{i,j} \qquad i \in HP, j \in CP, k \in \mathrm{ST} \quad (11)$$

$$\Omega_{i,j} = max\left\{0, thin_i - tcin_j, thout_i - tcin_j, thin_i - tcout_{\max}, thout_i - tcout_{\max}\right\} \qquad i \in HP, j \in CP \quad (12)$$

The exchanger area between hot stream i and cold stream j at stage k is :

$$Ar_{i,j,k} = q_{i,j,k} \cdot \left(h_w^{-1} + h_i^{-1}\right) \cdot \left[dt_{i,j,k} \cdot dt_{i,j,k+1} \cdot 0.5 \cdot \left(dt_{i,j,k} + dt_{i,j,k+1}\right)\right]^{-1/3} \quad (13)$$

Energy consumption of fan depends on volume flowrate of air V_{air} and pressure drop through the exchanger inside the tower. In Eq.(14), N is the number of bundle, B represents the amount of spraying water and v_{air} is air velocity. The volume flowrate of air depends on velocity of air as well as exchanger frontal area A_{fa}. And power of fan is obtained through Eq.(16).

$$\Delta P = 2.7 \cdot N \cdot B^{0.12} \cdot (v_{air} \cdot \rho_{air})^{1.54} \quad (14)$$

$$V_{air} = v_{air} \cdot A_{fa} \quad (15)$$

$$P_{fan} = \Delta P_{fan} \cdot V_{air} / \eta_{fan} \quad (16)$$

Eq.(17-20) gives the capital cost of components of cooling tower. (Zalewski et al. 2000)

$$CC_{fan} = (21.56 + 20.15 \cdot P_{fan}^{0.5})^2 / 1.16 \quad (17)$$

$$CC_{bath} = (599.9 + 1101.5 \cdot A_{fa}) / 1.16 \quad (18)$$

$$CC_{exchanger} = 310 + 170 \cdot A_{tower} \quad (19)$$

$$CC_{pump} = 2000 + 5 \cdot (B \cdot A_{fa} \cdot \Delta p / \rho)^{0.68} \quad (20)$$

The cooling tower has to provide cooling water for process hot streams as well as the condenser in refrigeration cycle. The heat load of cooling tower is obtained through Eq.(21).*ttin* denotes the inlet water temperature of cooling tower, and $tcin_1$ represents the outlet temperature of cooling tower, it is also the cooling water temperature inlet of the heat exchanger network. fc_1 and fc_2 are the flowrate of cooling water and chilled water. q_{con} is the heat load on condenser.

$$q_{tower} = (ttin - tcin_1) \cdot cp \cdot (fc_1 + fc_2) + q_{con} \quad (21)$$

Eq.(22) is used to calculate the cooling tower inlet temperature *ttin*. $Tcout_1$ and $Tcout_2$ are the return temperature of cooling water and chilled. f_{con} and $tout_{con}$ are the flowrate and outlet temperature of cooling water provided for condenser in refrigeration cycle.

$$ttin = (fc_1 \cdot tcout_1 + fc_2 \cdot tcout_2 + f_{con} \cdot tout_{con}) / (fc_1 + fc_2 + f_{con}) \quad (22)$$

Heat load of evaporator is obtained through Eq.(23)

$$q_{evap} = fc_2 \cdot cp \cdot (tcin_1 - tcin_2) \quad (23)$$

Energy balance around the evaporator gives flowrate of refrigerant. In Eq.(24), ΔH_{evap} is the difference between inlet and outlet enthalpy of evaporator.

$$fm_{ref} = q_{evap} / \Delta H_{evap} \quad (24)$$

The capital cost and power required of compressor are given by Eq.(25) and Eq.(26). COP is coefficient of performance and η_{isen} is isentropic efficiency of compressor. And the heat load of condenser is given by Eq.(27). (Jain et al. 2016)

$$CC_{compressor} = 573 \cdot fm_{ref} \cdot (0.8996 - \eta_{isen})^{-1} \quad (25)$$

$$W_{compressor} = Q_{evap} \cdot COP_{ref}^{-1} \cdot \eta_{isen}^{-1} \quad (26)$$

$$q_{con} = q_{evap} + w_{compressor} \quad (27)$$

Finally, the objective function is expressed as Eq.(28), which includes capital and operation cost of cooling tower and refrigeration cycle, as well as capital cost of coolers in heat exchanger network.

$$TAC = \begin{array}{l} +Af \cdot \{CC_{exchanger} + CC_{fan} + CC_{bath} + +CC_{pump} + CC_{compressor} + CC_{evaporator} + CC_{condenser}\} \\ +P_{fan} \cdot e \cdot h + P_{pump} \cdot e \cdot h + W_{compressor} \cdot e \cdot h + Cost_{makeup-water} \end{array} \quad (28)$$

3. Case Study

A case study is employed to verify the effectiveness of proposed model. MINLP model is implemented in the software GAMS, solvers DICOPT is used to solve the MINLP problem. Table 1 shows the data of hot streams. F_{cp} is heat capacity flowrate, h is film transfer coefficient.

Table 1. Hot stream data

Stream	Tin(°C)	Tout(°C)	Fcp (kW/h)	h (kW/m^2 °C)
1	55	42	200	0.854
2	58	32	150	1.743
3	85	70	60	0.720
4	105	65	100	1.352
5	102	36	80	0.750
6	85	30	120	0.785

In this study, minimum approach temperature difference of exchanger is 10°C, Maximum allowable outlet temperature of cooler is 55 °C. Wet bulb temperature is 26°C. Cooling water specific heat capacity is 4.18 kJ/(kg °C), cooling water film heat transfer coefficient is 2.5 kw/(m^2 K). Annualized factor is 0.298. The Electricity price *e* is 0.1 $/kWh. Annual operation time is 8600 hr. For closed cooling tower, the inlet and outlet temperature of wet air is 27.0°C and 32.0°C. In the base case study, we first formulated cooling water network. The temperature of tower outlet water is specified as 30°C which is 4°C higher than wet bulb temperature. Since the minimum temperature difference of exchanger is 10°C, cooling water could only cool down hot streams to 40°C. Hot stream 2,5,6 are cooled down by cooling water to 40°C and chilled water is required to cool them down to target temperature. With the fixed heat load, the refrigeration cycle was optimized sequentially. Then, we optimized cooling water system and refrigeration cycle simultaneously. The outlet temperature of cooling tower is an optimized variable. The optimal heat load distribution between two systems was obtained. Figure 2 and Figure 3 show two configurations obtained by different methods. Table 2 is the comparison of two configurations.

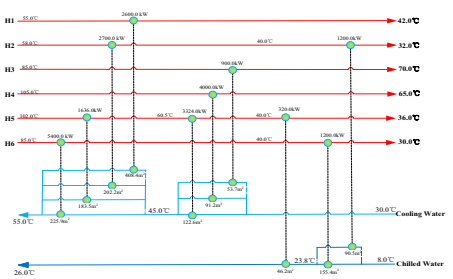


Figure 2. Sequentially optimized.

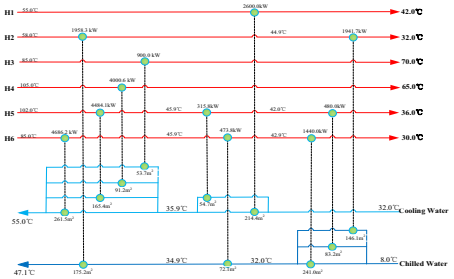


Figure 3. Simultaneously optimized.

Table 2. Comparison of two structures obtained by two different methods

Items	Sequential optimization	Simultaneous optimization
Total flowrate of cooling water	196.1 kg/s	176.1 kg/s
Total flowrate of chilled water	36.0 kg/s	38.3 kg/s
Cooling tower outlet temperature	30.0°C	32.0°C
Refrigeration cycle outlet temperature	8.0°C	8.0°C
Return temperature of cooling water	55.0°C	55.0°C
Return temperature of chilled water	26.0°C	47.1°C
Heat load of cooling tower	(20560+3385)kW	(19419+4805) kW
Heat load of refrigeration cycle (evaporator)	2720 kW	3861 kW
Power required for compressor	665kW	944kW
Heat load of condenser	3385kW	4805kW
Temperature approach of cooling tower	7.3°C	9.7°C
Capital cost of exchangers	$ 332871	$ 328686
Cost of cooling tower	$ 2811618	$ 2122941
Cost of refrigeration cycle	$ 836805	$ 1303482
Total annual cost	$ 3981294	$ 3755109

As shown in table.2, in comparison with sequentially optimized structure, simultaneously optimized structure has lower cooling water flowrate in the system. The minimum temperature difference is 10°C, therefore the return temperature of chilled water in base case cannot be higher than 26°C. However, in the simultaneously optimized structure, the chilled water was integrated with cooling water and return temperature is 47.1°C. This is because the chilled water can be used to cool the hot stream at high temperature. However, in base case, the chilled water can only be used to cool the streams that are under 40°C. In simultaneously optimized structure, the heat load of chilled water is 6295 kW, however, the heat load of evaporator in refrigeration cycle is only 3861kW. This is because in the cascade cooling and refrigeration system, the cooling tower first cools down the chilled water to 32°C, a part of heat load of chilled water is on cooling tower. Different from the structure in base case, the outlet temperature of cooling tower in cascade system is an optimized variable. The temperature increase from 30.0°C to 32.0°C, and the temperature difference inside cooling tower increase from 7.3°C to 9.7°C. Since the heat transfer driving force of cooling tower increase, the corresponding cost of tower reduced. After integrating cooling and refrigeration system into one, the cost of refrigeration system increased. However, the cost of cooling tower reduced significantly owning to the decrease of flowrate of cooling water and increase of heat transfer driving force. Finally, 5.7% reduction on TAC was achieved. This result demonstrated that the proposed method is capable to capture the trade-off between cooling and refrigeration system. The optimal heat load distribution between cooling tower and refrigeration cycle can be obtained by proposed method.

4. Conclusions

This paper presents a model for simultaneously optimizing cooling water and refrigeration system. The optimal distribution of heat load between two systems can be obtained by proposed method. In the simultaneously optimized structure, outlet temperature of cooling water was set as an optimized variable, which gives the cooling tower higher efficiency. In comparison with the method that formulates cooling water and chilled water network separately, simultaneously optimized structure has higher refrigeration cost but much lower cooling tower cost. Since cooling tower cost reduced significantly, the total annual cost of whole system decrease 5.7 %. It identifies that integrating two systems into one is an effective method to save on the cost.

References

Jain, V., G. Sachdeva, S. S. Kachhwaha and B. Patel (2016). "Thermo-economic and environmental analyses based multi-objective optimization of vapor compression–absorption cascaded refrigeration system using NSGA-II technique." Energy Conversion & Management 113: 230-242.

Kim, J. K. and R. Smith (2001) . "Cooling water system design. " Chemical Engineering Science 56(12): 3641-3658.

Ponce-Ortega, J. M., M. Serna-González and A. Jiménez-Gutiérrez (2007). "MINLP synthesis of optimal cooling networks." Chemical Engineering Science 62(21): 5728-5735.

Yin, X., X. Wang, S. Li and W. Cai (2016). "Energy-efficiency-oriented cascade control for vapor compression refrigeration cycle systems." Energy 116: 1006-1019.

Zalewski, W., B. Niezgoda-Żelasko and M. Litwin (2000). "Optimization of evaporative fluid coolers." International Journal of Refrigeration 23(7): 553-565.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50169-5

Integrated Ionic Liquid and Process Design involving Hybrid Separation Schemes

Yuqiu Chen*, John Woodley, Georgios Kontogeorgis, Rafiqul Gani

CAPEC KT Consortium, Department of Chemical Engineering, Technical University of Denmark, Building 229, DK-2800 Lyngby, Denmark
yuqch@kt.dtu.dk

Abstract

In petrochemical and chemical processes, most separations involving mixtures with low relative volatilities as well as CO_2 removal (recovery) are energy intensive. In bio-processes, the downstream separations are difficult because of product recovery from dilute solutions. Ionic liquids (ILs) based separation is emerging as a potentially sustainable and low energy consuming technology because of their non-volatility, good solubility and selectivity properties. That is, the removal of the solute and recovery of the IL-solvent requires very low energy and therefore, has low carbon foot-print. In this work, a method that combines group contribution based property prediction models for computer aided molecular (IL) design (CAMD) with process design is presented. Case studies involving CO_2 capture from natural gas and separation of azeotropic mixtures are highlighted.

Keywords: Ionic liquids, CAMD, Azeotrope separation, CO_2 capture, UNIFAC-IL model.

1. Introduction

The challenges related to energy, water, food and environment is encouraging us develop new sustainable and innovative process designs. Consequently, integration of new technologies into existing process designs, especially those involved within energy intensive processes, is of growing importance to industry. Downstream separations from bioreactors on the other hand, need to operate reliably, efficiently and at low cost due to the relatively small amounts of products in large amounts of reactants and carriers such as water. Separation of close-boiling, azeotropic mixtures, and CO_2 separation are examples of high energy consuming processes. Therefore, it is advantageous to investigate new separation technologies that allow efficient energy intensive operation.

Among many emerging separation technologies, use of ILs as solvents are being considered because of their non-volatility property and therefore low energy consuming solvent recovery operations. Also, unlike organic solvents that may also be classified as volatile organic chemical (VOCs), which would escape into the atmosphere for their high volatilities, ILs pose attractive features such as almost negligible vapor pressure, low-melting point, and high thermal and chemical stability. Moreover, ILs have been found to provide good solubility and selectivity, which are important for separation processes. Therefore, ILs provide promising alternatives for the replacement of VOCs in many separation processes (Roughton et al., 2012).

Works regarding the use of ILs as designer solvents have been in focus, mainly for CO_2 separation and IL-based extraction. Chong et al., (2015) proposed an approach based on

the visualization of high dimensional problems into two or three dimensions for design of ILs for the purpose of CO_2 capture, and then Chong et al., (2016) introduced disjunctive programming to identify optimal operating conditions of the carbon capture process while also solving the IL design problem. For IL-based extraction, Roughton et al., (2012) proposed an UNIFAC based method to simultaneously design IL and the process for separation of azeotropes. Fang et al., (2016) presented a molecular design method using the COSMO-SAC model to select ILs for extractive distillation. A study combing the UNIFAC-IL model and other group contribution (GC) based property models for CAMD and process simulation has been reported by Chao et al., (2017) to identify promising ILs for separation of n-hexane and methylcyclopentane by extractive distillation. An important deficiency with these methods is that the designed ILs have been selected based on their implicit separation task related properties but they are not actually verified through detailed process synthesis-design. Therefore, the actual energy consumption or the equipment design or the capital cost have not been considered.

Also, finding optimal ILs for specific separation tasks by the usual trial-and-error approach can be time consuming and expensive due to numerous ILs that may be considered as potential solvents. On the other hand, techniques such as CAMD is ideally suited as tailor-made ILs can be generated by adjusting the cation, anion, and side chains on the cation (Plechkova and Seddon, 2008). Thus, the "generate and test" approach of generating candidate ILs by tailoring their properties and testing them on the desired separation task is suitable for design of ILs as well as design-verification of their ability to perform specific separation tasks. In this work, such an approach where the CAMD and separation process synthesis design problems are solved simultaneously, is presented. The application of the developed method on two case studies is also highlighted.

2. Methodology

2.1. Framework

The overall framework of the proposed method is illustrated in Figure 1, where two option for problem solution and needed associated tools are shown.

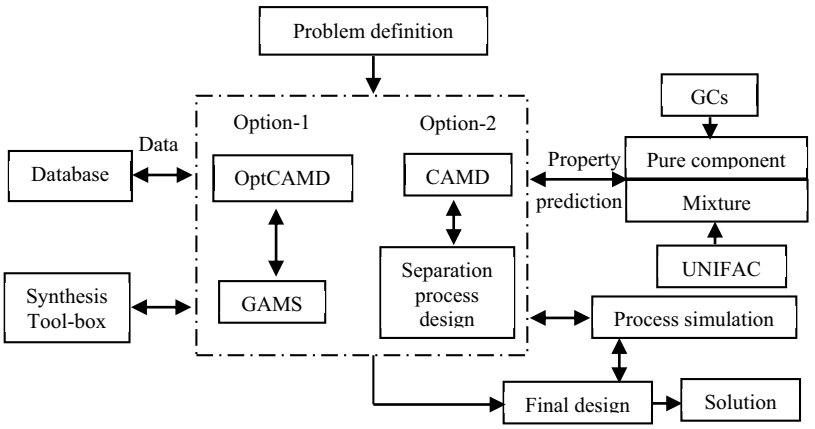


Figure 1: Framework of the integrate IL solvent and process design

2.2. Tools

A comprehensive database has been created by collection of physical properties of ILs from numerous literature sources and also by apriori generation of new ILs. The database has 4960 ILs, out of which around 300 have been reported to exist. There are 7 cations and 16 anions.

In this work, two types of property models: GC-based pure component property (θ_i) models and GC based mixture property (θ_{mix}) models are considered, as given by Eq. (1) and Eq. (2).

$$\theta_i = \sum_{k=1}^{N} n_{k,i}\psi_k \tag{1}$$

$$\theta_{mix} = \sum_{i=1}^{M} x_i\theta_i \tag{2}$$

Where ψ_k is the property contribution of group k for pure component property θ_i. $n_{k,i}$ is the number of group k in component i and N is the total number of groups representing the IL. x_i and M are the mole fraction of component i and the total number of components in mixture, respectively.

The property model library includes IL GC-based models for density, viscosity, melting point, surface tension, heat capacity, heat of vaporization and the Hildebrand solubility. For solubility and phase equilibrium predictions, the UNIFAC-IL model is included due to its high prediction accuracy (Lei et al., 2014). Note that these physical properties of ILs directly or indirectly impact the separation process where the IL is employed as the separation agent.

The integrated IL and process design problem is formulated as a mixed-integer nonlinear programming (MINLP) optimization problem given by Eqs. (3) - (8). In this work, the MINLP problem formulation from Burger et al., (2013) has been adopted.

$$max_{z,y}\, f(z,y) \tag{3}$$

$$s.t. \quad g_p(z,y) = 0 \tag{4}$$

$$g_t(z,y) \leq 0 \tag{5}$$

$$g_m(y) \leq 0 \tag{6}$$

$$z \in R^w \tag{7}$$

$$y \in U^q \tag{8}$$

where the m-dimensional vector z represents continuous variables involving mixture composition, physical properties, operating conditions, equipment sizes and process variables; the q -dimensional vector of integer and binary variables y denoting the molecular (IL) structure; f is the objective function, typically is an economic

performance metric. g_p and g_t are sets of equality constraints and inequality constraints representing the process model and the thermodynamic model, respectively. Molecular (IL) structure feasibility and valency rules are represented by constraints g_m.

3. Case studies

In this work, the four most widely studied cation cores, anions and two organic groups are retrieved from the IL database as the building blocks for IL design (Table 1).

Table 1: Molecular building blocks selected for IL design

Type	k	Groups	Type	k	Groups
Organic groups	1	CH_3	Anions	7	$[BF_4]^-$
	2	CH_2		8	$[Tf_2N]^-$
Cation cores	3	$[Im]^+$		9	$[PF_6]^-$
	4	$[MIm]^+$		10	$[DMP]^-$
	5	$[Py]^+$			
	6	$[MPyr]^+$			

The formulated MINLP problem of both examples are solved using deterministic global optimization solvers (LINDOGLOBAL) in the modeling system GAMS 24.4.6 on an Intel(R) Core(TM) i5-6200U 2.40 GHz PC running Windows 10 system. Information of the formulated MINLP problems is given in Table 2. Detailed results of the case studies can be obtained from the authors.

Table 2: Information of the formulated MINLP problems

	z	y		$g_p(z,y)$	$g_t(z,y)$	$g_m(y)$	Total constraints
		Binary	Integer				
Case 1	270	13	8	154	117	7	277
Case 2	213	12	62	124	71	7	201

3.1. CO_2 capture from natural gas

For this case study, the process flowsheet and cost models used by Pereira et al., (2011) and Burger et al., (2015) are adopted, where the net present value (NPV) of the process in 15 years is the objective function, as given by Eq. (9).

$$NPV = PV_{gas-sales} - TCI - TOC \tag{9}$$

where $PV_{gas-sales}$ is the total revenue from the natural gas sales, TCI and TOC are the total capital investment and the total operating cost of the separation process in 15years, respectively.

The best IL molecular structure and the optimal flowsheet configuration are simultaneously identified by solving the formulated MINLP problem. A selection of the computational results are given in Table 3.

In this case, 1-butyl-2, 3-dimethylimidazolium bis(trifluoromethylsulfonyl)amine ([C_5MIm][Tf_2N]) is found to be the best IL as solvent with an economically attractive NPV of 1764×10^6 US$. Burger et al., (2015) and Pereira et al., (2011) using the same process model reported the use of different organic solvents, including n-alkanes, poly(oxyethylene)dimethylethers and poly(oxymethylene)dimethylethers. For these three organic solvents, optimization yielded NPV values of 871×10^6 US$, 1577×10^6 US$ and 1721×10^6 US$, respectively. Clearly the best IL as solvent for this CO_2 capture process has a higher economic performance than the compared organic solvents.

Table 3: Optimal values of selected variables for CO_2 capture from natural gas

Variable	Optimization value
IL structure	[C_5MIm][Tf_2N]
Solvent flowrate	0.439 kmol/s
Pressure in absorption unit	7.50 Mpa
Temperature in absorption unit	304.9 K
Temperature in flash unit	452.1 K
Volume of the flash unit	82 m^3
Column height	42.6 m
Column cross section	13.0 m^2
NPV	1764×10^6 US$

3.2. Ethanol –water separation

For the case study of ethanol-water separation process, fixed process parameters used in Roughton et al., (2012) and cost models used in Zhou et al., (2015) are adopted, and the minimum of the total annual cost (TAC) of the distillation column is employed as the objective function, as given in Eq. (10).

$$TAC = CI_{column} + CI_{reb} + CI_{cond} + UC_{reb} + UC_{cond} \quad (10)$$

where CI_{column},CI_{reb} and CI_{cond} are the capital investment of the column, the reboiler and the condenser, respectively, while UC_{reb} and UC_{cond} denote the utility cost of the reboiler and the condenser, respectively.

Fixed process parameters (i.e., column operating pressure, composition of distillate, entrainer flow rate….) and the optimization results, including the best IL molecular structure and the optimal distillation column design variables are simultaneously obtained by solving the formulated MINLP problem (given in Table 4).

In this case, 1-methyl-2, 3-bimethylpyridinium hexafluorophosphate ([C_2MPy][PF_6]) is found to be the best IL as solvent with a minimum total annual cost of 423495 US $/year. Two solvents, 1, 3-dimethylimidazolium tetrafluoroborate ([C_1MIm] [BF_4]) and 1, 3-dimethylimidazolium dimethylphosphate ([C_1MIm] [DMP]) investigated by Roughton et al., (2012) are considered for comparison. They gave total costs of 549943 US $/year and 606125 US $/year, respectively. Clearly the optimal IL obtained in this work has a better economic performance than those considered earlier.

Table 4: Parameters and optimization results of the ethanol-water separation process

Fixed parameters	Value		
Column operating pressure	1 atm		
Feed rate and composition	200 kmol/h, (0.7 C_2H_5OH, 0.3 H_2O)		
Distillate flow rate	140 kmol/h		
Composition of distillate	0.998 C_2H_5OH		
Entrainer flow rate	40 kmol/h		
Optimization Variable	This work	Solvent 1	Solvent 2
IL structure	$[C_2MPy][PF_6]$	$[C_1MIm]$ $[BF_4]$	$[C_1MIm]$ [DMP]
Cation	$[Py]^+$	$[Im]^+$	[Im] +
Anion	$[PF_6]^-$	$[BF_4]^-$	$[DMP]^-$
Valence of the cation base	3	2	2
Number of CH_3 in side 1	1	1	1
Number of CH_3 in side 2	1	0	0
Number of CH_3 in side 3	1	1	1
NT (number of trays)	11	16	18
Reflux ratio R	0,307	0,653	0,875
TAC (US $/year)	423495	549943	606125

4. Conclusions

A systematic method combining GC based property models, CAMD and process design-simulation to simultaneously determine the optimal IL as a separating agent and the corresponding optimal flowsheet configuration has been developed. Case studies involving CO_2 capture from natural gas and ethanol-water separation have been presented, highlighting the improvements in the optimization results obtained in this work. Current and future work is extending the model library for IL properties as well as wide range of IL-based separation processes.

References

B.C. Roughton, B. Christian, J. White, K.V. Camarda, R. Gani, 2012, *Computers & Chemical Engineering,* 42**,** 248-262.

F.E. Pereira, E. Keskes, A. Galindo, G. Jackson, C.S. Adjiman, 2011, *Computers & Chemical Engineering,* 35**,** 474-491.

F.K. Chong, D.C. Foo, F.T. Eljack, M. Atilhan, N.G. CHEMMANGATTUVALAPPIL, 2015, *Clean Technologies and Environmental Policy,* 17**,** 1301-1312.

F.K. Chong, F.T. Eljack, M. Atilhan, D.C. Foo, N.G. CHEMMANGATTUVALAPPIL, 2016, *Computers & Chemical Engineering,* 91**,** 219-232.

H. Chao, Z. Song, H. Cheng, L. Chen, Z. Qi, 2017, *Separation and Purification Technology*.

J. Burger, V. Papaioannou, S, Gopinath, G. Jackson, A. Galindo, C.S. Adjiman, 2015, *AIChE Journal,* 61**,** 3249-3269.

J. Fang, R. Zhao, W. Su, C. Li, J. Liu, B. Li, 2016, *AIChE Journal,* 6*,* 60, 716-729.

N.V. Plechkova, K.R. Seddon, 2008, *Chemical Society Reviews,* 37, 123-150.

T. Zhou, K. Mcbride, X. Zhang, Z. Qi, K. Sundmacher, 2015, *AIChE Journal,* 61, 147-158.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50170-1

Integrated Solvent and Process Optimization Using PC-SAFT for Continuous Crystallization with Energy-intensive Solvent Separation for Recycling

Jiayuan Wang, Richard Lakerveld*

Department of Chemical and Biological Engineering, The Hong Kong University of Science and Technology, Hong Kong, China
kelakerveld@ust.hk

Abstract

Solvents are a major source of waste in the pharmaceutical industry. Anti-solvent crystallization is an important unit operation in many pharmaceutical processes. The trend towards continuous manufacturing in pharmaceutical industry offers better opportunities for anti-solvent recycling. However, the selection of solvents for a continuous process involving crystallization and anti-solvent separation and recycling is nontrivial due to the influence of solvents on the performance of various unit operations and the discrete nature of solvent selection. This work focuses on the simultaneous solvent selection and optimization of a continuous process involving crystallization and an energy-intensive separation for the anti-solvent recycle. A recently developed optimization method for continuous crystallization based on the PC-SAFT equation of state and continuous mapping is extended to include energy balances for the integrated optimization to address processes with energy-intensive separation methods. The extended method is illustrated with a case study that involves continuous crystallization of paracetamol with distillation for anti-solvent recycling.

Keywords: solvent design, crystallization, PC-SAFT

1. Introduction

Crystallization is an important purification and separation technology for pharmaceuticals. The current paradigm shift in pharmaceutical industry towards continuous manufacturing has renewed the interest in continuous crystallization and offers new opportunities to reduce solvent usage via recycling. However, the selection of an optimal solvent system for a continuous anti-solvent crystallization process including anti-solvent separation and recycling is complicated by possible trade-offs between optimal solvent properties for the various unit operations and the discrete nature of solvent selection. Furthermore, optimal process conditions depend on selected solvents. Therefore, there is a need for integrated design approaches that consider solvent selection and process design simultaneously for unit operations in a continuous process.

A systematic approach has recently been developed for the simultaneous optimization of process conditions and solvent selection for continuous crystallization including solvent recycling (Wang and Lakerveld, 2017). A unified PC-SAFT model (Gross and Sadowski, 2001) is used to calculate all thermodynamic properties for mixtures of solvents including solute solubility (Ruether and Sadowski, 2009) for crystallization. The integrated solvent and process optimization inherently leads to a mixed integer nonlinear programming (MINLP) problem due to the discrete nature of solvent selection, which is difficult to

solve. Therefore, the continuous mapping procedure (Stavrou et al., 2014) is applied as the basic solution strategy, which converts the MINLP problem into a NLP problem and provides an approximate solution for the original MINLP problem. The great promise of this approach has recently been demonstrated for cases involving process models based on material balances only (Wang and Lakerveld, 2017). However, the possible recycling of an anti-solvent may rely on an energy-intensive separation method such as distillation. Therefore, to account for realistic operational costs, energy balances should be included in the integrated solvent and process optimization of continuous crystallization involving anti-solvent separation and recycling.

The objective of this work is to develop a simultaneous solvent selection and process optimization approach for continuous anti-solvent crystallization using an energy-intensive separation method for anti-solvent separation and recycling, which considers variables from both the energy and material balances in the objective function. The unified PC-SAFT framework will be used to predict all relevant thermodynamic properties, which will be combined with a process model. The continuous mapping procedure will be used to convert the MINLP problem into an NLP problem. The approach enables the identification of promising solvent candidates for an economically optimized process with minimal experimental efforts, which ultimately can lead to faster process development and reduced solvent waste in the pharmaceutical industry.

2. Approach

2.1. Process model

The studied process configuration is illustrated in Figure 1. A solution that is saturated with an active pharmaceutical ingredient (API) is fed into a continuous crystallizer. Supersaturation is generated by the addition of recycled and fresh antisolvent. After filtration, the antisolvent is partially recovered from the mother liquor by distillation and recycled back to the crystallizer. A distillation column with a single tray, total condenser and partial reboiler is used to demonstrate the approach for a simple yet sufficiently rich case that includes key features of a rigorous distillation. Therefore, the method can be extended in a straightforward manner to, for example, include more trays.

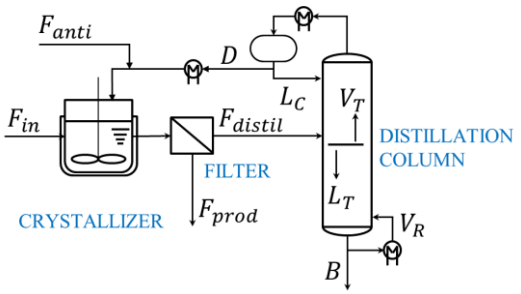


Figure1. Schematic illustration of continuous anti-solvent crystallization process with antisolvent separation and recycling.

The process is described by material balances, equilibrium conditions, summation equations and heat (enthalpy) balances (i.e., MESH equations).

The material balances over the crystallizer and the filter as a single system are given by:

$$F_{in} \cdot x_{i,in} + D \cdot x_{i,C} + F_{anti} \cdot x_{i,anti} = F_{distil} \cdot x_{i,distil} + F_{prod} \cdot x_{i,prod}, \quad (1)$$

and over the condenser, tray and reboiler of the distillation column, respectively by:

$$\left(D + L_C\right) \cdot x_{i,C} = V_T \cdot x_{i,V_T}, \quad (2)$$

$$F_{distil} \cdot x_{i,distil} + V_R \cdot x_{i,R} + L_C \cdot x_{i,C} = L_T \cdot x_{i,L_T} + V_T \cdot x_{i,V_T}, \quad (3)$$

$$L_T \cdot x_{i,L_T} = B \cdot x_{i,B} + V_R \cdot x_{i,R}. \quad (4)$$

F_{in}, F_{anti}, F_{prod}, F_{distil}, V_T, V_R, L_T, L_C, D and B denote the molar flowrate of each stream (see Figure 1). x_i denotes the molar fraction of compound i (i =solvent, antisolvent, API). The following assumptions are made: a) make-up antisolvent is pure; b) the solid product consists of pure API; c) no liquid is entrained with the product d) the API is neglected in the distillation column due to the low concentration and negligible vapor pressure.

Both the API feed and outlet of the crystallizer are assumed to be saturated at different temperatures. The solid-liquid-equilibrium (SLE) is described as follows:

$$x_{API}\gamma_{API} = \exp\left[\frac{\Delta H_m^f}{R}\left(\frac{1}{T_m} - \frac{1}{T}\right) - \frac{\Delta C_p}{R}\left(\ln\frac{T_m}{T} - \frac{T_m}{T} + 1\right)\right], \quad (5)$$

where, γ is the activity coefficient of the API in solution. ΔH_m^f, T_m, ΔC_p are the characteristics of the API, which denote the enthalpy of fusion, melting point and heat capacity difference between the hypothetical super-cooled liquid form and the crystalline form, respectively. Vapor-liquid equilibrium (VLE) is assumed to be present at the feed tray and in the partial reboiler. The VLE criterion is derived from the fugacity of a compound in the vapor and liquid phase:

$$\varphi_{i,V} \cdot x_{i,V} = \varphi_{i,L} \cdot x_{i,L}, \quad (6)$$

where, φ is the fugacity coefficient.

Summation equations (not explicitly listed here) close the material balances. The temperature in the crystallizer is constant, therefore, an energy balances is considered for distillation only:

$$F_{distil} \cdot H_{distil} + V_R \cdot H_R + L_C \cdot H_C = L_T \cdot H_{L_T} + V_T \cdot H_{V_T}, \quad (7)$$

$$V_T \cdot H_{V_T} + Q_C = \left(D + L_C\right) \cdot H_C, \quad (8)$$

$$L_T \cdot H_{L_T} + Q_R = B \cdot H_B + V_R \cdot H_{V_R}, \tag{9}$$

$$D \cdot H_C + Q_D = D \cdot H_{recyc}, \tag{10}$$

where, H denotes the molar enthalpy of each stream; Q_C, Q_R, Q_D denote the heat duty for condenser, reboiler and heat exchanger in the recycle, respectively.

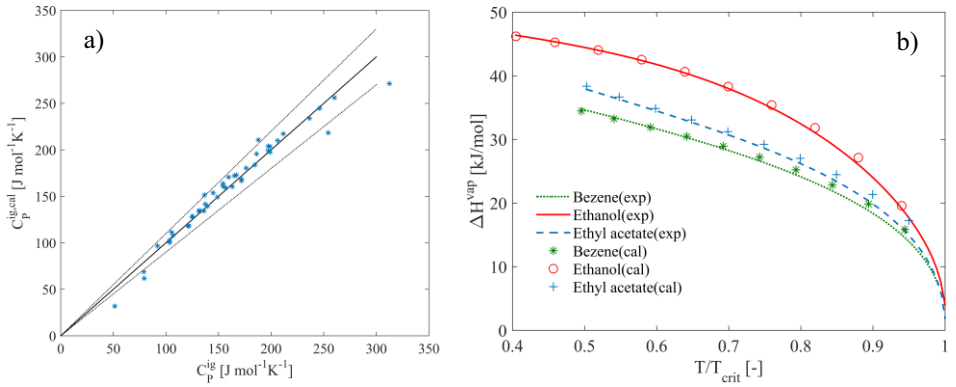


Figure2. a) Parity plot for ideal gas heat capacity at 400 K (dashed lines indicate ±10 % relative error); b) Predicted and experimental values of heat of vaporization for three presentative solvents: benzene representing a non-polar non-associating solvent, ethanol representing an associating solvent and ethyl acetate representing a non-associating polar solvent. All the reference values are taken from the NIST database.

2.2. Thermodynamic model

The PC-SAFT model (Gross and Sadowski, 2001) is used to calculate the reduced Helmholtz free energy $\tilde{a}^{res}$. The PC-SAFT parameterization schemes for APIs and different types of solvents and the solvent database with 48 candidate solvents are adopted from earlier work (Wang and Lakerveld, 2017). The molar enthalpy H consists of an ideal contribution H^{ig} and a residual contribution H^{res}, which can be derived from $\tilde{a}^{res}$ as follows:

$$H = H^{ig} + H^{res} = \left[T \sum_i x_i C_{p,i}^{ig} \right] + \left[(Z-1) - T \left(\frac{\partial \tilde{a}^{res}}{\partial T} \right)_{\rho, x_i} \right]. \tag{11}$$

The required ideal gas heat capacity can be derived from PC-SAFT pure component parameters as follows:

$$C_p^{ig} = -44.66 + 17.29 \left(m \frac{\varepsilon}{kT} \right) + 1.52 \left(m\sigma^3 \right) - 0.57 \left(m\sigma^3 \frac{\varepsilon}{kT} \right). \tag{12}$$

A comparison between predicted and experimentally measured ideal gas heat capacities and a comparison between reference values from the NIST database and calculated values for the heat of vaporization for three different types of solvents show that a satisfied prediction on solvent caloric properties can be obtained for relevant compounds for the considered process using the PC-SAFT model (Figure 2).

2.3. Case study and optimization problem formulation

Paracetamol is selected as the model compound in the case study and the experimental values for calculating the SLE from Eq. (5) are as follows: $\Delta H_m^f = 27.1\ kJ \cdot mol^{-1}$, $\Delta C_p = 99.8\ J \cdot mol^{-1} \cdot K^{-1}$, $\mathrm{T_m} = 443.6\ K$ (Granberg and Rasmuson, 1999). The optimization problem is formulated as:

$$\max_{\bar{p}_i, x_{i,distil}, B, D} \quad \frac{\varphi_1 \cdot F_{prod} - \varphi_2 \cdot F_{anti} - \varphi_3 \cdot B - \varphi_4 \cdot (Q_C + Q_D) - \varphi_5 \cdot Q_R}{F_{in}} \qquad (P1)$$

$$\text{s.t.} \qquad Eq.(1)-(10) \qquad (P1.1)$$

$$P_{process} = 1\,\text{bar} \qquad (P1.2)$$

$$T_{cry} = 298\,\text{K} \qquad (P1.3)$$

$$\frac{1}{x_{sol,distil}} + \frac{\partial \ln \gamma_{sol,distil}}{\partial x_{sol,distil}} \geq 0 \qquad (P1.4)$$

$$A \cdot \bar{p} \leq b \qquad (P1.5)$$

Table 1. Values for cost parameters in the objective function

	Symbol	Price	Source
Paracetamol	φ_1	0.9 US $ / mol	lepro pharma compass opc private limited (www.pharmacompass.com)
Solvent	φ_2	0.4 US $ / mol	icis company (www.icis.com)
Incineration	φ_3	0.2 US $ / mol	-
Cooling water	φ_4	6.5E-7 US $ / kJ	Biegler et al. (1997)
Hot steam	φ_5	3.2E-6 US $ / kJ	Biegler et al. (1997)

The economic objective function is a linear combination of the product revenue ($\varphi_1 \mathrm{F_{prod}}$) with penalties on the use of make-up antisolvent ($\varphi_2 \mathrm{F_{anti}}$), waste solvent incineration ($\varphi_3 \mathrm{B}$), and cooling and heating utilities ($\varphi_4(\mathrm{Q_C} + \mathrm{Q_D})$, $\varphi_5 \mathrm{Q_R}$) normalized by the inlet flowrate ($\mathrm{F_{in}}$). The used values for cost parameters are provided in Table 1. The free variables for the optimization are the solvent and antisolvent PC-SAFT parameters ($\bar{p}_i$), crystallizer outlet solvent composition ($\mathrm{x}_{i,disitl}$), distillate flowrate (D), and distillation bottom flowrate (B). The operating pressure for the integrated system is fixed as 1 bar (P1.2). The crystallizer inlets temperature is fixed as 298 K (P1.3). The solvent and antisolvent should be a miscible pair, which is enforced by constraint P1.4. Finally, the solvent parameter search space is restricted by a convex hull (P1.5). P1 is an NLP problem, which is derived from the original MINLP problem by treating the solvent parameters as continuous. Therefore, the solution provides an upper bound for the objective function value of the original MINLP. After optimization of P1, a Taylor

expansion is used to map the hypothetical PC-SAFT parameters to real solvent parameters from a database and generate a list of optimal solvent pairs. Finally, a process optimization is conducted for the fixed optimal solvent pair to further improve the objective function. The optimization problem is solved using the GAMS/CONOPT solver with multiple initial guesses.

3. Results

The optimal objective function value is 0.015 US $ / mol with the corresponding optimal solvents and process operating conditions as listed in Table 2. The optimal solvent/antisolvent pair enables a drastic reduction in paracetamol solubility from 0.228 to 0.002, which favors a high production rate. Due to the penalty on make-up antisolvent, the optimal F_{anti}/F_{in} remains small and most of the antisolvent is recycled. The antisolvent and solvent can be separated sharply by distillation due to the high relative volatility.

Table 2. Optimization results for the paracetamol case study.

Solvent	Dimethyl sulfoxide	B/F_{in} [-]	0.801
Antisolvent	Isopropyl acetate	D/F_{in} [-]	7.720
F_{prod}/F_{in} [-]	0.210	$(Q_C+Q_D)/F_{in}$ [kJ/mol]	130.74
F_{anti}/F_{in} [-]	0.030	Q_R/F_{in} [kJ/mol]	378.54
$x_{API,in}$ [-]	0.228	$x_{API,distil}$ [-]	0.002
$x_{anti,C}$ [-]	0.984	$x_{anti,B}$ [-]	0.037

4. Conclusions

An approach for the simultaneous optimization of process conditions and solvent selection using continuous mapping and the PC-SAFT model to calculate SLE, VLE, and caloric properties for energy balances is developed for continuous anti-solvent crystallization including recycling. The method allows for process performance criteria related to material balances and energy balances to be combined in a single objective function to minimize waste and energy consumption. A case study based on continuous anti-solvent crystallization of paracetamol demonstrates the effectiveness of the approach.

References

L. Biegler, I. Grossmann, A. Westerberg, 1997, Systematic Methods of Chemical Process Design, Prentice Hall, New Jersey, United States.

R. Granberg, Å. Rasmuson, 1999, Solubility of paracetamol in pure solvents, Journal of Chemical & Engineering Data, 44, 6, 1391-1395.

J. Gross, G. Sadowski, 2001, Perturbed-chain SAFT: An equation of state based on a perturbation theory for chain molecules, Industrial & engineering chemistry research, 40, 4, 1244-1260.

F. Ruether, G. Sadowski, 2009, Modeling the solubility of pharmaceuticals in pure solvents and solvent mixtures for drug process design, Journal of pharmaceutical sciences, 98, 11, 4205-4215.

M. Stavrou, M. Lampe, A. Bardow, J. Gross, 2014, Continuous molecular targeting-computer-aided molecular design (CoMT-CAMD) for simultaneous process and solvent design for CO2 capture, Industrial & Engineering Chemistry Research, 53, 46, 18029-18041.

J. Wang and R. Lakerveld, 2017, Integrated solvent and process design for continuous crystallization and solvent recycling using PC-SAFT, AIChE Journal, 10.1002/aic.15998.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50171-3

Optimal Matches with Load Shifting Strategy in Hybrid Power System Considering Varied Price of Outsourced Electricity

Yinghua Jiang, Lixia Kang, Yongzhong Liu*

Department of Chemical Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, P.R. China
yzliu@mail.xjtu.edu.cn

Abstract

For hybrid power systems (HPS), an energy management method with load shifting strategy under the varied price of outsourced electricity is presented, where both electricity dispatching and appliance scheduling are involved. In order to minimize the total outsourced electricity (TOE) and total cost of outsourced electricity (TCOE) of HPS, a mixed integer linear programming (MILP) formulation is developed. It features that the optimal load shifting strategy and exact matches between power demands and supplies are determined simultaneously. The application and effectiveness of the proposed model are verified and illustrated through a case study of a photovoltaic-wind-biomass-battery HPS. Results show that the TOE and TCOE of the HPS with load shifting strategies are lower than those without load shifting strategies. The effects of the load shifting strategies on the optimal matches of HPS are further analysed and discussed.

Keywords: hybrid power system; load shifting; varied price of outsourced electricity

1. Introduction

Due to increasing electricity demand, hybrid power systems (HPS) are confronted with high pressure of power supply (Rozali et al., 2016). Load shifting strategy that relocates power demands from peak to off-peak hours in terms of expectations determined by energy utilization and economic judgements is effective to maximize the potential benefits of HPS (Wimmler et al., 2017).

Recently, graphical analysis and mathematical programming methods have been applied to obtain the load shifting strategy for HPS (Theo et al., 2017). For graphical analysis methods, Alwi et al. (2013) presented a graphical tool to visualize and assess the effect of load shifting on the HPS. Mohammad Rozali et al. (2015) formulated a set of strategies to guide load shifting and optimized the electricity cost. For mathematical programming methods, Lujano-Rojas et al. (2012) proposed a load management strategy with load shifting to minimize the electricity supplied by diesel and battery in a wind-battery-diesel HPS. Yang and Xia (2017) developed a power control model of a household photovoltaic-battery HPS with load shifting under the time of use electricity tariff, aiming to minimize the energy cost and consumption. However, these methods cannot simultaneously obtain the optimal load shifting strategy and exact matches between power demands and supplies. In addition, the mutual interactions between the load shifting strategies and the operations of HPS are usually ignored.

To this end, a mixed integer linear programming (MILP) model of HPS with load shifting strategy considering the varied price of outsourced electricity is presented in this work to address these problems.

2. Problem description

Figure 1 shows a schematic diagram of an HPS, where the power demands, including non-flexible and flexible ones, and power supplies are connected to AC and DC buses. The electricity generated by power supplies are used to meet the power demands and/or will be stored in a battery energy storage system (BESS) for later use. The deficit of electricity will be supplied by the grid. The aim of this work is to propose a method so that the optimal matches and load shifting strategy can be determined at the same time.

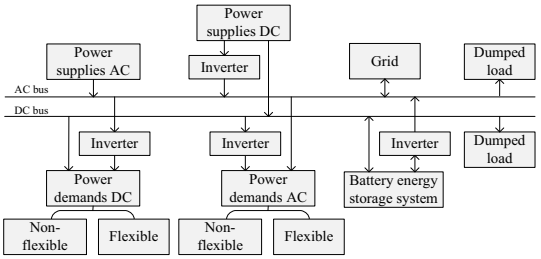


Figure 1 A schematic diagram of an HPS

3. Mathematical programming model of the HPS

3.1. Objective Function

(1) Energy objective. The TOE of the HPS is the electricity consumed by power demands and the BESS from the grid. It can be written as

$$\min \mathrm{TOE} = \sum_{k\in\mathbf{K}}\sum_{j\in\mathbf{J}} E^{\mathrm{D}}_{j,k} + \sum_{k\in\mathbf{K}} E^{\mathrm{B}}_{k} \tag{1}$$

where E^{D} and E^{B} denote the electricity transmitted from the grid to the power demands and BESS, respectively. **K** and **J** are the sets of the time intervals and power demands. The subscripts, k and j denote the time interval k and the power demand j.

(2) Cost objective. The TCOE of the HPS depends on the TOE and electricity price, ρ, which can be expressed as

$$\min \mathrm{TCOE} = \sum_{k\in\mathbf{K}}\sum_{j\in\mathbf{J}} \rho_k E^{\mathrm{D}}_{j,k} + \sum_{k\in\mathbf{K}} \rho_k E^{\mathrm{B}}_{k} \tag{2}$$

3.2. Constraints

(1) Power supply. Electricity generated by each power supply is consumed by the power demands and the BESS. If surplus electricity remains, it will be dumped. The electricity balance of each power supply at the time interval k can be expressed as

$$P_{i,k}\Delta t = \sum_{j\in\mathbf{J}} E_{i,j,k} + E^{\mathrm{SB}}_{i,k} + E^{exp}_{i,k}, i\in\mathbf{I}, k\in\mathbf{K} \tag{3}$$

where P is the power, and Δt denotes the duration of each time interval. E^{SB} and E^{exp} represent the electricity from power supplies to BESS and the electricity being dumped. **I** is the set of the power demands and subscript i denotes the i-th power supply.

(2) Power demand. The electricity required by each power demand can be satisfied by power supplies, the BESS and the grid, which can be written as

$$\begin{cases} P_{j,k}\Delta t = \sum_{i\in\mathbf{I}_\mathrm{A}} E_{i,j,k} + \sum_{i\in\mathbf{I}_\mathrm{D}} \eta^{\mathrm{DC/AC}} E_{i,j,k} + \eta^{\mathrm{DC/AC}} E_{j,k}^{\mathrm{BD}} + E_{j,k}^{\mathrm{D}}, j\in\mathbf{J}_\mathrm{A}, k\in\mathbf{K}\ (\mathrm{AC}) \\ P_{j,k}\Delta t = \sum_{i\in\mathbf{I}_\mathrm{A}} \eta^{\mathrm{AC/DC}} E_{i,j,k} + \sum_{i\in\mathbf{I}_\mathrm{D}} E_{i,j,k} + E_{j,k}^{\mathrm{BD}} + \eta^{\mathrm{AC/DC}} E_{j,k}^{\mathrm{D}}, j\in\mathbf{J}_\mathrm{D}, k\in\mathbf{K}\ (\mathrm{DC}) \end{cases} \tag{4}$$

where E^{BD} is the electricity from BESS to power demands. $\eta^{\mathrm{AC/DC}}$ and $\eta^{\mathrm{DC/AC}}$ denote the conversion efficiencies of AC/DC and DC/AC units. The subscripts A and D denote AC and DC.

Flexible power demands include interrupted and non-interrupted ones. To facilitate the explanation, we define the set of operating time intervals, $\mathbf{K}_j^{oper}$, and that of start time intervals, $\mathbf{K}_j^{star}$, as follows.

$$\begin{cases} \mathbf{K}_j^{oper} = \left\{k \mid k_j^{earl}, ..., k_j^{last}\right\}, j\in\mathbf{J} \\ \mathbf{K}_j^{star} = \left\{k \mid k_j^{earl}, ..., k_j^{last} - d_j\right\}, j\in\mathbf{J}_2 \end{cases} \tag{5}$$

where d_j denotes the elapsed time intervals of the power demand j. k^{oper}, k^{earl} and k^{last} are the operating time interval, the earliest time interval to start the operation and latest time interval to end operation. $\mathbf{J}_2$ is the set of non-interrupted power demands.

The total operating time intervals of each flexible power demand can be expressed as

$$\begin{cases} P_{j,k} = P_j Dz_{j,k}, \sum_{k\in\mathbf{K}_j^{oper}} Dz_{j,k}\Delta t = d_j, j\in\mathbf{J}_1\ (\text{interrupted}) \\ \sum_{k=k_j^{star}}^{k_j^{star}+d_j/\Delta t} P_{j,k} = z_{k_j^{star}} d_j P_j, \sum_{k_j^{star}\in\mathbf{K}_j^{start}} z_{k_j^{star}} = 1, j\in\mathbf{J}_2, k_j^{star}\in\mathbf{K}_j^{start}\ (\text{non-interrupted}) \end{cases} \tag{6}$$

where Dz is binary variable to indicate the states of flexible demand. $Dz_{j,k}$=1 when the flexible demand j operates at the time interval k. $z_{j,k}$ is binary variable to denote whether the non-interrupted demand j starts to operate at the time interval k. $\mathbf{J}_1$ is the set of the interrupted power demands.

(3) BESS. The electricity of the BESS, B, is determined by the remaining electricity of the last time interval, charged electricity, B^c , and discharged electricity, B^d of the current time interval, where B^c is the sum of the electricity supplied by power supplies and grid, and B^d equals to the sum of the electricity consumed by power demands.

$$B_k = (1-\delta)B_{k-1} + \eta^c B_k^c - B_k^d / \eta^d, k\in\mathbf{K} \tag{7}$$

$$\begin{cases} B_k^c = \sum_{i\in\mathbf{I}_\mathrm{A}} \eta^{\mathrm{AC/DC}} E_{i,k}^{\mathrm{SB}} + \sum_{i\in\mathbf{I}_\mathrm{D}} E_{i,k}^{\mathrm{SB}} + \eta^{\mathrm{AC/DC}} E_k^{\mathrm{B}}, k\in\mathbf{K} \\ B_k^d = \sum_{j\in\mathbf{J}} E_{j,k}^{\mathrm{BD}}, k\in\mathbf{K} \end{cases} \tag{8}$$

where δ is the self-discharge rate. η^c and η^d are the efficiencies of charging and discharging of BESS.

The other constraints for BESS are given as follows.

$$\begin{cases} B_k^c \le \Gamma \cdot zc_k, B_k^d \le \Gamma \cdot zd_k, zc_k + zd_k \le 1, k \in \mathbf{K} \text{ (charge and discharge contraints)} \\ C(1-DOD) \le B_k \le C, k \in \mathbf{K} \text{ (electricity contraints)} ; \\ B_0 = B_{NOK} \text{ (periodicity)} \end{cases} \tag{9}$$

where Γ is the maximum charge/discharge power. *C* and *DOD* are capacity and depth of discharge of the BESS. *NOK* is the number of time intervals. *zc* and *zd* are binary variables to indicate the charging and discharging states of the BESS.

4. Case study

4.1. Fundamental data

The fundamental data, taken from the reference (Rozali et al., 2013), is listed in Table 1.

Table 1 Fundamental data of power supplies and demands in HPS

Power supply/demand	Type	Operating time interval			Power (kW)
		From	To	Interval	
P1 (wind)	AC	3	10	8	50
P2 (biomass)	AC	1	24	24	70
P3 (solar)	DC	9	18	10	60
D1 (appliance 1)	DC	1	24	24	30
D2 (appliance 2)	AC	9	18	10	50
D3 (appliance 3)	DC	1	24	24	20
D4 (appliance 4)	AC (flexible)	9	18	10	50
D5 (appliance 5)	DC (flexible)	9	20	12	40

In this case, one day is equally divided into 24 time intervals. The capacity of BESS is assumed to be 500 kWh, and its maximum charge/discharge power takes 20 % of its capacity. The transfer efficiencies between AC and DC are 95 %, and the efficiencies of charge and discharge are 90 %. The self-discharge rate is 0.004 %·h^{-1}, and the DOD is 80 %. The electricity price in the peak hour (at the time intervals 9-11 and 20-21) is 0.16 \$·kWh^{-1}. The flat price is 0.05 \$·kWh^{-1} at time intervals 8, 12-19 and 22-23, and the off-peak price is 0.03 \$·kWh^{-1} at the time intervals 1-7 and 24.

To facilitate the description, three scenarios are assumed. (1) S1: the demands, D4 and D5 are non-flexible. (2) S2: the demands, D4 and D5 are interrupted. (3) S3: the demands, D4 and D5 are non-interrupted. The possible operating time intervals of D4 and D5 are assumed to be 6-18 and 1-24, respectively. Note that the initial electricity of BESS is provided by the grid at the off-peak price, and one day is taken as the time horizon. All calculations are implemented on the GAMS platform, with the CPU of Intel(R) Pentium(R) G630@2.70 GHz. CPLEX is selected to solve the MILP problem.

4.2. Optimal matches and load shifting strategies for TOE and TCOE reduction

The operating parameters of HPS with load shifting strategies are listed in Table 2, when the minimum TOE and TCOE are taken as the optimization objectives, respectively.

In Table 2, the values of the optimization objectives in S2 and S3 are lower than those in S1. For example, the minimum TOE in S2 and S3 are 205 kWh·day^{-1} and 231 kWh·day^{-1}, while that in S1 is 318 kWh·day^{-1}. It indicates that load shifting strategies can effectively reduce the TOE and TCOE of HPS. Taking the results in S2 as an example, the TOE and TCOE are 205 kWh·day^{-1} and 8.2 \$·day^{-1} for the minimum TOE, while 223 kWh·day^{-1} and 6.7 \$·day^{-1} for the minimum TCOE. These differences mainly result from the trade-off between the electricity price and outsourced electricity. In addition, the

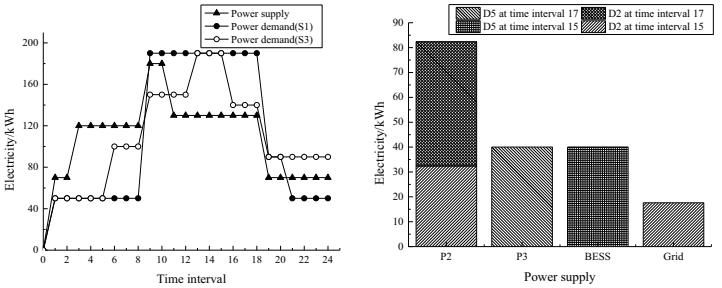
effects of the load shifting strategies on energy and economic efficiencies in S2 are much more remarkable than those in S3. This is because the interrupted power demands (in the scenario S2) are more flexible than the non-interrupted ones (in the scenario S3).

The results in Table 2 also show that the time intervals using outsourced electricity vary with the scenarios and optimization objectives. The outsourced electricity time intervals corresponding to the minimum TOE are at peak and flat hours, while those corresponding to the minimum TCOE are at off-peak and flat hours. It can be inferred that the outsourced electricity time intervals prefer a lower electricity price to a higher one when the electricity prices are considered.

Table 2 The operating parameters of three scenarios with the minimum TOE and TCOE

Operating parameters		Values		
		S1	S2	S3
Minimum TOE	TOE (kWh·day^{-1})	318	205	231
	TCOE ($·day^{-1})	14.1	8.2	11.2
	Outsourced electricity time intervals	11-18	14-18	11-15
	Operating time intervals of D4	9-18	6-10;14-18	6-15
	Operating time intervals of D5	9-20	3-6;9-13;19;22;24	13-24
Minimum TCOE	TOE (kWh·day^{-1})	323	223	260
	TCOE ($·day^{-1})	13.5	6.7	8.1
	Outsourced electricity time intervals	12-18	1-2; 6-7; 24	6-7;15;24
	Operating time intervals of D4	9-18	6-12;14;16;18	6-15
	Operating time intervals of D5	9-20	1-7;9-10;21;23-24	13-24

The optimal load shifting strategies of the HPS corresponding to the minimum TOE and TCOE are listed in Table 2. Taking the results in S3 as an example, the effect of load shifting strategy on the results of HPS is illustrated in Figure 2 (a).



(a) Comparison of electricity curves (b) Matches between supplies and demands

Figure 2 The results of HPS with load shifting corresponding to the minimum TCOE (S3)

Figure 2 (a) is the electricity curves of the supplies and demands. It shows that the differences between the electricity curves of supplies and demands in S3 are smaller than those in S1 when taking TCOE as the optimization objective. This mainly attributes to the utilization of load shifting strategy.

The optimal matches between supplies and demands of the HPS can be determined when the minimum TOE and TOCE are respectively taken as the optimization objectives. In this work, the optimal matches between supplies, and D2 and D5 at the time intervals 15 and 17 in S3 corresponding to the minimum TCOE are used as an example to give the matches in detail, as shown in Figure 2 (b).

It can be seen from Figure 2 (b) that, the required electricity of D2 at time interval 15 is taken from P2 and the grid, while that at time interval 17 is totally taken from P2. The electricity consumed by D5 at time interval 15 is supplied by BESS, while that at time interval 17 is supplied by P3. It implies that, for a given power demand, the electricity providers alter with the time intervals. Likewise, for the same time interval, the matches between the electricity providers and demands are distinct. In addition, the AC power demands prefer the AC power supplies, and the DC power demands prefer the DC power supplies so that the power losses of the electricity transfer between AC and DC are avoided to its minimum.

5. Conclusions

A MILP model is developed to determine the optimal matches between power demands and supplies, aiming to minimize the TOE and TCOE of the HPS. In the proposed method, the load shifting strategies under the varied price of outsourced electricity are considered and the effects of the load shifting strategies on energy and economic efficiencies of HPS are investigated. The application and effectiveness of the proposed model are illustrated through a case study of a photovoltaic-wind-biomass-battery HPS. The results show that the load shifting strategies under varied electricity price can effectively reduce the TOE and TCOE of HPS. The optimal load shifting strategy and exact matches between power demands and supplies can be obtained simultaneously by solving the proposed model. Based on the proposed model, the components capacities of the HPS and environmental impacts can also be evaluated. Of course, it deserves further effort in the future.

Acknowledgments

The authors gratefully acknowledge funding by the projects (No. 21676211 and No.21376188) sponsored by the National Natural Science Foundation of China (NSFC).

References

S.R.W. Alwi, O.S. Tin, N.E.M. Rozali, Z.A. Manan, and J.J. Klemeš, 2013. New graphical tools for process changes via load shifting for hybrid power systems based on Power Pinch Analysis. Clean Technologies and Environmental Policy. 15, 459-472.

J.M. Lujano-Rojas, C. Monteiro, R. Dufo-López, and J.L. Bernal-Agustín, 2012. Optimum load management strategy for wind/diesel/battery hybrid power systems. Renewable Energy. 44, 288-295.

N.E. Mohammad Rozali, S.R. Wan Alwi, Z.A. Manan, and J.J. Klemeš, 2015. Peak-off-peak load shifting for hybrid power systems based on Power Pinch Analysis. Energy. 90, 128-136.

N.E.M. Rozali, S.R.W. Alwi, Z.A. Manan, and J.J. Klemes, 2016. Process Integration for Hybrid Power System supply planning and demand management - A review. Renewable & Sustainable Energy Reviews. 66, 834-842.

N.E.M. Rozali, S.R.W. Alwi, Z.A. Manan, J.J Klemeš, and M.Y. Hassan, 2013. Process integration of hybrid power systems with energy losses considerations. Energy. 55, 38-45.

W.L. Theo, J.S. Lim, W.S. Ho, H. Hashim, and C.T. Lee, 2017. Review of distributed generation (DG) system planning and optimisation techniques: Comparison of numerical and mathematical modelling methods. Renewable and Sustainable Energy Reviews. 67, 531-573.

C. Wimmler, G. Hejazi, E. de Oliveira Fernandes, C. Moreira, and S. Connors, 2017. Impacts of Load Shifting on Renewable Energy Integration. Energy Procedia. 107, 248-252.

F. Yang, and X. Xia, 2017. Techno-economic and environmental optimization of a household photovoltaic-battery hybrid power system within demand side management. Renewable Energy. 108, 132-143.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50172-5

Design of refinery hydrogen network with near-zero hydrogen emission

Meiqian Zhu[a], Chun Deng[a*], Dominic Chwan Yee Foo[b], Xiao Feng[c]

[a]*State Key Laboratory of Heavy Oil Processing, College of Chemical Engineering, China University of Petroleum, Beijing, 102249, China*
[b]*Department of Chemical and Environmental Engineering/Centre of Excellence for Green Technologies, University of Nottingham, Malaysia Campus, Broga Road, 43500 Semenyih, Selangor, Malaysia*
[c]*School of Chemical Engineering & Technology, Xi'an Jiaotong University, Xi'an, 710049, China*
chundeng@cup.edu.cn

Abstract

At present, the shortage of hydrogen resources in refineries is severe and part of the refinery off-gas is used as fuel gas, in which hydrogen and other high value-added components are not fully utilized, resulting in a serious waste of resources. In order to improve the utilization efficiency of resources and minimize the flowrate of hydrogen discharged to the fuel system, we firstly proposed a conceptual approach to design a hydrogen network with near - zero hydrogen emission with the introduction of hybrid separation processes to recover hydrogen and other high value-added components in refinery off-gas streams. The hydrogen emission rate is defined to measure the degree of waste hydrogen discharged to fuel system. The hydrogen emission rate of optimized hydrogen network is only 2.5 % and the near - zero hydrogen emission is achieved conceptually.

Keywords: hydrogen network; multi-contaminant; refinery off-gas; gas separation.

1. Introduction

The quality of crude oil is getting inferior and environmental regulations on clean oil are becoming more stringent. In order to fulfil the environmental regulations on sulphur content (i.e. 10 ppm) for clean fuels, more hydrotreating and hydrocracking processes are needed in refineries. Those processes increase the consumption of hydrogen which is used for desulphurization, aromatics and olefin saturation, as well as the improvement of product quality. At present, most part of the hydrogen - rich refinery off-gas is used as fuel gas. Since the heat value of hydrogen is lower than that of other light hydrocarbon components, excessive hydrogen discharged to the fuel system may cause combustion instability in the furnace. On the other hand, high-valued light hydrocarbons (ethylene, propylene, and LPG) contained in refinery gas can be reused as feedstocks for chemical and plastic products and their demand has been increasing.

The methodologies for the synthesis of hydrogen network include pinch analysis and mathematical programming approaches. Its state-of-the-art can be found in the recent literature review (Marques et al., 2017). Most research on the synthesis of hydrogen network only consider reuse and recycling of hydrogen stream, and few work has been conducted on the recovery of components in the hydrogen stream. Liu et al. (2011)

proposed an evolutionary method for the design of hydrogen networks with multi-impurity which refers to H_2S and other components. Zhou et al. (2012) proposed a mathematical optimization model embedded light hydrocarbon removal module. Jia and Zhang (2011) pointed out the necessity for recovery of other high value-added components of refinery gas. The light hydrocarbon production and integrated flash calculation are incorporated into the hydrogen consumer model. Wu et al. (2014) proposed a model with hydrogen purification units in which C3+ hydrocarbons was regarded as an impurity required to remove.

Hydrogen-rich tail gases are typical by-products of refinery plant, and hydrogen recovery is the key to efficient use of refinery gases. Liu and Zhang (2004) proposed a methodology to select appropriate purifiers for recovering hydrogen from refinery off-gases between the three most common purification, namely, pressure swing adsorption unit (PSA), membrane processes and cryogenic processes. Liu et al. (2014) adopted absorption–hydration hybrid method to separate and recover H_2 from CH_4/H_2 gas mixtures under hydrate formation conditions. In addition, several techniques, such as absorption of intermediate oil, rectification (desulfurization, decarbonization and deoxygenation etc.), can be used to remove and recover other components in the refinery gas and it leads to more efficient utilization of hydrogen.

With the introduction of hybrid separation processes, this paper aims to design the refinery hydrogen network with near - zero hydrogen discharged to the fuel system. Along with gas streams from hydrogenation units, off-gas streams from crude oil distillation unit (CDU), fluidized catalyst cracking unit (FCC), delayed coking (DCU), are taken into consideration. The conceptual design is illustrated via one industrial case.

2. Conventional refinery hydrogen network

A conventional refinery hydrogen network design is shown in Figure 1. High-pressure separator gas (HP gas) from the hydrogenation units can be recycled back into the hydrogenation reactors. Refinery off-gas from CDU, FCC and DCU, low-pressure separator gas (LP gas) and dry gas from hydrocracking diesel and naphtha hydrotreating, as well as the desorption gas from PSAI and PSAII, are purged to the fuel system in conventional design.

Tables 1 and 2 show the extracted data that gas streams are discharged to the fuel system from a conventional industrial refinery plant as shown in Figure 1.

Table 1: Data for dry gas and LP gas streams from hydrogenation units

Purge	Variable and units	Hydrocracking	Diesel hydrotreating	Naphtha hydrotreating
LP Gas	Flowrate (Nm^3/h)	900	900	530
	H_2 Purity (vol %)	80	69	79
	C3+ (vol %)	7	16	5
Dry gas	Flowrate (Nm^3/h)	300	1,700	230
	H_2 Purity (vol %)	45	7	16
	C3+ (vol %)	35	80	72

Table 2: Data for refinery off-gas and desorption gas streams from non-hydrogenation units

	Dry gas from FCC	Dry gas from DCU	Dry gas from CDU	Desorption gas from PSAI	Desorption gas from PSAII
Flowrate (Nm^3/h)	8,000	8,900	690	1,240	6,870
Component (vol %)					
H_2	27	26	49	48	23
C1	26	45	26	18	16
C2	31	26	24	14	0
C3+	0	2	1	20	0
CO_2	4	1	0	0	50

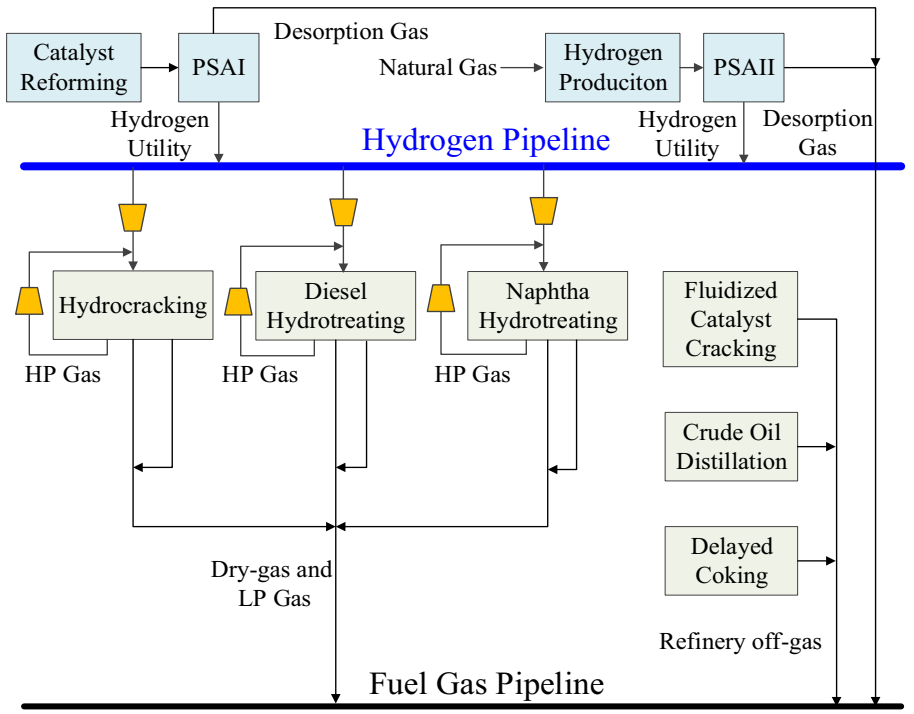


Figure 1: Conventional refinery hydrogen network

In order to quantify the amount of hydrogen released to fuel system, the hydrogen emission rate is defined as Eq. (1).

$$\eta = \frac{\sum_{s} F_{s,f}^{out} y_{s,H_2}^{out}}{F_{HU} y_{HU,H_2}} \times 100\% \quad (1)$$

where η represents the hydrogen emission rate. $F_{s,f}^{out}$ and y_{s,H_2}^{out} denote hydrogen flowrate and hydrogen purity of gas stream discharged to the fuel system, respectively. F_{HU} and y_{HU,H_2} denote the flowrate and hydrogen purity of hydrogen utility. In this case, the flowrate of hydrogen utility (F_{HU}) is 19,100 Nm^3/h with a hydrogen purity (y_{HU,H_2}) of 99.9 % (vol %).

The hydrogen emission rate of conventional refinery hydrogen network can be calculated as 41.4 % via solving Eq. (1) based on the data as shown in Tables 1 and 2.

3. Design of near zero-emission hydrogen network

The components contained in the refinery gas can be recovered by several separation processes, such as distillation, pressure swing adsorption, membrane separation, cryogenic separation, hydrate separation, absorption - adsorption coupling process, etc. The hydrogen contained in refinery off-gas is recovered to the maximum extent. Once the high value-added components such as light hydrocarbons in refinery gas are recovered, hydrogen is enriched and can be allocated to hydrogen consumers.

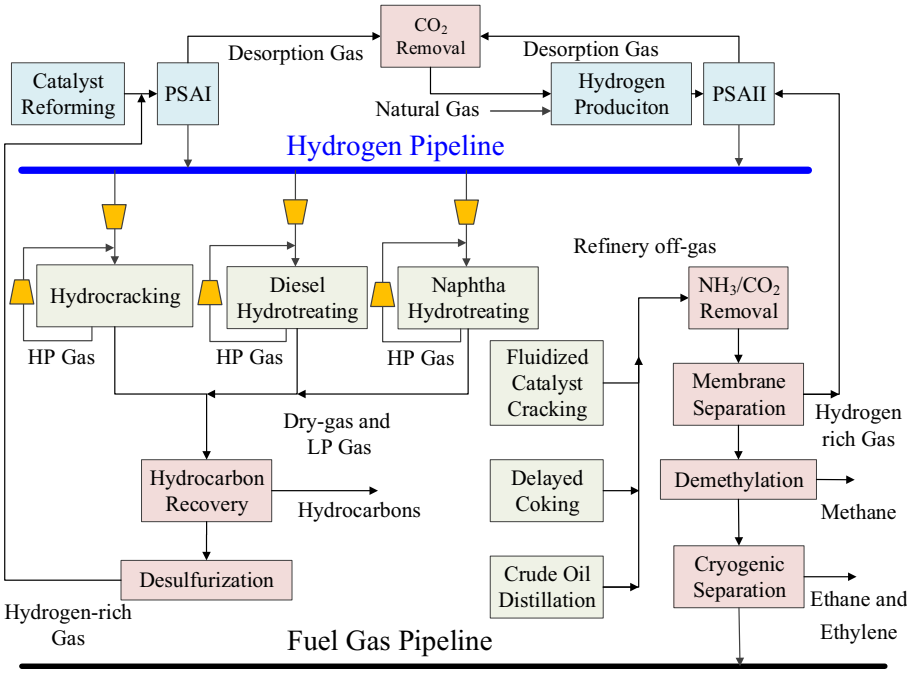


Figure 2: Improved refinery hydrogen network with near - zero hydrogen emission

As shown in Figure 2, LP gas and dry gas with light hydrocarbons (C3+) from hydrogenation units can be delivered to light hydrocarbon recovery unit. We can determine the accurate flowrate of recovered light hydrocarbons. The residual hydrogen rich stream could be sent to a PSA for further purification. The product stream of PSA is fed to hydrogen utility pipeline. With the CO_2 removal for the residual desorption gas,

the components H_2, CO, CH_4 are left in desorption gas and it is suitable raw material for hydrogen production.

The recovery flowrate for each component by different separation treatments can be calculated as Eq. (2).

$$F_{t,c}^{\text{out}} = R_{t,c} \sum_{\text{s}} F_{\text{s}}^{\text{out}} y_{\text{s},c}^{\text{out}} \qquad \forall c \tag{2}$$

where $F_{t,c}^{out}$ represents the total flowrate of recovery component *c*. F_s^{out} denotes the flowrate of the stream being fed in the hybrid separation system and $y_{s,c}^{out}$ denotes the purity for component *c* in this stream. $R_{t,c}$ denotes recovery ratio of the separation treatment for component *c*.

Light hydrocarbons (C3+) recovery is taken as an example to determine the recovery flowrate for C3+. Comparing Figures 1 and 2, LP gas and dry gas originally discharged to the fuel system are delivered to light hydrocarbon recovery. Flowrates and purity of C3+ for LP gas and dry gas can be seen in Table 1. The net flowrate of C3+ contained in the stream can be obtained by the product of the flowrate and purity of C3+ for each stream (i.e. for LP gas from hydrocracking, $F_s^{out} = 900\ \text{Nm}^3/\text{h}$, $y_{s,\text{C3+}}^{out} = 7\ \%$,900 ×7 %=63 Nm^3/h). The net flowrate of C3+ for each LP gas and dry gas stream can be obtained similarly. Then we can get the total flowrate of C3+ fed in the recovery unit by summing up the net flowrate for each stream (1,864.1 Nm^3/h). The recovery ratio of light hydrocarbon recovery is assumed to be 90%. The flowrate of recovered light hydrocarbons (C3+) can be calculated as 1,678 Nm^3/h via solving Eq. (2) ($F_{t,c}^{out} = 1{,}864.1 \times 0.9 = 1{,}678\ \text{Nm}^3/\text{h}$). Next, the hydrogen-rich stream could be sent to PSAI for further hydrogen purification.

As shown in Table 2, off-gas streams from CDU, FCC and DCU contain considerable amount of Methane, Ethane and Ethylene. Since hydrogen purity is much lower in these refinery off-gases, membrane separation is applied to recover hydrogen after the removal of acidic components (NH_3, CO_2). The remaining gas can go through the demethanizer tower to recover high-purity methane. Cryogenic separation technology or hydrate separation technology can be used to recover ethane and ethylene (C2), which are good feed material for ethylene production. Finally, the residual gas is flared or discharged to the fuel system. The recovery ratios of all separation processes are assumed to be 90%. Similar to the light hydrocarbons recovery, we can obtain the flowrates of methane (5,638 Nm^3/h) and C2 (4,464 Nm^3/h) via solving Eq. (2). And the flowrate of fresh hydrogen purified by PSA can be calculated as 6,182 Nm^3/h with the purity of 99.9 %. After recovering hydrogen and high value-added products, the flowrate of residual gas discharged to fuel system is 3,157 Nm^3/h, where the flowrate of pure hydrogen is 481.21 Nm^3/h.

The near zero-hydrogen network is constructed, and the hydrogen emission rate of improved hydrogen network can be calculated as 2.5 % via solving Eq. (1). Since the recovery ratio of separation processes can not reach 100 %, hydrogen cannot be fully recovered. However, the amount of waste hydrogen discharged to the fuel system is already tiny. The hydrogen emission efficiency decreases to be acceptable limit and it can be regarded as a near-zero emission refinery hydrogen network.

4. Conclusions

In this paper, a hydrogen network with zero hydrogen emission is designed by the introduction of hybrid separation processes to recover hydrogen and other valuable components from refinery off-gas streams. Compared with the conventional refinery hydrogen network, the flowrates of recovery hydrogen, methane, ethane and ethylene (C2), light hydrocarbon recovery (C3+) reach 6,182 Nm^3/h, 5,638 Nm^3/h, 1,678 Nm^3/h and 4,464 Nm^3/h, respectively in the improved hydrogen network. In addition, a quantitative index, hydrogen emission rate, is defined to measure the degree of waste hydrogen discharged to fuel system. The hydrogen emission rate of optimized hydrogen network is only 2.5 %, which is much lesser than 41.4 % for the conventional hydrogen network and it can be a near-zero hydrogen emission network. This conceptual approach has been proven to strengthen hydrogen utilization. Detailed simulation and optimization will be followed by the conceptual designed hydrogen network. The recovery processes for refinery gas from FCC, CDU and DCU have been well applied in the refinery. This paper combined with the optimization of refinery hydrogen system with refinery off-gas recovery system

5. Acknowledgment

Financial support provided by the National Natural Science Foundation of China (No. 21576287) are gratefully acknowledged. The research is also supported by Science Foundation of China University of Petroleum, Beijing (No. 2462018BJC003, 2462015BJB02 and 2462015YQ0305).

References

N. Jia, N. Zhang, 2011. Multi-component optimisation for refinery hydrogen networks. Energy 36, 4663-4670.

B. Liu, H. Liu, B. Wang, J. Wang, C. Sun, X. Zeng, Chen, G., 2014. Hydrogen separation via forming hydrate in W/O emulsion. Fluid Phase Equilibria 362, 252-257.

H. Liu, L. Mu, B. Wang, B. Liu, J. Wang, X. Zhang, C. Sun, Chen, J., Jia, M., Chen, G., 2013. Separation of ethylene from refinery dry gas via forming hydrate in w/o dispersion system. Separation and Purification Technology 116, 342-350.

G. Liu, M. Tang, X. Feng, C. Lu, 2011. Evolutionary Design Methodology for Resource Allocation Networks with Multiple Impurities. Industrial & Engineering Chemistry Research 50, 2959-2970.

F. Liu, N. Zhang, 2004. Strategy of purifier selection and integration in hydrogen networks. Chemical Engineering Research & Design, 82, (A10), 1315-1330.

J.P. Marques, H.A. Matos, N.M.C. Oliveira, C.P. Nunes, 2017. State-of-the-art review of targeting and design methodologies for hydrogen network synthesis. International Journal of Hydrogen Energy, 42 (1), 376-404.

S. Wu , Y. Wang, X. Feng, 2014. Unified model of purification units in hydrogen networks, Chinese Journal of Chemical Engineering, 22, 730-733.

L. Zhou, Z. Liao, J. Wang, B. Jiang, Y. Yang, 2012. Hydrogen sulfide removal process embedded optimization of hydrogen network., International Journal of Hydrogen Energy,37,18163-18174.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50173-7

Optimal Synthesis of Multi-Component Refinery Hydrogen Network

Chun Deng[a*], Meiqian Zhu[a], Yeyang Zhou[a], Xiao Feng[b]

[a]*State Key Laboratory of Heavy Oil Processing, College of Chemical Engineering, China University of Petroleum, Beijing, 102249, China*
[b]*School of Chemical Engineering & Technology, Xi'an Jiaotong University, Xi'an, 710049, China*
chundeng@cup.edu.cn

Abstract

Nowadays, the demand for hydrogen in refineries has been increasing rapidly and it is reported that the cost of hydrogen has become the second biggest cost for refineries. As a result, the optimal synthesis of refinery hydrogen network can effectively save hydrogen resources and alleviate the deficit of hydrogen in refinery. In this paper, a superstructure-based mathematical model for the synthesis of refinery hydrogen network with multiple components is constructed. All the possible connections between those units are taken into consideration. Four components, namely H_2, H_2S, C3+ and other components are included, and it leads to a multi-component model. One industrial case is analysed to illustrate the feasibility and applicability of the proposed mathematical model. 9.8 % of reduction on the flowrate of shift gas in optimal hydrogen plant can be achieved.

Keywords: hydrogen network, mathematical programming, multi-component, optimization.

1. Introduction

The processing ratio of inferior crude oil is increasing year by year. At the same time, environmental regulations on the clean fuel are becoming more stringent. In order to meet the standard requirements for sulphur content in the product clean fuel, modern refineries need to improve the ratio of hydrotreating and hydrocracking processes. It leads to the consumption of substantial amounts of hydrogen. Nowadays, synthesis of hydrogen networks is widely accepted as a useful tool to improve hydrogen utilization in refineries. Generally, the methodologies developed for the synthesis and retrofit of refinery hydrogen networks can be classified into pinch analysis and optimization-based mathematical approaches and they are reviewed in the latest article (Marques et al., 2017). Hallale and Liu (2011) firstly proposed a mathematical model incorporated with compressors and pressure swing adsorption for hydrogen network in refineries, which aims to maximize the amount of hydrogen recovered from the whole plant. Jia and Zhang (2011) developed a method of modeling and optimization considering multi-component to overcome the shortcomings of previous work in which light hydrocarbon generation and flash evaporation calculation is included. Liu and Zhang (2004) introduced the shortcut model of PSA and investigated the impacts of design conditions. The optimization of hydrogen network is integrated with hydrogen sulfide removal process (L. Zhou et al., 2012), hydrogen header (C. Deng et al., 2014), flash calculation

model and developed short-cut hydro-treating models (B. Umana et al., 2014), conditioning equipment such as fuel cells (Y. Chiang et al., 2014).

This paper aims to construct a general optimization model of multi-component hydrogen network in refinery, which includes external hydrogen sources (catalytic reforming, hydrogen production units, hydrogen header), purifier (pressure swing adsorption, desulfurization tower, light hydrocarbon recovery units), hydrogenation units, and fuel system. The multiple components consisted of hydrogen streams is assumed as four components, H_2, H_2S, C3+ and other components. One simplified industrial case is utilized to solved and analysed to illustrate the advantages of the optimal synthesis model.

2. Problem statement

The problem of hydrogen network optimization in general refinery can be described as follows. The external hydrogen sources ($u \in U$) of the refinery include catalytic reforming units and hydrogen production units. In general, the catalytic reforming and hydrogen production units are all integrated with a pressure-swing adsorption (PSA) unit. The hydrogen-enriched gas from catalytic reforming should be sent to PSA in the same plant, and shift gas produced by methane steam reforming in hydrogen plant should be sent to its own PSA. The product stream of PSA is fed to hydrogen header pipeline ($h \in H$), and the residual gas is discharged to the fuel system. Hydrogenation units ($k \in K$), mainly include hydrocracking, hydrogenation or hydrotreating units. Each process hydrogen sink is specified by its required flow rate (F_k^{in}), lower bound of hydrogen purity (Y_{k,H_2}^{min}), upper bound of inlet concentration of impurities ($Y_{\mathrm{k,c}}^{\max}, \forall c \neq \mathrm{H}_2$).The outlet of hydrogenation unit mainly includes high-pressure separator gas (HP gas) and low-pressure separator gas (LP gas). The flowrate of HP gas is denoted as F_k^{HP} with the hydrogen purity of Y_{k,H_2}^{HP} and the flowrate of LP gas is denoted as F_k^{LP} with the hydrogen purity of Y_{k,H_2}^{LP} . Most of the HP gas is reused as recycle hydrogen after desulfurization and the residual could be sent to PSA, other hydrogenation units or fuel system. LP gas should be firstly delivered to desulfurization tower to removal acidic gas. Then it could be delivered to PSA, other hydrogenation units or fuel system with or without light hydrocarbon recovery.

The purpose of this paper is to construct a model for hydrogen network optimization in general refinery to reduce the consumption of feed natural gas for hydrogen production.

3. Mathematical Model

Steam-methane reforming (SMR) for hydrogen production includes reforming, shift conversion, and pressure swing adsorption (PSA). The flow rate of shift gas is strongly associated with hydrogen utility and cost of natural gas. Therefore, the objective function can be formulated as Eq. (1).

$$\min fu\_\mathrm{shiftgas} = f_u \qquad \forall u = \mathrm{HPlant} \tag{1}$$

where f_u denotes the flowrate of shift gas produced by methane steam reforming. The flowrate of natural gas can be converted by that of syngas. The conversion coefficient λ

is defined as the amount of natural gas required to produce each unit of shift gas shown as Eq. (2).

$$fu\_ng = \lambda \cdot fu\_\text{shiftgas} \tag{2}$$

where λ denotes 0.2282, which is taken from the case of a petrochemical refinery in northeast China.

The constraints of the objective function include mass balance constraints, capacity constraints and logic constraints. For the compactness of the article, the elaborate constraints do not be listed in this paper.

4. Case study

The case is simplified industrial case from a refinery in northeast China. Current hydrogen network is shown in Figure 1. The current process data for this example is listed in Table 1 to Table 4. As shown in Figure 1, the flowrate of natural gas allocated to the hydrogen plant is 6,067 Nm^3/h. There are two hydrogen consuming processes, namely, hydrocracker unit (HC), and diesel hydrotreater (DHT). The inlets for those hydrogen consumers serve as process hydrogen sinks and their outlets would be considered as process hydrogen sources. In addition to meeting the flow rate requirements, the purity of hydrogen at the inlet of hydrogenation units should be greater than the minimum, and the purity of H_2S should be less than the maximum.

In the current hydrogen network, HP gas of the hydrogenation units is recycled back into the hydrogenation reactors, and LP gas is sent directly to PSA inlet after desulphurization without light hydrocarbon recovery. The purity and flowrate up bound requirements for the sinks are listed in Table 1 and the flowrate, purity and components for the internal sources are listed in Table 2. HP gas and LP gas requires desulfurization before recycling. The ratios of desulfurization tower are both assumed to be 80 %. It is assumed that the light hydrocarbon recovery unit can recover all C3+. Outlet gas stream of the recovery units does not contain C3+. The data of external hydrogen sources, including catalytic reforming (CCR) and hydrogen plant (HPlant), is shown in Table 3. In order to further reduce the consumption of hydrogen utility, the purification unit (PSA) is incorporated into the hydrogen network to upgrade the quality of internal hydrogen sources. The data for two PSAs respectively located in the catalytic reforming and hydrogen plant is listed in Table 4. The recovery ratios of hydrogen for PSAI and PSAII are assumed as 84.38 %. Besides, the fuel system would receive unused process hydrogen sources and other refinery off-gas.

Table 1: Data for inlet requirements of hydrogenation units

Hydrogenation units	Inlet flowrate (Nm^3/h)	Minimum purity for hydrogen (mol %)	Maximum purity for sulfide (mol %)
hydrocracking (HC)	3.681×10^5	94.24	0.1
Diesel hydrogenation (DHT)	7.7568×10^4	91.82	0.1

Table 2: Data for Outlet streams of hydrogenation units

Gas streams	Flowrate (Nm^3/h)	Components (mol %)			
		H_2	H_2S	C3+	others
HP gas from HC	3.281×10^5	93.64	0.1	1.6	4.66
HP gas from DHT	7.435×10^4	91.5	0.1	0.1	8.3
LP gas from HC	7,003	86.61	0.45	5.8	7.14
LP gas from DHT	190	85.67	0.45	0.36	13.52

Table 3: Data of external hydrogen sources

External hydrogen sources	Maximum flowrate (Nm^3/h)	Components (mol %)			
		H_2	H_2S	C3+	others
Hydrogen-rich stream in CCR	28,000	90.83	0	3.92	5.25
Shift gas in HPlant	100,000	74.99	0	0	5.01

Table 4: Data of pressure swing adsorption

PSA	Recovery ratio of hydrogen	Components for product gas (mol %)			
		H_2	H_2S	C3+	others
PSA1	84.38 %	99.98	0	0	0.02
PSA2	84.38 %	99.98	0	0	0.02

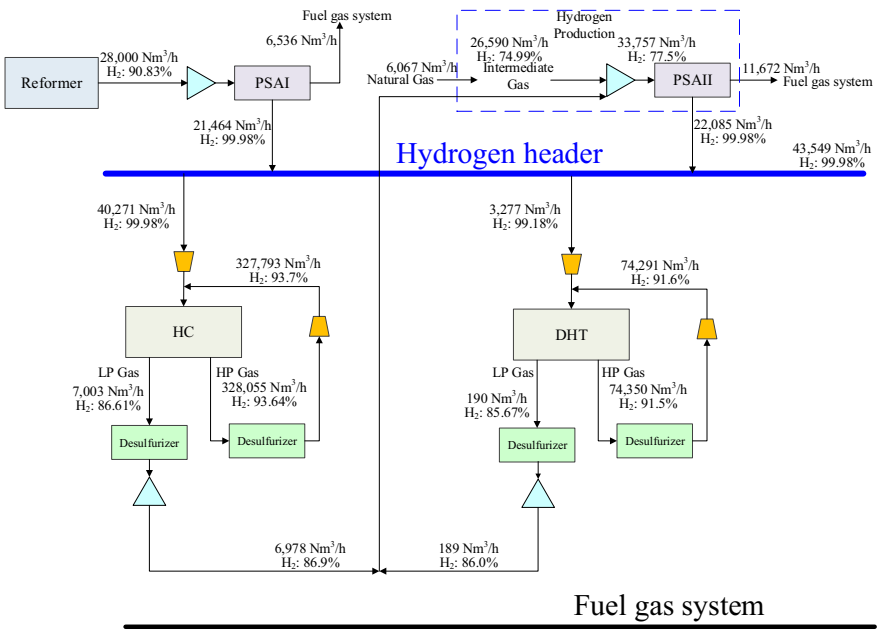


Figure 1: Current hydrogen network

The data from Table 1 to Table 4 were substituted into the optimization model for general refinery hydrogen networks. It is solved in GAMS software (Rosenthal, 2010) using KNITRO for NLP (based on the PC specification: Intel D CPU 3.00 GHz, 4 GB RAM). GAMS software version is 24.2, and the solution time is 1.123 CPUs. By optimization calculation, we can get the minimum flowrate of shift gas (23,986 Nm^3/h), corresponding to flowrate of natural gas (5,473 Nm^3/h). The optimized refinery hydrogen network is shown in Figure 2.

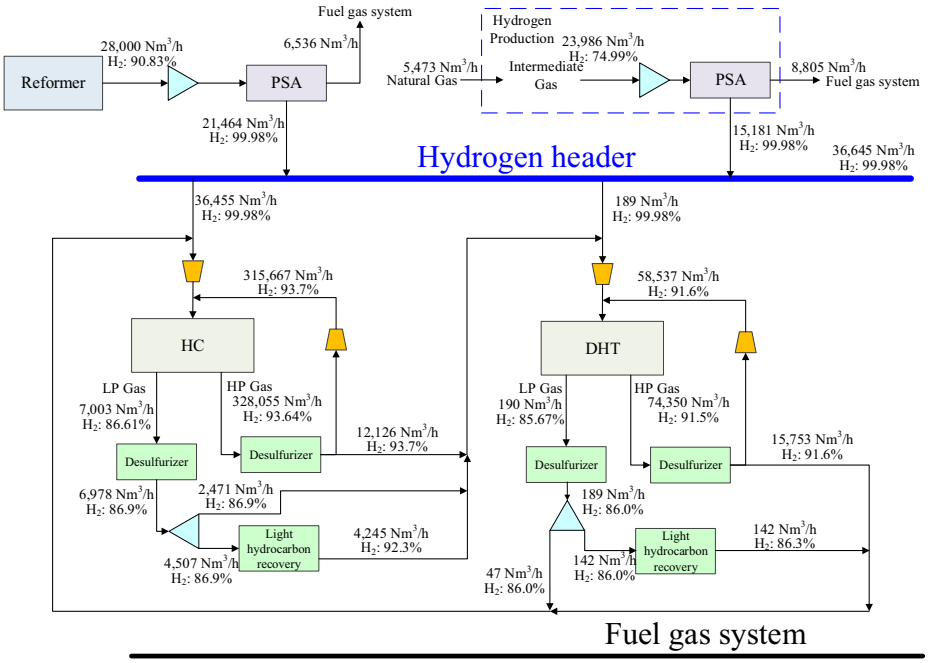


Figure 2: Optimal retrofitted hydrogen network

In the optimized hydrogen network, HP gas of the hydrogenation units is mostly recycled back into the hydrogenation reactors and the small portion is sent to the inlet of the other hydrogenation units directly. The LP gas from hydrogenation units does not be sent to PSA. LP gas is firstly allocated to the desulfurization tower, then one part of them is delivered to the inlet of the other hydrogenation unit and the other part is delivered to light hydrocarbon recovery unit. After recovery of light hydrocarbons, it is allocated to the inlet of the other hydrogenation unit. Compared with the current hydrogen network, the flowrate of shift gas in the hydrogen plant is 23,986 Nm^3/h, with a decrease of 2,604 Nm^3/h, and the consumption of natural gas is 5,473 Nm^3/h, with a decrease of 594 Nm^3/h.

5. Conclusions

In this paper, an optimization model for multi-component hydrogen network model is constructed, which mainly includes four components, namely H_2, H_2S, C3+ and other components. The desulfurization and light hydrocarbon recovery are considered in the optimal model. In addition, referring to the design of actual hydrogen network in refinery, fixed connections between different units are applied as constrains. The high-

pressure separator gas, low-pressure separator gas and dry gas from the hydrogenation units can be sent to PSA, the light hydrocarbon recovery unit, other hydrogenation units or fuel systems. Compared with the current hydrogen network, the flowrate of shift gas in optimal hydrogen plant is 23,986 Nm^3/h and the consumption of natural gas is 5,473 Nm^3/h, with a decrease of 9.8 %. This model can be used for the optimal design of refinery hydrogen network.

6. Acknowledgment

Financial support provided by the National Natural Science Foundation of China (No. 21576287) are gratefully acknowledged. The research is also supported by Science Foundation of China University of Petroleum, Beijing (No. 2462018BJC003, 2462015BJB02 and 2462015YQ0305).

References

Y. Chiang, C. Chang, 2014, Single-objective and multi-objective designs for hydrogen networks with fuel cells, Ind Eng Chem Res, 53, 6006-6020.

C. Deng, H. Pan, J. Lee, D. Foo, X. Feng, 2014, Synthesis of hydrogen network with hydrogen header of intermediate purity. International Journal of Hydrogen Energy, 39, 13049-13062.

N. Hallale and F. Liu, 2001, Refinery hydrogen management for clean fuels production, Adv Environ Res, 6, 81-98.

N. Jia, N. Zhang, 2011, Multi-component optimisation for refinery hydrogen networks. Energy 36, 4663-4670.

F. Liu, N. Zhang, 2004, Strategy of purifier selection and integration in hydrogen networks. Chemical Engineering Research & Design, 82, (A10), 1315-1330.

J.P. Marques, H.A. Matos, N.M.C. Oliveira, C.P. Nunes, 2017, State-of-the-art review of targeting and design methodologies for hydrogen network synthesis. International Journal of Hydrogen Energy,42 (1), 376-404.

R. Rosenthal, 2010, GAMS - A User's Guide, GAMS Development Corporation, Washington, DC.

B. Umana, A. Shoaib, N. Zhang, R. Smith, 2014, Integrating hydroprocessors in refinery hydrogen network optimization, Applied Energy,133,169-182.

L. Zhou, Z. Liao, J. Wang, B. Jiang, Y. Yang, 2012, Hydrogen sulfide removal process embedded optimization of hydrogen network. International Journal of Hydrogen Energy, 37, 18163-18174.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50174-9

Optimal operation of seawater desalination system based on load rolling prediction

Aipeng Jiang*, Haokun Wang, Quannan Zhang, Jian Wang, Shu Jiangzhou

School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China
jiangaipeng@163.com

Abstract

In this work, a real time optimization method integrating freshwater production and supply water load was presented based on hourly water load rolling prediction and sequential solution of the optimization problem. Firstly, based on historical data of supply water, a new load rolling prediction method with better prediction accuracy was proposed. Then, the optimal operation problem was formulated and the real time optimization strategy including water load rolling prediction, discretization of DAEs and rolling update of optimal operation profile was given. Lastly, case study based on history data of water load was carried out as illustration. Computing result shows that the proposed operational optimization method can practically bring more than 25% energy cost saving than the conventional method.

Keywords: operational optimization; seawater desalination; rolling prediction; DAEs

1. Introduction

The seawater reverse osmosis (SWRO) technique is becoming the most attractive seawater desalination technique because of its economy and convenience [1-2]. However, the energy cost for the production process is still quite high because of its high pressure driven character. To reduce the operational cost of the SWRO system, Kim et al. conducted a comprehensive study on the RO process, particularly on the minimization of product costs using the system engineering method[3]. Sassi et al. evaluated the optimal operation of SWRO system through minimizing the specific energy consumption [4]. With the consideration of thermodynamic restriction, Zhu studied the energy cost optimization problem. Among which factors such as stages, energy recovery efficiency and so on was discussed to get the optimal operation condition [5]. Li studied the optimal plant operation of brackish water reverse osmosis (BWRO) desalination to reduce specific energy consumption (*SEC*)[6]. With the consideration of the variable operating and the time-of-use electricity tariff rate schedule, Ghobeity et al. conducted the optimal operation of SWRO system to reduce the electricity charges[7]. The results show the significant electricity and production cost saving potentials with proposed operation method. Kim et al. [8] also discussed the optimization operation of the RO system with the fluctuation of seawater demand and temperature. In his study, storage tank was used as the buffering unit between water production and water supply to add the flexibility of operation. But his result was based on fixed load demand and known temperature profile in the future 24 hours. In fact, large load demand and temperature fluctuations are always unavoidable, which would cause infeasibility of the so call optimal operation. To overcome this problem, in this study an optimal operation framework including rolling prediction of water demand based on history data is presented. With better water demand prediction and correction, and rolling updating of

the optimal control trajectories obtained by solution of the operational optimization problem, more practical energy cost saving was aimed to achieve.

2. Problem modelling and formulation

The RO process model [9-10] is listed in Table 1. From these equations, permeate flowrate Q_p and salt concentration C_p can be obtained.

Table 1 the equations for RO process

Equations	Notes
$Q_p = Q_f - Q_r$	(1)
$Q_f C_f = Q_r C_r + Q_p C_p$	(2)
$Q_p = n_{PV} n_l W \int_0^L Jv dz$	(3)
$Jv = A_w \times FF \times (P_f - P_d - P_p - \Delta\pi)$	(4)
$Js = B_s (C_m - C_{sp})$	(5)
$P_b = P_f - P_d$	(6)
$\Delta P = (P_b - P_p)$	(7)
$\Delta\pi = RT(C_m - C_p)$	(8)
$\phi = \frac{C_m - C_p}{C_b - C_p} = \exp(\frac{Jv}{k_c})$	(9)
$Sh = k_c d_e / D_{AB} = 0.065 Re^{0.875} Sc^{0.25}$	(10)
$Re = \rho V d_e / \mu$	(11)
$Sc = \mu / (\rho D_{AB})$	(12)
$Js = Jv * C_p$	(13)
$\lambda = 6.23 K_\lambda Re^{-0.3}$	(14)
$\mathrm{d}V / \mathrm{d}z = -2Jv / h_{sp}$ z=0, $V=V_f$; z=L, $V=V_r$	(15)
$\mathrm{d}P_d / \mathrm{d}z = -0.5\lambda\rho V^2 / d_e$ z=0, $P_d = 0$; z=L, $P_d = P_f - P_r$	(16)
$\mathrm{d}C_b / \mathrm{d}z = 2Jv(C_b - C_p) / (h_{sp} V)$ z=0, $C_b = C_f$; z=L, $C_b = C_r$.	(17)
$A_w = A_{w0} \exp(\alpha_1 (T - 273) / 273 - \alpha_2 (P_f - P_d))$	(18)
$B_s = B_{s0} \exp(\beta_1 (T - 273) / 273)$	(19)
$R_{ec} = Q_p / Q_f$	(20)
$SEC = (P_f Q_f / \eta_{HP} - P_r Q_r \eta_{PX}) / Q_p$	(21)
$Sp = C_p / C_f \times 100\%$	(22)
$Ry = (1 - C_p / C_f) \times 100\%$	(23)

The dynamic behaviour of the storage tank can be described as:

$$\mathrm{d}H_t / \mathrm{d}t = (Q_p - Q_{out}) / S_t \tag{24}$$

$$\mathrm{d}C_{p,out} / \mathrm{d}t = Q_p (C_{sp} - C_{p,out}) / S_t H_t \tag{25}$$

Here S_t and H_t denote area and water level of storage tank, respectively. Q_{out} is water load demand, $C_{t,out}$ is salt concentration of supply water to users, and C_{sp} is the average salt concentration from permeate water. The operational optimization problem within 24 hours can be formulated as follow:

$$\underset{Q_f, P_f, H_t}{\mathrm{Min}} \sum_{i=1}^{24} E_p(i) SEC(i)\, Q_p(i) \tag{26}$$

Equality constraints: Eq.(1)-Eq.(25)

Inequality constraints: $\left| H_t(0) - H_t(24) \right| \le \Delta$

Bound constraints: $P_{flo} \le P_f \le P_{fup}$; $Q_{flo} \le Q_f \le Q_{fup}$; $C_{plo} \le C_p \le C_{pup}$

$$T_{lo} \le T \le T_{up}; \qquad H_{t\min} \le H_t \le H_{t\max}; \ \phi \le \phi_{\max}$$

In the inequality constraints, the Δ denotes the allowable deviation of water level to make sure the safety and adjustable margin of storage tank. Solution of the problem leads to optimal trajectories of feed pressure and feed flowrate in the future hours. Prediction of Q_{out} in the future hours is critical to the optimal trajectories, and conventional MPC control method has difficulty in dealing with the terminal inequality constraints in so long time zone.

3. Prediction of water load demand

The moving average method or the auto-regressive moving average method can be used for water demand prediction [11]. But it is difficult to achieve the desired accuracy when the water fluctuation is large, and the cumulative effect of error on water level will cause safe problem. Based on the above considerations, this paper uses the principle of predictive control to reduce the prediction error by rolling prediction and rolling correction. After obtaining the predicted curve for the second day, the hourly water load demand profile is updated and corrected each hour. The characteristics of history data from the 20th hour of the first day to the 5th hour of the second days was used to get and correction the prediction profile in a few hours ahead. The prediction method is as follows:

1) Obtain the hourly water load profile template with history data by moving average method as below:

$$F_{i,m} = \left(x_{i,t} + x_{i,t-1} + \cdots + x_{i,t-n+1} \right) / n \qquad i=1,2,\ldots,24. \tag{27}$$

2) With the hourly water load of the first day, predict hourly water demand of the first several hours by Eq.(28). Then obtain the initial water load prediction profile of the whole second day.

$$\mathrm{Y}_{i+1} - Y_i = \phi_1 (Y_i - Y_{i-1}) + \phi_2 (Y_{i-1} - Y_{i-2}) + \cdots + \phi_n (Y_{i-n+1} - Y_{i-n}) \tag{28}$$

Here $\phi_1 \ldots \phi_n$ denotes the auto-regressive coefficient obtained by nonlinear least square. With Eq.(28) the hourly water load of the first 4-5 hour $Y_{25} \ldots Y_{29}$ is obtained sequentially. Then $Y_{30} \ldots Y_{48}$ is calculated by:

$$Y_j = F_{j-24,m} + G_{ap} \ , j = 30 \ldots 48 \tag{29}$$

$$G_{ap} = \mathrm{Min} \sum_{\bar{j}=25}^{j-1} (Y_{\bar{j}} - F_{\bar{j}-24,m})^2 \tag{30}$$

3) Correct the water load of the i^{th} hour after it is measured, and then update the hourly load from the $(i+1)^{\text{th}}$ hour to the 24^{th} hour by Eq.(29)-Eq.(30).

4. Integrated solution strategy

With known water load prediction, solution of the operational optimization problem leads to optimal control trajectories of feed pressure and feed flowrate in the next 24 hours. Since the water load prediction is hourly updated in to reduce prediction errors, the optimal problem accordingly should be solved sequentially as the prediction profile changes. Since the problem is comprised with a set of differential equations and algebraic equation, it is not easy to be solved in direction way based on necessary optimality conditions. Simultaneous method based on finite element collocation[12] was used to discretize the optimal problem and transfer it into nonlinear program. Then large scale solver IPOPT was used for its efficient solution, and simulation based strategy was used to get reasonable state variable value. With fixed value of control variables and initial state value on the first finite element, the simulation was carried out one by one finite element until all the state variable values obtained. The following monomial basis representation[12] was used for the discretization of the differential equations:

$$w(z) = w_{i-1} + h_i \sum_{q=1}^{K} \Omega_q \left(\frac{z - z_{i-1}}{h_i} \right) \frac{\mathrm{d}w}{\mathrm{d}z_{i,q}} \tag{31}$$

Here, w_{i-1} is the value of the differential variable at the beginning of element i; h_i is the length of element i; $\mathrm{d}w/\mathrm{d}z_{i,q}$ denotes the value of its first derivative in element i at the collocation point q; and Ω_q is a polynomial of order K, satisfying:

$$\Omega_q(0) = 0 \qquad q = 1 \ldots, K \tag{32}$$

$$\Omega_q^{'}(\sigma^r) = \delta_{q,r} \qquad q = 1 \ldots, K \tag{33}$$

The continuity of the differential profile is enforced by:

$$w_i = w_{i-1} + h_i \sum_{q=1}^{K} \Omega_q(1) \frac{\mathrm{d}w}{\mathrm{d}z_{i,q}} \tag{34}$$

5. Case study

A single-stage SWRO system was used for case study. The RO unit consists of 55 parallel pressure vessels with 8 SW30HR-3XX membrane modules in series in each pressure vessel. One storage tank is used to connect the freshwater production and water supply process. Hourly electricity price is shown in Fig.1. Fig.2 shows the comparison of water load demand prediction based on moving average method and the actual value,

from which it can be seen that the conventional prediction method has quite big deviation. Fig.3 shows the rolling prediction of water load demand. Though the first prediction is not good enough, the rolling correction and update reduced the errors along time, just as shown in Fig.4.

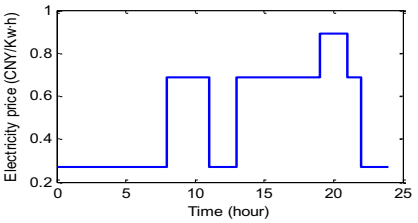


Fig. 1 Curve of hourly electricity price

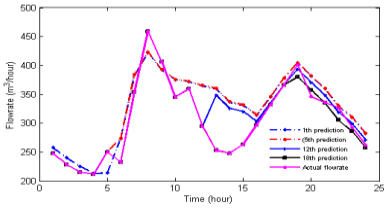


Fig.3 Rolling prediction of hourly load

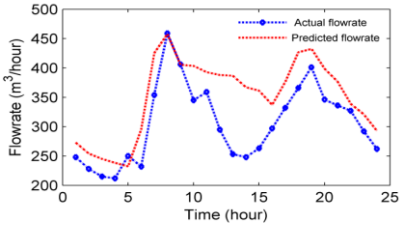


Fig. 2 Prediction of water load demand

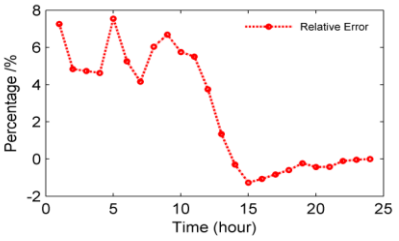


Fig. 4 Total error change along tim

To solve the optimal problem and ensure the discrete accuray, 20 finite elements with 3 Radau collocation points are used to discretize the differential equations of RO. 24 finite elements with 3 collocation points are allocated for the dynamic equation of storage tank. IPOPT under the platform of GAMS24.0 is used as solver. A 3.06GHz computer with Intel Core2 Duo T9900 processor and 5GB of memory is used.

Table 2 Computing results of the cases

No.	Solution time (s)	Objective (10^4CNY/day)	Cost save (%)	Final *Ht* (m)
Case 1	34.375	0.917	0	7
Case 2	285.892	0.671	26.83	6.68

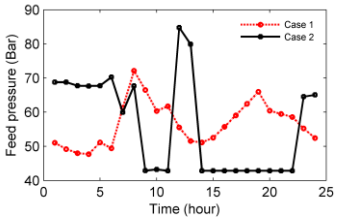


Fig. 5 Optimal feed pressure of the cases

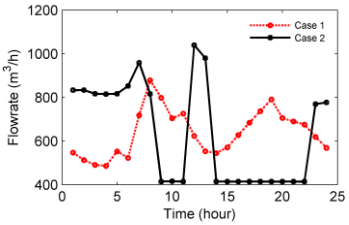


Fig. 6 Optimal feed flowrate of the cases

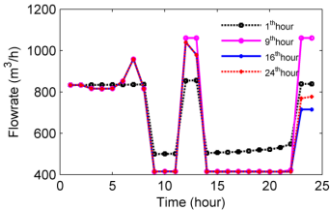


Fig. 7 Rolling update of feed flowrate

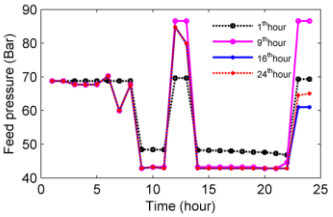


Fig. 8 Rolling update of l feed pressure

Results of our method(Case 2) is compared with conventional method which just optimize the operation at current water load (Case 1). Table2 shows the objective value and computing information. It can be seen that though more computing efforts was used, the energy cost is reduced from 9170 CNY/day to 6710 CNY/day, about 26.83% energy cost saving potential achieved. The computing time 285.892 second is short enough for our rolling optimal operation. Fig.5 and Fig.6 show the final profile of optimal operation in different cases. Since in our method, the profile of optimal operation is rolling updated, we can get 24 optimal control trajectories. Fig.7 and Fig.8 show the rolling change of these optimal control trajectories. Here we must mention that, as the water demand changes, the final terminal constraint in Case 1 can't be satisfied to ensure the safety of water level.

5. Conclusion

Optimal operation of SWRO system has significant potential of cost saving, especially when storage tank and hourly electricity were used as supply water load changes. In this paper, an optimal operation of SWRO system based on water demand rolling prediction was studied. Utilizing the history data, water load characters and rolling update method, better water load prediction accuracy was achieved, and sequentially solution of the operational optimization problem as water load prediction changes lead to better control trajectories. Compared with conventional method, about 26.83% energy cost saving potential can be achieved, and at the same time, terminal constraint of water level can also be guaranteed.

Acknowledgments:

This work was supported by the NNSF of China (61374142) and the NSF of Zhejiang Province (LY16F030006).

Reference

[1] L. Greenlee, D. Lawler, B. Freeman, et al. (2009). Reverse osmosis desalination: water sources, technology, and today's challenges. Water Research, 43(9): 2317–2348

[2] T. Yun, C. Gabelich, M. Cox, et al. (2006). Reducing costs for large-scale desalting plants using large-diameter, reverse osmosis membranes, Desalination 189 (1-3): 141-154

[3] Y. Kim, S. Kim, Y. Kim, S. Lee, I. Kim, J. Kim(2009). Overview of systems engineering approaches for a large-scale seawater desalination plant with a reverse osmosis network, Desalination 238(1–3):312–332.

[4] K. M. Sassi, I. M. Mujtaba (2010). Simulation and Optimization of Full Scale Reverse Osmosis Desalination Plant. Computer Aided Chemical Engineering, 28, 895-900.

[5] A. Zhu, P. D. Christofides, Y. Cohen (2009). Effect of thermo- dynamic restriction on energy cost optimization of RO membrane water desalination. Ind. Eng. Chem. Res. 48, 6010–6021

[6] M. H. Li (2012). Optimal plant operation of brackish water reverse osmosis (BWRO) desalination. Desalination, 293: 61–68

[7] A. Ghobeity, A. Mitsos (2010). Optimal time-dependent operations of seawater reverse osmosis, Desalination, 26376–88

[8] K. M. Sassi, I. M. Mujtaba(2013). Optimal operation of RO system with daily variation of freshwater demand and seawater temperature, Computers & Chemical Engineering, 59(4): 101-110

[9] S. Senthilmurugan A. Ahluwalia and et al. (2005). Modeling of a spiral-wound module and estimation of model parameters using numerical techniques, Desalination, 173(3):269-286

[10] J. Marriott, E. Sorensen (2003). A general approach to modeling membrane modules, Chemical Engineering Science, 58(22): 4975-4990

[11] Z. Jiang (2012). Research and development on optimal operation and scheduling system in desalination, Hangzhou: Hangzhou Dianzi University.

[12] L. T. Biegler (2010). Nonlinear Programming: Concepts, Algorithms and Applications to Chemical Processes, Society for Industrial and Applied Mathematics (Cambridge University Press), Pittsburgh.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50175-0

Modelling and Optimisation Approach of an Integrated Oil Refinery and a Petrochemical Plant

Elham Ketabchi*, Evgenia Mechleri, Sai Gu, Harvey Arellano-Garcia

Department of Chemical and Process Enginerring,University of Surrey, Guildford, GU2 7XH, United Kingdom
e.ketabchi@surrey.ac.uk

Abstract

An optimised integration approach connecting a conventional oil refinery with an ethylene production plant is investigated. Using the intermediate materials produced as the connection between the two plants, the use of internally provided feedstocks and blending options removes, at least partially, the reliance on external sourcing. This is also beneficial in terms of increasing profit margins and quality for both production systems. Thus, a mathematical model has been developed and implemented in this work to model the oil refinery and the ethylene production plant while considering their integration as an MINLP problem with the aim of optimising the integrated plants. This work considers the optimisation of each plant individually and later the final integration by modelling the interconnection between the oil refinery and the ethylene production plant. Moreover, a case study using practical data was carried out to verify the feasibility of the integration for an industrial application.

Keywords: Integration, Oil Refinery, Petrochemical Plant, Ethylene Production plant, mixed-integer non-linear programming (MINLP)

1. Introduction

Petroleum refining and refineries play an important role in our lives, mainly because most forms of transportation are powered by the products this industry produces. Accordingly, the price of crude oil has a big impact on this industry as its main feed is crude oil. The rise in this price encourages the industry to develop new ways to increase the profit margin. This requires all resources to be utilised in a highly efficient way that can be done through adequate integration. Other factors, apart from the feedstock price, that need to be considered are the strict environmental regulations that have raised the cost of producing clean fuels (Fahim, Alsahhaf and Elkilan, 2010).

Most of the technological changes in refineries in the last 20 years were due to the environmental concerns that lead to the change and alteration of the existing processes. This “cause and effect” response to environmental mandates have had a negative impact on the refineries’ economic performance, causing increased costs (Leffler, 2008). As mentioned before, an increasing crude oil price has a responsive price increase in all products that are a result of its refining. To mitigate this issue, integrating this plant with other processes can be a more profitable and worthwhile plan.

The idea of this study is to be able to connect complex plants in a way to achieve a beneficial outcome for both integrated plants, in this case, an oil refinery and an ethylene

production plant. This study aims to give the economic benefit in terms of providing an alternate source of feedstock and intermediate material required for both production plants. The integration is conducted in a way to optimise each plant individually while considering the integration. Thus, modelling the interconnection between the oil refinery and the ethylene production plant in such a way that the suitable interaction between the plants is determined by efficient material utilisation. A mathematical model is constructed and presented in this paper to model the oil refinery and the ethylene production plant considering their integration as an MINLP problem. A schematic of the integration can be seen in Figure 1.

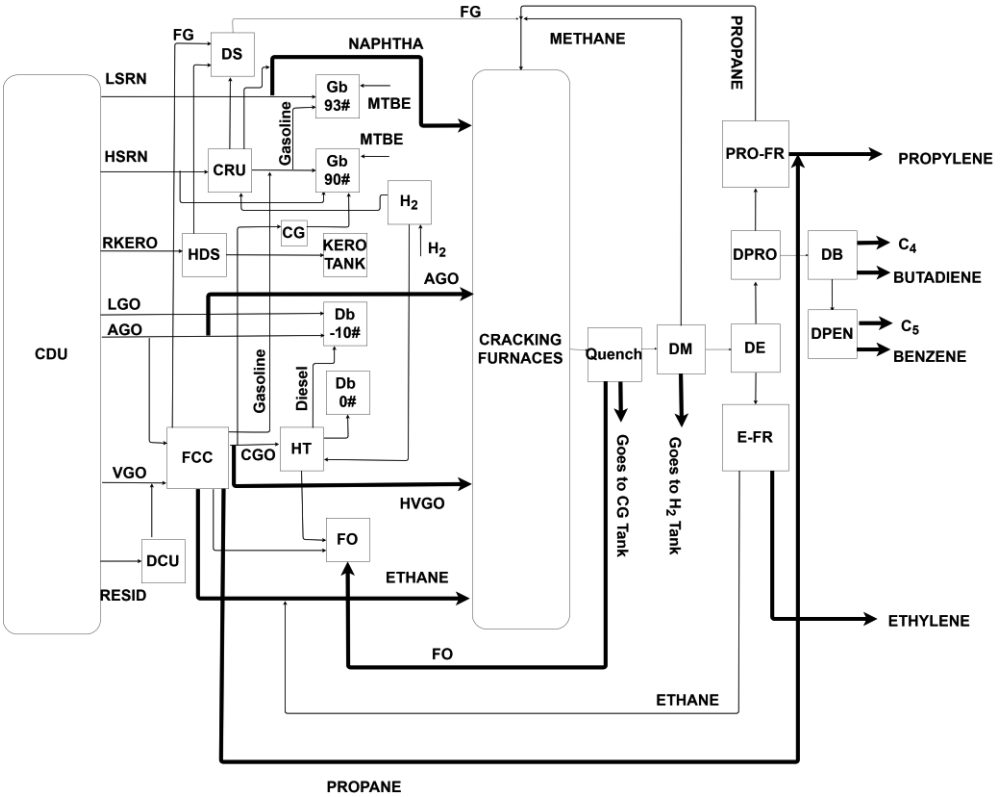


Figure 1. Schematic representation of a possible route of integration between an oil refinery and an ethylene production plant. (DM-Demethaniser, DE-Deethaniser, E-FR-Ethylene Fractionator, DPRO-Depropaniser, PRO-FR-Propylene Fractionator, DB-Debutaniser, DPEN-Dependentaniser)

2. Problem Statement and Mathematical Formulation

In this section, the processes of each plant will be briefly explored and modelled including avenues of potential integration between the two facilities identified and considered as candidates to achieve the final objective of beneficial integration of both the oil refinery and ethylene plant. The process of refining crude oil in an oil refinery consists of converting crude oil to its final products through many processing units. The units chosen for the refinery are: The Crude Distillation Unit (CDU), Fluid Catalytic Cracker (FCC), Delayed Coking Unit (DCU), Catalytic Reforming Unit (CRU), Hydro-Treating unit (HT), Gas Desulfurization unit (DS), and blenders. The distillation process within the

Crude Distillation Unit (CDU) produces, lightest to heaviest: Light Straight-Run Naphtha (LSRN), Heavy Straight-Run Naphtha (HSRN), Raw Kerosene (RKERO), Light Gas Oil (LGO), Atmospheric Gas Oil (AGO), Vacuum Gas Oil (VGO), with heaviest product, residue (RESID), entering the DCU to produce more VGO. This, alongside AGO, is then fed to the FCC producing Cracked Gas Oil (CGO), Cracked Gasoline (CG) and FCC Fuel Oil (FFO) while also generating by-products such as propylene, ethane and Fuel Gas (FG). The ethane finds use in the cracking furnaces of the ethylene production plant, while propylene is stored in a tank as a final product in the aforementioned plant. The Fuel Oil (FO) and FG are used internally to fuel the utility system, though their initially high sulphur content requires these to be desulphurised in the DS prior to use.

The CGO is fed to the HT that produces diesel (DIE) and HT Fuel Oil (HFO) and by-products such as FG. The DIE produced along with portions of LGO and AGO are blended in the diesel blender (DB) producing −10# diesel, while blending DIE with portions of CGO will result in 0# diesel product. RKERO is then consumed in the HDS that produces kerosene (KERO) as a final product that is then stored in a product tank.

The CRU produces Reformer Gasoline (RG), FG and naphtha. RG is blended with LSRN and Methyl tert-butyl ether (MTBE) in the gasoline blender (GB) to produce 90# gasoline. Meanwhile, HSRN and CG are blended with MTBE to produce 93# gasoline. The gasoline and diesel blending processes have requirements in terms of octane number and pour point, for example, the octane number of gasoline 90# and 93# should be higher than 90 and 93, respectively, while diesel −10# and 0# should have a pour point lower than −10◦C and 0◦C, respectively. The most desired products of the refinery plant consist of two kinds of gasoline (90# gasoline and 93# gasoline), two kinds of DIE (-10# diesel and 0# diesel), KERO and FO. One of the main points of this integration is to maintain a balance of material usage. So, the least valuable material in one plant can be utilised in the other, especially in this case as ethylene and ethane that are wasted through heating and boiling in the refinery, could otherwise be utilised in the ethylene production plant. Additionally, the propylene produced by the FCC can be used in the ethylene plant to produce a wide range of chemicals.

The process of ethylene production considered in this study consists of a series of parallel cracking furnaces that utilise ethane, naphtha, AGO and Hydrocracking Vent Gas Oil (HVGO), which are acquired from the oil refinery. Following cracking, quenching and compression occur before entering the separation units that consist of a demethaniser, deethaniser, depropaniser, debutanizer, and dependaniser together with two rectification towers that further separate ethane and propane from ethylene and propylene, respectively. These separations also produce hydrogen, butadiene, benzene, C4 and C5 products. The produced hydrogen is more valuable to the refinery plant than the ethylene plant; a portion of hydrogen can be transferred to the oil refinery, as well as FO and CG from the furnace output.

This optimised integration is beneficial in many ways, implementing changes in requirements for each plant that should lead to mutual improvements, while also increasing production levels using available process technologies. Each plant is mathematically formulated separately as well as specifying their connection to minimise

externally purchased feed material, with the overall model based on the work of Zhao et al.(2017). The simplification of the model involved the removal of the time limits that many variables and parameters considered. Additionally, using literature values instead of a range, assisted in streamlining the problem. Another simplification method was assigning pre-calculated values to some variables that were involved in equations to derive coefficients, temperature, etc. for units such as the furnace. This act makes the model less complex with a lesser amount of equations to be solved. Through these alterations to their model, we developed a simpler, more-straight forward expression of this integration leading to a novel approach.

The integrated models' objective function maximizes the profit by considering the revenue of refinery and ethylene plant sales, raw material cost for both plants, inventory costs, equipment and maintenance costs. This objective function is detailed in Eq. (1).

$$\begin{aligned} Profit = {} & \sum_t \sum_{c \in CP(C)} pri(c) * SC(c,t) + \sum_t \sum_{c \in CE(C)} pri(c) * \\ & SC(c,t) + \sum_t \sum_{c \in CU(C)} pri(c) * SC(c,t) - \sum_{c \in RR(C)} \sum_t pri(c) * \\ & PC(c,t) - \sum_t \sum_{c \in CU(C)} IN(c,t) * INC(c) - \sum_t \sum_{u \in UP(U)} FR(u,t) * OPC(u) - \\ & \sum_{c \in RE(C)} \sum_t pri(c) * SM(c,t) - \sum_t \sum_{c \in RE(C)} PEN(c) * INC(c) - \\ & \sum_t \sum_{u \in US(U)} FU(u,t) * OPC(u) - \sum_t \sum_f Z(u,t) * SEC(f(u)) \end{aligned} \quad (1)$$

The first term in Eq. (1) depicts the revenue of product sales from the oil refinery, $pri(c)$ being the price of material and $SC(c,t)$ the amount of material sold. The second and third terms are the revenue of the ethylene plant products and intermediates, respectively. The fourth, fifth and sixth terms take the costs for raw material, inventory and processing into account with $PC(c,t)$ being the amount of material purchased, $IN(c,t)$ the material inventory, $INC(c,t)$ inventory cost, $FR(u,t)$ the flow rate and $OPC(u)$ the price of unit operation. The last four terms are demonstrating the cost for the supply for the ethylene plant, penalty and material switching costs as well as the cost of operating the separation columns. Additional equations are implemented for this model, which are not presented here due to space limitations, that depict non-linear algebraic constraints for the amount of commodities sold, market demand, unit capacity, flowrates, operation limitations, material balance for each unit, inventory economic penalty, and inventory capacity.

3. Results

The proposed mathematical model, implemented in GAMS, is applied to a typical oil refinery and ethylene production plant in the U.K over a 24 hour period. It requires a number of parameters such as price, capacity, properties, supply, demand, product specifications and process yield, which are not calculated but defined using (BOC, 2017), (Deloitte, 2012) and (Oil and Gas Journal, 2014). The simplified model contained a total of 40 equations with 23 variables while considering the demand of material to be equal to the amount of material sold as well as using only one type of crude oil for the oil refinery section, assuming the maximum capacity of 200,000 bbl/day in the CDU. Furthermore, the physical and chemical properties of the materials and their intermediates from the blending processes were considered, however, the utility costs for both the oil refinery and ethylene plant, were not.

The main units implemented in GAMS for the oil refinery were the CDU, FCC, CRU, DS and blenders having the products and intermediates of gasoline, diesel, kerosene, fuel oil, fuel gas, ethane and propylene. Meanwhile, the main units modelled for the ethylene production plant are the series of furnaces and separation units. These units produce ethylene, propylene, butadiene, benzene, C4 and C5.

The feed required for the ethylene plant is sourced from the refinery products having the values obtained from the optimisation of the oil refinery presented below:

Table 1. Amount of material produced in the oil refinery after optimisation that is sent to the ethylene production plant

Material	Amount (t)
Ethane	10
Naphtha	30
AGO	345
HVGO	380

Table 2.The capacity of units chosen by the solver in order to optimise the plant.

Unit	Capacity (t)
CDU	77,2048
FCC	85,824
HDS	85,824
CRU	80,460
GB	237,551
DB	237,551

All three models were formulated as an MINLP problem solved with the BARON solver (Tawarmalani and Sahinidis, 2005) and using GAMS version 24.8.5 with a compilation time of 0.312 seconds.

The table below represents the profit of each plant individually followed by their integration.

Table 3. Demonstrating the profit of the oil refinery, ethylene production plant, and their integration after implementing the proposed model in GAMS

Plant	Profit (million UK pounds)
Oil Refinery	1.28
Ethylene Production	0.475
Integrated Oil Refinery and Ethylene Production Plant	13.46

4. Conclusions

In this work, a mixed-integer nonlinear programming model has been proposed for the integration of an oil refinery and an ethylene production plant. Each plant was modelled separately and then integrated by adding additional equations followed by their implementation in GAMS. The aim was to use simplified models to solve the problem in a timely manner and thus resulting in a higher revenue for the plants. Using data from a typical oil refinery and ethylene plant in the UK, the models were implemented in GAMS producing the results demonstrated in Table 3.

The challenges encountered for this problem are mainly due to the simplification in certain units resulting in a possibility of less realistic modelling which leaves room for improvement as well as the issue of restraining the model. This has resulted in infeasibilities since less bounds are involved and fixed numbers are presented for simplicity. This has been tackled through implementing different relaxation techniques such as assigning artificial variables as well as using certain options in the GAMS coding to identify the infeasibilities so as to overcome them.

In conclusion, through this integration there is an increase in the profit for both plants, showing the integration to be highly beneficial. The high increase in profit after the integration is due to merging the two plants and using the intermediates from the oil refinery as opposed to purchasing them from an outside source. Not only does the plant benefit from reducing the external feed purchases, it also benefits from upgrading the quality of the final products. Comparing the results, this proposed study aiming to connect the two plants as efficiently as possible, demonstrates the potential of profit increase for both plants compared to the state of operating them individually, which is also depicted in Table 3. This method also gives the opportunity of optimising each plant individually, increasing the profit for each plant, which can be seen in Tables 1 and 2 by choosing the best option of operation regarding the unit capacity, minimising costs. Through this implementation, the best options of materials and units are chosen to have the most beneficial impact on the integration and the end profit. The model proposed is also very versatile as it could be used for further integration regarding similar combinations of production plants.

Acknowledgement. The authors thank the financial support of the EPSRC grant EP/R001588/1.

References

BOC (2017) Industrial gases UK, Propylene Surcharge Monthly Update. Available at: www.boconline.co.uk (Accessed: 10 September 2017).

Deloitte (2012) Study of the UK petroleum retail market. London. Available at: http://www.ukpia.com/Libraries/Download/UK_Petroleum_Retail_Market_Study_Final_Report_v3_STC.sflb.ashx.

Fahim, M. F., Alsahhaf, T. A., & Elkilan, A. E. (2010). Fundamentals of Petroleum Refining. (1st, Ed.). Oxford: Elsevier.

Leffler, W. L. (2008) Petroleum Refining in Nontechnical Language. 4th edn. Tulsa, Oklahoma: PenWell.

Oil and Gas Journal (2014) 'Worldwide Report', Oil and Gas Journal, December, p. 45.

Tawarmalani, M. and N. V. Sahinidis, A polyhedral branch-and-cut approach to global optimization, Mathematical Programming, 103(2), 225-249, 2005

Zhao, H. et al. (2017) 'Integrated model of refining and petrochemical plant for enterprise-wide optimization', Computers and Chemical Engineering. Elsevier Ltd, 97, pp. 194–207. doi: 10.1016/j.compchemeng.2016.11.020.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50176-2

Integration of Optimal Cleaning Scheduling and Flow Split Control for Crude Oil Fouling Mitigation in the Operation of Refinery Heat Exchanger Networks

Federico Lozano Santamaria[a*], Sandro Macchietto[a]

[a]*Department of Chemical Engineering, Imperial College London, SW7 2AZ, UK*
f.lozano-santamaria16@imperial.ac.uk

Abstract

Fouling in heat exchangers, a slow dynamic process, significantly reduces their thermal and hydraulic performance. For heat exchanger networks (HENs) in particular, mitigation alternatives are needed to restore a profitable and safe operation. Controlling the flow rate distribution in parallel branches of the network is one option, periodically cleaning selected heat exchangers another. These two strategies have been traditionally addressed separately or sequentially, but there is a strong interaction between them. We propose i) a general formulation to model heat exchanger networks under fouling, suitable for ii) the simultaneous optimal control and optimal cleaning scheduling. The resulting large scale dynamic optimization problem with binary variables (MINLP) is reformulated as a mathematical program with complementarity constraints (MPCC) which is solved efficiently for networks of industrially relevant size. The benefit of integrating the two decision layers is demonstrated in a case study for a crude oil preheat train, where the simultaneous solution of the two problems leads to a 25 % savings in operational cost.

Keywords: heat exchanger networks; energy integration; optimisation, MPCC; scheduling and control.

1. Introduction

Improving the energy efficiency of chemical processes is a primary objective in the context of sustainable and profitable operations. Large heat exchanger networks play an important role in energy recovery of chemical processes, where the hot outlet streams of the process are used to heat the cold inlets. Oil fractioning is an energy intensive operation due to the large amount of crude oil processed, and around 65% of the energy necessary for fractioning can be recovered in the pre-heat train (Yeap et al. 2004), with the rest provided by a furnace. However, the efficiency of heat exchangers decreases over time because of the undesired progressive deposition and accumulation of material over the exchange surfaces (fouling). This deposit increases the thermal resistance and reduces the available flow area, which increases pressure drops. The combined effect leads to higher furnace duty and operational cost (Muller-Steinhagen and Zettler 2011).

Fouling mitigation in heat exchanger networks is necessary to maintain the desired operation for long periods, and to satisfy process constraints such as limits on pressure, temperature, and energy consumption at the furnace. Periodical cleaning of a heat exchanger can restore its efficiency (completely or partially, depending on the method), but choosing which exchanger to clean at any specific time, the cleaning sequence, and

the cleaning method is not an easy task. Moreover, when an exchanger is being cleaned, flows, pressure drops, temperatures, deposition and energy recovery in the network will shift significantly. After a clean unit is restored, the flow rate distribution can be changed to enhance heat transfer and recover more energy by preferentially using the clean(er) unit. Thus, the heat exchanger network operates dynamically, and significant changes can be implemented after the cleanings to improve the efficiency of the process.

In principle, solving simultaneously the optimal control and optimal cleaning scheduling problems for HENs should reduce energy consumption by a larger fraction than using only one strategy. However, the integrated problem is very challenging due to its large scale, nonlinearities, large number of binary variables, and different time scales involved. This paper addresses fouling mitigation for the hot end of a refinery pre-heat train using optimal control and optimal cleaning scheduling strategies, with sufficiently detailed models. It further proposes to reformulate the MINLP problem as a mathematical program with complementarity constraints (MPCC) to solve the integrated problem.

2. Problem formulation and case study

The following formulation is general for any HEN undergoing fouling. The HEN is modelled as a directed multigraph network in which each stream type corresponds to a graph, and the connection between two nodes define an arc. The nodes of the network are heat exchangers, sources, sinks, mixers, and splitters, but it is only at the heat exchanger nodes that the streams of two different graphs interact transferring heat. The heat transfer model considered for a heat exchanger is a lumped parameter model subject to fouling based on the definition of P efficiency and number of transfer units (NTU).

To define a consistent mathematical model for the optimal cleaning scheduling of any network of heat exchangers the following sets and indexes are defined:

- $HEX = \{1,2,\dots,n_{HEX}\}$. Set of heat exchangers in the network.
- $Sp = \{1,2,\dots,n_{Sp}\}$. Set of splitters in the network
- $Mx = \{1,2,\dots,n_{Mx}\}$. Set of mixers in the network
- $So = \{1,2,\dots,n_{So}\}$. Set of source nodes
- $Si = \{1,2,\dots,n_{Si}\}$. Set of sink nodes
- $Nodes = HEX \cup Sp \cup Mx \cup So \cup Si$. Set of all the nodes in the network
- $Stream\ types = \{1,2,\dots,n_F\}$. Set of stream types in the network, where each type is a different fluid
- $Arcs = \{(i,j,k) | \exists (i,j,k) \in Nodes \times Nodes \times Stream\ types\}$. Set of arcs that define the connection between nodes for a given stream type.
- $Cleaning\ types = \{1,2,\dots,n_{cl}\}$. Set of distinct types of cleanings that can be used.

The cleanings of the heat exchangers are modelled by introducing bypasses around the shell and the tube sides. The time horizon is discretized in a fix number of periods, but the length of each period is variable. If a heat exchanger state is “being cleaned” at a specific period the length of the period is equal to the cleaning time, while if the exchanger state is “operating” the length of the period is a decision variable with a lower and an upper bound. To increase the accuracy of the integration each period has a fixed number of internal discretization points. Here, orthogonal discretization in finite elements with Radau points is used. This approach of dividing the time horizon in periods, with additional discretization points within the periods allows capturing all the dynamics of the process, from fast ones (e.g. changes of flow rates, valve openings) to slow ones (e.g.

fouling, deposit growth, cleanings). It also guarantees accuracy of the solution in all the time scales involved in the problem.

The following set of equations define the equality constraints in the overall model:

- Mass balances at each node.
- Energy balances at each node.
- Fix inlet temperature at source nodes.
- Assignment of each exchanger state in the network (here, "being cleaned" or "operating").
- Time representation. Non-decreasing starting time of events and events duration.
- Pressure drop equations for the tube side.
- Change in flow diameter of the tubes due to the deposition of material.
- Disjunction for the definition of the length of the period.
- Disjunction between the heat transfer model (P-NTU) for the operating mode and no heat transfer for the cleaning mode.
- Disjunction between the fouling (Ebert-Panchal) and ageing models (first-order kinetics) for the operating mode, and the restoration of the fouling resistance and deposit composition for the cleaning modes. These two models allow predicting the deposit thermal resistance based on its composition (ageing), on the thermal conductivity of the components, and on the deposit thickness. It is assumed that the fresh deposit and the aged deposit do not mix, but form two distinct layers.

The fouling rate and the ageing rate are differential equations with slow dynamics that over long time scales cannot be ignored. In addition, the cleanings introduce discontinuities in the differential equations that have to be handle appropriately. Therefore, changes in operation mode are only allowed between two periods.

The objective function selected, Eq. 1, is the minimization of the total operation cost, which is composed by the loss of throughput, furnace fuel cost, CO2 emission cost, and cleaning cost over the operation horizon, τ_p (Coletti and Macchietto, 2011).

$$\min J = \sum_{p \in Periods} \tau_p \int_0^1 \left[P_{kg} m_{p,l}^{prod} + P_{fuel} \frac{Q_{f,p,l}}{\eta} + P_{CO_2} m_{CO_2} \frac{Q_{f,p,l}}{\eta} \right] d\hat{t} + \sum_{i \in HEX} \sum_{p \in Periods} P_{cl,p,i} \tag{1}$$

There are continuous decisions, related to the flow rate distribution in the network controlled by the inlet flow rates and split fractions, and binary decisions for the assignment and timing of the cleanings. Therefore, the dynamic optimization problem after discretization is defined as a large scale MINLP. To solve the problem the binary variables are relaxed and bounded between zero and one, and complementarity constraints and slack variables are introduced, resulting in a mathematical problem with complementarity constraints, MPCC. Complementarity constraints are a type of constraints between two variables that forces one to be at its bound and the other to be free, thus it is an alternative to model the logical OR operator. Eq. (2) – (4) show the complementarity model for the optimal control and cleaning scheduling problem, where $y_{i,p}$ is a variable used to assign a cleaning to a heat exchanger *i* at a specific period *p*. The

symbol $\perp$ indicates complementarity between the two variables, and in this case the bilinear product is used as complementarity function.

$$0 \leq y_{i,p} \perp s^{b}_{i,p} \geq 0, \qquad \forall i \in HEX, p \in Periods \tag{2}$$

$$0 \leq \left(1 - y_{i,p}\right) \perp s^{a}_{i,p} \geq 0, \qquad \forall i \in HEX, p \in Periods \tag{3}$$

$$0 \leq s^{a}_{i,p}, s^{b}_{i,p} \leq 1, \qquad \forall i \in HEX, p \in Periods \tag{4}$$

The MPCC problem formulation for the optimal control and cleaning scheduling of heat exchanger networks is treated as a NLP. However, the complementarity constraints define an optimization problem which is not well posed because it does not satisfy any constraint qualification criteria (Herty and Steffensen 2012). The complementarity constraints are relaxed by defining an inequality and a ϵ parameter for regularization, $Reg(\epsilon)$. The formulation is illustrated in Eq. (5). The problem is solved in a sequence, decreasing the value of ϵ such that when $\epsilon \to 0$ the solution of the MPCC problem is equivalent to the solution of the original MINLP problem.

$$\begin{aligned} Reg(\epsilon): \min_{x,y,z} & f(x,y,z) \\ s.t. & \; h(x,y,z) = 0 \\ & g(x,y,z) \leq 0 \\ & x_i y_i \leq \epsilon, \;\; i = 1,2,\dots,q \\ & x \geq 0, y \geq 0 \\ & x \in X \subseteq \mathbb{R}^n, z \in Z \subseteq \mathbb{R}^n \\ & 0 \leq y \leq 1 \end{aligned} \tag{5}$$

For a more comprehensive discussion of the detailed dynamic model of heat exchangers, continuous time formulation, time discretization, and MPCC reformulation, the reader is referred to Lozano Santamaria and Macchietto, 2017.

The problem analysed here is a typical refinery network structure in which the crude oil is preheated in heat exchangers in a series configuration, with the flow split in two branches. Figure 1 displays the networks structure and the streams involved. All the exchanger are tube and shell heat exchangers, and the operational conditions and design specifications are taken from previous works that are based on relevant industrial information and plant data (Coletti and Macchietto 2011).

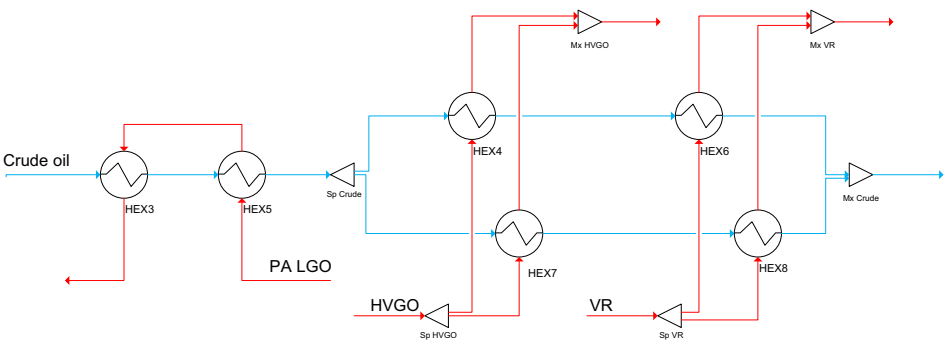


Figure 1. Case study of a refinery preheat train with parallel branches

3. Results and discussion

The regularization parameter, ϵ, is reduced from an initial value of 10 by a factor of 0.1 to a value of 1x10-6 when the optimality and integrity conditions are deemed satisfied. The MPCC problem is modified when the regularization parameter is updated, and solved again. Each modified problem is initialized from the previous solution, Lagrange multipliers and gradients to facilitate the convergence. This way the problem bounds are continuously tightened and the MINLP relaxation gets closer to the original problem.

Four scenarios are analysed: i) no mitigation (NM), with split fractions and flow rates fixed and no cleanings, ii) optimal control (Ctrl), where only the dynamic flow rate distribution in the network is optimized, iii) optimal cleaning scheduling (Sch), where number and timing of cleanings are optimised for fixed flow rates, and iv) simultaneous optimal control and cleaning scheduling (Sch + Ctrl), where both flow distribution and cleanings are optimised. In all cases the starting point is a clean network and horizon 365 days. Figure 2 shows the profiles of two important network performance indicators, the furnace duty (the additional energy required in the furnace to satisfy a fixed distillation column inlet temperature), and the coil inlet temperature (CIT), the temperature before the furnace. In this case, the optimal control solution exhibits almost the same behaviour as the NM (no mitigation) case. Changing the flow distribution alone is not sufficient to mitigate fouling. Scheduling alone with fixed split fractions results in 5 cleanings and a furnace duty well below that of the no mitigation case during almost all the operating window. The combination of scheduling and control results in 6 cleanings and similar profiles. The changes are occasionally more drastic and responses significantly different from the Sch optimal solution, but give an economically significant improvement of about 10 °C in the CIT temperature at the end of the period.

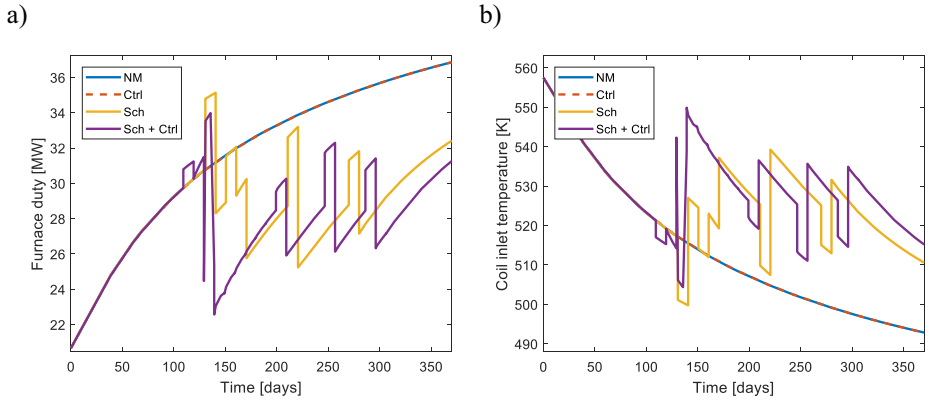


Figure 2. Case study furnace duty (a) and CIT (b) for the optimal cleaning schedule.

Table 1 summarizes the results for the four scenarios analysed in terms of computational time, cleaning schedule, and economic and energy savings relative to the no mitigation case. Introducing flow split control changes the cleaning scheduling significantly: one decision affects the other and vice versa. The simultaneous scheduling and control solution has one additional cleaning compared to case when only the schedule is optimised. When the control variables are available for manipulation over time, the additional cost of an extra cleaning is justified because the changes of flow distribution allow recovering more energy and reducing the operational cost.

Table 1. Computational results and optimal cleaning scheduling for large HEN.

	NM	Optimal control	Optimal scheduling	Simultaneous solution
CPU secs	28.11	197.8	8,407.9	2,466.76
Cleaning schedule (Time of cleaning in days, HEX no.)	X	X	(131, HEX4) (151, HEX7) (161, HEX6) (211, HEX8) (270, HEX7)	(109, HEX7) (119, HEX8) (129, HEX4) (199, HEX6) (247, HEX4) (286, HEX6)
Energy savings [MWd]	0.0	4.70	1,186.0	1,481.7
Savings [MM-USD]	0.0	0.003	0.864	1.080

Savings with optimal control of flow split alone are just $ 0.003 million compared to $ 0.864 million when optimizing the cleaning schedule alone, over 1 year of operation. However, when combined the two approaches have a strong synergistic effect that increases savings by a further 25 % to $ 1.08 million. By optimizing these two decision problems simultaneously, operation of heat exchanger networks under fouling can be improved, the operational cost of the process reduced, and the operational time of the network increased. The operating policies obtained can be practically implemented. Computation times are quite acceptable. All solutions are possibly local optima.

4. Conclusions

Flow distribution control and the cleaning scheduling are two fouling mitigation alternatives that individually can improve the performance of heat exchanger networks. Optimising them simultaneously can significantly reduce the operational cost and the energy consumption of the process. The formulation of the integrated problem presented is general, and its efficient solution strategy based on complementarity constraints allows solving large scale dynamic optimization problem with discrete variables involving different time scales with a high fidelity model, in a reasonable computational time, even for the network of realistic size and complexity considered.

References

Coletti, F, and S Macchietto. 2011. “Refinery Pre-Heat Train Network Simulation Undergoing Fouling: Assessment of Energy Efficiency and Carbon Emissions.” *Heat Transfer Engineering* 32 (3–4). Taylor & Francis: 228–36. doi:10.1080/01457632.2010.495606.

Coletti, F, and S Macchietto. 2011. “A Dynamic, Distributed Model of Shell-and-Tube Heat Exchangers Undergoing Crude Oil Fouling.” *Industrial & Engineering Chemistry Research* 50 (8). American Chemical Society: 4515–33. doi:10.1021/ie901991g.

Herty, M, and S Steffensen. 2012. “MPCC Solution Approaches for a Class of MINLPs with Application in Chemical Engineering.”

Lozano Santamaria, F and Macchietto, S. 2017. *Integration of optimal cleaning scheduling and control using a MPCC formulation for mitigating fouling in heat exchangers*. Submitted for publication.

Muller-Steinhagen, H, and H.U Zettler. 2011. *Heat Exchanger Fouling. Mitigation and Cleaning Technologies*. 2do ed. Essen: PP Publico.

Yeap, B.L., D.I. Wilson, G.T. Polley, and S.J. Pugh. 2004. “Mitigation of Crude Oil Refinery Heat Exchanger Fouling Through Retrofits Based on Thermo-Hydraulic Fouling Models.” *Chemical Engineering Research and Design* 82 (1): 53–71. doi:10.1205/026387604772803070.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50177-4

Simultaneous Controllability Analysis and Cyclic Scheduling of Grade Transitions for the Continuous Multi-product Chemical Processes

Yue Li, Jiawen Wei, Zhihong Yuan*

Department of Chemical Engineering, Tsinghua University, 100084 Beijing, China
zhihongyuan@mail.tsinghua.edu.cn

Abstract

The smart production of high-end polymers with agility and flexibility calls for the optimal production sequences and high-efficient control strategy for grade transitions with desirable dynamics. Conducting controllability analysis can not only guarantee the desirable dynamics but also relieve the burden of detailed controller. Therefore, this paper addresses a framework to integrate the controllability analysis and the scheduling of grade transitions for the continuous multi-product chemical process. Zero dynamics is used to evaluate the controllability over the wide design/operating zone. The series of limit points of the bifurcation analysis for zero dynamics divides the whole zone into Non-minimum and Minimum phase ones. The changing trajectories of limit points and the grade transition time are extracted as the function of key design/operating variables according to our proposed machine learning method, respectively. These extracted functions are then introduced to the scheduling framework. By solving the resulted MINLP model, the production time, production sequence, and key design/operating parameters can be optimized to maximize the economic benefit while roughly guaranteeing the desirable dynamic behavior during transitions. A polymerization of methyl methacrylate is studied to demonstrate the availability of the proposed method.

Keywords: Scheduling; Controllability; Data-driven model, MINLP

1. Introduction

The conventional sequential approach for scheduling problem takes economic benefits as the objective function to obtain the optimal production sequence (Sahinidis and Grossmann, 1991). For example, the traditional sequential approach for scheduling and controlling grade transition of polymerization usually generates very long transition time or undesirable transition curve. The basic issue is that the inherent behaviors such as limited cycle, unstable characteristics, and non-minimum phase (NMP) behavior are ignored. To treat such problem, a new approach which integrates scheduling and control has been extensively investigated (Baldea and Harjunkoski, 2014; Dias and Ierapetritou, 2017). Such approach usually formulates a computationally-expensive mixed integer dynamic optimization (MIDO) model. MIDO model can be transferred to a large-scale mixed integer nonlinear programming (MINLP) model. MIDO model therefore might be inapplicable in complex production process such as polymerization. Furthermore, many inherent undesirable behaviors such as input multiplicity cannot be tackled by conventional controller and even nonlinear model predictive controllers.

As known, the control performances are heavily determined by the above inherent behaviors, and these inherent behaviors are basically determined by process design. It

should be pointed out that modifications of a process design itself, such as changing inputs or outputs, operating points, values of design parameters, can sometimes affect the dynamics of the process significantly more than changes in the controllers. Therefore, this paper addresses a framework to integrate the controllability analysis and the scheduling of grade transitions for the continuous multi-product chemical process. The aim of the proposed framework is to determine those key design/operating parameters which would roughly avoid the undesirable inherent characteristic (NMP behavior) while guaranteeing the economic performance through the optimal scheduling.

In our work, zero dynamics is used to evaluate the phase behavior over the wide design/operating zone (unstable zero dynamics denote NMP behavior, while stable zero dynamics mean MP behavior). The series of limit points of the bifurcation analysis for zero dynamics divide the whole zone into NMP and MP ones. The changing trajectories of limit points, the transition time among different grades are then extracted as a function of key design/operating variables, respectively. These obtained functions are added to the scheduling framework as inequality constraints. By solving the resulted MINLP model, the production time, production sequence, and key design/operating parameters are optimized to maximize the economic benefit while roughly guaranteeing the desirable dynamic behavior during transitions. A polymerization of methyl methacrylate is finally studied to demonstrate the availability of the proposed method.

2. Method Description

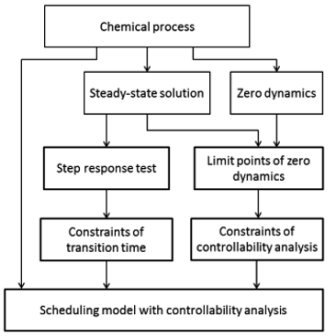


Figure 1. A schematic diagram of integrated scheduling and controllability analysis

2.1. Zero Dynamics and its Bifurcation Points

Controllability is a dynamic characteristic that describes the achievable control performance. Given a chemical process that is described as differential algebraic equations, one can easily extract its zero dynamics (Yuan et al., 2012). Unstable zero dynamics denotes the non-minimum phase behavior implies which means poor controllability. Stable zero dynamics means that the original chemical process exhibits minimum phase behavior. The bifurcation analysis for zero dynamics can divide the whole operating zone into NMP and MP ones. Based on our proposed machine learning method named hybrid symbolic regression and mathematical programming (Wei and Yuan, 2018), the changing trajectories of limit points will be extracted as a function of key design/operating variables.

2.2. Integration of scheduling and controllability analysis

The framework that integrates the scheduling and controllability analysis for a continuous multi-product chemical process is described by Figure 1.

(1) Set-up differential algebraic equations to describe dynamic chemical process.
(2) Solve the differential algebraic equation to obtain the steady-state solutions at different value of operating/design parameters via homotopy continuation.
(3) Extract the zero dynamics of the chemical process and obtain the distribution (changing trajectories) of the limit points of zero dynamics through bifurcation analysis.
(4) Build the function that describes the relationship between changing trajectory of limit points and key design/operating variables.
(5) Set the function that bridges the transition time among different grades and key design/operating variables.
(6) Introduce the functions obtained in steps (4) and (5) into the scheduling framework. Solve the MINLP model to get the optimal schedules and design/operating parameters.

2.3. Mathematical programming model for integrated scheduling and controllability analysis

The major assumptions in the model:

A1. Cyclic production schedule is desired

A2. Demand rate of each kind of product is constant

The following are the parameters of conventional model:

c_i = Production cost of product i per unit mass
d_i = Demand rate of product i
m_i = Unit inventory holding cost for product i
r_i = Production rate of product i
s_{ij} = Transition cost when production is changed from product i to j
τ_{ij} = Transition time when production is changed from product i to product j

The following are the variables:

H = Length of cycle
T_k = Length of time slot k
W_{ik} = Amout of product i produced during time slot k
X_{ik} = Binary variable to denote assignment of product i and slot k

The binary variables defined as follows:

$$X_{ik} = \begin{cases} 1 \text{ if in time slot } k \text{ the product } i \text{ is assigned} \\ 0 \text{ otherwise} \end{cases}$$

$$\min = \sum_i \sum_k c_i r_i \frac{[T_{ik} - \sum_{j \neq i} \tau_{ji} X_{jk-1}] X_{ik}}{H} + \sum_i \sum_j \sum_k s_{ij} \frac{X_{ik-1} X_{jk}}{H} + \sum_j \sum_k m_i W_{ik} \tag{1}$$

$$\sum_i X_{ik} = 1, \quad \forall k, \tag{2}$$

$$H = \sum_k T_k, \tag{3}$$

$$W_{ik} = r_i \left[T_k - \sum_{j \neq i} \tau_{ji} X_{jk-1} \right] X_{ik}, \quad \forall i \, \forall k, \tag{4}$$

$$\sum_k \frac{W_{ik}}{H} = d_i, \tag{5}$$

$$X_{ik} = \{0,1\}, \quad T_{ik} \geq 0, \quad W_{ik} \geq 0, H > 0 \tag{6}$$

$$f_i(u_{i1}, u_{i2} \ldots u_{ir}) = 0, \tag{7}$$

$$c_i = f_{ci}(u_{i1}, u_{i2} \ldots u_{ir}), \tag{8}$$

$$r_i = f_{ri}(u_{i1}, u_{i2} \ldots u_{ir}), \tag{9}$$

$$\tau_{ij} = f_{\tau ij}(u_{i1}, u_{i2} \ldots u_{ir}), \tag{10}$$

$$S_{ij} = f_{sij}(\tau_{ij}), \tag{11}$$

$$g((u_{i1}, u_{i2} \ldots u_{ir}) \geq 0, \tag{12}$$

$$U_{rMin} \leq u_{ir} \leq U_{rMax}, \tag{13}$$

Since the demand at unit time of each product is constant, so the selling revenue is constant at unit time. Thus the objective function as shown in Equation (1) is to minimizing the inventory cost, transition cost and feedstock cost to maximize the profit at unit time. Equations (1) ~ (6) actually describe the conventional scheduling model of multi-product chemical process. Equation (7) is constraint of operating variables on production point to ensure that the chemical process produces product i. Equation (8) denotes the relationship between operating parameters and production cost. Equation (9) is the functional relationship between operating parameters and production rate. Equation (10) is the constraint of transition time. Equation (11) represents the functional relationship between transition cost and transition time. Equation (12) denotes the constraints of controllability analysis.

3. Case Study and Discussion

We take the free-radical polymerization of methyl methacrylate (MMA) (Daoutidis P, 1990) as an example to illustrate the efficiency of our proposed framework. The temperatures of three operating points regarding products 1, 2, 3 are assumed to be 350K, 360K, 370K, respectively. Our previous contribution demonstrated that the open-loop steady-state multiplicity and the stability/controllability of the carried case are heavily depended on the flow rate of initiator F_I and the cooling water flowrate F_{cw}(Yuan Z, 2011). The objective of our novel scheduling is not only to seek the optimal production sequence but also guarantee desirable controllability properties. In addition to the short transition time, the optimal decision through the controllability analysis can enable F_I as the only operating variable to fulfil the desirable transitions. In other words, F_{cw} will keep constant. Through taking the extensive multiplicity analysis around each temperature, the specific relationship between F_I and F_{cw} can be extract as an explicit function $F_I = f(F_{cw})$. So in the present work, F_{cw} will be optimized to meet all of the above requirements.

As shown in Figure 2, the red zone means the undesirable zone with non-minimum phase behaviour, the green zone means the desirable zone with good controllability properties.

Through the hybrid symbolic regression and mathematical programming method, the distribution of bifurcation point can be extracted as the function of F_I and F_{cw} as mentioned in Equation (12). For the conventional scheduling, for each point, F_{cw} is as constant, F_I is also given. Through solving MINLP model shown by Equations (1) ~ (13), the results and the comparisons with conventional scheduling are discussed as follows.

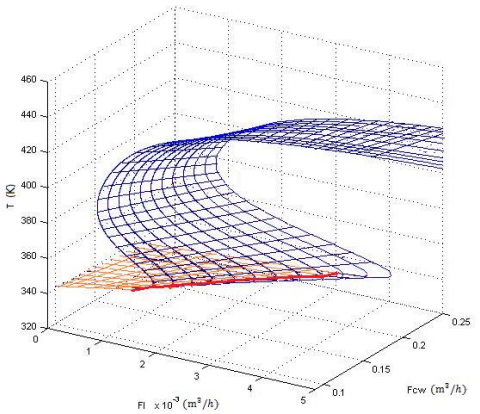


Figure 2. Zone classification of Phase Behavior for space surface of T-F_I-F_{cw}.

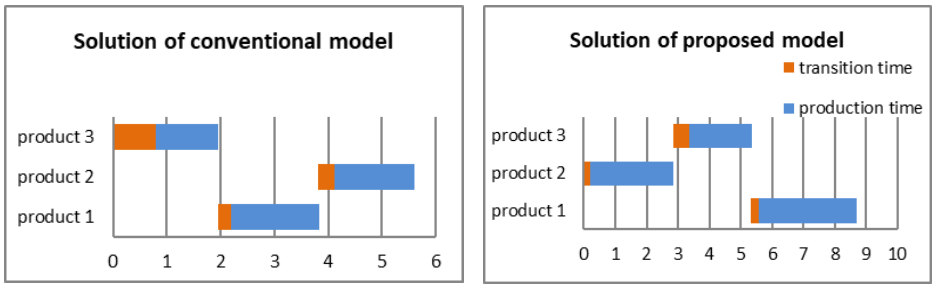


Figure 3. Comparisons on the Gantt chart of Production Sequences.

Figure 3 gives the Gantt Chart of Production Sequences. Table 1 outlines the detailed value of relevant inventory cost, transition cost, and production cost. Clearly, the production sequence from conventional scheduling and our scheduling with controllability is totally different (3-1-2 v. s. 2-3-1). From Table 1, it can be easily found that the optimized F_{cw} is lower than the value given in conventional scheduling. Due to the lower F_{cw} and then the resulted F_I, the production time of each grade is longer than the conventional scheduling. The pair of F_I and the optimized F_{cw} enables the system exhibits good controllability properties such as shorter total transition time. The conventional scheduling enjoys a shorter cyclic time at the cost of the higher inventory cost due to the longer transition times among different grades. The lower F_{cw} can further help the scheduling reduce the production cost. Thus the total cost of conventional scheduling is higher.

It should be mentioned that the uncertainties (accurate of the surrogates) have not been considered in our work. Clearly, the accurate of the surrogates reflects the distribution of the limit points. In the future, the proposed method will be tested by more complex process while also considering the above mentioned uncertainties.

Table 1. Detailed Results of Costs.

	Conventional model	New model
$F_{cw}(m^3/h)$	0.1588	0.117
$F_l(m^3/h)$	0.00175-0.00425-0.00275	0.00200-0.001375-0.002375
Cyclic time (h)	5.597	8.732
Production sequence	3-1-2	2-3-1
Time slot (h)	1.960-1.844-1.793	2.851-2.504-3.377
Production time (h)	1.152-1.615-1.48	2.639-1.992-3.134
Transition time (h)	0.808-0.229-0.313	0.212-0.512-0.243
Total cost ($/h)	152.148	130.665
Inventory cost ($/h)	13.515	21.039
Transition cost ($/h)	38.745	14.528
Production cost ($/h)	99.889	95.099

4. Conclusion

A new method to deal with scheduling problem while considering controllability properties for continuous multi-product chemical processes was proposed. By integrating the constraints which were represented by data-driven surrogate models, the scheduling model was equipped with the relationships between the changing trajectories of limit points. Taking the polymerization of methyl methacrylate as an example, the efficiency of the proposed scheduling model was demonstrated. Also, this work successfully promoted the switch from quantitative controllability analysis to qualitative controllability assessment.

Acknowledgement: The financial support from National Scientific Foundation of China (No. 21706143) is gratefully acknowledged.

References

Sahinidis N V, Grossmann I E. 1991, MINLP model for cyclic multiproduct scheduling on continuous parallel lines. Computers & Chemical Engineering, 15: 85-103.

Baldea M, Harjunkoski I. 2014, Integrated production scheduing and process control: A systematic review. Computers & Chemical Engineering, 71: 377-390.

Dias L, Ierapetritou M. 2017, From process control to supply chain management: An overview of integrated decision making strategies. Computers & Chemical Engineering, 106: 826-835.

Yuan Z, Zhang N, Bingzhen C, Zhao J. 2012, Systematic controllability of chemical processes. AIChE Journal, 58: 3096-4009.

Wei J, Yuan Z. Novel symbolic regression-mathematical programing based predictions of the molecular Cetane Number. Submitted to PSE 2018.

Daoutidis P, Soroush M, Kravaris C. 1990, Feedforward/feedback control of multivariable nonlinear processes. AIChE Journal, 36(10): 1471–1484.

Yuan Z, Chen B., Zhao J. 2011, Phase behavior analysis for industrial polymerization reactors. AIChE J. 57: 2795-2807.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50178-6

Co-current Partial Condensation of a Multicomponent Vapor Flowing Down Vertical Tubes – Cooling and Purification of a Reactive Mixture

Yihui Tom Xu

DuPont, 200 Powder Mill Road, Wilmington, Delaware 19803, USA
tom.xu@dupont.com

Abstract

Reactive multicomponent vapor condensation inside vertical tubes involves simultaneous transfer of mass and energy between vapor and condensed thin liquid film. Among the modeling tools most popular is the film theory where diffusional interactions, particularly in the vapor phase, are of critical importance for condensable components. In this work, linearization of the Krishna-Standart matrix based manipulations was made by extending the binary models of Colburn and Drew to multicomponent systems using Ackermann's correction factor in the diffusional mass flux calculations. Preliminary results indicate sweeping effect of the light species from the vapor phase bulk to the interface by the heavy component near the tube entrance. The light product compound can condense and react even if its physical driving force for condensation is negative throughout the tube.

Keywords: multicomponent reactive condensations, simultaneous momentum, mass and heat transfer, reactive separations, film theory based non-equilibrium model.

1. Introduction

Many industrially important vapor phase reactions are equilibrium controlled and highly exothermal in nature which leads to low conversions and high reactor exit temperatures. Jiang et al. (1997) developed tools for condensed mode cooling of polyethylene reactor gas in a tubular heat exchanger before it is recycled back to the reactor to shift the reaction forward. Sometimes the product needs to be purified or recovered by condensation of less volatile heavier by-products. Botsch and Stephan (1997) presented dynamic design tools for condensing hydrocarbons from a volatile non-condensable gas. Reactions often occur in the condensed liquid film under reduced temperatures so the end-product purity, yield and final temperature are determined by the relative rates of mass and heat transfer, and of reactions in the liquid film. While studying combined processes of NH_3 recovery from CO_2 containing waste water, Mackowiak et al. (2009) and Weiland and Taylor (1982) concluded that although one can usually simplify diffusional interactions and liquid reactions in absorption and stripping operations, one should not do so for condensations where the fluxes and liquid concentrations of the transferred species are no longer low.

In practice, the vapor mixture could contain non-condensable species, and condensable components which condense in the form of two immiscible liquid phases. This operation's broad importance in industry has brought sustained efforts with theories and tools for its analysis. Ackermann (1937) and Colburn and Drew (1937) developed rate based film model for binary mixtures, followed by the generalization of Fick's Law

applying to film models in the linearized theory of Toor (1964) and Stewart and Prober (1964). A classical general matrix method of implicit solutions to the Maxwell-Stefan equations was developed by Krishna and Standart (1976). This theory was later simplified by Krishna (1981), and Taylor and Smith (1982) to be explicit in computing molar fluxes. This was investigated and summarized in a HTRI document by Sardesai (1994), and in a design manual by Webb (2011). Davis et al. (1979) and Villadsen and Nielsen (1986) studied film reactors of simple first order reactions with rigorous transport models. Higher order and reversible reaction enhancement factors were also studied based on dilute solute assumption for absorption processes (Yue et al. 2013 and Meldon et al. 2007). However, publications on multicomponent condensations in which the condensed species undergo complex reversible reactions in liquid phase are scarce. The situation gets even more complicated when desorption is also involved due to the high volatility of some of the formed reaction intermediate products. This paper presents and solves a rigorous model combining kinetically controlled reversible reactions in the non-dilute liquid condensate film with simultaneous interacting multicomponent condensation and evaporation.

2. Process Description

Figure 1 shows schematics of a vertical shell-and-tube exchanger and falling film in tube for partial condensation of solvents S2, S1, and S3 mixed with product A with increasing volatility, and impurities Imp1 and Imp2 that could condense but do not take part in liquid reactions. For the sake of general applicability, species are not specified but rather, their thermodynamic and transport properties and mutual reactivities are defined. Although very volatile, product A does not have to act as an inert gas in this system. T_V, T_i, T_L, T_W, T_C are temperatures of gas, interface, liquid, wall, and coolant respectively. P, F, Y are system pressure, vapor flow rate and mole fraction respectively. $\delta(z)$ is the film thickness (mm) as function of downflow distance. $U_z(x)$ is the film axial velocity as function of the film thickness (m/s). N_{CV} and N_{VZ} are condensing and downflow gas fluxes respectively.

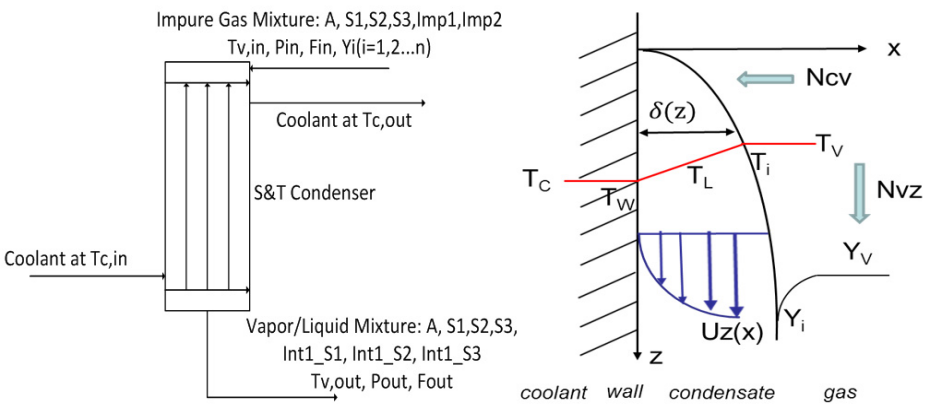


Figure 1. Configurations of a vertical shell-and-tube exchanger and liquid film in tube.

3. Condensate Film Energy and Momentum Balances

Depending on cooling capacity of the tubes the condensate flow can be in the laminar or turbulent regimes. Therefore, uniform or varied temperature and velocity profiles across the film thickness could be generated. Holdup time of a laminar film closest to the tube wall should be longer than that at the gas-liquid interface. If the drag force predominates

over the gravitational force the radial profiles of the axial velocity will be nearly linear with thickness, rather than quadratic. Rates of reactions will be impacted under all these scenarios. Dukler (1959) and Bird et al. (1960) added turbulent transport terms to the standard momentum and energy equations so that transitions from laminar to turbulent flow regimes can be realized in a simple and elegant manner. Charef et al. (2018) also presented rigorous transport equations solving single component tube condensations. The momentum and energy balance equations and their boundary conditions are listed below:

$$\frac{1000}{\delta(z)} * \frac{d\tau_z}{dx} + \rho_{ML} * g = 0 \qquad \text{Motion (g is acceleration due to gravity)} \tag{1}$$

$$\frac{1000}{\delta(z)} * \frac{dq_z}{dx} = 0 \qquad \text{Energy} \tag{2}$$

$$U_z = 0 \qquad @ \text{ x = 0 (no slip at tube wall)} \tag{3}$$

$$\tau_z = \tau_{drag} + P \qquad @ \text{ x = 1 (total stress = drag stress plus momentum)} \tag{4}$$

$$T = T_w \qquad @ \text{ x = 0 (contacting tube wall)} \tag{5}$$

$$T = T_i \qquad @ \text{ x = 1 (gas-liquid interface)} \tag{6}$$

where the terms and variables in the above equations are defined below:

$$\tau_z = 0.001 * \left(\nu_{film} + \nu_{eddy}\right) * \frac{1000}{\delta(z)} * \frac{dU_z}{dx} \qquad \text{Stress} \tag{7}$$

$$q_z = -0.001 * \left(\lambda_L + \nu_{eddy} * c_{pL}\right) * \frac{1000}{\delta(z)} * \frac{dT}{dx} \qquad \text{Heat flux} \tag{8}$$

$$\tau_{drag} = 0.5 * f_{gas} * \rho_{Mv} * (u_{gas} - U_z(@x = 1))^2 \qquad \text{Drag of gas on liquid} \tag{9}$$

$$P = N_{CV} * MW_{cond} * u_{gas} \qquad \text{Momentum transferred by condensing gas} \tag{10}$$

$$\nu_{eddy} = 1000 * \rho_{ML} * n_{Deissler} * U_z * x \left(\frac{\delta(z)}{1000}\right)(1 - exp(-n_{Deissler} * U_z * x * (\frac{\delta(z)}{1000})/(0.001 * \nu_{film})/\rho_{ML})) \qquad \text{Eddy viscosity per Deissler} \tag{11}$$

where $\boldsymbol{n_{Deissler}}$ is a constant. $\boldsymbol{\lambda_L}$, $\boldsymbol{\nu_{film}}$, $\boldsymbol{\rho_{ML}}$ and $\boldsymbol{c_{pL}}$are liquid thermal conductivity, viscosity, density and heat capacity respectively. $\boldsymbol{u_{gas}}$ and $\boldsymbol{\rho_{Mv}}$ are gas axial velocity and density respectively. MW_{cond} is the molecular weight of the condensing vapor.

4. Multicomponent Mass and Heat Transfer

The process being examined here involves large fluxes of the heavy species early in the entrance region which greatly influence the fluxes of the more volatile species and could sweep some of them toward interface. The simplified treatment of gas absorption and reaction into dilute solvent system should not apply here because the condensates are concentrated and the impact of overall flux to individual fluxes could not be ignored. Simplifications of Krishna-Standart method are made in this work by extending the binary models of Colburn and Drew to multicomponent systems using Ackermann's correction

factor in the diffusional flux calculations. Zero-flux vapor mass transfer coefficients are calculated from heat transfer correlations through Colburn-Chilton analogy (1934). All components are condensable and reactions consuming condensed species enhance liquid film mass transfer, so the liquid side mass transfer is not important and the model sets the liquid film resistance to zero. The analysis of Ackermann (1937) is adopted in the bulk vapor sensible heat transfer calculations. All equations for radial and vertical mass and heat balance can be found in the detailed HTRI report authored by Sardesai (1994).

5. Phase & Chemical Equilibrium and Reaction Kinetics

All preferentially condensed solvents react with the more volatile product A to form by-products (intermediates) in the liquid phase. Each individual solvent reacts with A forming first intermediate between A and the respective solvent. Although A could react with the first intermediates to form higher intermediates, only the first reactions have significant rates and only the first intermediates have sizable equilibrium concentrations. These reversible reactions in series can be written as the following:

$$A + S_j \overset{Kx_1}{\Longleftrightarrow} Int_1 - S_j \quad j = 1,\ 2,\ 3\ \textit{(solvents)} \tag{12}$$

$$A + Int_i - S_j \overset{Kx_{i+1}}{\Longleftrightarrow} Int_{i+1} - S_j \quad i = 1,\ 2, \dots, n\ (steps)\ , j = 1, 2, 3 \tag{13}$$

where Kx_{i+1} is the chemical equilibrium constant of the reaction forming respective intermediate, chemical and phase equilibrium equations can be written as the following:

$$K_{x_i - S_j} * a_A * a_{Int_{i-1}-S_j} = a_{Int_i - S_j} \quad i = 2,\ \dots, n+1;\ j = 1, 2, 3 \tag{14}$$

$$y_i P = a_i P_i^{sat} exp(\int_{Psat}^{P} \frac{v_{i,liq}}{RT} dP) \quad i = all\ species\ in\ the\ system \tag{15}$$

note that Poynting pressure correction term in Eq. (15) is not important under moderate pressure, the highest vapor pressure component is product A which is still moderately condensable under the operating pressure. Activity coefficients in $a_i = \boldsymbol{\gamma_i x_i}$ are regressed from experimental total pressure data or estimated from the UNIFAC method. For components showing low concentrations in liquid phase temperature dependent infinite dilution activity model can be used for non-ideal interactions. Reversible reaction rate expressions for reaction step i and solvent j can be written as:

$$R_{x_i - S_j} = k_{i-S_j}(x_A * x_{Int_{i-1}-S_j} - \frac{x_{Int_i - S_j}}{K_{x_i - S_j}}) \quad i = 2,\ \dots, n+1;\ j = 1, 2, 3 \tag{16}$$

$$R_{A-S_j} = -\sum R_{x_i - S_j} \quad i = 1,\ \dots . n \quad \text{for } j = 1, 2, 3 \tag{17}$$

$$R_{S_j} = -R_{x_1 - S_j} \quad \text{for } j = 1, 2, 3 \tag{18}$$

activation energy, equilibrium constants $K_{x_i - S_j}$ and rate constants k_{i-S_j} are all determined from experiments of individual reactions targeting specific solvent S_j.

6. Preliminary Results and Discussions

The combined set of above equations is solved along the incremental length of the tube using Method of Lines. Useful species physical properties are obtained from Aspen Plus® Databank (2015) by working in the environment of Aspen Custom Modeler® version 8.8. Figures 2 and 3 show the mole fractions of species in bulk vapor, the interface and major condensed and formed species in the liquid phase. S2 is the main and heaviest solvent so it condenses first and reacts with condensed A to form Int1_S2. But S1 and S3 are more volatile and react faster with A in liquid and their formed first intermediates show peak

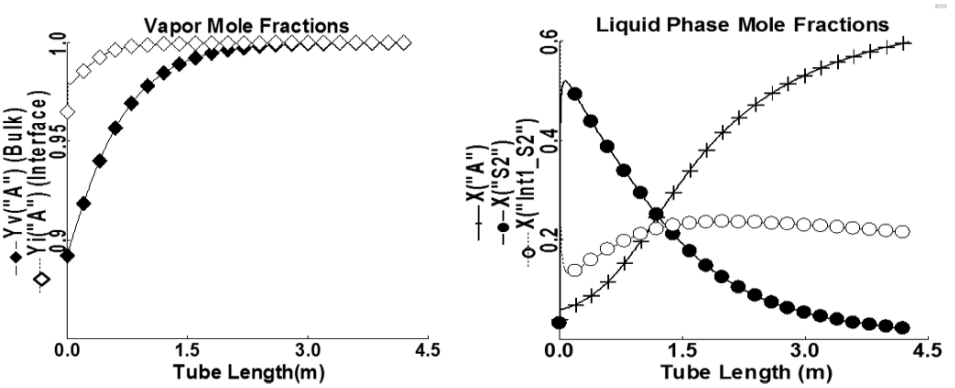


Figure 2. Consistent negative condensing flux of A and liquid profiles of S2 and Int1_S2.

behaviour in the liquid phase, forming higher intermediates by reacting further with A. The initial high S2 flux into liquid film sweeps the lighter species to the interface causing initial higher interface than bulk vapor concentrations for A, S1 and S3. Whereas S1 and S3 fluxes quickly reverse back to positive into liquid due to faster reactions, the physical driving force for A is always negative in the vapor film while it is condensing from the beginning. A's purity increases and the recovery drops as more of it condenses and reacts.

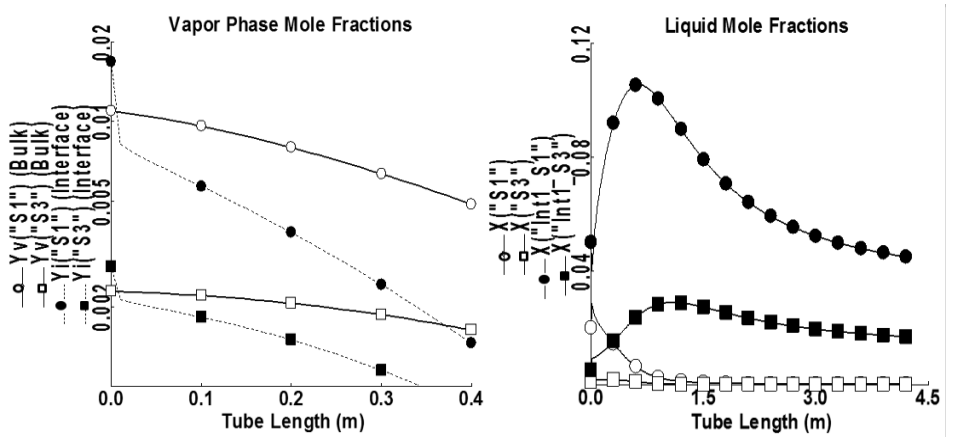


Figure 3. Initial negative condensing fluxes of S1 and S3 (left) and liquid species (right).

7. Conclusions

Rigorous kinetic equations of parallel reversible reactions between several species are introduced into a multicomponent condensation and evaporation system. Preliminary

analysis shows that there are strong interactions between the diffusion fluxes of condensing species, and between mass and heat transfer rates and reaction kinetics. Species could still condense when physical driving force is negative which is the cause of reduced recovery for product A. Strong enhancements on condensation by reactions under the current relative rates of reaction and transport are caused by large concentration driving force and reasonably good solubility of the components in liquid film.

8. References

1. Jiang, Y., McAuley, K. B., and Hsu, C. C., "Nonequilibrium Modeling of Condensed Mode Cooling of Polypropylene Reactors." AICHE J., 43, 13, 1997.
2. Botsch, T. W., and Stephan, K., "A Model to Describe the Steady-State and Dynamic Behavior of Vertical Condensers Fed by Mixtures." Rev. Gen.Therm. (1997) 36, 481-491.
3. Mackowiak, J. F., Gorak, A., and Kenig, E. Y., "Modeling of combined direct-contact condensation an dreactive absorption in packed columns." Chem. Eng. J.,146 (2009), 362.
4. Weiland, R. H., and Taylor, R., "Mass Transfer." Chem. Eng. Education, XVI, 4, 1982.
5. Ackermann, G., "Heat Transfer and Molecular Mass Transfer in the Same Field at High Temperature and Large Partial Pressure Differences." V. D. I. Forschungsheft 8, 1-10. 1937.
6. Colburn, A. P., and Drew, T. B., "The Condensation of Mixed Vapors." Trans. Amer. Inst. Chem. Eng., 33, 197, 1937.
7. Toor, H. L., "Solution of the Linearized Equations of Multicomponent Mass Transfer." AICHE J., 10(4), 460, 1964.
8. Stewart, W. E., and Prober, R., "Matrix Calculation of Multicomponent Mass Transfer in Isothermal Systems." Ind. Eng. Chem. Fundam., 3, 224, 1964.
9. Krishna, R., and Standart, G. L., "A Multicomponent Film Model Incorporating a General Matrix Method of Solution to the Maxwell-Stefan Equations." AICHE J., 22, 383, 1976.
10. Krishna, R., "An Alternative Linearized Theory of Multicomponent Mass Transfer." Chem. Eng. Sci., 36, 219, 1981.
11. Taylor, R., and Smith, L. W., "On Some Explicit Approximate Solutions of the Maxwell-Stefan Equations for the Multicomponent Film Model." Chem. Eng. Comm., 14, 361, 1982.
12. Sardesai, R., G., "Film Theory Methods for Condensation of Multicomponent Vapors or Vapor-Gas Mixtures." HTRI Report CG-7, 1994.
13. Webb, D., "Condensation of Multicomponent Vapors." THERMOPEDIA™, 2011.
14. Davis, E. J., Van Ouwerkerk, M., and Venkatesh, S., "An Analysis of the Falling film Gas-Liquid Reactor." Chem. Eng. Sci., 34, 539, 1979.
15. Villadsen, J., and Nielsen, P. H., "Models for Strongly Exothermic Absorption and Reaction in Falling Films." Chem. Eng. Sci., 41, 1655, 1986.
16. Yue, J., Nijhuis, T. A., and Schouten, J. C., "Enhancement Factor for Gas Absorption in a Finite Liquid Layer." Chem. Eng. Technol. 2013, 36, No. 4, 611-626.
17. Meldon, J. H., Olawoyin, O. O., and Bonanno, D., "Analysis of Mass Transfer with Reversible Chemical Reaction." Ind. Eng. Chem. Res. 2007, 46, 6140-6146.
18. Dukler, A. E., "Dynamics of Vertical Falling Film Systems." Chem. Eng. Prog. 55, 10, 1959.
19. Bird, R. B., Stewart, W. E., & Lightfoot, E. N., Tables 3.4-2, 3.4-5, 10.2-2, P160-161, 1960.
20. Charef, A., Feddaoui, M., Najim, M., and Meftah, H., "Liquid Film Condensation from Water Vapor Flowing Downward along a Vertical Tube." Desalination, 409 (2018) 21-31.
21. Chilton, T. H. and Colburn, A. P., "Mass Transfer Coefficients; Prediction from Data on Heat Transfer and Fluid Friction." Ind. Eng. Chem. 26, 183, 1934.
22. Aspen Plus® V8.8, Aspen Technology, Inc. Bedford, Massachusetts, USA, 2015.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50179-8

Adjusting the Heating Value of LNG using a Superstructure for Hydrocarbon Recovery

Arnab Dutta, Iftekhar A. Karimi*, Shamsuzzaman Farooq*

Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117585
cheiak@nus.edu.sg chesf@nus.edu.sg

Abstract

LNG is expected to play a major role in global natural gas trade. Implementing a heating value adjustment unit within regasification terminal allows importing LNG from diverse sources around the world, thus increasing the energy security of the importing country. In this study, we investigate the profitability of heavy hydrocarbons removal (HHR) as a strategy to adjust the heating value of LNG to meet desired specifications. In this context, we have developed a superstructure for the HHR process, which incorporates LNG cold utilization, operating and capital costs of the process, and the revenue generated from coproducts produced. The proposed superstructure is then solved within a simulation-based optimization paradigm. The optimal HHR process configuration not only adjusts the heating value of LNG by partially recovering heavy hydrocarbons, but also utilizes LNG's cold energy to reliquefy the methane-rich stream before pumping to send out pressure and produces ethane and LPG as co-products. Our results show that the additional revenue generated by selling these co-products at current market prices yields profit after offsetting the investment for the HHR process.

Keywords: LNG, Heating value adjustment, Hydrocarbons recovery, Simulation-based optimization

1. Introduction

Riding on the shale gas boom, we are witnessing a paradigm shift towards natural gas (NG) as the major energy source. The cleanest among fossil fuels, consumption of NG is projected to increase at a rate of 1.4 – 1.6% per year from 2008 to 2035 (Lim et al., 2013). Like other fossil fuels, NG is also a traded commodity. Countries endowed with surplus NG reserves export it either via pipeline (Piped Natural Gas or PNG), as CNG (Compressed Natural Gas), or as LNG (Liquefied Natural Gas) in specially designed cryogenic tankers. For transportation over a long distance, typically above 3500 km, LNG has been the preferred method for economical, technical, safety-related, and political reasons. The LNG trade is projected to have a higher growth rate than the PNG trade (Lim et al., 2013).

Imported LNG is stored in an insulated cryogenic tank at the receiving terminal and must be re-gasified before it is sent out to the end consumers. The send out gas must meet certain specifications, and one such specification is its heating value. LNG is composed of methane along with other heavy hydrocarbons like ethane, propane, butane, and traces of pentane. The heating value depends on the composition of LNG, which depends on the country of origin. Unless the regasification terminal is equipped with facilities to adjust the heating value of the send out gas, the imported LNG must

match the desired pipeline specification, which may severely restrict the choice of import LNG sources. An LNG importing terminal equipped to adjust the heating value of LNG before regasification and send out will allow LNG import from diverse sources around the world, thus increasing the energy security of the importing country. It would also allow taking advantage of reduced LNG spot prices, when available. The heating value of LNG can be adjusted by adding nitrogen to the LNG stream. This will not only require nitrogen but will also increase the percentage of inert in the send out NG. According to pipeline specifications, the percentage of inert in the send out gas should not exceed 3 mol % (Fahmy et al., 2015). Therefore, the composition range of LNG amenable to heating value adjustment by diluting with nitrogen is also limited. A less restrictive strategy for lowering the heating value of LNG is to partially recover these heavy hydrocarbons from the LNG feed. These hydrocarbons recovered can be further separated to obtain value-added products like ethane and LPG (Liquefied Petroleum Gas) (Gao et al., 2011). Thus, unlike nitrogen dilution, the heavy hydrocarbon removal (HHR) process not only adjusts the heating value of any LNG feed but also produces sellable co-products that will generate additional revenue for the regasification terminal. However, the revenue generated by selling these co-products at current market prices must recover the additional investment required for the HHR process.

Thus, the objective of this study is to analyze the profitability of the HHR process as a strategic measure to adjust the heating value of any given LNG feed. We present a superstructure for the HHR process and solve it using a simulation-based optimization algorithm to obtain a process configuration for a given LNG composition that maximizes the profit for the regasification terminal.

2. HHR process superstructure

The HHR process involves distillation as the separation technique to separate the heavy hydrocarbons from LNG to reduce its heating value. The LNG stream is first pressurized and then fed to a demethanizer column, where methane is completely recovered as the top product, and a hydrocarbon stream rich in ethane, propane, and butane is recovered as the bottom product. The top product is primarily methane vapor. The operating pressure of the demethanizer column is usually lower than the required send out pressure of natural gas. As a result, the methane rich top vapor product from the demethanizer is either compressed using a compressor to the desired send out pressure or the cold energy of LNG is used to reliquefy it so that it can be pumped to the desired send out pressure instead of using the energy intensive compression route. The methane rich vapor product from the demethanizer can be completely or partially liquefied using a multi-stream heat exchanger using the pressurized subcooled LNG stream. If the available cold LNG stream can only partially liquefy the vapor stream, then the liquefied fraction is pumped and the remaining vapor needs to be compressed to the send out gas pressure. The cold LNG stream gains heat in the process and passes through a flash drum. The liquid stream from the flash drum is fed to the demethanizer, and the methane–rich vapor stream is sent back to the multi-stream heat exchanger for liquefaction. The bottom product of the demethanizer is then sent to the deethanizer column to separate ethane from LPG. It is advantageous to store these co-products as liquid at a low pressure. Therefore, the bottom stream from the demethanizer column is depressurized before being fed to the deethanizer. The liquid bottom stream from the demethanizer can either be throttled or it can be expanded using an expander to generate power to partially offset the electricity consumption of the HHR process. However, to

implement an expander the stream must be superheated to avoid any liquid at the exit of the expander. In this study, steam is used for superheating the liquid bottom product from the demethanizer before expansion. The reboilers of the demethanizer and the deethanizer use either steam or sea water as the hot utility depending on the operating pressures of the distillation columns. The cold energy associated with the LNG stream not only liquefies the top product of the demethanizer but also satisfies the condenser duty of the deethanizer. The processed LNG stream at the send out pressure is finally vaporized using sea water to meet the required send out gas temperature.

In this study, we have adopted a comprehensive approach to incorporate these options within our proposed superstructure for the HHR process. The superstructure is then solved within a simulation-based optimization paradigm to obtain process configuration and operating parameters that maximize the profitability of the HHR process for an LNG stream of a given composition. The proposed superstructure for the HHR process is illustrated in Figure 1. The decision variables pertaining to the optimization of the HHR process include: pressures of stream 2 and feed of the deethanizer; pressure differences across pumps 2 and 4; temperature differences across streams 3-2, 6-5A, and 4-3V; number of stages for the demethanizer and the deethanizer columns; feed location for the deethanizer column; split ratios between streams 5A:5B and 10A:10B; minimum temperature approach across the heat exchanger (HEX) and the degree of superheating for stream 11.

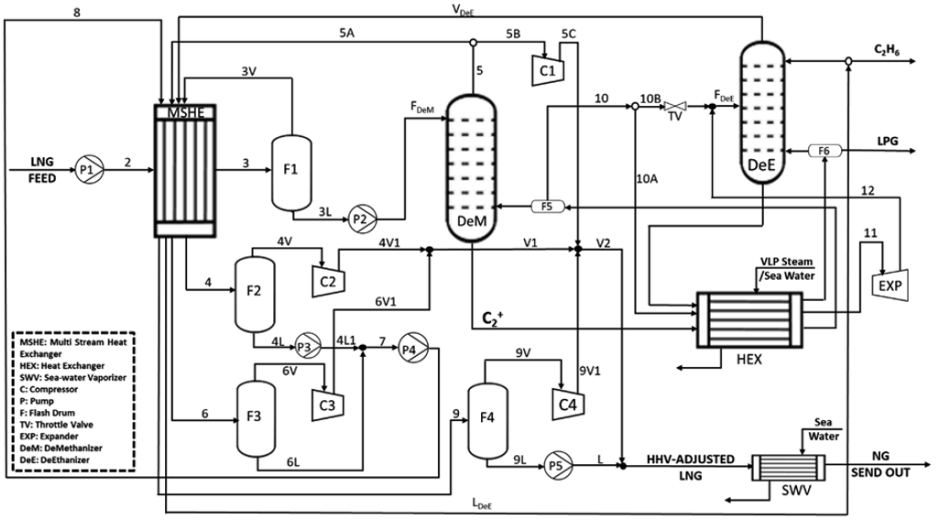


Figure 1: Proposed superstructure for the HHR process.

3. Simulation-based optimization paradigm

Here we adopt a simulation-based optimization paradigm where we combine the benefits of a process simulator to seamlessly perform all thermodynamic calculations and an external platform to perform rigorous optimization. In this study, Aspen HYSYS is used as the process simulator and it is interfaced with MATLAB as the optimization platform. The decision variables obtained at each iteration of the optimization are sent to HYSYS to perform the process simulation and the process data obtained from HYSYS is sent back to MATLAB to update the decision variables for the next iteration.

The superstructure optimization is solved with the following process constraints: the minimum temperature approach for the multi-stream heat exchanger must be at least 3 ^{0}C, the stream exiting the expander (i.e. stream 12 in Figure 1) must be in vapor phase, and the higher heating value (HHV) of the send out natural gas must meet the desired heating value specification used in this study (38.5 MJ/Nm^3). The objective is to maximize the profitability of the HHR process. The objective function is given by equation (1). The capital cost for each equipment is calculated using the relevant cost correlations as given in Turton et al. (2012). The capital investment is annualized using the annualizing factor (AF) given by equation (2). The parameters used for process simulation are highlighted in Table 1.

$$\max_x \ Profit = \left(\sum Revenue_{NG,\ Ethane,\ LPG} - OPEX - AF * CAPEX\right) \quad (1)$$

$$AF = \frac{i(1+i)^n}{(1+i)^n - 1} \quad [i = 10\%, n = 20\ years] \quad (2)$$

Table 1: Summary of process simulation parameters

	Parameters
Fluid package	Peng-Robinson
LNG compositions (mol %):	
Methane	87.83
Ethane	8.30
Propane	2.98
i-Butane	0.40
n-Butane	0.48
Nitrogen	0.01
LNG HHV (MJ/Nm^3)	42.23
NG send out specifications:	
Pressure (kPa)	8000
Temperature (^{0}C)	15
HHV (MJ/Nm^3)	38.50
Flowrate (t/h)	100
Efficiency: Compressor, Pump, Turbine	75 %
Pressure drop per stage in distillation column (kPa)	0.7
Pressure drop on each side of heat exchangers (kPa)	20
Methane recovery	99.99 %
Ethane recovery	99.90 %
Propane recovery	99.90 %
Sea water: Inlet temperature (^{0}C)	25
Outlet temperature (^{0}C)	22
Inlet pressure (kPa)	500
Costs of co-products: Ethane ($/t)	290
LPG ($/t)	580
Utility costs: Electricity ($/kWh)	0.15
Steam ($/t)	21.9

3.1. Optimization algorithm

In this study, Particle Swarm Optimization (PSO) algorithm is coded in MATLAB to optimize the proposed superstructure for the HHR process. The PSO algorithm implemented in this study is based on the gbest type, where every particle is attracted to the best solution obtained by any particle within the population thus, accelerating the rate of convergence. Initialization of PSO particles are done using the Sobol sampling technique. PSO effectively handles exploration and exploitation using three parameters: a cognitive component, a social component, and inertia weight (Eberhart and Kennedy, 1995). In our PSO code, the cognitive component is linearly decreased, and the social component is linearly increased with each iteration of PSO. The inertia weight is adapted based on the total number of particles that improve its solution in each iteration as a feedback parameter. Lower and upper bounds of the decision variables confine the positions of particles within the multi-dimensional search space. The velocities of particles are also restricted by suitable bounds specified in each dimension. When the position or velocity of a particle violates the upper or lower bound in one dimension, it is fixed at the respective bound. If a particle is at its upper bound with a positive velocity or if the particle is at its lower bound with a negative velocity, then the direction of velocity is reversed to impede the movement of the particle in the wrong direction. To handle process constraints, particles are compared according to the following rules: a feasible particle is always preferred over an infeasible particle; among two feasible particles, one having the better objective function value is preferred; among two infeasible particles, one having the smaller constraint violation is preferred. The PSO algorithm is also combined with a mesh adaptive direct search optimization technique (NOMAD) available from the OPTI Toolbox (Currie and Wilson, 2012). Feasible PSO particles are used as initial guesses for NOMAD, and feasible solutions obtained from NOMAD are then used to replace infeasible PSO particles followed by PSO particles with worse objective function values compared to the feasible NOMAD solutions. However, replacement of a PSO particle is allowed only if the feasible NOMAD solution is unique in the decision space w.r.t. all other PSO particles. A distribution-based criterion based on the difference between the best and the worst objective function value for PSO particles is used to terminate our optimization algorithm. The parameters pertaining to the PSO algorithm are given in Table 2.

Table 2: Parameters for the PSO algorithm

No. of particles (N)	Cognitive component (c_{10})	Social component (c_{20})	Inertia weight (w_{Min}, w_{Max})	Maximum iterations (itermax)
60	2.5	1.5	[0, 1]	100

4. Results and discussion

The proposed superstructure for the HHR process is solved using the optimization algorithm within a simulation-based optimization paradigm. The average computational time for solving the optimization problem is about 30 h, which includes repeated evaluations of the HHR process flowsheet within HYSYS during the optimization. The amount of LNG to be processed to maintain a desired send out gas rate depends on the extent of heating value adjustment required. Our results show that the available cold energy of LNG is sufficient to liquefy the various vapor streams in the HHR process, thus avoiding any vapor compression in the optimal HHR process configuration. This not only reduces the power consumption (thus the operating cost) but also the capital

investment of the HHR process. The topology corresponding to the expansion via throttle valve is chosen over the use of an expander, as the power generated using an expander is not adequate to compensate the additional capital investment. As seen from Table 3, the revenue generated from the co-products (ethane and LPG) at their respective current market prices yields a profit of about 9.4 % for the LNG regasification terminal. Thus, the HHR process turns out to be a profitable strategy to adjust the heating value of the imported LNG to meet the desired pipeline specification.

Table 3: Summary of results for the HHR process

LNG Processed (t/h)	Ethane Produced (t/h)	LPG Produced (t/h)	TAC (million$/y)	Profit (million$/y)	Power Consumption (kWh)	Steam Consumption (t/h)
122	10.65	11.31	374.80	35.40	1473	5.25

5. Concluding remarks

In this study, we have demonstrated that in an LNG receiving terminal, heavy hydrocarbons removal prior to regasification and send out is a techno-economically viable option to diversify LNG sources and increase the energy security of the importing country. We have presented a superstructure for the heavy hydrocarbons removal process and solved it within a simulation-based optimization paradigm. The HHR process configuration utilizes the cold energy of LNG thus, resulting in an energetically efficient process that not only adjusts the heating value of the LNG feed but also produces value added co-products i.e. ethane and LPG. The additional revenue generated by selling these co-products yields a net profit of about 9.4 % for the regasification terminal for the representative rich LNG composition as given in Table 1.

The simulation-based optimization framework developed in this study can readily evaluate the techno-economic feasibility of a HHR process to adjust the heating value of any given LNG feed within a regasification terminal.

Acknowledgement

This work is funded by the National University of Singapore through a seed grant (R261-508-001-646/733) for CENGas (Centre of Excellence for Natural Gas) and Energy Innovation Research Programme Award (EIRP) NRF2014EWTEIRP003-008, administrated by the Energy Market Authority (EMA).

References

Lim, W.; Choi, K.; Moon, I. Current Status and Perspectives of Liquefied Natural Gas (LNG) Plant Design. *Ind. Eng. Chem. Res.* **2013**, *52* (9), 3065.

Fahmy, M. F. M.; Nabih, H. I.; El-Rasoul, T. a. Optimization and Comparative Analysis of LNG Regasification Processes. *Energy* **2015**, *91*, 371.

Gao, T.; Lin, W.; Gu, A. Improved Processes of Light Hydrocarbon Separation from LNG with Its Cryogenic Energy Utilized. *Energy Convers. Manag.* **2011**, *52* (6), 2401.

Turton, R.; Bailie, R. C.; Whiting, W. B.; Shaewitz, J. A.; Bhattacharya, D. *Analysis Synthesis and Design of Chemical Processes*; **2012**.

Eberhart, R.; Kennedy, J. A New Optimizer Using Particle Swarm Theory. In *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*; IEEE, **1995**; pp 39–43.

Currie, J.; Wilson, D. I. OPTI, Lowering the Barrier Between Open Source Optimizers and the Industrial MATLAB User. *Found. Comput. Process Oper.* **2012**, No. January 2012.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50180-4

Simultaneous optimization of solvent composition and operation parameters for sulfolane aromatic extractive distillation processes

Q. Wang, B.J. Zhang, C. He, Q.L. Chen*

School of Chemical Engineering and Technology, Guangdong Engineering Center for Petrochemical Energy Conservation, Sun Yat-Sen University, No. 135, Xingang West Road, Guangzhou, 510275, China
chqlin@mail.sysu.edu.cn

Abstract

BTXs (benzene, toluene and xylenes), used widely as basic chemical intermediates, are separated generally from petroleum naphtha, coking naphtha, and pyrolysis gasoline. These oil fractions contain aromatic and non-aromatic hydrocarbons between C_5 to C_8 in which certain components have approximate boiling points and form azeotropes. Extractive distillation (ED) is one of the most attractive approaches for separating azeotropic mixtures with approximate boiling points. The aromatic ED technique aided by sulfolane used widely in recent years because of its simple process flowsheet, small equipment investment and high energy efficiency. The solvent and operating parameters are the cores of ED process in the economical design. Some results were published to rapidly screen the solvents as well as process parameter optimization for ED separating process, but few contributions have been reported to optimize an extractive distillation from the view of simultaneous optimization of the solvent composition and operation parameters for process economics. In this work, we propose a systematic optimization approach for energy saving of sulfolane aromatic ED processes based on process simulation and genetic algorithm (GA) optimization. The NRTL property method and rigorous unit models in Aspen Plus are used to simulate the ED processes. In particular, considering multi-variable and their interaction, a coordinating strategy is proposed to optimize the solvent composition and significant operating parameters for the ED processes by combining Aspen Plus with MATLAB. Under given separation specifications, the energy consumption and separation efficiency are investigated and optimized by GA optimization. The results show that the through the above optimization, the minimum heating utility of the ED process is significantly reduced by 29.2%.

Keywords: Extractive distillation; Sulfolane; Process simulation; Optimization; Energy saving.

1. Introduction

BTXs (benzene, toluene and xylene isomers) are extremely important basic chemicals for the oil, gas and petrochemical industries with global productions above 40 million of metric tonnes per year for benzene and xylenes and roughly the half for toluene. BTXs are popularly produced from petroleum naphtha, coking naphtha, and pyrolysis gasoline through hydrotreating and continuous catalytic reforming (CCR) processes and used as starting points of derived downstream

products. CCR processes produce high-quality oil streams that include aliphatic, alicyclic and aromatic hydrocarbons between C_5 to C_8 [1]. Certain components in the oil streams have approximate boiling points and form azeotropes. Hence, it is difficult and energy-intensive to separate aromatics from the oil streams via regular distillation. Some improved distillation processes including extractive distillation (ED) process are adopted to separate the mixture of aromatic and non-aromatic hydrocarbons[2]. The aromatic ED technique has been widely used in recent years because of its simple production process, small equipment investment and high energy efficiency. As an effective method, ED with a separating agent, especially a polar solvent[3-5], has been used to separate the slightly polar aromatics from the reformate, pyrolysis gasoline or coke-oven light oil. Thus, the light components can be drawn out from the top of the column as a product; the heavy components can be drawn out from the bottom of the column as another product. The product containing the entrainer can be further separated through regular distillation. Among these available entrainers, sulfolane is very popular because of the highest selectivity, higher dissolving capacity, higher boiling point and higher thermal stability[6].The sulfolane process licensed by UOP[7] is currently the most popular technology.

ED uses an entrainer to increase the relative volatility of the light and heavy components. Thus, the light components can be drawn out from the top of the column as a product; the heavy components can be drawn out from the bottom of the column as another product. This implicates additional distillation steps to separate the entrainer from both the extract and raffinate phases, and to purify the solvent, with consequently, additional investments and energy consumption. With an increasing awareness of the environmental impacts and practical limitations on energy supply, researchers aim to increase the efficiency of energy utilization. Entrainer optimization and process optimization are the useful approaches to save energy as well as increase the efficiency of the current energy utilization for aromatic ED process. In the research of Pete et al.[8], a preliminary solvent screening tool for two-column extractive distillation flowsheets is described whose results can be used to initialize process simulators. The tool uses shortcut approximations to extractive distillation that accurately predict the relative column performance of alternative solvents with rigorous simulations using the Aspen Plus process simulator. It was found that staging requirements for the flowsheets could be approximated using 4 times the number of stages predicted using the Fenske equation when it was applied over each column section. Using these approximations resulted in a shortcut method for evaluating and comparing alternative solvents and solvent families, predicting the staging requirements and quickly determining the feasibility of completely heat integrating a flowsheet. Ionic liquid, co-solvents are also reported as promising solvents for the separation of this close-boiling point mixture by extractive distillation[9,10]. The operating parameters are the cores of an existing ED process as well as solvent in the economical design. Process simulation and optimization also studied in the aromatic ED process. Choi et al.[11] established a model and optimization system for the sulfolane extraction plant based on process simulation. Optimum recycle rates were found using the sequential quadratic programming (SQP) algorithm. It was found that almost a 10%

increase of the purity of benzene could be achieved by the application of the optimization system.

An effective solution with high selectivity and solubility as well as process operating parameters is crucial to the energy requirement and economic performance for an ED process. As mentioned above, some results were published to rapidly screen the solvents as well as process parameter optimization for ED separating process, but few contributions have been reported to optimize an extractive distillation from the view of simultaneous optimization of the solvent composition and operation parameters for an existing ED process economics aided by co-solvent. In this paper, we propose a systematic optimization approach for energy saving of sulfolane aromatic ED processes based on process simulation and genetic algorithm (GA) optimization. Process simulation and energy analysis were implemented for an existing aromatic ED process aided by the solvent sulfolane. The NRTL property method and rigorous unit models in Aspen Plus are used to simulate the sulfolane ED processes. In particular, considering multi-variable and their interaction, a coordinating strategy is proposed to optimize the solvent composition and significant operating parameters for the ED processes by combining Aspen Plus, EXCEL and MATLAB. Under given separation specifications, the energy consumption and separation efficiency are investigated. And the significant parameters of the process including the temperature, feed location and composition of the lean solution are optimized by GA program.

2. Problem statement

A typical flow sheet for the aromatic ED process is shown in Fig.1, which typically contains an extractive distillation column (EDC), a nonaromatic rectifying column(NRC) and a solvent recovery column (ERC). The EDC uses an entrainer stream (S12, S) to separate the crude feed stream (S1, aromatic and nonaromatic) into a raffinate stream (S3, nonaromatic) and an extract stream (S3, aromatic + solvent, rich solvent). In EDC, the solvent is fed at a location different from the crude feed, resulting in an extra extractive section, as depicted in Fig.1, compare to a conventional column with stripping and rectifying sections. Nonaromatics which have a fainter interaction with the entrainer molecular is recovered from the distillate of EDC and the extract stream is withdrawn from the column bottom. The extract stream is fed into the ERC sequentially to separate aromatics from the rich solvent. The ERC removes aromatics from the top and recycles the entrainer from the bottom. The recycled entrainer is cooled before going back to the EDC. A fresh entrainer compensation (S13, Makeup) is needed to replenish the entrainer lost along with product streams.

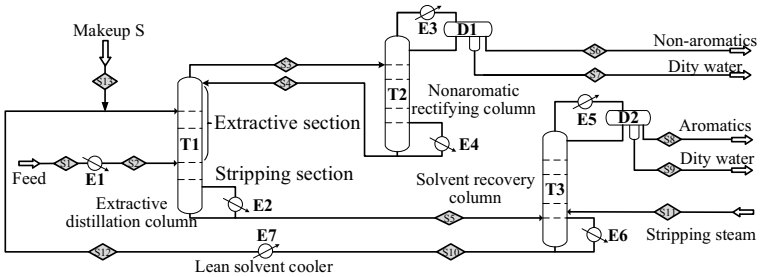


Fig. 1. Flowsheet of the sulfolane extractive distillation process

Solvents used in the ED process is highly important for achieving good process economics and low energy consumption. In other words, optimal process design should also include the optimal design of solvents composition, especially for a co-solvent process. In this work, we propose a systematic optimization approach for energy saving of an existing sulfolane aromatic ED processes based on process simulation and genetic algorithm (GA) optimization. We first conduct a process simulation of an existing sulfolane aromatic ED process in China. Sequentially, a systematic approach for solvent composition and process operation parameters optimization was proposed by taking the minimum hot utility as the objective function based on GA model.

3. Process simulation

The Aspen Plus software (V8.8) is used for the rigorous process simulations carried out in this work. The columns are simulated using the rigorous distillation model RADFRAC. The rigorous and "strongly non-ideal liquid" option is used in the distillation block for columns. The convergence method for tear streams used Wegstein, and the tolerance is set to 1.0×10^{-4}. In particular, the operating pressures of EDC and ERC are 0.07 & -0.05Mpag. The feed location of S12 must be set as 1th when the EDC without condenser. Table 1 lists the real operating data extracted from an existing sulfolane ED plant and the simulation results obtained by Aspen Plus. The relative deviation between the real operating data and simulation results is less than 3%, indicating that the simulation result is reliable.

Table 1. Comparison of plant data and simulation result

	EDC			NRC			ERC		
Item	$T_{top,EDC}$	$T_{bot,EDC}$	D_{EDC}	$T_{top,NRC}$	$T_{bot,NRC}$	D_{NRC}	$T_{top,ERC}$	$T_{bot,ERC}$	D_{ERC}
Unit	°C	°C	kg·h^{-1}	°C	°C	kg·h^{-1}	kg·h^{-1}	°C	kg·h^{-1}
Plant data	114.0	167.0	13120	94.0	140.0	12790	63.0	176.0	42930
Simulation	115.8	172.0	13293	95.9	140.0	12726	61.6	176.0	43027
AAD%	1.58	2.99	1.32	2.02	0	0.5	2.22	0	0.23

4. Process optimization

4.1. Variable and objective function

Process optimization plays a crucial role in process analysis and evaluation. Generally, lower dissolvability and selectivity resulted in a larger circulation of solution which required greater design sizes of process units and higher energy consumption. Co-solvents are good choices to adjust the dissolvability and selectivity of solvents. In this section, the fraction of sulfolane, hydrocarbons and water in the solvent are analyzed to obtain the minimum circulation volume of solution. In order to obtain the minimum energy requirement of the ED process, the energy consumption and separation efficiency are optimized through adjusting the key operating parameters of the ED process under given separation specifications and an optimal design procedure is proposed by GA program. The adjustable operating variables are summarized as the following four aspects: 1) the operating parameters of column feeding stages, such as feeding location of the crude feed (N_{S2}) and rich solution (N_{S5}), 2) the operating parameters of feeding temperatures, such as the temperature of the crude feed (T_{S2}) and lean solution (T_{S12}), 3) the pressure of columns and 4) the reflux ratio of column T3 (R_{T3}). The solvent composition variables include the hydrocarbon concentration ($x_{HC,S12}$) and water ($x_{H2O,S12}$) in lean solution.

4.2. Aspen plus – MATLAB integrated model

The overall framework of this integrated model is illustrated in Fig. 2. The program use Component Object Model (COM) technologies available in the Microsoft Windows operating system to provide interoperability of COM-aware applications as well as a highly effective approach for process simulation and optimization. The steady-state simulation is implemented in Aspen Plus. A calculation procedure of the minimum heating utility ($Q_{h,min}$) is coded in Excel.

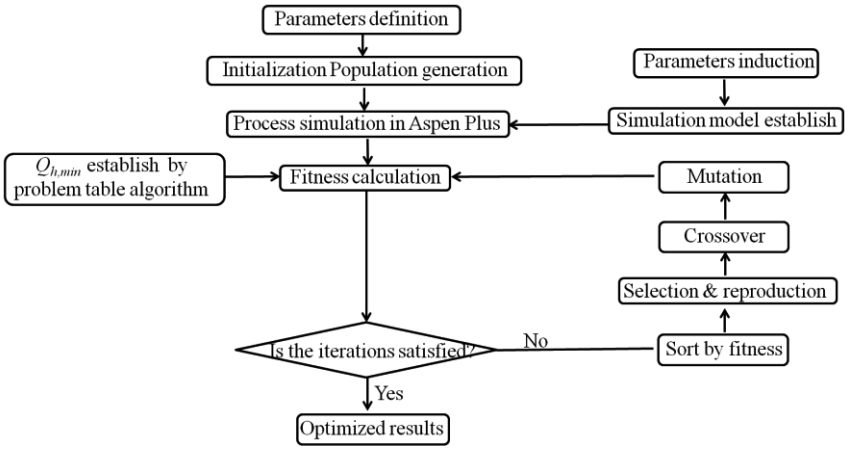


Fig. 2. The overall framework of the integrated program

5. Results and conclusion

A genetic algorithm has been used to optimize both the solvent composition and process operating parameters. As illustrated in Table 2, a set of results was obtained for the ED process when optimized on the basis of minimizing $Q_{h,min}$. Grand composite curves of the existing and optimal process are depicted in Fig.3.

Table 2. Comparison of existing process and optomal process

Plant	EDC			NRC			
Item	$T_{top,EDC}$	$T_{bot,EDC}$	Q_{REDC}	$T_{top,NRC}$	$T_{bot,NRC}$	Q_{RNRC}	Q_{CNRC}
Unit	°C	°C	GJ/h	°C	°C	GJ/h	GJ/h
Orig.	114.8	172.0	49.3	95.9	140.0	10.7	3.0
Opt.	108.8	176.7	29.4	97	145	10.3	3.1

Plant	ERC			
Item	$T_{top,ERC}$	$T_{bot,ERC}$	Q_{RERC}	Q_{CERC}
Unit	kg/h	°C	GJ/h	GJ/h
Orig.	61.6	176.0	60.2	29.81
Opt.	57.7	176	33.2	11.6

The results show that co-solvents have higher adjustability than single solvents. Based on the existing ED process, a compressive simulation model and optimizing strategy of the solvent composition and operating parameters are presented sequentially. Moreover, considering multi-variables and their interaction, a co-ordinative optimization strategy is proposed based on the GA optimization by taking $Q_{h,min}$ as the objective function. Through the above optimization, $Q_{h,min}$ of the ED process is significantly reduced by 29.2%.

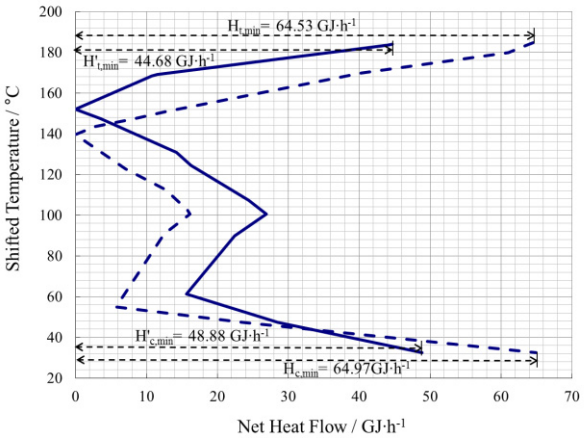


Fig.3. Grand composite curves of the existing and optimal process

Acknowledgement

Financial supports from the National Natural Science Foundation of China (No. U1462113) and the Science and Technology Planning Project of Guangzhou city (No. 201704030136) is gratefully acknowledge.

References

[1] Chauvel A, Lefebvre G. *Petrochemical Processes*. Paris: Editions Technip; 1989.
[2] Lei ZG, Li CY, Chen BH. Extractive distillation: A review. *Sep. Purif. Technol.* 2003;32:121−213.
[3] Wang ZR, Xia SQ, Ma PS. (Liquid + liquid) equilibria for the ternary system of (N-formylmorpholine + ethylbenzene + 2,2,4-trimethylpentane) at temperatures (303.15, 313.15, and 323.15) K. *Fluid Phase Equilib.* 2012;328:25-30.
[4] Vakili-Nezhaad GR, Modarress H, Mansoori GA. Solvent extraction of aromatic components from lube-oil cut by N-methylpyrrolidone (NMP). *Chem. Eng. Technol.* 1999;22:847-53.
[5] Nicolae M, Oprea F, Fendu EM. Dipropylene glycol as a solvent for the extraction of aromatic hydrocarbons. Analysis and evaluation of the solvency properties and simulation of the extraction processes. *Chem. Eng. Res. Des*. 2015;104:287-95.
[6] Chen WK, Lee KJ, Ko JW, Chang CMJ, Daina H, Lee LS. Vapor-liquid equilibria and density measurement for binary mixtures of toluene, benzene, o-xylene, m-xylene, sulfolane and nonane at 333.15K and 353.15K. *Fluid Phase Equilib*. 2010;287:126-33.
[7] Universal Oil Prod Co (UNVO-C). Aromatic hydrocarbons spen-from mixts with non-aromatics by solvent extraction and extractive distillation. U.S. Patent FR2141889-A, August 01, 1977.
[8] And PGT, Brunt VV. Extractive Distillation Solvbetweement Characterization and Shortcut Design Procedure for Methylcyclohexane-Toluene Mixtures. *Ind. Eng. Chem. Res*. 2000; 39(6):1614-23.
[9] Buchbender F, Onink F, Meindersma W, et al. Simulation of aromatics extraction with an ionic liquid in a pilot-plant Kühni extractor based on single-drop experiments. *Chem. Eng. Sci.* 2012; 82:167-176.
[10] Quijada-Maldonado E, Meindersma G W, Haan A B D. Pilot plant study on the extractive distillation of toluene–methylcyclohexane mixtures using NMP and the ionic liquid [hmim][TCB] as solvents. *Sep. & Pur. Tec.* 2016; 166:196-204.
[11] Choi YJ, Cho KW, Cho BW, Yeo YK. Optimization of the Sulfolane Extraction Plant Based on Modeling and Simulation. *Ind. Eng. Chem. Res*. 2002;41:5504-9.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50181-6

Multi-objective Optimization of Membrane-based CO_2 Capture

Miguel A. Zamarripa, John C. Eslick*, Michael S. Matuszewski,
David C. Miller

National Energy Technology Laboratory, 626 Cochran Mills Road, Pittsburgh, PA, USA
John.Eslick@netl.doe.gov

Abstract

This work presents a multi-objective optimization framework that uses a nonlinear programming (NLP) mathematical model for the optimal design and operation of membrane-based post combustion CO_2 capture plants. The proposed approach provides advanced decision-making tools for the power generation industry, involving the simultaneous optimization of alternative process configurations and operating conditions while analyzing different objectives such as cost of electricity and carbon capture target. The proposed approach was demonstrated with a case study from the literature and the mathematical model has been validated against a rigorous simulation model. Results show how the optimal flowsheet changes with different objectives.

Keywords: superstructure optimization, multi-objective optimization, post-combustion capture, membrane systems.

1. Introduction

Post-combustion capture (PCC) is an energy intensive and expensive process that is the subject of considerable research using a variety of technologies, including solvents, sorbents and membranes. Membrane based PCC systems exploit differences in partial pressure across a selective membrane layer to separate CO_2 from other components in the flue gas. The separation is particularly challenging due to the low concentration of CO_2, large volume of flue gas to process, and the high purity requirement of the resulting CO_2 stream. Research efforts include developing new membrane materials as well as new system configurations. Recent work (Budhathoki et al., 2018) reports efforts to simultaneously optimize potential membrane materials and process conditions for a fixed design configuration.

Merkel et al. (2010) demonstrated the importance of the membrane system configuration and integration with the CO_2 generating system by developing a multistage membrane process with multiple recycles, including an air sweep which removes CO_2 from the exit gas and recycles it to the combustor. Ironically, no reported superstructure-based optimization model has enabled the identification of a configuration of that sophistication. Arias et al. (2016) presented superstructure optimization model, which considers multi-stage configurations, including recycles; however, it does not enable recycle streams to be integrated with the CO_2 generating system. Hasan et al. (2012) showed the effect of CO_2 source concentration on overall system cost for simple membrane configurations, and showed how optimal capture technologies are dependent on process requirements. There remains a need for more complex superstructure models to adequately cover the potential optimal configurations that could emerge as new materials are developed,

various CO_2 source concentrations are identified, and different CO_2 capture levels are required.

We present a multi-objective superstructure optimization framework to determine the optimal configuration of membrane-based carbon capture systems under varying CO_2 source concentrations and under various capture levels. The framework is based on rigorous models and incorporates all the innovations proposed by Merkel et al. (2010), including liquefaction and vacuum pumps, multi-stage membranes, a detailed compression train, and a simplified boiler model to capture the effects of recycle streams integrated with the CO_2 source. The multi-objective framework can identify the Pareto front to show the trade-off among objectives such as Cost of Electricity (COE) and capture rate, while also identifying how the optimal process configuration changes.

2. Mathematical model

The proposed framework is formulated as an NLP model that categorizes decisions into two decision types: discrete and continuous. Typically, discrete decisions, such as determining the plant design, are modelled using discrete variables and/or binary variables allowing the model to install or eliminate units (also referred as on-off decisions). However, in this process when selecting the optimal number of stages, a bypass variable (continuous variable) is used instead, avoiding the use of binary variables for this purpose. While formulating a superstructure problem as an NLP is not general, NLP formulations have been used for separation problems, e.g., membrane systems (Hasan et al., 2012) and distillation columns (Dowling, 2015), where a partial bypass would lead to purifying a stream later mixed with an impure stream. Other continuous decisions are flows, temperatures, concentrations, pressure ratios, and sweep feed.

Figure 1 shows a schematic representation of the superstructure framework. We propose a heterogeneous superstructure framework, in which part of the flowsheet is fixed (M1, M2, M3, liquefier column, and compressors with intercooling as these always exist), while the number of membrane stages in series that can be installed is not fixed.

The main objective is to determine the Pareto front showing the optimal tradeoff between COE and CO_2 capture rate. COE is estimated as follows [Eq. (1)]:

$$COE = \frac{(TOC)(CCF) + OC_{fix} + (CF)(OC_{var})}{(CF)(MWH)} \tag{1}$$

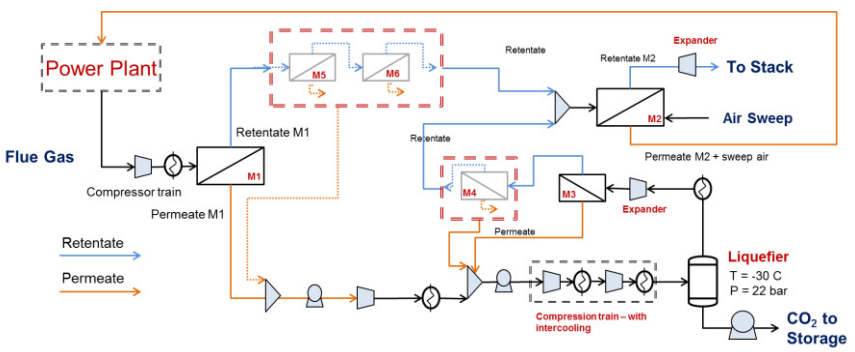


Figure 1. Superstructure optimization framework

where, TOC is the total overnight capital expenses, OC are the operating costs (fixed and variable), net power output is calculated based on the power produced by the plant minus the power used by the PCC process. CF is the plant capacity factor and CCF is the capital charge factor (NETL/DOE, 2011).

2.1. Membrane model

Each membrane stage has been modeled as a counter-current system in which the inlet flow to the membrane is equal to the outlet flow plus/minus the flux change (J_i is the flux moving from the retentate side of the membrane to the permeate side of the membrane). Eq.(2) and Eq.(3) show the material balances for the retentate (component molar flowrate - FR_i) and permeate (component molar flowrate - FP_i) sides of the membrane, respectively.

$$0 = -\frac{\partial FR_i}{\partial x} - J_i \tag{2}$$

$$0 = -\frac{\partial FP_i}{\partial x} - J_i \tag{3}$$

A finite difference method has been applied to Eq. (2) and Eq. (3) (forward and backward, respectively). After the discretization, $J_{i,n}$ represents the flux moving from one side to the other one for component i, at sub-compartment n (a fixed length is discretized over a finite number of elements), and is given by the following equation [Eq. (4)]:

$$J_{i,n} = 2\pi r_{Fo} n_F \frac{Q_{CO_2}}{\alpha_i}\left(PR_n xr_{i,n} - PP_n xp_{i,n}\right) \tag{4}$$

where, Q_{CO_2} is the permeance of the membrane (considered fixed for this study), α_i is the membrane selectivity of component i (relative to CO_2), n_F is the number of fibers/tubes in the membrane, r_{FO} is the external radius of the tube, PR_n is the retentate pressure and PP_n is the permeate pressure in the n^{th} compartment, and $xr_{i,n}$ and $xr_{i,n}$ are the component molar fractions at the retentate and permeate sides, respectively.

The pressure drop at the retentate side behaves linearly, thus, a linear regression has been considered. While, the pressure drop for the permeate side was modeled based on the Hagen-Poiseuille relation along with ideal gas assumptions (dynamic viscosity, ideal gas law, etc.). (Morinelly and Miller, 2011)

2.2. Condenser

The condenser has been modeled as a flash tank; using a mix of surrogate models and material balances, we developed a square formulation of the column. The inlet gas (F, P, T, z_i) and flash temperature are the optimization variables of the problem. A set of surrogate models have been developed for the vapor outlet (F_V, y_i, P, T) and the heat duty of the column, while the material balances estimate the liquid outlet (F_L, x_i).

The Framework for Uncertainty Quantification, Optimization and Surrogates (FOQUS, Miller et al., 2016) was used: (i) to prepare the data set by interfacing with the Aspen process simulation, (ii) to generate the surrogate model using ALAMO (Cozad et al., 2015), and (iii) to validate the surrogate model. The predictions from the surrogate model

showed a good fit when compared to the rigorous flash models from Aspen Plus. All the surrogate models presented have an R^2 value between 97% and 99%.

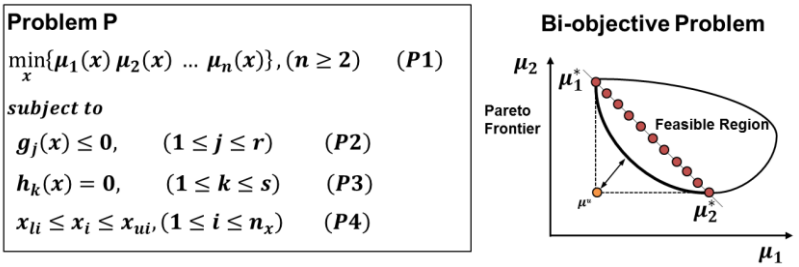


Figure 2. Multi-objective optimization approach.

2.3. Simplified power plant representation

Typically, a simple model for a thermal power plant consists of a boiler, bag house, desulfurization unit and steam cycle. For this work, a representation of the power plant has been developed to account for the potential benefit of recirculating the "sweep" air used in the membrane system, consequently a simplified boiler model was used that considers a stoichiometric balance for the O_2 in the system, while also accounting for recycled CO_2. This model includes a preheater and a desulfurization unit.

2.4. Multi-Objective Optimization (MOO) approach

This work implements the epsilon-constraint method for decision making under multiple objectives (Messac et al., 2003). This method solves a set of single objective optimization problems to determine the Pareto front, from which a solution can be selected based on the Utopia-tracking approach. The Epsilon constraint algorithm for a bi-objective problem consists of the following steps (see Figure 2):

1. Minimize μ_1 (obtaining μ_1*)
2. Minimize μ_2 (obtaining μ_2*)
3. Draw a line between μ_1* and μ_2*
4. Divide the line into m_k points
5. For all m_k
   a. Constraint $\mu_2 \leq \mu_2$ (at m_k)
   b. Minimize μ_1
6. Draw Pareto frontier
7. Select the solution closer to the Utopia point

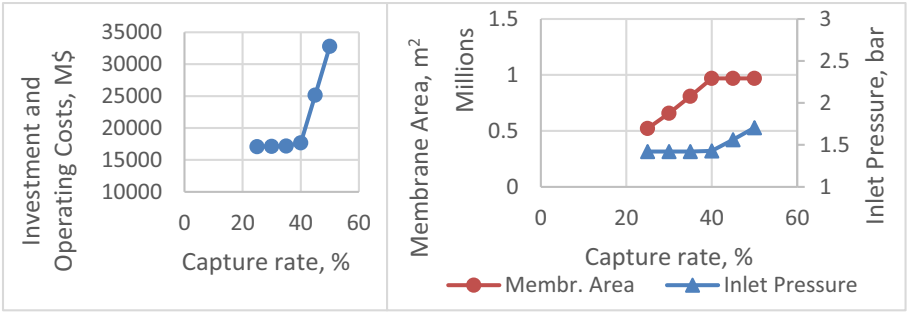


Figure 3. Single unit MOO analysis (M1 – membrane 1), costing method - Hasan et al. (2012).

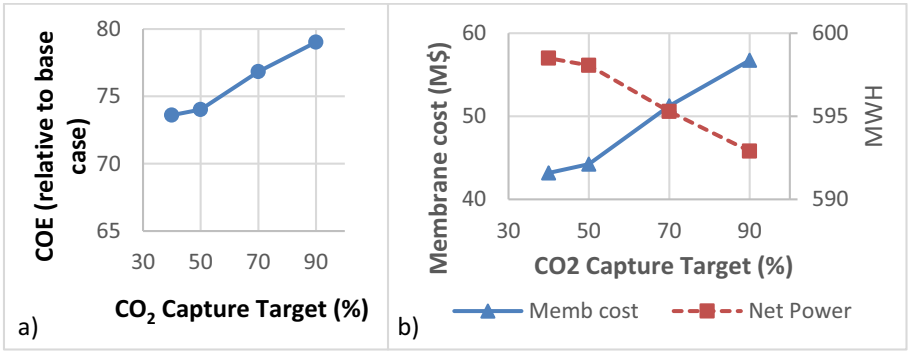


Figure 4. MOO results

3. Results

The membrane system model was tested using a case study from the Carbon Capture Simulation Initiative (CCSI) toolset (Morinelly and Miller, 2011). A 650MW power plant has been used as a basis to analyze the proposed framework. The flue gas from the coal plant is preconditioned to 50°C and 1.013 bar and with a molar fraction of CO_2 12 % (79% N_2, 1.8% H_2O, 5.1% O_2, balance Ar). As discussed above, membrane stage 1, 2, and 3 are installed, while membranes 4 to 6 can be installed or not. The membrane properties are: permeance, Q_{CO2} = 0.12047 kgmol/m2·bar·hr; membrane selectivity, $\alpha_i(CO_2/N_2, CO_2/O_2, CO_2/Ar)=100$; and $\alpha_i(CO_2/H_2O) = 0.5$. The superstructure model has been implemented in GAMS and solved using IPOPT and CONOPT. (IPOPT was used to initialize the problem, and CONOPT was used to solve the optimization problem.) The predictions from the proposed model were compared to the CCSI Aspen model, and the results show similar performance with a 3% of error in the COE and less than 1% of error for the retentate flowrate, permeate flowrate, and permeate pressure.

Regarding the multi-objective analysis, this work focuses on the trade-off between the COE and the CO_2 capture rate (CO_2 out/CO_2 in). At a single unit level, the trade-off is very clear [see Eq. (4)]. A larger membrane area and larger pressure difference are required for a higher capture; however, this also increases the final costs (investment and operating). Figure 3 shows the multi-objective analysis of membrane 1 (M1), in which the trade-off between cost and capture rate for a fixed flue gas is observed. For a higher capture rate (between 40% and 50%), the model increases the inlet pressure instead of the membrane area. At a lower CO_2 capture rate, the model reduces the membrane area maintaining the inlet pressure fixed (25% to 40% capture rate).

Figure 4 shows a few candidate solutions (40%, 50%, 70%, and 90% capture target) from which it can be observed that there is a clear trade-off between these two objectives. A change in the trend is evident from 40% to 70% capture versus 70% to 90% capture. The model predicts the installation of four membrane stages (M1-M3 and M5 see Figure 1) to achieve 70% and 90% capture target, whilst only three stages are installed for 40% and 50% capture target (M1-M3 see Figure 1). In all cases, the optimal membrane area (Memb. Cost = Area ·\$50/m^2) and operating conditions are different.

4. Conclusions

The proposed model has been compared against a rigorous process simulation model predicting a COE with less than 3% error. This work has demonstrated that the optimal process configuration changes at different capture rates, and that advanced PSE tools can help to identify the optimal configurations that result in minimal COE as CO_2 capture rate changes. Additionally, conceptual design and PSE tools should be used to complement the decision-making process, specifically for post-combustion capture. These tools allow the analysis of potential process configuration and the model could be used to analyze alternative membrane technologies/arrangements.

Acknowledgments and Disclaimer

This research was supported in part by an appointment to the National Energy Technology Laboratory Research Participation Program, sponsored by the U.S. Department of Energy and administered by the Oak Ridge Institute for Science and Education. This project was conducted as a part of the Carbon Capture Simulation for Industry Impact ($CCSI^2$) program. This paper was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

References

A.M. Arias, M.C. Mussati, P.L. Mores, N.J. Scenna, 2016, Optimization of multi-stage membrane systems for CO_2 capture from flue gas, International Journal of Greenhouse Gas Control, 53, 371-390.

S. Budhathoki, O. Ajayi, J.A. Steckel, C.E. Wilmer, 2018, High-throughput computational prediction of the cost of carbon capture using mixed matrix membranes, submitted to Energy and Environmental Science

A. Cozad, N.V. Sahinidis, D.C. Miller, 2015, A combined first-principles and data-driven approach to model building, Computers and Chemical Engineering, 73, 116-127.

A.W. Dowling, 2015, An Equation-based Framework for Large-Scale Flowsheet Optimization and Applications for Oxycombustion Power System Design, Dissertations, Paper 568.

M.M. Hasan, R.C. Baliban, J.A. Elia, C.A. Floudas, 2012, Modeling, Simulation, and Optimization of Postcombustion CO_2 Capture for Variable Feed Concentration and Flow Rate. 1. Chemical Absorption and Membrane Processes, Industrial and Engineering Chemistry Research, 51, 15642-15664.

T.C. Merkel, H. Lin, X. Wei, R. Baker, 2010, Power plant post-combustion carbon dioxide capture: An opportunity for membranes. Journal of Membrane Science, 359, 126-139.

A. Messac, A. Ismail-Yahaya, C.A. Mattson, 2003, The normalized normal constraint method for generating the Pareto frontier, Structural and Multidisciplinary Optimization, 25, 86-98.

D.C. Miller, D. Agarwal, D. Bhattacharyya, J. Boverhof, Y.W. Cheah, Y. Chen, J. Eslick, J. Leek, J. Ma, P. Mahapatra, B. Ng, N.V. Sahinidis, C. Tong, S.E. Zitney, 2016, Innovative computational tools and models for the design, optimization and control of carbon capture processes, Computer Aided Chemical Engineering, 38, 2391-2396.

J.E. Morinelly, D.C. Miller, 2011, Gas Permeation Carbon Capture – Process Modeling and Optimization, International Pittsburgh Coal Conference Proceedings.

NETL/DOE, 2011, Quality Guidelines for Energy System Studies: Cost Estimation Methodology for NETL Assessments of Power Plant Performance.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50182-8

Simultaneous optimization of power consumption and heat exchange network in refinery hydrogenation system

Yu Zhang[a], Jing Li[a] and Qiao Zhang[a*]

[a]*School of Chemical Engineering and Technology, Xi'an Jiaotong University, Xi'an 710049, China*
q.zhang1986@stu.xjtu.edu.cn

Abstract

As energy consumption in chemical industry is intensive, energy conservation in is of great significance. In refinery hydrogenation operations, hydrogen streams are not only pressurized by compressors which consuming power to increase their pressures, but also cooled down or heated up by other streams, heating or cooling utility before or after compression to adjust their temperatures for hydrogenation reactions. For hydrogen compressor, in addition to inlet pressure, the inlet temperature of hydrogen stream will also affect its power consumption. Meanwhile, for a hydrogen stream, the outlet temperature of compressor also serves as the inlet temperature of heat exchanger and then will simultaneously affect the utility consumption of heat exchange network. Therefore, for each compressor there are optimal inlet and outlet temperatures that can minimize the total energy including compression work, heating and cooling utilities. This paper establishes a nonlinear programming (NLP) model to integrate the hydrogen streams entering and outflowing from the compressors and heat exchangers, optimizing the inlet and outlet temperatures of compressors for simultaneous optimization of compression work and utilities of heat exchange network. Case study shows that this method can further achieve the energy conservation than individual integration of HEN.

Keywords: Heat exchange network; Compression work; Mathematical Programming; simultaneous optimization

1. Introduction

The hydrogenation operation of refinery is of great importance and its products are used in chemical industries. In order to produce qualified products, hydrogenation processes have strict requirements for hydrogen stream properties, such as pressure and temperature, consuming intensive energy to pressurize and heat up or cool down.

As for compression work conservation in refinery hydrogenation processes, Hallale et al. (2001) established a superstructure and set up a mathematical optimization model accounting pressure constraints. Yu et al. (2011) carried out the exergy standard to unify and optimize the consumption of hydrogen and compression work. Wu et al. (2012) developed a mathematical model of hydrogen network to minimize the total exergy consumption including fresh hydrogen and power consumption.

The research for heat exchanger network (HEN) is progressive since it was found. Linnhoff et al. (1978) proposed a thermodynamically oriented pinch analysis for the synthesis of HEN and solved the problem in two stages. First, identify the maximum

heat recovery network and then set the preliminary network as starting point and design the allocation network. Cerda et al. (1983) formulated the minimum utility calculation method for HEN synthesis using a "transportation problem" model by linear programming(LP). Onishi et al. (2015) optimized a liquefied natural gas production process considering the pressure recovery of streams during the HEN design and realized the minimization of the retrofit total annualized cost. Then, an effluent recovery superstructure contains HEN synthesis operator, pressure operator and separation operator was introduced to maximize the total annualized revenue by Liao et al. (2017).

For refinery hydrogenation operation, hydrogen streams are generally flow through heat exchangers and compressors in certain sequence, so the inter-unit temperature of them will influence on both the utility consumption of HEN and compression work demand. Thus, there is energy conservation potential for the two kinds of energy are considered simultaneously when such temperature is optimized. However, those methods stated above are not able to deal with such case. This paper establishes a nonlinear programming(NLP) model to minimize both utilities and power consumption of refinery hydrogenation systems. Case study results show its advantages and application.

2. Theory

The HEN of the refinery hydrogenation system is different from common ones as heat transfer and compression of hydrogen streams are series-wound. Before entering the reactor, hydrogen streams need to be compressed under low temperature and then heated up by heat exchanger. After reaction, they should be cooled down and then recycled to compressors again. The compressor inlet temperature of a hydrogen stream has upper limit and meanwhile the outlet temperature affects the HEN utilities consumption, so for hydrogen stream from each compressor there are optimal inlet and outlet temperatures that can minimize the total consumption of compression work and utilities. Aiming to such goal, a mathematical model is established below.

2.1 Superstructure

Figure 1 shows the superstructure of this method. *HR* and *HS* represent heat source and sink, and the solid lines connecting all sources and sinks represent that heat source and heat sink can transfer heat among all others. Spacer lines indicate the hot streams flowing into compressors after cooling while dotted lines denote cold streams flowing out compressors and then need to be heated up for hydrogenation units. F_{im} and F_{jm} are flow rates of hot and cold streams, respectively, T_{im}^{in} is the inlet temperature of hot stream that cooled down to T_{im}^{out} and then pressurized by compressor. After compression, the stream is cold stream with temperature T_{jm}^{in} and then heated up to T_{jm}^{out} by heat exchanger.

2.2 The mathematical model

Based on the above superstructure, a NLP model aiming at minimizing consumption of compression work and utilities of HEN is established. Meanwhile, the mass balance, heat balance, and temperature requirements for compressors and hydrogenation units are set as the constraints for this model.

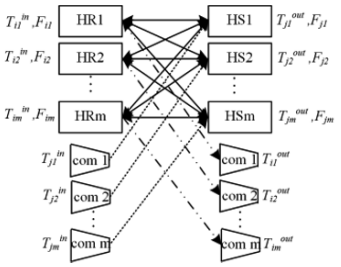


Figure 1 The superstructure of the simultaneous optimization network

$$\begin{aligned} \min \quad & gw = W_{i,j} + Q_c + Q_h \\ \text{s.t.} \quad & m(x) = 0 \\ & h(x) = 0 \\ & t(x) \le const \\ & W = f(t, p, F) \end{aligned}$$

In the NLP model, the objective function *gw* is the sum of compression work, heating and cooling utilities. Restraint *m(x)* is the mass balance of hydrogen streams flowing through the compressors and heat exchangers. Similarly, heat balance for all hydrogen streams is restricted as *h(x)*. Besides, all temperature constraints are simplified as *t(x),* where *const* is a constant, generally taken as 40℃ for inlet temperature of compressor. For the heat exchange process, the minimum heat transfer temperature difference is taken as 10℃ in this model. The compression work W can be calculated using Eq.(1). The cooling and heating utility can be calculated using Eq.(2) and Eq.(3), respectively.

$$W_{i,j} = \frac{C_{pi} \cdot T_i^{out}}{\eta} \cdot \left[\left(\frac{p_j}{p_i} \right)^{\frac{(\gamma-1)}{\gamma}} - 1 \right] \cdot F_{i,j} \tag{1}$$

W is the compressor work needed in the process from stream i to unit j, which is decided by the flow rate $F_{i,j}$, specific heat capacity at constant pressure C_{pi} and the inlet temperature of the compressor T_i^{out} and its efficiency, η . The γ is the ratio of specific heat capacity of gas at constant pressure to that at constant volume, and for hydrogen stream it is generally taken as 1.4. The T_i^{out} is related to compressor and HEN and thus it is an important variable to be optimized in this model.

$$Q_C = \sum_{i=1}^{m} \left(T_i^{out1} - T_i^{out} \right) \cdot Cp_i \tag{2}$$

$$Q_h = \sum_{j=1}^{k} \left(T_j^{out} - T_j^{out1} \right) \cdot Cp_j \tag{3}$$

For a hydrogenation system with m heat exchangers and n compressors. Q_c and Q_h are cooling and heating utilities that hydrogen streams consumed for hydrogenation unit requirements after they heat exchange with process streams. T_j^{out1} and T_i^{out1} are that of cold and hot streams after the heat exchange, and similarly, T_i^{out} and T_j^{out} are outlet temperatures of hot and cold streams after using utilities, respectively.

3. Case study

A practical refinery case in China is employed to demonstrate the application of this method. The schematic diagram is shown in Figure 2 and Table 1 shows the basic data of hydrogen streams in hydrogenation system. There are pressures, flow rates, inlet and outlet temperatures of hydrogen streams, and these are target temperatures of heat exchangers and compressors. Generally, the inlet and outlet temperatures for compressor are taken as 40℃ under the hypothesis that the compressor is in isothermal compression, which is illogical and unrealistic. The case will be optimized by both HEN integration and the proposed model in this paper to demonstrate its advantages and application. The NLP model is coded in GAMS 22.3 and solved by Baron solver. the computational studies are performed on a Dell Vostro 260s with Intel(R) Pentium(R) CPU G630 @ 2.70GHz and 4 GB RAM, using windows 7, 64 bit operating system.

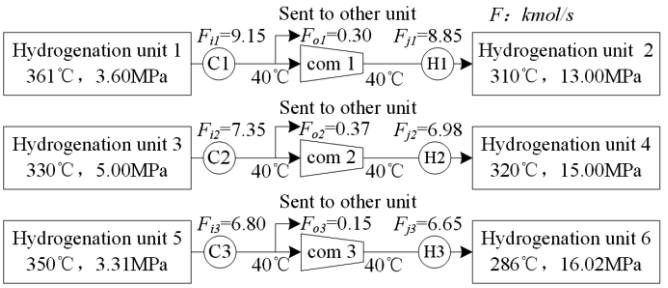


Figure 2 The schematic diagram of case

Table 1 The basic data of the case

Streams	Pressure(MPa)	T^{in}(℃)	T^{out}(℃)	F(kmol/s)
i1	3.60	361	40	9.15
i2	5.00	330	40	7.35
i3	3.31	350	40	6.80
j1	13.00	40	310	8.85
j2	15.00	40	320	6.98
j3	16.02	40	286	6.65

3.1 The optimized network with HEN

Table 2 shows the resulting data after the HEN optimization, where F is flow rate of each stream. For each hot/cold stream, T^{in} represents the inlet temperature of heat exchanger. After interception and heat exchange, all divided streams are mixed back to their original stream, whose temperature is T^{x}. This temperature of i1, i2, i3 are 84.0℃ ,150.8℃,88.7℃ in Figure 3. For hot streams, T^{x} is cooled down to T^{out} using cooling utility while heating utility is used for cold streams. T^{out} is the outlet temperature of heat exchangers and meanwhile the inlet temperature for compressor, such as the T^{out} of stream *i1* is the optimizing variable T_{i1}^{out} in the model. U is the utility consumption and W means the compressor work needed for the whole system. Optimized system with only heat integration is shown in Figure 3, and it can be seen that after HEN optimization, there are 6 heat exchangers, 3 coolers and 1 heaters. The demand for cooling utility is 14534.01kW, 301.76kW for heating utility and 4990.64kW for compression work, respectively.

Table 2 The results after optimization of HEN

Streams	F(kmol/s)	T^{in}(℃)	T^{x}(℃)	T^{out}(℃)	U(kW)	W(kW)
i1	9.15	361.0	84.0	40.0	3779.61	1908.05
i2	7.35	330.0	150.8	40.0	7645.84	1865.14
i3	6.80	350.0	88.7	40.0	3108.56	1167.14
j1	8.85	40.0	310.0	310.0	0.00	0.00
j2	6.98	40.0	315.4	320.0	301.76	0.00
j3	6.65	87.9	286.0	286.0	0.00	0.00
Total:Heating utility: 301.76kW, Cooling utility:14534.01kW, Compression work: 4990.64 kW						

3.2 Simultaneous optimization for compression work and utility of HEN

Different from the results obtained by HEN optimization, the simultaneous optimization achieved the minimization for compression work and heating and cooling utilities simultaneously, and an further optimized network is shown in Figure 4. Table 3 shows the data after simultaneous optimization.

Table 3 The results after simultaneous optimization of heat and compression

Streams	F(kmol/s)	T^{in}(℃)	T^{x}(℃)	T^{out}(℃)	U(kW)	W(kW)
i1	9.15	361.0	74.3	28.0	3976.19	1332.48
i2	7.35	330.0	140.9	31.0	7580.57	985.94
i3	6.80	350.0	76.9	35.0	2675.20	1588.06
j1	8.85	28.0	308.7	310.0	108.24	0.00
j2	6.98	31.0	317.0	320.0	196.59	0.00
j3	6.65	76.9	286.0	286.0	0.00	0.00
Total: Heating utility: 304.83kW, Cooling utility: 14231.96kW, Compression work: 3906.48 kW						
Saved: Utility: 298.98kW, Compression work: 1084.16 kW						

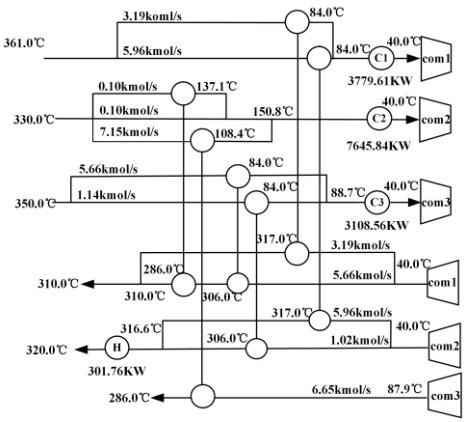


Figure 3 The HEN optimized network

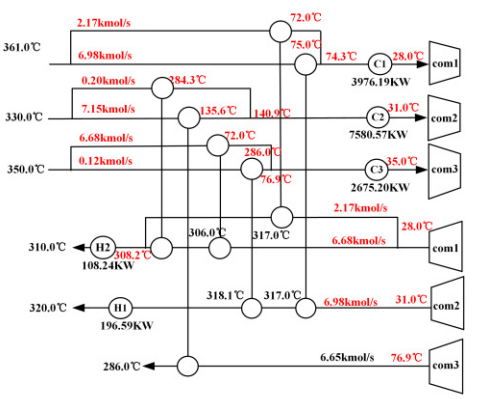


Figure 4 The simultaneous optimized network

The optimal inlet temperature of the compressor for hot streams are 28.0℃, 31.0℃ and 35.0℃, respectively. The minimum cooling utility is 14231.96kW, the heating utility is 304.83kW, and the compression work is 3906.48kW. In summary, the total energy consumption, *gw,* which is the sum of *U* and *W,* was saved about 6.97% of its original consumption, about 1383.14kW.

Compared the two networks, simultaneous optimized network has 6 heat exchangers, 3 coolers and 2 heaters, so only one heat exchanger is added for optimization of T_i^{out}, and T_j^{in}, simultaneously for compressors. In Figure 4, the red numbers are difference to that in Figure 3. This model optimizes the inlet temperatures for compressor, which is also the outlet temperature for hot streams, so that the compression work and HEN utilities can be reduced after optimization to further improve the energy utilization performance. Therefore, this method is contributory on both theory and application aspects.

4. Conclusions

Focused on the holistic energy conservation of refinery hydrogenation system, a method for simultaneous optimization of compression and HEN to minimize total energy consumption is proposed, and then a NLP model is formulated and demonstrated by a practical refinery case. Results show that the simultaneous optimization method can further conserve more energy than that obtained with only heat integration.

Acknowledgement

Financial support from the National Natural Science Foundation of China under Grant 2 1736008 is gratefully acknowledged.

References

Cerda J., Westerberg A W., Mason D., et al, 1983, Minimum utility usage in heat exchanger network synthesis A transportation problem [J], Chemical Engineering Science, 38(3):373-387.

Ding Y.,Feng X.,Chu K H., 2011, Optimization of hydrogen distribution systems with pressure constraints, Journal of Cleaner Production, 19, 204−211.

Kumar A., 2010, Hydrogen Distribution in the refinery using mathematical modeling, Energy, 35, 3763 – 3772.

Hallale N., Liu F., 2001, Refinery hydrogen management for clean fuels production, Advances in Environmental Research, 6(1), 81 – 98.

Linnhoff B., Flower J R., 1978, Synthesis of heat exchanger networks: I. Systematic generation of energy optimal networks[J], Aiche Journal, 24(4):633-642.

Liao Z., Jiang B., Tu G., et al., 2017, Optimal process design for recovering effluent gas at subambient temperature[J], Journal of Cleaner Production, 144.

Onishi V C., Ravagnani M A S S., Caballero J A., 2015, Retrofit of heat exchanger networks with pressure recovery of process streams at sub-ambient conditions[J], Energy Conversion & Management, 94(C):377-393.

Wu S, Liu G, Yu Z, et al., 2012, Optimization of hydrogen networks with constraints on hydrogen concentration and pure hydrogen load considered[J], Chemical Engineering Research & Design, 90(9):1208-1220.

Zemiao Y U., Xiao F., 2011, Hydrogen distribution networks targeting minimum exergy consumption[J], Chinese Journal of Chemical Engineering, 62(7):1951-1956.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50183-X

Procurement Planning of Refinery Integrated with Production System by Quantifying the Effect of the Gap Between Supply and Demand on the Production System's Performance

Yechan Choi[a], Joohyun Shin[a], Kyungseok Noh[a,b], Jay H. Lee[a*]

[a] *Chemical and Biomolecular Engineering Department, Korea Advaced Institute of Science and Technology, Daejeon, Korea*
[b] *Global Technology, SK Innovation CO., Ltd., 325 Expo-ro, Yuseong-gu, Daejeon 305-712, Korea*
jayhlee@kaist.ac.kr

Abstract

This study addresses the problem of procurement planning and its integration with production planning of a refinery. Though a complete integration of the optimization models of crude procurement planning and refining process operation is desirable, it is currently computationally problematic due to the very large size of the resulting optimization problem. An alternative to the fully integrated optimization is to determine the optimal crude demand by solving the refinery process optimization problem first, and then to solve the procurement planning problem to minimize the deviation of actual crude supply from the optimized demand. However, such a separation can result in a significant sub-optimality or even an infeasible refinery operation since deviations of same size between supply and demand can bring very different effects on the refining margin and operation depending on the type of deviation and the market situation. Such drawbacks can be alleviated by quantifying the effects of deviations and reflecting them on the weight parameters in the optimization problem of procurement planning. The developed approach is tested on a case study involving an actual industrial refinery. It is shown that the quantified weight factors help reduce feasibility problems and loss from the sub-optimality, compared to the conventional approach without such weight factors.

Keywords: Procurement planning, Oil refining industry, Mixed integer non-linear programming, Process Integration

1. Introduction

A petroleum refinery plant comprises diverse sub-systems with various value chains and time scales, and there exists a significant difference in time scale between the planning stage (usually in months) and the scheduling stage (usually in days). Recently, as the environmental regulation and product specification become tighter and crude price and product demand fluctuate, great economic potentials are there for the refining industry if the gaps between different decision-making stages can be reduced by integrating them more tightly (Grossman, 2012).

Although the importance of integrating planning and scheduling has recently been brought up to attention, few researches have addressed the integration in the context of oil refineries (Erdirik-Dogan et al., 2008). The major difficulty in integrating different

decision-making stages having different time scales is a large size of a completely integrated optimization problem, which must be formulated in the fine timescale not to lose the relevant details. Such problems are computationally challenging despite all the recent advances in computing hardware and numerical methods. To handle the complexity of the problem, Kelly et al. (2017) proposed a solution strategy for a mixed-integer nonlinear programming (MINLP) problem in a petroleum refinery by decomposing it into an MILP and a NLP and reducing the gap between the two problems. Menezes et al. (2015) developed a solution algorithm called phenomenological decomposition heuristic (PDH), which divides the original MINLP problem into 2 sub-models and solve them in an iterative way until the two models converge.

In this research, an alternative to the complete integration approach is proposed, where the integration is achieved by minimizing the gap between the demand and supply profiles of various crude types, which are obtained from production planning and procurement scheduling, respectively. That is, an optimal crude selection profile is first calculated using a refinery optimization model, and then, procurement scheduling is done by a MINLP model aimed at supplying crude oils to the refinery with as small a deviation as possible from the optimal demand for each crude type. However, such a simple approach can result in sub-optimal, or even infeasible operation in the refinery system due to the gap. Note that a gap between supply and demand for different crude types at different times can have different effects on the refining margin and operation, depending on the physical and chemical properties of various crude oils and the market situation.

Motivated by this, this paper explores the possibility of reducing the above-mentioned sub-optimality and infeasibility in the two-step optimization approach by assigning an weight parameter to the difference between demand and supply for each individual crude type. The value of weight parameter is chosen based on the marginal value of deviation for each crude type calculated by giving a perturbation to the optimal crude profile.

The content of this paper is organized as follows. In Section 2, the studied refinery system is described and the optimization problem is cast as an MINLP problem. A method of estimating the weight parameter for a deviation between optimal demand and actual input is introduced in Section 3. In Section 4, a case study involving a refinery operation for a 6-months period (From May 2015 to October 2015) is performed with real price data of crude oil and product and the results are presented. Finally, Section 5 concludes the paper.

2. Problem Description & MINLP Model Design

The graphical diagram of the studied refinery procurement and production system is presented in Figure 1. The purpose of the procurement planning is to supply crude oils to the refining system by meeting the crude demand profile requested by the refining process. In practice, it is not always possible to do this exactly and some deviation of supply from desired demand is unavoidable. Optimization can help to reconcile such differences. In the optimization of procurement planning, crude oil purchase and inventory management decisions are determined as well as supplying crude profile into the processing unit. On the other hand, the refinery optimization calculates the optimal profile of crude oil to maximize the gross margin of the refining system. In such a context, the procurement system would use the optimal crude profile obtained from the refining process optimization as input parameters, whereas the refining process optimization would need the optimal input crude profile calculated by the procurement model as input parameters to achieve the highest possible final gross operational margin.

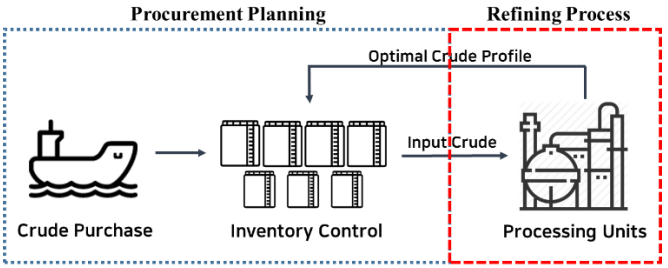


Figure 1. Refinery procurement and production system

In this study, the refinery optimization model developed by Noh et al. (2017) is used for determining one-period (5 days) optimal crude demand. In the multi-period optimization problem of procurement planning, the required parameters such as the number of tanks and their capacity, and purchase cycle and delivered time of crude are chosen to reflect the practical refinery situation. The general procedure for the procurement is as follows: 1) Order decision of which crude oil will be purchased is made every 5 days; delivered period is one month from ordering. 2) Arrived crude is unloaded to the port and distributed to the inventory tanks. 3) The crude from each inventory tank is gathered into buffer tanks. 4) The crude is finally put into the processing unit (the crude distillation unit). The specific amounts of crude flows from the purchase to the processing unit are determined by optimization.

The corresponding optimization problem takes a form of MINLP due to the nonlinear constraints related to the crude composition in the inventory tank through various types of crude oil mixing and the binary variables for the crude purchase decisions.

Positive or negative deviation of crude supply profile from the optimal demand (calculated from a separate optimization of refinery operation without considering constraints from the supply side) is expressed as in Eq.(1), where $P_{o,i,t}$ is the proportion of crude oil *o* in tank *i* at time *t*, and $FO_{i,t}$ is the amount flow put into the CDU from tank *i* at time *t*. Thus, $\sum_{i\in I}\left(P_{o,i,t}\cdot FO_{i,t}\right)$ represents the total amount of crude *o* going to CDU at time *t*. And $D_{o,t}$ is the optimal crude demand of crude *o* at time *t*, which is determined from the production optimization. $DVP_{o,t}$ represents the absolute value of positive deviation of crude *o* at time *t*, whereas $DVN_{o,t}$ represents the absolute value of negative deviation of crude *o* at time *t*.

$$\sum_i\left(P_{o,i,t}\cdot FO_{i,t}\right)-\ D_{o,t}=\ DVP_{o,t}-DVN_{o,t}\ \ for\ \forall o,t \tag{1}$$

When the amount of oil *o* going to CDU at time *t* exceeds the optimal demand, a positive deviation occurs and the value of $DVP_{o,t}$ equals to the value of $\sum_{i\in I}\left(P_{o,i,t}\cdot FO_{i,t}\right)-\ D_{o,t}$ while $DVN_{o,t}$ is zero. On the other hand, when the amount of oil *o* going to CDU at time *t* is below the optimal demand, a negative deviation occurs and the value of $DVN_{o,t}$ equals to the value of $D_{o,t}-\sum_{i\in I}\left(P_{o,i,t}\cdot FO_{i,t}\right)$ while $DVP_{o,t}$ is zero. This relation between positive and negative values of the deviations are incorporated into the objective function of the procurement model, which is to minimize the sum of positive or negative deviations of the crude supplies from the optimal demands requested by the production system, as shown in Eq.(2).

$$obj=\sum_o\sum_t(DVP_{o,t}+DVN_{o,t}) \tag{2}$$

Further detailed formulations are omitted here due to the length constraint, but one can refer to the model used in Oddsdottir et al. (2013).

Despite the reduction in the problem size achieved through the hierarchical separation of the two systems, the resulting MINLP is still very difficult to solve exactly within reasonable time. 'PROcurement-NO-DIScrepancies (pronodis)' approach (Oddsdottir et al., 2013) is used as a relaxation method of MINLP, where the MINLP is decomposed into MILP and NLP to avoid solving the MINLP directly.

3. Estimation of weight parameter for the deviance between the crude demand and supply for individual crude type

Separate implementations of the production planning and the procurement planning (Chapter 2) result in sub-optimal, or even infeasible operation of the refining system. In this study, such sub-optimality and infeasibility are reduced by estimating an appropriate weight parameter for deviation between the crude oil demand and supply for each individual crude type. To quantify the effect of supply deviation from optimal demand, the marginal value is computed for each crude type by giving a perturbation to its optimized demand value. Then the calculated values are set as the weight parameters of the deviations in the objective function of Eq. (2).

The procedure for the weight parameter estimation method is as follows: 1) A set of tested crude profile scenarios is generated by giving perturbations to the optimal crude profile for each crude type. Since too large a size of perturbation results in an infeasible operation in the production planning, the size of perturbation is chosen to be as large as possible, ~ 0.02 % of the maximum crude availability. Here, the sum of the total amount of crude supplied to the processing unit is kept constant in each scenario. 2) Among the scenarios generated from 1), physically unrealizable cases are eliminated where the amount of supplied crude is negative or exceeds the allowed availability. 3) For all feasible scenarios, loss in the operating margin is measured by solving the refinery optimization problem. 4) For each negative or positive deviation in the supply of a crude type, scenarios that contain the matching deviations are picked. Summing up the losses in the operational margin for the selected scenarios and dividing it by the number of scenarios picked, an average effect of each deviation on the gross margin is quantified. The average marginal values are then directly used as a $WP_{o,t}$ and $WN_{o,t}$. Then, estimated values of positive weight parameters $WP_{o,t}$ and positive weight parameters $WN_{o,t}$ are multiplied to each deviation in the objective function of procurement planning as in Eq.(3)

$$obj\ w/weight = \sum_o \sum_t (WP_{o,t} * DVP_{o,t} + WN_{o,t} * DVN_{o,t}) \quad (3)$$

Note that this estimation method is based on the assumption that the relationship between the gross margin and deviation of crude supply from an optimal value is linear and the effect of each crude type is independent from those of others.

4. Case Study

In order to validate the MINLP model of the procurement planning system and effectiveness of the weight estimation method, a case study involving a real industrial refinery is presented, based on its operational record for a six-month period, from May to October of 2015. Based on the historical price data for crude oil and refinery products for the 6 months period, the refining process model selects optimal crude oil profiles that maximize the gross refining margin. (Table 1)

Table 1. Optimal crude portfolio

Optimal Demand (kbbl/day)	Crude 1	Crude 2	Crude 3	Crude 4	Crude 5	Crude 6	Crude 7	Crude 8	Crude 9	Crude 10
May 2015	29.88	0.52	20.50	50.00	1.63	50.00	0.00	0.00	47.46	0.00
June 2015	29.88	0.52	20.50	50.00	1.63	50.00	0.00	0.00	47.46	0.00
July 2015	0.00	45.60	50.00	50.00	3.16	50.00	0.00	0.00	1.24	0.00
August 2015	37.49	50.00	50.00	45.43	0.00	0.00	0.00	0.00	17.08	0.00
September 2015	37.49	50.00	50.00	45.43	0.00	0.00	0.00	0.00	17.08	0.00
October 2015	50.00	50.00	37.61	43.55	0.00	0.00	0.00	0.00	18.84	0.00

Table 2. Estimated weight parameters for positive (upper) and negative (lower) deviance

	Crude 1	Crude 2	Crude 3	Crude 4	Crude 5	Crude 6	Crude 7	Crude 8	Crude 9	Crude 10
May 2015	43.05	35.90	45.57	0.00	63.45	0.00	62.90	61.63	65.26	81.99
June 2015	41.15	34.21	43.61	0.00	60.34	0.00	59.96	58.67	62.16	77.92
July 2015	60.17	34.79	0.00	0.00	64.09	0.00	73.91	78.36	114.04	123.64
August 2015	38.89	0.00	0.00	41.29	61.04	57.33	66.46	58.78	55.14	67.50
September 2015	42.24	0.00	0.00	45.81	68.86	64.31	74.97	65.01	60.46	73.31
October 2015	0.00	0.00	42.04	43.83	66.44	62.09	72.58	63.06	60.02	72.26
	Crude 1	**Crude 2**	**Crude 3**	**Crude 4**	**Crude 5**	**Crude 6**	**Crude 7**	**Crude 8**	**Crude 9**	**Crude 10**
May 2015	66.97	74.35	64.56	60.09	50.91	54.66	0.00	0.00	48.94	0.00
June 2015	63.68	70.81	61.33	57.30	48.61	52.11	0.00	0.00	46.62	0.00
July 2015	0.00	110.83	83.09	81.20	82.57	77.75	0.00	0.00	44.07	0.00
August 2015	62.00	59.02	57.73	60.12	0.00	0.00	0.00	0.00	53.48	0.00
September 2015	69.26	65.51	63.79	66.70	0.00	0.00	0.00	0.00	59.62	0.00
October 2015	62.64	64.02	66.60	65.41	0.00	0.00	0.00	0.00	57.20	0.00

Then, the weight parameters for the positive and negative deviations of each crude supply from the optimal demand are estimated at each month (Table 2). The weight parameters of positive deviations from the maximum available (50 kbbl) are set as zero since there is no case where the actual input exceeds the maximum amount of availability, both the optimization problem the and actual system. In a similar manner, the weight parameters of negative deviations from the optimal demand of zero are also set as zero.

Then, MINLP problem is decomposed into MILP and NLP using ProNoDis relaxation, and solved for 2 cases, with or without the weight parameters. The MILP and NLP are solved by CPLEX and NLP respectively using the GAMS software. The resulting solution from the IPOPT solver is a local optimum, which is common in dealing with large size non-convex systems.

Due to the large size of the complete problem involving the whole 6 months period, the rolling horizon method is used in this paper where it solves the scheduling problem with a 2 month period after an update of the problem at each month.

Finally, the refinery optimization is re-solved using the input crude profile, which are determined in the procurement planning as input parameters. The resulting refining margin are compared with the optimal margins (Figure 2). The zero value of gross margin in Figure 2 indicates an infeasible operation in the refining system where it was impossible to calculate a feasible solution with given input to the system due to a violation in the product specification. The total gross margin when the weight parameters are applied is 37.83 million dollars, which is 1.55 times higher than the gross margin achieved when the weight parameters are not applied. In addition, the operation result with the supplies determined with the weight parameters was feasible for 14 periods (38.89 %), compared to just 9 periods (25 %) for the case without using the weight parameters. This result demonstrates the effectiveness of varying the weight parameters of the deviations in terms of ensuring feasibility and increasing the refining margin.

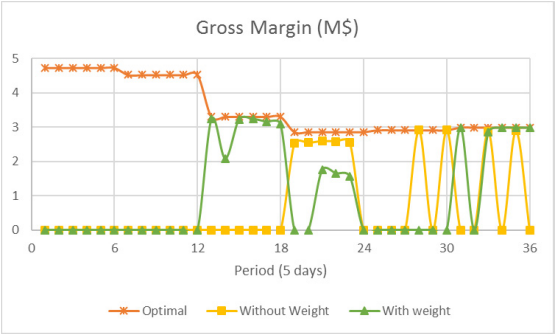


Figure 2. Refining margin for optimal operation, and actual operation with and without weight parameters

5. Conclusions

This paper discussed the integration of procurement planning and production process in petroleum refinery. This study sought to integrate the two systems by reducing the gap between the supply and demand by the two systems rather than by solving the combined mathematical programming problem. The MINLP problem, which aims to minimize the difference between the optimal crude portfolio calculated by the production system and the actual input was formulated for procurement planning. However, simply minimizing the value of deviance of crude profile resulted in substantial sub-optimality and sometimes even infeasibility. In order to resolve these drawbacks, weight parameter for each crude type was estimated at each time period. As a case study, the MINLP model and weight estimation were conducted for an actual industrial petroleum refinery based on historical data of crude and refinery product price information from May to October 2015. The performance of weight parameter estimation scheme was verified in simulation using these data and the results were significantly improved: the gross margin was 1.55 times higher and feasible region was 1.55 times longer compared to the operation without the weight estimation.

6. References

I. E. Grossman, 2012, Advances in mathematical programming models for enterprise-wide optimiztion, Computers & Chemical Engineering, 47, 2-18

M. Erdirik-Dogan, I E.Grossmann, 2008, Simultaneous planning and scheduling of single-stage multi-product continuous plants with parallel, Computers & Chemical Engineering, 32(11), 2664-2683

J. D. Kelly, B. C. Menzes, I. E. Grossmann, F. Engineer, 2017, Crude-oil Blend Scheduling Optimization of an Industrial-Sized Refinery: a Discrete-Time Benchmark, In Foundations of Computer Aided Process Operations: Tucson, United States, coference paper

B. C. Menzes, J. D. Kelly, I. E. Grossmann, 2015, Phenomenological Decomposition Heuristic for Process Design Synthesis in the Oil-Refining Industry, Computer Aided Chemical Engineering, 37, 1877-1882

K. Noh, J. Shin, J. H. Lee, 2017, An optimization based strategy for crude selection in a refinery with lube hydro-processing, Computers & Chemical Engineering, DOI

T. A. Oddsdottir, M. Grunowb, R. Akkerman, 2010, Procurement planning in oil refining industries considering blending op erations, Computers & Chemical Engineering, 58, 1-13

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50184-1

Integration of the Biorefinery Concept for Development of Sustainable Processes for the Pulp and Paper Industry

Ghochapon Mongkhonsiri[a], Rafiqul Gani[b], Pomthong Malakul[c], Suttichai Assabumrungrat[a,*]

[a]*Center of Excellence in Catalysis and Catalytic Reaction Engineering, Department of Chemical Engineering, Faculty of Engineering, Chulalongkorn University, Bangkok 10330, Thailand*
[b]*PSE for SPEED, Skyttemosen 6, DK 3450 Allerod, Denmark*
[c]*Petroleum and Petrochemical College, Chulalongkorn University, Bangkok 10330, Thailand*
Suttichai.a@chula.ac.th

Abstract

This work aims at developing sustainable processes for pulp and paper industry by integration of the biorefinery concept to an existing pulp and paper process. A systematic methodology employing a superstructure-based process synthesis approach is used with support from computer-aided tools to determine potential pathways for long-term sustainable growth. A superstructure of the multi-product biorefinery process network for the pulp and paper industry is developed. It is divided into three sub-networks, a chemical pulping section, a biochemical production section and a black liquor utilization section. Lignocellulosic biomass extracted from eucalyptus and bagasse is applied as feedstock for the network. Potential products from the biochemical production section include ethanol, lactic acid (LA) and succinic acid (SA) while those from black liquor utilization section include dimethyl ether (DME) and lignin apart from its conventional utilization as low-heating value energy for the pulp and paper process. Superstructure optimization is performed with the objective to maximize profit to determine optimal integrated networks for three scenarios. The obtained results provide useful insights for further development of optimal networks as sustainable integrated biorefinery combined with pulp and paper mills.

Keywords: Pulp and paper industry, Biorefinery, Process integration, Process synthesis, Superstructure optimization

1. Introduction

The pulp and paper industry today has the world's largest woody biomass utilization system and is a large user and producer of bioenergy. However, the pulp and paper production require large consumption of chemicals, energy and water, and is characterized by low efficiency of raw material and energy utilization, low innovation development and high product price volatility. Moreover, the printing paper consumption has been diminishing because of increased use of digital media through computers, tablets and smart phones. The traditional pulp and paper industry needs to transform to respond to the current situation. Integration of biorefinery into an existing pulp mill has been identified as a potential pathway for long-term sustainable growth of

not only the paper industry but also biorefineries. The integrated process could transform a low-margin paper manufacturing production into a marketable multi-product production system involving new revenue streams through the production of biomaterials, bioenergy, biofuels and bio-chemicals and thereby achieve efficiency and profitability improvements (Rafione et al., 2014).

Superstructure-based process synthesis approach provides the overview and analysis needed for commercial development of biorefinery integrated with pulping processes. Here, the superstructure represents all possible pathways in an integrated network of processing options. Superstructure optimization can be performed to find out the most effective pathways in terms of the defined criteria (Zondervan et al., 2011). This study aims to identify a promising process pathway for a biorefinery integrated with an existing pulp mill considering different objectives (scenarios). The superstructure consists of three main sections, which are pulping section as the receptor, the biochemical production section and the black liquor utilization section. The three scenarios considered, include pulp sale and biochemical production as alternatives (Scenario I), biochemical co-production with pulp sale (Scenario II) and multiple biochemical production (scenario III). The appropriate integrated networks for each scenario are presented.

2. Synthesis Methodology

Process synthesis problems involving many alternative technologies for each processing step is complex and a systematic methodology supported by computer-aided tools are needed for their solution. Super-O (Bertran et al., 2017) with GAMS (GAMS Development Corporation, 2011) is applied for process synthesis and the four steps solution approach is employed in this paper.

2.1. Step I :Problem definition

First the synthesis problem is defined. A pulp and paper mill is selected as the receptor mill. Biorefinery platforms, involving biochemical production, and black liquor utilization are integrated into the selected receptor mills to share raw material, by-products and some utilities. The objective function here maximizes profit of the integrated network. The profit is calculated as the difference of income from product sale and expenditure consisting of raw materials, chemicals added, utility and capital cost.

2.2. Step II :Superstructure generation

A superstructure represents all processing routes that are possible as alternatives in terms of processing steps and intervals (processing technologies for each step). Interval information on raw materials, main products, side products, reactions, chemical added, utilities and economic data such as product price, raw material cost, chemical cost and capital cost are collected from published articles and scientific literatures, available industrial data and databases. Receptor pulp mill configuration consists of Kraft process, where eucalyptus is a raw material, and Soda process with bagasse as a raw material. Overview of the superstructure is illustrated in Figure 1 and intervals for biochemical production are shown in Figure 2. The biochemical process includes productions of ethanol, lactic acid and succinic acid. Alternatives of pulp pathway are to market pulp for sale and as feedstock for biochemical production. Bark and pith can be alternatively applied for energy supplement with black liquor utilization. Three

technologies including Black Liquor Gasification Combined Cycle (BLGCC), Black Liquor Gasification for Dimethyl Ether Production (BLG/DME), and Lignin Extraction (LE) are considered as efficient processes to replace a conventional process, the Tomlinson boiler. For all black liquor utilization technologies, the generated steam and electricity is supplied to meet pulping process demand. The excess electricity is sold to grid while any deficit electricity is purchased. Regarding DME production, the process can supply energy for the pulping section with DME production for sale. Lignin extraction is designed to supply extracted lignin as biofuel instead of fuel oil from petroleum in the lime kiln operation. Incomes from extra pulp productivity due to lignin extraction are also considered.

2.3. Step III :*Mathematical model*

Each process interval is represented by corresponding mathematical models (Bertran et al, 2017). The mathematical programming problem, process constraints, logical constraints and the variable bounds are formulated as a mixed-integer (non)linear programming problem or MI(N)LP model (Quaglia et al., 2015).

2.4. Step IV :*Problem Solution*

The optimization problem can be solved with GAMS through the user interface Super-O. The generic model for the intervals, the user defined objective function, process interval constraints, superstructure flow constraints and logic constraints are all defined through Super-O. The necessary data to solve the synthesis problem is given as input through a database. The problem is solved to find the optimal processing route with associated generated output data.

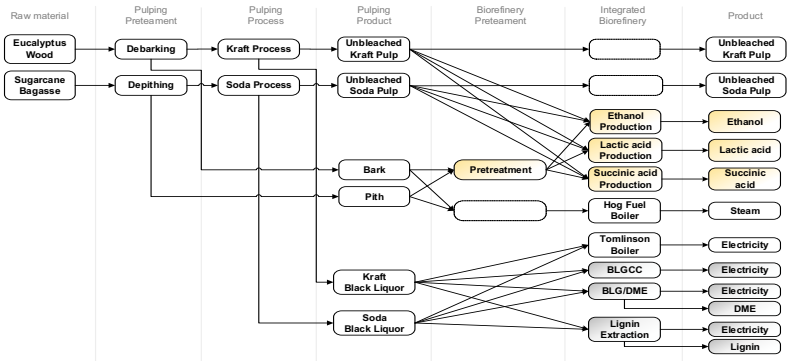


Figure 1. Overview of superstructure for integrated biorefinery network

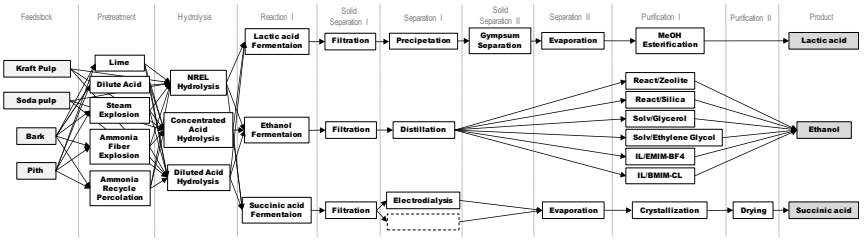


Figure 2. Superstructure of biochemical production

3. Results and Discussion

The superstructure optimization was carried out to determine optimal integrated networks for three scenarios. For all cases, raw material flow rate is maintained at 100,000 tons per year and project life is assumed to be 20 years.

Table 1. Statistics of the superstructure optimization problem (one scenario)

Problem	NI	71
	NF	2
	NP	15
	NC	51
	NR	55
Model and Solver	NEQ	825,633
	NV	812,856
	Problem type	MIP
	Solver	CPLEX

(Abbreviations as follows: NI is the number of processing intervals; NF is the number of feedstocks; NP is the number of products)note that NPI includes NF and NP(; NC is the number of components; NR is the number of reactions; NEQ is the number of equations; NV is the number of variables)

3.1. Scenario I: Biochemical production and pulp sale as alternatives

For economic decision on selecting one of the possible alternatives of conventional pulp sale and production of a biochemical product from unbleached pulp, superstructure optimization is performed to maximize profit (scenario 1). Economic results of the optimal networks are given in Table 2. The optimal network is Soda process with succinic acid production and black liquor gasification for DME production that produce 31,227 tons per year of succinic acid and 3,950 tons per year of DME from 100,000 tons per year of bagasse. Electrodialysis with crystallization is a suitable method for succinic acid purification. Ammonia recycle percolation is used for pith pretreatment with steam explosion for bark. Concentrated acid hydrolysis is employed before fermentation. Integrated biorefinery network without pulp sale increases revenue to 60 M$/year, which is higher than the conventional route with Soda process (48 M$/year) when integrated with the Kraft process. Optimization results for scenario I illustrate that top value-added biochemical products like succinic acid, lactic acid and DME with the existing pulp mill provide higher profit than the traditional pathway, whereas, the one with ethanol product is uneconomical. Instead of conventional recovery system, black liquor gasification with DME production is obviously an optimal black liquor utilization technology with Soda process that can also supply energy for the pulping process with high value biofuel production. However, lignin extraction is selected in case of Kraft process due to benefit of extra pulp production with lower capital cost.

Table 2. Economic results of optimal networks of overall superstructure and integrated biorefinery with Kraft process for scenario I.

Network	Profit (M$/y)	Sale (M$/y)	Raw material Cost (M$/y)	Chemical Cost (M$/y)	Utility Cost (M$/y)	Capital Cost (M$/y)
Soda pulping process with SA production and BLG/DME	67.68	96.36	2.30	2.48	22.21	1.69
Kraft pulping process with SA production and LE	63.28	89.30	2.90	1.83	21.20	0.09

3.2. Scenario II: Biochemical co-production with pulp sale (20, 50, 80 and 100% of the pulp sale)

In scenario II, the amount of pulp sale is specified at different percentages of total unbleached pulp production; i.e. 20, 50, 80 and 100% (Total pulp sale). Results of optimal networks are given in Table 3. The integrated network with Soda process shows the highest profit for all cases. With Soda pulp sale, succinic acid and DME production is an optimal pathway but the lactic acid process provides higher profit with 100% pulp sale) with pith employed as feedstock for lactic acid production. Although the integrated biorefinery with Kraft pulp sale offers 7% lower profit than with the Soda pulp, it provides rather high benefit with succinic acid production and lignin extraction. The profit of network decreases by about 21% with 60% increase of pulp sale due to the reduction of biochemical production but it can respond to the market need of paper pulp.

Table 3. Economic results of the optimal networks for scenario II

Optimization case	20 % pulp sale	50 % pulp sale	80 % pulp sale	100 % pulp sale
Pulp/SA or LA Flow (t/y)	7,128/26,857	17,820/20,022	28,512/12,824	35,640/10,388
Product sale (M$/y)	85.40	68.10	49.71	37.27
Profit (M$/y)	59.78	47.26	34.47	25.14

3.3. Scenario III: Multiple biochemical production integrated with pulping process

Production of multiple biochemical products has the flexibility to respond to fluctuating markets. According to scenarios I and II, succinic acid is an optimal product for the integrated biochemical plant with pulping process. Scenario III presents optimal pathways and economic data of integrated multiple biochemical production with pulping process along with co-production with pulp sale. The optimization in scenario III was performed with varying succinic acid productivity in terms of conversion of sugar to succinic acid: 30, 50 and 80 %. Multi-biochemical co-production with pulp sale is considered by maintaining the succinic acid production to around 10,000 tons per year, based on current succinic acid demand surveyed in Thailand. Table 4 gives product flow and economic results of integrated network with Soda process for scenario III. Multiple biochemical production integrated with Soda process gives the highest profit with succinic acid, lactic acid and also DME productions. Integrated multi-production with Kraft process is also a potential network, even if it has lower profit. The results of different succinic acid productivity indicate that the increase of succinic acid leads to a profit drop but only 3-8 %. Pulp split for sale makes profit dwindle to about 10-50 % depending on % of pulp sale. However, multiple biochemical streams have

great potential to improve economic benefit along with supplying pulp according to market need and especially to respond to the volatile market.

Table 4. Economic results of the optimal networks for scenario III

Optimization Case	30 % sugar division	50 % sugar division	80 % sugar division	20 % pulp sale	50 % pulp sale	80 % pulp sale
SA/LA flow (t/y)	9,368/ 25,195	15,613/ 17,997	24,982/ 7,199	10,144/ 19,141	10,149/ 11,363	10,065/ 3,142
Product sale (M$/y)	88.74	90.92	94.19	79.28	64.61	48.66
Profit (M$/y)	61.85	63.28	65.42	55.04	44.49	33.45

4. Conclusions

Superstructure optimization has been performed to generate optimal integrated network alternatives considering three scenarios that indicate promising future trends for pulp and paper industry transformation. From the results, top value-added biochemicals have been identified, such as succinic acid and lactic acid but not ethanol to improve the profitability of pulp mill as well as black liquor gasification with DME production that can supply bioenergy and biofuel. Soda pulping mill is a suitable receptor for the integrated biorefinery. However, Kraft process implemented with succinic acid and lignin extraction is a feasible network even though it has lower profit. Also, the cases with integrated multiple bio-product with/without pulp sale illustrate that they provide higher profit than the conventional pathway, only pulp sale, together with responding the fluctuated market. Superstructure-based process synthesis approach supported by Super-O as a user-friendly software interface with GAMS solver is an effective systematic methodology for the synthesis of integrated biorefinery networks with pulp and paper industry. The case studies solved so far have demonstrated the applicability of the synthesis methodology through Super-O, which can manage large and complex problems with a fast problem formulation, robust solution and efficient data management.

Acknowledgement

This work is financially supported by the Thailand Research Fund (DPG5880003) and the Royal Golden Jubilee PhD Program from the Thailand Research Fund and Chulalongkorn University.

References

T. Rafione, M. Marinova, L. Montastruc, J. Paris, 2014, Applied Thermal Engineering, 73, 74–81.

E. Zondervan, M. Nawaz, A.B. de Haan, J.M. Woodley, R. Gani, 2011, Computers and Chemical Engineering, 35, 1752–1766.

M. Bertran, R. Frauzem, A. Sanchez-Arcilla, L. Zhang, J.M. Woodley, R. Gani, 2017, Computers and Chemical Engineering, 106, 892–910.

GAMS Development Corporation, 2011. General Algebraic Modelinng System (GAMS).

A. Quaglia, C.L. Gargalo, S. Chairakwongsa, G. Sin, R. Gani, 2015, Computers and Chemical Engineering, 72, 68–86.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50185-3

Multi-period Optimization of Hydrogen Networks with Sulfur Content Variation

Jing Li[a] and Qiao Zhang[a,*]

[a]*School of Chemical Engineering & Technology, Xi'an Jiaotong University, Xi'an, Shaanxi Province, China, 710049*
q.zhang1986@stu.xjtu.edu.cn

Abstract

Crude oil is becoming heavier and lower-quality worldwide, especially the increase of sulfur content, which significantly influence on hydrogen consumption and products quality. At present, mathematical programming for hydrogen network integration can effectively conserve hydrogen resource. However, existing model hardly consider sulfur variation of crude oil. In this paper, a mixed integer nonlinear programming (MINLP) model is established to address hydrogen networks with feed oil sulfur content variation in multi-period. Hydroprocessing units are classified as upstream and downstream ones and new connections are added to perform integration. Fresh hydrogen minimization are employed as objective to optimize the hydrogen networks. Case study indicates that the novel model can achieve optimal strategy for sulfur content variation.

Keywords: Hydrogen network; MINLP model; sulfur content variation; Multi-period

1. Introduction

In recent years, daily demand for petroleum oil products is increasing rapidly. The problem of processing much heavy crude oil become more serious (Wang & Zhang 2005). Because of it, sulfur, as an important impurity in crude oil which has greatly influence on the hydrogen consumption of hydrogenation process, whose content in some areas is lower than 0.5%, while others are 2% or even higher (Xiang et al., 2006). Hydrogen network integration can effectively reduce the waste of hydrogen resources. Mathematical programming method based on superstructure is widely used in research and application. It was early proposed by Hallale and Liu (2001) to address hydrogen networks with basic constraints. Zhou et al. (2012) considered desulfurization to improve hydrogen resource utilization. Liang et al. (2016) discussed flexible and debottlenecking of multi-period hydrogen networks by considering possible operational conditions fluctuation to minimize total annual cost. By releasing the total flow rate and concentration normalization constraints, relative concentration basis was first put forward in hydrogen in 2016a (Zhang et al.) to synthesize single impurity hydrogen networks and then generalized to multi-impurity case (2016b).

Although some mathematical programming based literature aforementioned can address multi-period hydrogen networks, sulfur content variation is not considered. When sulfur content of source crude oil varies, hydroprocessing units in upstream and downstream will be both affected. To offset such variation, probably more fresh hydrogen is needed to guarantee production. For a certain sulfur content of crude oil, the hydrogen network should be optimized holistically and thus for variational sulfur content multi-period hydrogen network is the solution to refinery hydrogen network integration. In order to

minimize additional fresh hydrogen demand, this paper proposed a MINLP model to address multi-period hydrogen networks with sulfur content variation. Results show that the total fresh hydrogen with this method can be reduced compared to that obtained traditional methods.

2. Theory

2.1. Problem illustration

There are many hydrogenation units in petroleum processing, and they locate in system with upstream and downstream relationship, which is typically shown in Figure 1. The source crude oil contains sulfur and such impurity will be removed in upstream and downstream hydroprocessing units.

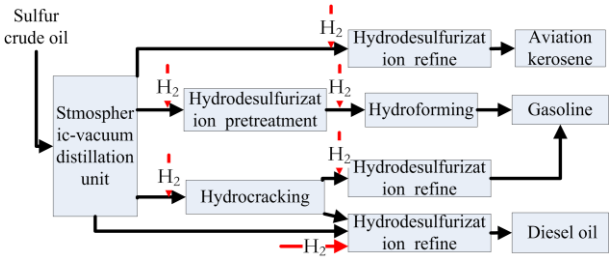


Figure 1: Upstream and downstream sites of hydrodesulphurization in petroleum processing

Therefore, when the crude oil sulfur content varies, the sulfur removal amount and hydrogenation performance in following upstream and downstream units will change accordingly and thus both the hydrogen supply and product oils quality cannot be guaranteed with constant hydrogen allocation network. This paper aims to optimize hydrogen network under sulfur content variation circumstance for minimizing fresh hydrogen consumption and the caused influence on the production of whole system.

2.2. Superstructure

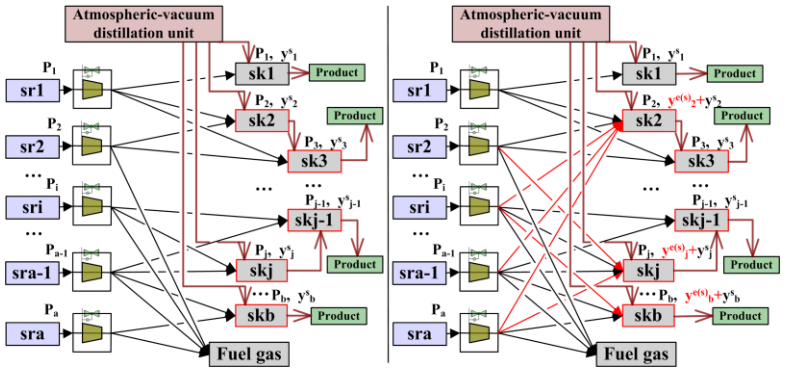


Figure 2: Original structure (left) and superstructure (right)

In Figure 2, *sr* represents hydrogen source; *sk* is hydrogen sink; the left one is a hydrogen network in steady state, where y^s_j is sulfur content of oil processed in *skj*. Sinks in red frame line, whose hydrogen demand affected by the sulfur content of the crude oil. In particular, units with upstream and downstream relationship like *sk2* and

sk3, if the influence of sulfur content variation is eliminated in *sk2*, *sk3* will not be influenced. Because of the heavier crude oil, if there is sulfur content change ($y^{e(s)}_j$) in those units, the superstructure on the right of Figure 2 is built to optimize hydrogen network. To achieve the optimal hydrogen network without adding compressor, only cases that hydrogen sources with enough pressures for sinks are considered to save hydrogen by only adding pipelines (red lines).

2.3. Mathematical model

Relative concentration basis in this paper refers to Zhang et al. (2016b).

$$RC_{k,sr/sk} = \frac{y_{k,sr/sk}}{y_{H,sr/sk}} \tag{1}$$

Relative concentration is used to quantify the impurity concentration while hydrogen flow rate is employed to quantify total flow rate.

Generally, data of the processing throughput (V_j) and its hydrogen demand (H_j) about different hydrogenation reactors is provided by refineries. However, when sulfur content of the oil processed by hydrogen sink changes $y^{e(s)}_j$, the hydrogen consumption need to change $H^{e(s)}_j$ at least to eliminate the influence. This paper establishes a direct proportional relationship between sulfur content and hydrogen consumption, and q is the coeffcient of proportionality. q is influenced by heuristic factors like origin of crude oil, while generally, there is an empirical value from refineries.

$$H^e_j = H_j + H^{e(s)}_j \tag{2}$$

$$H^{e(s)}_j = q \cdot y^{e(s)}_j \cdot V_j \tag{3}$$

About the constraints of sk_j, Eq.(4) means the amount of hydrogen provided by all sources should meet the demand of sk_j; Eq.(5) indicates that the average impurity concentration ($RC_{k,i}$) should be no larger than the upper limit of sk_j ($RC_{k,j}$).

$$\sum_{i=1}^{a} H_{i,j} + \sum_{i=1}^{a} H^{new}_{i,j} \geq H^e_j \tag{4}$$

$$\sum_{i=1}^{a} H_{i,j} \cdot RC_{i,k} + \sum_{i=1}^{a} H^{new}_{i,j} \cdot RC_{i,k} \leq RC_{j,k} \cdot (\sum_{i=1}^{a} H_{i,j} + \sum_{i=1}^{a} H^{new}_{i,j}) \tag{5}$$

For the adjustment of hydrogen supply from sources, hydrogen can be supplied according to the original state. In Eq.(6), $d_{i,j}$ is current allocation network matrix.

$$H_{i,j} \leq H^{up}_{i,j} \cdot d_{i,j} \tag{6}$$

In order to add new pipelines ($H^{new}_{i,j}$) described in superstructure, binary variables, $Bin1_{i,j}$ and $Bin2_{i,j}$ have been set.

$$If \quad P_i > P_j \ and \ d_{i,j} = 0 \Rightarrow Bin1_{i,j} = 1 \tag{7}$$

$$Bin2_{i,j} \leq Bin1_{i,j} \tag{8}$$

$$H^{new}_{i,j} \leq H^{up}_{i,j} \cdot Bin2_{i,j} \tag{9}$$

Main constraint for *sr* is availability of hydrogen resource: the sum of $H_{i,j}$ and $H^{new}_{i,j}$ (hydrogen from sr_i to sk_j) should be less than the total hydrogen flow rate of *sri* (H_i).

$$H_i = F_i \cdot y_{H,i} \tag{10}$$

$$\sum_{j=1}^{b} H_{i,j} + \sum_{j=1}^{b} H_{i,j}^{new} \le H_i \tag{11}$$

Fresh hydrogen is highest in the total cost. So in this paper, objective function is minimum the consumption of fresh hydrogen. In Eq.(12), *sr1* is fresh hydrogen.

$$Min{:}FSR1 = \frac{\sum_{j=1}^{b} H_{sr1,j}}{y_{H,sr1}} \tag{12}$$

This mathematical model consists of Eq.(1) ~ (12). There are endogenous variables $H_{i,j}$, $H^{new}_{i,j}$, and binary variables $Bin1_{i,j}$ and $Bin2_{i,j}$, so it is a mixed integer nonlinear programming (MINLP) model.

3. Case study

The data of hydrogen network comes from a refinery in China. Simplified by concept of hydrogen source and sink, data has been collated in Table 1 and 2.

Table 1. Data of hydrogen sources.

	F (Nm^3/h)	$y_{H,i}$	*P* (MPa)	$y_{k,i}$ (%)		
				H_2S	N	C
sr1	<20000	0.9999	2.2	0	0.01	0
sr2	17100	0.872	17.5	1.0	2.3	9.5
sr3	26015	0.884	8.0	1.3	1.5	8.8
sr4	13320	0.942	7.2	1.7	1.5	2.6
sr5	11790	0.932	4.0	0.6	5.1	1.1
sr6	22390	0.914	5.0	2.3	3.2	3.1
sr7	14630	0.897	4.0	4.2	2.6	3.5

Table 2. Data of hydrogen sinks.

	V (t/h)	$y_{H,j}$	H_j (Nm^3/h)	*P* (MPa)	$y_{k,j}$ (%)		
					H_2S	N	C
sk1	17.6	0.884	15621	18.4	2.0	2.8	6.8
sk2	60.4	0.913	22141	13.0	1.2	2.4	5.1
sk3	34.2	0.951	12087	8.0	1.4	1.9	1.6
sk4	38.6	0.939	17076	6.5	0.7	4.6	0.8
sk5	59.4	0.847	18982	4.8	3.0	1.2	5.1
sk6	29.7	0.908	13875	3.3	1.8	3.3	4.7

This hydrogen network has 7 sources and 6 sinks. There are three impurities in hydrogen streams, and they are sulfur, nitrogen and carbon, respectively. The original oil of the refinery is low sulfur content crude oil with 0.4% sulfur mass fraction. While due to the heavier crude oil, the variation range of sulfur content is large. Through the analysis of the categories and quantities of sulphide in this case, as the upstream hydrogenation units, sk5 and sk6 and *q* is 1388.9 Nm^3/t. The variation of sulfur content in crude oil is random, in order to optimize hydrogen networks, this paper takes a group of sulfur content to perform optimization. Meanwhile, the total fresh hydrogen

consumption obtained by this method is compared with that by existing methods for same sulfur content crude oils. Results are summarized as follows.

Table 3. Total hydrogen consumption calculation at the upstream hydrogen sinks 5,6

Sulfur content (%)	Fresh hydrogen consumption (Nm^3/h)		
	Original strategy	New allocation strategy	New allocation strategy with new connection
0.4	19311.2	19311.2	19311.2 (None)
0.8	19806.2	19424.7	19424.7 (None)
1.2	20361.2	19542.2	19542.2 (None)
1.6	20841.2	19663.2	19663.2 (None)
2.0	21351.2	19793.3	19786.6 (sk6-sr3)
2.4	21861.2	19991.8	19971.9 (sk6-sr3)
2.8	22381.2	20200.1	20068.9 (sk6-sr3)
SUM	145913.4	137926.5	137768.7

In Table 3, Original strategy is variation of sulfur content eliminated only by adding fresh hydrogen directly; column of New allocation strategy is adding fresh hydrogen with redistribution of hydrogen resources without new pipelines; the last column, with the reallocation of hydrogen network, sometimes requires additional pipelines. Results show that, redistribution of hydrogen resources can save lots of fresh hydrogen. Specifically described as, the amount of sulfur content increases by 0.4%, and the average of hydrogen saving is 385 Nm^3/h based on New allocation strategy without new pipelines; new pipelines can be further economical on this basis.

Analysis of the whole period of sulfur content variation, only by adding fresh hydrogen, the total fresh hydrogen consumption is 145913.4 Nm^3/h; original structure but new hydrogen allocation strategy, the total fresh hydrogen consumption is 137926.5 Nm^3/h, 7986.9 Nm^3/h less than the Strategy 1; new hydrogen allocation strategy with few new connections, the total fresh hydrogen consumption is the least, 137768.7 Nm^3/h. The results of three strategies are illustrated in Figure 3.

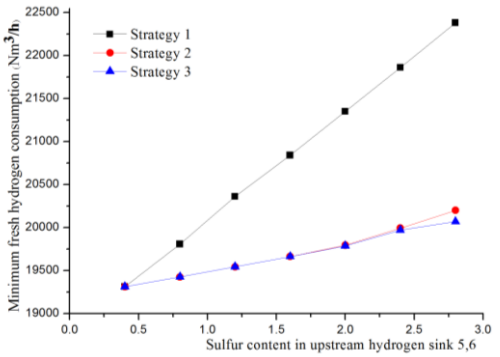


Figure 3: Fresh hydrogen consumption in three strategies

Figure 4 shows the hydrogen allocation networks of minimum and maximum sulfur content optimized by strategy 3. By comparison, the network is slightly changed only with one pipeline addition and none of new compressor equipments. Therefore, such retrofit is very economical for refineries to improve its hydrogen resource utilization.

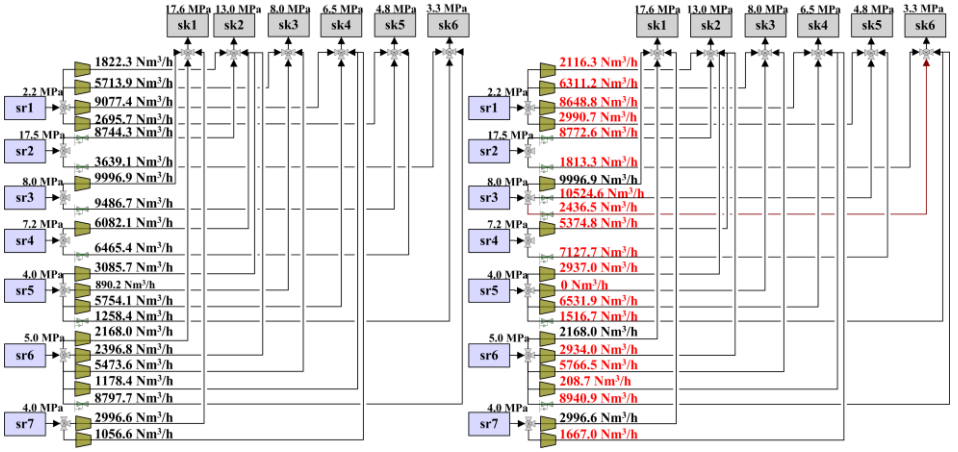


Figure 4 Optimal network of 0.4 % sulfur content (left) and 2.8% sulfur content (right)

4. Conclusions

According to sulfur content variation in refining process, a superstructure and MINLP model is proposed in this paper to optimize hydrogen resources distribution strategy under different states. Results prove that, during the same period of sulfur content variation, this model can effectively save fresh hydrogen consumption and eliminate the influence of impurity variation with very small change of the original structure and the hydrogen allocation network.

Acknowledgements

Financial support from the National Natural Science Foundation of China under Grant 21506169 is gratefully acknowledged.

References

L. Zhou, Z. Liao, J. Wang, B. Jiang, & Y. Yang, 2012, Hydrogen sulfide removal process embedded optimization of hydrogen network, International Journal of Hydrogen Energy, 37(23), 18163-18174.

N. Hallale, F. Liu, 2001, Refinery hydrogen management for clean fuels production, Advances in Environmental Research, 6(1), 81-98.

Q. Zhang, M. Yang, G. Liu, & X. Feng, 2016a, Relative concentration based pinch analysis for targeting and design of hydrogen and water networks with single contaminant, Journal of Cleaner Production, 112, 4799-4814.

Q. Zhang, H. Song, G. Liu, & X. Feng, 2016b, Relative concentration-based mathematical optimization for the fluctuant analysis of multi-Impurity hydrogen networks, Industrial & Engineering Chemistry Research, 55(39), 10344-10354.

W. Baoyi, Z. Baosheng, 2005, Analysis of china's crude oil supply and demand in the year 2010 and the year 2020, Petroleum Science (3), 89-92.

X. Liang, L. Kang & Y. Liu, 2016, The flexible design for optimization and debottlenecking of multi-period hydrogen networks, Industrial & Engineering Chemistry Research, 55(9).

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50186-5

Energy Recovery in Heat Exchanger Networks in a Dynamic, Big-data World: Design, Monitoring, Diagnosis and Operation

Sandro Macchietto[a,b], Francesco Coletti[b,c] and Emilio Diaz Bejarano[b]

[a]*Department of Chemical Engineering, Imperial College London, SW7 2AZ, UK*
[b]*Hexxcell Ltd., Innovation Hub White City, 80 Wood Lane, London W12 0BZ*
[c]*College of Engineering, Design and Physical Sciences, Brunel University London, Uxbridge, UB8 3PH, UK*
s.macchietto@imperial.ac.uk

Abstract

Heat exchanger networks (HENs) are often analysed with simplified models which are not sufficiently predictive in the presence of fouling, in particular for operations monitoring, diagnosis and support. A comprehensive approach is presented, with application to oil refining, that moves substantially beyond the usual simplifying assumptions and incorporates: i) far more detailed models of the physics involved ii) detailed dynamics for both slow and fast changes; iii) detailed reaction engineering models of the deposition of fouling deposits on thermal surfaces as well as deposits removal in chemical and mechanical cleanings; iv) the ability to easily generate multi-scale, detailed models of individual exchangers and whole networks; v) consideration of the complex thermal and hydraulic interactions between exchangers and effects on performance; vi) exploiting the abundant plant data available (historical and current) in conjunction with the more sophisticated models; vii) solution of all problems within an easy to use engineering software environment.

Keywords: heat exchanger networks; energy integration; high-fidelity modelling; design and retrofit; scheduling and control; monitoring and data analysis.

1. Introduction

Energy recovery in integrated processes may be achieved through a complex network of heat exchangers (HEN), aimed at transferring as much as possible of the energy from a set of hot streams to cold streams. In crude oil refining, particularly relevant are the extensive heat exchanger network and energy recovery in the crude distillation units (CDU) at the front end of every refinery, which process essentially the entire world crude oil. Indeed, it may be argued that the whole fields of Pinch Analysis, HEN synthesis and HEN optimisation resulted from the exceptional economic motivation in this area, generating much research (e.g. Furman and Sahinidis, 2002), innovation and benefits over the last 50 years. In spite of the many advances, the drive towards additional efficiencies in the design and operation of such HENs is continuing, with strong economic incentives. To achieve such improvements, one must focus on aspects which have not been (sufficiently) considered, simplifying assumptions which may be limiting, the nature of residual efficiencies, and tools required to exploit them. In particular, the simplified exchanger models often used in Pinch analysis, HEN synthesis and optimization have been found in practice to poorly represent and support actual operations.

In this paper, a comprehensive modelling and software system is presented that moves substantially beyond the usual simplifying assumptions and incorporates the following main aspects: i) a far more detailed description of the physics involved in the process (from characterisation of the fluids and equipment geometry, to description of the heat transfer mechanisms and hydraulics and all aspects that affect them); ii) moving away from a steady-state or pseudo steady-state description of the process to properly include dynamics in detail, not just in a simplified way; iii) detailed reaction engineering models of the deposition of fouling deposits on thermal surfaces, considering deposit composition and coking (ageing) and their key effects on thermal and hydraulic performance, as well as the removal of such deposits in chemical and mechanical cleaning activities to restore thermal/hydraulic performance. This enables a seamless simulation of operation/cleaning cycles; iv) the ability to easily generate individual exchangers and networks from multi-scale models in such detail; v) proper consideration of the interactions between exchangers, and very complex shifts of thermal and hydraulic performance within a network which depend on its structure; vi) exploiting the abundant plant data available (historical and current) in conjunction with the more sophisticated models to analyse and improve operations and verify design retrofits in realistic scenarios that include variable feed and other conditions; vii) efficient solution of the resulting multi-scale, large PDAE equation systems within an easy to use engineering software environment.

2. Heat exchanger and Network models

The model framework used for individual (shell and tube) exchangers (Coletti and Macchietto, 2011, Diaz-Bejarano *et al*., 2016) is schematically shown in Figure 1. Basically, 2D distributed (axial and radial) dynamic mass and energy balances are developed for each tube. A detailed, moving boundary deposit layer model accounts for its growth and (2D distributed) properties, as a function of local deposition (and possibly removal) rates at the oil/deposit boundary. These depend on temperature and shear stress, hence fluid velocity (hence available flow area, which depends on deposit thickness). This dictates the temperatures profile (hence ΔT and heat transferred) and contribution to pressure drop at each axial position. The temperature history of each point in the deposit affects its coking (ageing), hence thermal conductivity, hence the temperature profiles. Physical and transport properties of the fluids are also a function of temperature, as well as the nature of the fluids. Tube side models are coupled with shell side models and a whole exchanger model is assembled, accounting for geometry, co-current or counter-current configuration, single or multiple passes, etc. Of course, this results in quite a large number of coupled partial differential and algebraic equations (PDAEs), which nowadays can however be solved efficiently. As the extent, nature and location of the deposit and its effects are the sole difference between a clean and a fouled exchanger, capturing these highly complex interactions in detail is key to faithfully describing its operation and ensuring confidence in the results for simulation, design, monitoring, diagnosis of abnormal events, performance prediction and performance improvement purposes. A multiscale, high fidelity, thermo-hydraulic dynamic model of an entire HEN undergoing fouling is then simply developed by assembling its various heat exchanger units according to the network topology, with the addition of stream sources and sinks, mixers and dividers, by-passes, pumps (usually, just one), control elements, and (in refining applications) the final furnace (Figure 2). The Hexxcell Studio™ software by Hexxcell Ltd is used for this purpose and all other applications mentioned in this paper.

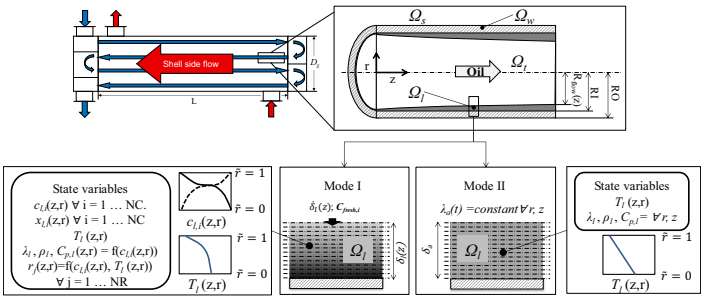


Figure 1. Schematic representation of exchanger modelling framework

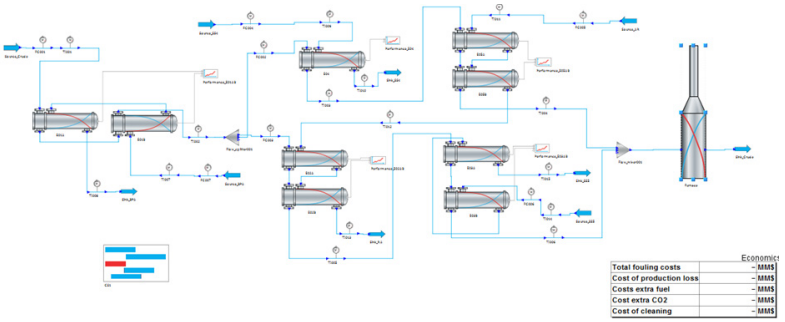


Figure 2. Preheat Train – Refinery X.

The fouling propensity of crude oils is notoriously difficult to predict (or even measure). Here, the *local* deposition rates are based on semi-empirical functional models of the Ebert and Panchal type. Typical fouling and ageing parameters for these models may be stored in a library, or estimated from primary plant data: flowrates, temperatures and, if available, pressure drops. This challenging dynamic parameter estimation problem is successfully performed for both individual and multiple exchangers, even with large data sets, making use of plant data are usually collected, stored and available but seldom used. They are usually noisy and there is no such thing as steady-state (Figure 3a). Extracting value from them is challenging. A Dynamic Data Analysis™ of such data (Figure 3b) allows "inverting" the above models and estimating an operational signature for each exchanger, including a wealth of normally unavailable quantities, e.g. time profiles of deposition rates, depth of deposits and their thermal conductivity, shear rates, etc. In a project carried out with a major US oil company, >120 year of operating data were analysed for 31 heat exchanger shells of diverse types in 7 refineries, operating a large variety of crudes in different conditions (Chunangad *et al*., 2016). The usual temperature and flowrate measurements were available for all the units, pressure drop measurements only for some. After parameter estimation, the model predictions were validated against plant data, confirming very good predictive ability (duty predicted within ±10% and ΔP within ± 0.2 bar) and parameters portability to other units in similar conditions. Refitting parameters may be needed when oil slates change significantly.

A new plot, the TH-Plot visualizes performance losses relative to a clean exchanger operating in the same dynamic conditions. Figure 4a) (and 4b for noisy data), shows the combined duty and pressure drop evolution (TH-line) in a heat exchanger against set operational reference lines (thermal limit, TL and hydraulic limit, HL). This unit is clearly heading to hit TH before TL. Using the models, additional reference lines may be

calculated and plotted, representing different degrees of deposit coking (Figure 4c). The TH-line of an exchanger allows monitoring its current vs. expected evolution, gives very useful early indication of deviations and provides diagnostic information about potential underlying causes. Figure 5a) shows the very distinct signatures of acute organic fouling and deposition of inorganic material (e.g. salts, oxides). The latter would not be detected using conventional monitoring of thermal fouling resistance. The TH-Plot of four years of operation of a problematic exchanger in a refinery hot end (Figure 5b), points to various periods of acute fouling and possible inorganics. A model-based analysis of the data estimated characteristics (thermal conductivity and inorganics fraction) of each layer of the deposit that matched surprisingly well a laboratory analysis of deposit samples collected from the exchanger tubes at the end of run (Figure 6), Diaz-Bejarano *et al*. 2017.

In Hexxcell Studio™ deposit removal is simulated in the same way as deposition. Chemical cleaning is incorporated using similar semi-empirical kinetic-type models. Its parameters may be fitted to actual cleaning performance of different chemicals and dosages. An entire campaign of operation (deposition) and cleanings (removal) may therefore be seamlessly simulated (Diaz-Bejarano *et al*., 2015), giving the ability to calculate, rather that assume, the initial state of a unit after a chemical clean. In Dynamic Data Analysis™ mode, the initial cleaning state may be back-estimated from a small number of data after the unit is returned to service.

Changing an exchanger design or operation to reduce fouling does not always deliver the expected improvement within a HEN. Increasing its duty changes temperatures and ΔTs elsewhere in the network, modifying its pressure drop shifts flowrates between parallel branches, all changing in time in ways hard to predict. A clear example was shown in a work with a European refiner (Diaz-Bejarano *et al*. 2017b). Validation for an existing exchanger confirmed model simulations were within the uncertainty of the measurement (±5% of the heat duty) for the entire 4.5-year period. A exchanger retrofit, checked on its own by re-running operating history (with the same inputs to the original design) for 4 years (Dynamic Retrofit Test™), confirmed its improved fouling mitigation effectiveness and higher energy recovery by 4.5 MW. With the same check for the network as a whole (Dynamic Network Retrofit Test™), due to complex interactions across parallel branches, only 60% of the 4.5 MW was recovered, however still reducing the fuel requirement at the furnace by $2.8 MM after 4 years, with high confidence in the results achieved.

Figure 3. Exchanger inlet temperatures and flowrates a); Dynamic Data Analysis of plant data b)

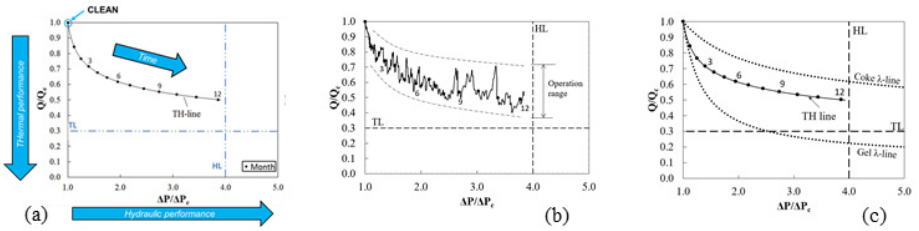


Figure 4. Thermo-hydraulic plot and reference lines are useful for monitoring and diagnosis (Diaz-Bejarano *et al* 2016b).

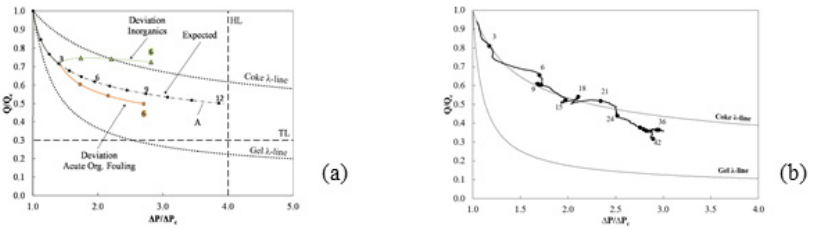


Figure 5. Monitoring and diagnosis of inorganics deposition a) (Diaz-Bejarano *et al* 2015) and TH-Plot for 4 years of operation in a refinery indicates period of acute fouling and inorganic deposition b).

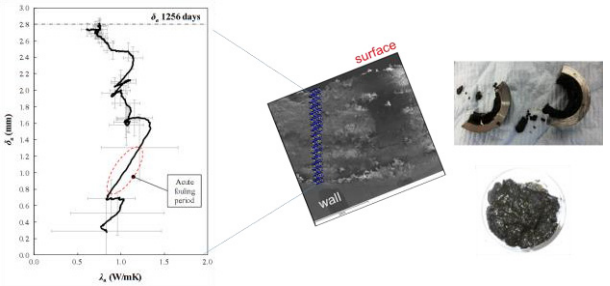


Figure 6. Model-based diagnosis of layering of deposit in a 4 years of operation in a refinery heat exchanger a) matched well with experimental composition analysis of the deposit b).

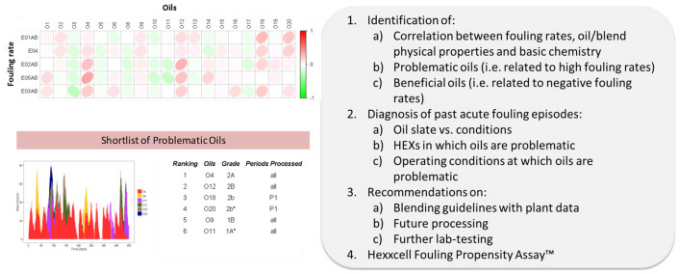


Figure 7. Model-based statistical analysis of large operational data sets identifies which crude oils are problematic, in which parts of the network, and in which operating conditions.

The ability to accurately predict the network performance and "soft-sensing" un-measured quantities, opens a further opportunity: a statistical analysis correlating fouling and performance indicators to inputs and operating conditions. With representative data (to protect client confidentiality) Figure 7 (top left) shows correlations of (soft-sensed)

fouling rates in 5 exchangers in a HEN to 20 crude oils used at different times and in different quantities in the oil blends fed to a refinery. The analysis identifies problematic or beneficial oils, in which combinations, and in which operating conditions. Scheduling and control aspects are covered in Lozano Santamaria and Macchietto (2017).

3. Conclusions

Numerous industrial applications have shown that integration of the predictive models described with the analysis of large scale sets of (already available) plant data delivers: i) excellent estimation of the current state of the exchangers and HEN overall ii) estimation of selected model parameters in the detailed dynamic models so as to fit the recent/current state iii) detailed monitoring and problems diagnosis of individual heat exchangers and the HEN overall iv) accurate long-range prediction of the HEN state and performance v) the ability to simulate and improve flow split and pressure driven control, cleaning schedules, and simultaneous control and scheduling vi) the use of such models in conjunction with plant data to statistically analyse the contribution of operational events and decisions to overall performance, in particular the effect of oil slates and "good" and "bad" oils in a blend vii) the ability to drive all above from economic as well as technical objectives. A shift in approach is therefore warranted from the simple models mostly used so far towards the more rigorous and predictive models described, analogous to moving from McCabe Thiele diagrams to rigorous models in distillation. The integration of advanced models with appropriate statistical analysis significantly increases the ability to extract useful information from data, predictive accuracy and improved economics.

References

Coletti, F. and S Macchietto, 2011, "A Dynamic, Distributed Model of Shell-and-Tube Heat Exchangers Undergoing Crude Oil Fouling", *Industrial & Engineering Chemistry Research* 50 (8). American Chemical Society, 4515-33.

Chunangad, K., R. Chang and R. Casebolt, 2017, "Evaluation and prediction of the thermal and hydraulic impact of crude oil fouling in exchanger design using pressure measurements", Int. Conf. Heat Exchangers Fouling and Cleaning, 11-16 Jun., Madrid, Spain

Diaz-Bejarano, E., F. Coletti and S. Macchietto, 2016, "A new dynamic model of crude oil fouling deposits and its application to the simulation of fouling-cleaning cycles", *AIChE J.*, 62 (1) 90–107.

Diaz-Bejarano, E., F. Coletti and S. Macchietto, 2016b. "Model-based monitoring of thermal-hydraulic performance of refinery heat exchangers undergoing fouling", *Computer Aided Chemical Engineering* 38, 1911-1916.

Diaz-Bejarano, E., F. Coletti and S. Macchietto, 2015, ESCAPE 25, Copenhagen Denmark.

Diaz-Bejarano, E., Coletti, F. and Macchietto, S., 2017a, "Organic and Inorganic Fouling in Heat Exchangers – Industrial Case Study: Analysis of Fouling State", *Applied Energy*, 206, 1250-1266

Diaz-Bejarano, E., M. Yugo Santos, M. García Dopico, L. Lanchas-Fuentes and F. Coletti, 2017b. "The impact of fouling on the optimal design of a heat exchanger network: an industrial case study", Proceedings, HEFC-12 Conference, June 11-16 2017, Madrid, Spain

Furman, K C and N V Sahinidis, 2002, "A Critical Review and Annotated Bibliography for Heat Exchanger Network Synthesis in the 20th Century." Industrial & Engineering Chemistry Research 41 (10) 2335-2370.

Lozano Santamaria, F and s. Macchietto, 2017 "Integration of optimal cleaning scheduling and control using a MPCC formulation for mitigating fouling in heat exchangers", Submitted for publication.

Macchietto S. *et al.*. 2011. "Fouling in Crude Oil Preheat Trains: A Systematic Solution to an Old Problem," *Heat Transf. Eng.*, 32 (3-4) 197–215.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50187-7

Design of Circular Economy Plants – The Case of Waste Textiles to Chemicals

Foteini Barla[a*], Athanassios Nikolakopoulos[a], Antonis Kokossis[a*]

[a]*National Technical University of Athens, Department of Chemical Engineering, 9 Heroon Polytechniou Street, Athens 15780, Greece.*
fbaria@mail.ntua.gr

Abstract

Circular Economy (CE) plants act as intermediate facilities, transforming waste resources produced from one industry to raw materials utilised by another. The main feedstock of such chemical plants is multi-component wastes. Currently, there is an abundancy of feasible chemical and biochemical pathways to achieve these transformations but their integration to result to an integrated CE plant is not trivial. In the course of a European Research Project, RESYNTEX, focusing to convert waste textile to chemicals, the task of process synthesis has raised various alternative process configurations, which are assessed by mathematical optimisation (MILP). This work investigates the most economically viable strategy to valorise waste textiles, comparing various configurations of CE plant (waste-to-chemicals) to what is currently practised (landfill and waste-to-energy). The results of this work suggest that waste-to-chemicals is the most feasible strategy due to chemicals' selling prices. Further, the study presents that the most economically favourable configuration is the one, where the waste textiles are not separated in fractions but treated as whole in cascaded process. The presented model can be utilised for examining how the uncertainty of landfill tariffs, the technology's efficiency and CAPEX can affect the selection of the optimal waste valorisation strategy and the CE plants' configuration.

Keywords: circular economy, waste textiles, biorefinery, process design, process synthesis

1. Introduction

The dual problem of waste management and scarcity of resources has sparked renewed interest in exploiting raw materials from waste resources. Circular Economy (CE) promotes the concept of interconnecting linear production lines into a closed loop. European Commission forecasts that by redesigning product lifecycles, there could be an increase up to 7% in GDP by 2035, resulting from creation of 580,000 new jobs and €600 billion savings for EU businesses. Additionally, there could be a reduction up to 450 million tonnes of carbon emissions (European Commission, 2015[1]). More than 10 EU research projects have been funded working towards circular economy, incorporating examples such as utilizing dairy byproducts to produce rich-protein feedstock, noise protection panels from recycled tyres, production of chemicals from end-consumer textile waste and others (European Commission, 2015[2]). Leading companies such as DuPont, Dell, HP, Levi Strauss and many others have also embraced the concept of circular economy by demonstrating in practise eco-innovation (Hower, 2016).

The textile and apparel industry is an ideal case, where the CE concept could be applied since it is among the most profitable industries, with a market value of €0.82 trillion and it strongly follows the linear business model dictated by the inexpensive and "fast fashion" trend. Furthermore, it is notoriously known for producing vast amounts of post-industrial and post-consumer textile wastes, with an average recycling portion of only 15%. The EU and the USA generated 10 and 14.3 million tonnes of waste textile respectively in 2016, attaining 5% of solid waste found in the landfills (EASME, 2015) (Leonas, 2017).

Currently, there are four main reclamation strategies for waste textiles. They are sold directly in (a) the used-clothing markets, they can (b) be down-cycled to stuffing for automotive components, building material, insulation, carpet underlays and others, they are (c) closed-loop recycled through the use of mechanical transformation to produce once again fibres and (d) open-loop recycled through chemical and biochemical processing to produce chemicals, materials and energy. Down-cycling is not a lucrative business line and thus little innovation has been achieved. Closed-loop recycling is still facing many obstacles in converting blended fibres due to lack of processing equipment. For open-loop recycling, multiple chemical and biochemical recycling routes have been developed at laboratory and pilot level to valorise only pure polymers found in textiles such as polyester (Paszun & Spychaj, 1997), cotton (Gallezot, 2012) and polyamide (Sethi, 2016). However, more than 50% of the global waste textile is blend-fibres (inseparable by mechanical means) and to the best of our knowledge, there is no real-world application that integrates the existing process pathways to valorize the waste textiles holistically other than the work performed in the RESYNTEX project.

RESYNTEX is a collaborative project that involves several industrial partners, research institutions, textile collectors, sorting companies and chemical industry end-users (RESYNTEX, 2015). Its goal is to transfer the lab and pilot chemistries into an integrated industrial scale process, promoting chemical/biochemical textile recycling, aiming to create a new circular economy concept for the textile and chemical industries. Some of the key challenges in the early stage of the process design are how to integrate and in what order to arrange different processes valorizing different part of waste textile? Which is the most feasible process configuration in a given policy frame for waste textile, along with a given technology readiness level (TRL)? The paper demonstrates a case study, intending to assess alternative process configurations of pre-specified chemistries (dictated by RESYNTEX), with the objective to attain maximum economic feasibility. The work further investigated how waste-to-chemicals strategy compares to the current practices of landfill and waste-to-energy.

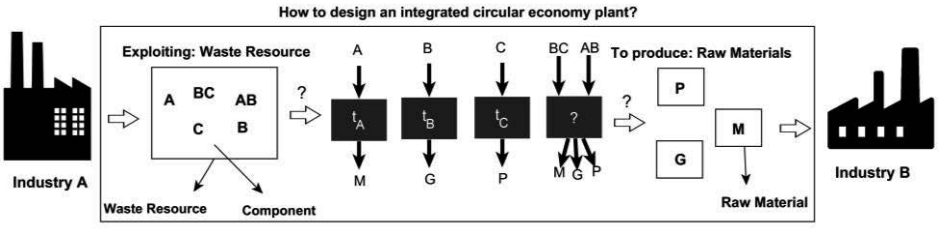


Figure 1: In what manner to combine waste-to-chemicals transformation processes in order to design a feasible integrated circular economy plant? A, B, C are pure materials, while AB and BC are blended, inseparable material.

2. Problem description

The problem assumes (given) a list of feasible, pre-specified chemistries demonstrated at lab and pilot, converting individual textile polymers to group of desired products such as wool: α-keratin (A) to wood-panel adhesives (M), cotton: cellulose (B) to glucose (G) and PET (C) to terephthalic acid (P). Figure 1 theoretically illustrates the transformation routes, linking two industries. It also assumes (given) fixed waste textile composition (percentage of polymers types, percentage of pure fibres and blended fibres). Furthermore, the problem assumes (given) fixed policies on the waste textile valosation within a specific time horizon, including tax breaks, landfill tariffs and mandatory waste textile collection. The problem is then to optimally determine:

(a) Waste valorisation strategy: no valorisation (landfill), Waste-to-Energy, Waste-to-chemicals or a combination of options.
(b) Feasible configuration of CE plant: to select among sorting waste to all separable components (A, B, C, AB, BC) and process individually, to process waste textile as whole in cascaded process or to sort to pure (A, B, C) and blend fibres (AB, BC) and treat in combined discrete-cascaded process configuration.

3. Methodology

3.1 Synthesis

The synthesis problem is represented using the *biomass bipartite graph* (BBR), presented in previous work (Kokossis et al., 2015). The transformation pathways are composed of flows (nodes) and enabling technologies (bar). The flows are divided in three sets, waste feedstock (inlet to a pathway), intermediate chemicals (component produced by a process and consumed by another) and final raw materials (not further processed). Figure 2 demonstrates the synthesis representation of a transformation pathway.

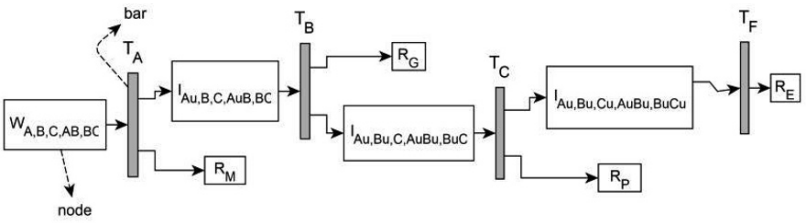


Figure 2: Synthesis representation of a transformation pathway.

3.2 Development of circular economy plant superstructure

To develop a superstructure, dedicated to wastes, a sequence of steps should be followed in prior: (a) any multi-component waste needs to be analyzed in terms of separable entities, composed of single-components (i.e. A, B, C) and inseparable components (i.e. AB, BC). (b) In case, two or more components are found in an inseparable mixture, then the order of the transformations should be identified as dictated by the governing phenomena of the processes such as reaction temperature and pressure, limiting alkali concentration, sensitivity of the material, etc. In order to map out the various waste valorization options:

(i) When no valorization: landfilled
(ii) When valorization: Waste-to-Energy
(iii) When valorization: Waste-to-Chemicals

And for the Waste-to-Chemicals strategy, the portfolio of configuration options is expanded:
CASE A: Waste feedstock is not sorted to its separable components and it is processed in cascaded configuration according to the order identified ($T^{all}_{A} \rightarrow T^{all}_{B} \rightarrow T^{all}_{C}$).
CASE B: Waste feedstock is sorted to all separable components and processed independently. For the inseparable mixture components, a cascaded process can be applied. (T^{pure}_{I}, T^{blends}_{I})
CASE C: Waste feedstock is separated to single-component (A, B, C) and mixture components (AB, BC). The configuration is processing the mixed-components in cascaded manner, whilst injecting pure component in the relevant process (T^{pnb}_{I}).

None of the Waste-to-Chemicals transformation technologies is converting 100% waste feedtstock to raw material. The unconverted components are denoted by subscript *u* and they are either delivered to another transformation process or sent to thermal valorisation and/or landfill. All options are depicted by Figure 3 and elucidated in Table 1.

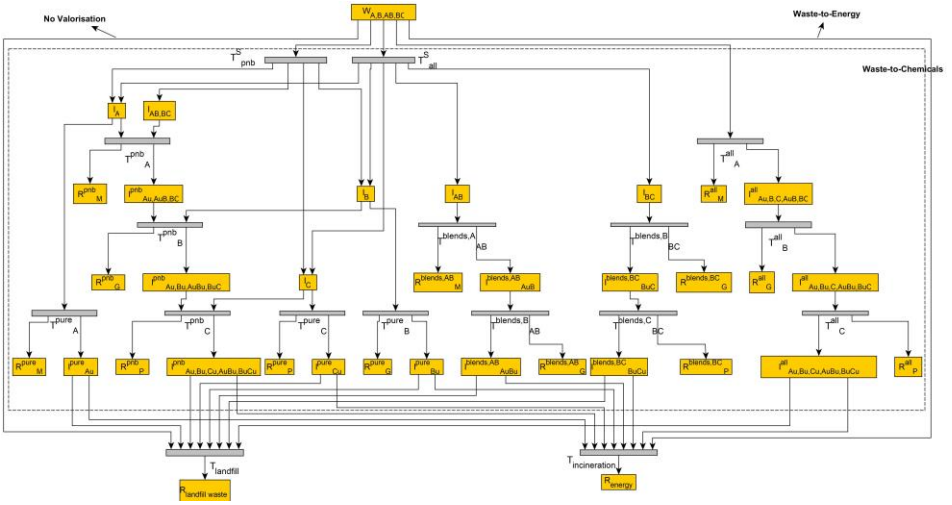


Figure 3: Development of circular economy plant: superstructure development indicating different valorization strategies and different plant configurations.

Table 1: Superstructure features of circular economy plant configuration

Sorting Process	**Transformation Processes**	**Description**
N/A	T^{all}_{A}, T^{all}_{B}, T^{all}_{C}	Processes that treat A, B, C respectively when no prior sorting is performed
Sorted all separable components- T^{S}_{all}	T^{pure}_{A}, T^{pure}_{B}, T^{pure}_{C} $T^{blends,A}_{AB}$, $T^{blends,B}_{AB}$, $T^{blends,B}_{BC}$, $T^{blends,C}_{BC}$	Processes that treat all separable components independently. These processes can only accept a well-sorted feedstock. They have high yields and not specialised separation equipment.
Sorted to blended and pure components (T^{S}_{pnb})	T^{pnb}_{A}, T^{pnb}_{B}, T^{pnb}_{C}	Cascaded process that can treat blended material with the corresponding injection of pure material.

3.3 Mathematical Formulation

The optimisation problem is formulated as Mixed Integer Linear Programming (MILP), following the work previously presented by Kokossis et al, 2015. In this work, the objective function is the maximisation of economic potential attained by the optimal circular economy process configuration. The optimisation problem is subject to mass balance, utility requirements (water & energy) as well as economic balance constraints. The mathematical formulation can account simultaneously for other objectives by scalarizing the problem.

4. Implementation

The presented methodology is applied in the main case study of RESYNTEX. The waste textile feedstock is mixture of pure and blended fibres, more specifically wool, cotton, PET, wool-cotton, cotton-PET denoted as A, B, C, AB, BC. The waste textile can be valorised according to strategy (i)-(iii) presented in section 3.2. For the textile waste and the chemistries dictated by the project, the order of the processes is set by the solvolysis' temperature of natural and synthetic fibres. The processes treating only pure materials ($T^{pure}_{A/B/C}$) have almost twice higher yields than the ones treating waste as a whole ($T^{all}_{A/B/C}$) and this is due to the immaturity of technologies available for separating the renewed raw materials from unconverted materials, when blended feedstock is processed. The CAPEX of processes treating blended waste ($T^{all}_{A/B/C}$) is higher by approximately 40% in comparison to the corresponding process treating pure. The OPEX for cascaded process is higher as more water is needed to wash out unreacted waste textile from liquid products. Although the yields, CAPEX and OPEX are more desirable for process treating pure materials, the pathway resulting to the highest economic potential is when waste is not sorted and it is processed in a cascaded manner due to the smaller total of processing steps.

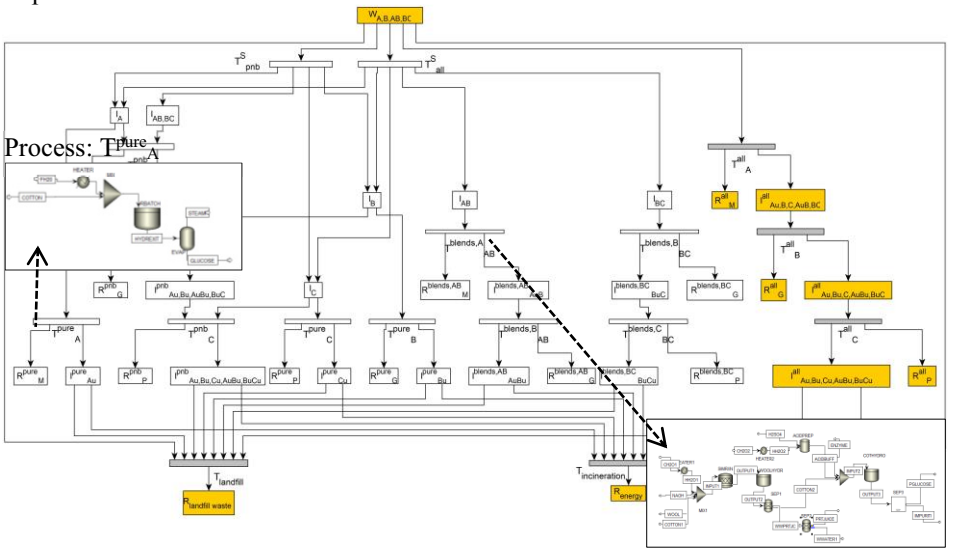


Figure 4: Circular economy process configuration achieving highest economic potential is the cascade process ($T^{all}_{A/B/C}$)

The unreacted product (Au, Bu, Cu, AuBu, BuCu) is given 70% to landfill while the rest 30% is thermally valorised. This choice is due to the relatively low landfill tarrif, $60/ton. In total, the cascaded process has a much lower capital expenses than discrete and discrete-cascaded configuration. It should be noted that sorting processes correspond to 30% of CAPEX when they are selected.

5. Conclusions

Circular economy promotes to interconnect linear production lines between industries so that no major raw materials are lost from our resource system. Textile industry is an ideal example, where circular economy can be introduced, since it has no standard strategy to collect and recycle its post-consumer waste. Evidence of chemical and biochemical chemistries exist for most individual textile polymers, however there is no real world application that combines existing processes to design an integrated circular economy plant other than the work performed in RESYNTEX project. The integration of chemistries to holistically recuperate waste is not straightforward task and there are many configuration options to select from. This work applied mathematical optimisation (MILP) to assess various process configurations aiding in the early stage process synthesis. Data from RESYNTEX project are used in the mathematical formulation. The most integrated and simplest configuration (cascaded process when no prior sorting is performed) is suggested to be the most technically and economically feasible in a specific time horizon. The model can be used to assess the impact of key parameters in the final plant configuration selection and it is planned as future work.

Acknowledgement

This project has received funding from the *European Union's Horizon 2020 research and innovation program* under grant agreement No 641942.

References

European Comission[1], Publications, Factsheets on the circular economy, An ambitious EU circular economy package, 2015. https://ec.europa.eu/commission/publications/ambitious-eu-circular-economy-package_en, last accessed on 29/11/2017.

European Comission[2], Executive Agency for SMEs, 2015. Closing the loop on the ground: 10 EU projects working towards a circular economy, https://ec.europa.eu/easme/en/news/closing-loop-ground-10-eu-projects-working-towards-circular-economy, last accessed on 29/11/2017.

EU Research Project RESYNTEX 2015, A new circular economy concept: from textile waste towards chemical and textile industries feedstock. Project ID: 641942. http://www.resyntex.eu/

P. Gallezot, 2012, Conversion of biomass to selected chemical products, Chemical Society Reviews, 41, 1538-1558.

M. Hower, Greenbiz, 2016. 8 Companies to watch in the circular economy, https://www.greenbiz.com/article/8-companies-watch-circular-economy, last accessed on 29/11/2017.

K. Leonas, 2017, The Use of Recycled Fibers in Fashion and Home Products, In: Textiles and Clothing Sustainability, DOI: 10.1007/978-981-10-2146-6, Springer, 55-77.

A.C. Kokossis, M. Tsakalova, K. Pyrgakis, 2015. Design of integrated biorefineries. Comput. Chem. Eng. 81, 40–56.

D. Paszun, T. Spychaj, 1997. Chemical Recycling of Poly(ethylene terephthalate), Industrial & Engineering Chemistry Research, 36, 1373-1383.

B. Sethi, 2016, Methods of Recycling, Recycling of Polymers: Methods, Characterisation and Application, Wiley-VCH.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50188-9

Total Site Utility Systems Structural Design Considering Electricity Price Fluctuations

Timothy G. Walmsley[a*], Petar S. Varbanov[a], Matthias Philipp[b], Jiří J. Klemeš[a]

[a]*Sustainable Process Integration Laboratory – SPIL, NETME Centre, Faculty of Mechanical Engineering, Brno University of Technology, Brno, Czech Republic*
[b]*Technische Hochschule Ingolstadt, Institute of new Energy Systems, Ingolstadt, Germany*
walmsley@fme.vutbr.cz

Abstract

This study aims to optimise the structural design of industrial central utility systems to while taking full advantage of intra-day electricity spot price fluctuations. Surges renewable electricity generation uptake has amplified the rises and falls between peak and off-peak electricity prices. Industrial Total Sites and Locally Integrated Energy Sectors can take maximum advantage of periods of both low and high electricity prices through appropriate technology investment. Using P-graph, a Utility Systems Planner superstructure is extended to apply a multi-period analysis to optimally select the fuels, energy conversion technologies, and the requires sizes to install for a representative case study. A key difference between the multi-period solution and the single-period solution is an Electric Boiler that operates when the electricity price is low. This results in a Total Annual Cost saving of 7.5 %.

Keywords: Process Integration; Total Site; utility systems; energy planning.

1. Introduction

Sustained growth in renewable electricity generation has changed the dynamics of many electricity markets. Large swings between high-peak and low-off-peak electricity prices resulting from the complex interplay between intermittent renewable energy supply, continuously fluctuating electricity demands, and market structure and policy (Aflaki and Netessine, 2017). To maintain a functional grid, electricity spot prices decrease to encourage greater consumption and increase to discourage. Locally Integrated Energy Sectors – LIES (Perry et al., 2008) and Total Sites (Klemeš et al., 1997), i.e. large industrial sites, can profit from periods of both low and high electricity prices through the right technology investments. For example, with Combined Heat and Power (CHP), a site may increase electricity exports when the price is high and, with heat pumps, a site may increase electricity imports when the price is low. The structural design of the central utility system with built-in degrees of freedom (e.g. spare capacity) holds the key to a system becoming responsive to grid dynamics.

Design and optimisation of Total Site central utility systems in literature for both retrofit and greenfield cases have chiefly focused on the operational optimisation of steam and gas turbine networks with some studies considering various uncertainties. For example, Sun et al. (2017) studied the multi-period operational optimisation of an existing utility system that consists of boilers, gas turbines, steam turbines, condensers, and let-down valves, with natural gas as the fuel, under both time- and probability-based uncertainties.

Zhu et al. (2016) applied Mathematical Programming (MP) to optimise a network of complex steam turbines. Luo et al. (2014) applied a multi-objective optimisation to determine a low-emissions utility system using similar technologies as Sun et al. (2017) without emissions abatement, fuelled by coal and natural gas. In other works, forecasting utility price increases over the entire lifetime of a project was considered by Nemet et al. (2015) as part of Total Site system optimisation. These studies, and other similar works, have focused on only a few fuels and energy conversion technologies in an optimisation with highly constrained structures and no consideration was given for how sites can benefit from a variable electricity spot price.

The aim of this study is to optimise the structural design of a Total Site (or LIES) central utility system, which cogenerates heating, cooling and power, considering electricity spot price fluctuations such that Total Annual Cost (TAC) is minimised. The study extends the work of Walmsley et al. (2018) that introduced a new Utility Systems Planner (USP) tool built as a P-graph superstructure. Using a multi-period approach in P-graph (Heckl et al., 2015), the USP is transformed to model electricity price fluctuations and to solve for the cost-optimal utility system structural design. The USP superstructure includes 5 energy sources, 16 energy conversion technologies, and 9 intermediate utility levels. A case study is presented to illustrate the extended USP tool for a LIES case study.

2. P-graph

P-graph is a combinatorial optimisation framework specifically for Process Network Synthesis (PNS) problems (Friedler et al., 1996), and the basis of P-graph Studio (2017). Underpinned by five fundamental axioms, P-graph has proven effective at reducing computational burden by very quickly identifying the sub-set of feasible structures for optimising PNS problems with high combinatorial complexity. Three other advantages over standard MP software are: (1) the use a graphical user interface for inputting superstructures and displaying results, (2) the locating of optimal and near-optimal solutions, and (3) the software is free. A key limitation of P-graph Studio is it only allows input of Mixed-Integer-Linear (MIL) type problems, where non-linear functions must be handled using a piece-wise linear approach (Ong et al., 2016). Recent applications of P-graph have included Total Site mass, heat and power integration (Ong et al., 2017) and multi-period optimization of sustainable energy systems (Aviso et al., 2017). P-graph is selected as the optimisation tool for this work due to its excellent visualisation of the problem and its solution as well as its inherent capability to reduce computational burden that helps solve the complex multi-period superstructure.

3. Utility Systems Planner - USP

Figure 1 presents the USP superstructure from P-graph Studio. Energy resources (circles) at the top of the superstructure, from left to right, are Electricity Import from the grid, Biomass, Coal, Natural Gas (NG), and Geothermal heat. Located at the bottom are the industrial and district energy users. Considered energy conversion technologies (rectangles) are: solid fuel and gas-fired boilers, NG Combined Cycle Gas Turbine (CCGT), NG Reciprocating Gas Engine (GE), Thermal Vapour Recompression (TVR), Mechanical Vapour Recompression (MVR), Organic Rankine Cycle (ORC), Heat Pump (HP), Absorption Chiller (ACH), Electric Chiller (EC), Steam Turbine, and Let-down Valves (LV). The intermediate energy levels (circles) are Electricity (EL), four steam levels (VHPS, HPS, MPS, LPS), High-Pressure Hot Water (HPHW), Hot Water (HW), Cooling Water (CW), and Chilled Water (CHW). Dark red lines represent electricity

flows, other shades of red and orange lines represent different hot utility levels, blue lines are cold utility, and light blue lines are condensate and water flows. The USP model requires specifying (i) user energy demands, (ii) energy conversion performances ratios, (iii) energy prices, (iv) capital and operating costs, and (vi) GHG emissions factors.

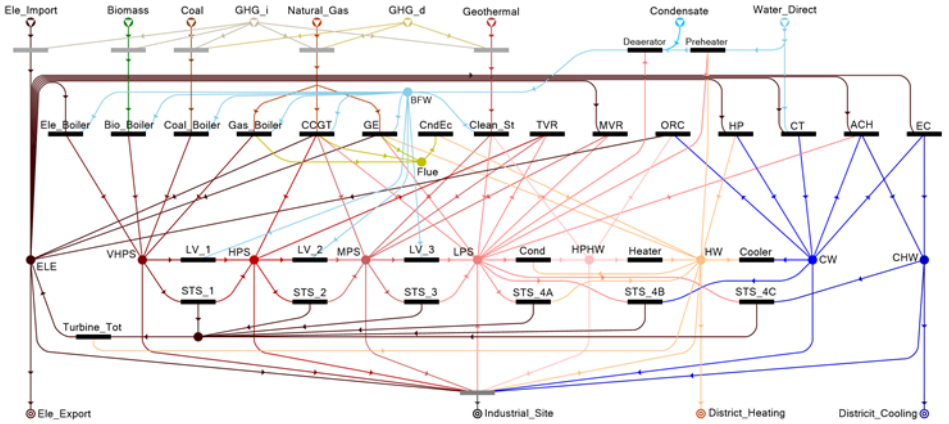


Figure 1: Utility Systems Planner tools – initial superstructure.

4. Method

The optimisation method in this study involved four stages: (1) Analyse historical electricity spot prices. This step included looking at the daily cyclic profile and the ordered value plots for the electricity prices. (2) Determine electricity prices for the multi-period analysis. Multi-period analysis is applied to approximate electricity price variations based on the ordered value plot of the electricity price. This step calculates the average price for each period. (3) Convert the USP superstructure into a multi-period equivalent to include electricity prices variations. This step uses the method of Heckl et al. (2015) to duplicate the initial superstructure (e.g. Figure 1) for n periods so that different electricity prices may be inputted for each period. Attention must be given to the multi-period superstructure to ensure capital cost is accounted appropriately. (4) Solve the P-graph superstructure for cost-optimal and near-optimal feasible structures. The objective function of the optimisation in P-graph Studio is to minimise TAC in Eq(1).

$$\mathrm{TAC} = \mathrm{UC} + \mathrm{CC} + \mathrm{OM} \tag{1}$$

Where UC is the utility cost, CC is the annualised capital cost, and OM is the operating and maintenance costs.

5. Locally Integrated Energy Sector - LIES Case Study

The considered LIES encompasses a representative industrial chemical processing site and the heating and cooling needs of the surrounding district. Figure 2 presents the Site Utility Grand Composite Curves for the industrial site and the entire LIES. The industrial site has a Utility System Pinch between the MPS and LPS levels, which extends to between the MPS and CW levels once the district heating/cooling demands are included for the LIES. Table 1 presents the prices for energy use while fresh water is 0.2 €/t and on-site GHG emissions are 5 €/t. For details on capital costs and performance ratios of each energy conversion process refer to Walmsley et al. (2018).

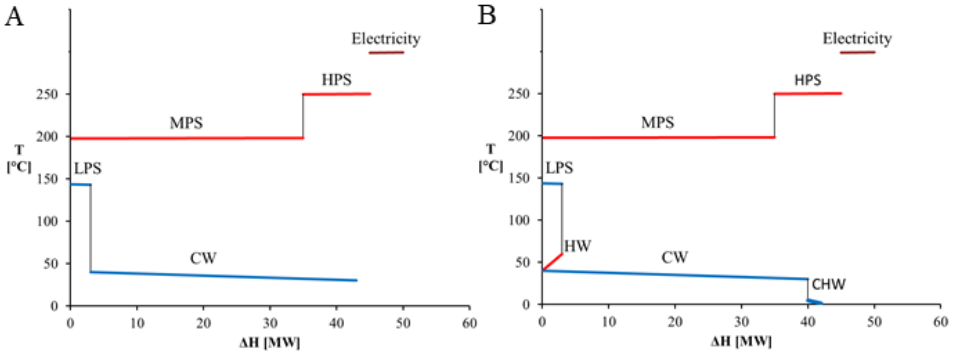


Figure 2: Site Utility Grand Composite Curves for the industrial site (A) and LIES (B).

Table 1: Industrial energy, GHG emissions and water prices and GHG emissions factors (EF).

	Electricity	Biomass	Coal	Natural Gas	Geothermal
Consumption (€/GJ)	22.3*	10.0	5.0	7.0	3.0
Generation (€/GJ)	16.7*	-	-	-	-
GHG emissions (kg/GJ)	-	-	91.4	55.2	1.3

*Average electricity prices (80.1 €/MWh), minus lines charge for sale (20 €/MWh)

The historical price fluctuations in wholesale electricity index for the EU-28 are presented in Figure 3. The median hourly electricity price trend for a daily cycle shows a low trough (4 a.m.) during the night, a small peak in the morning (8 a.m.) followed by a downturn in the afternoon before rising to a high peak in the evening (7 p.m.). The absolute electricity price is heavily influenced by the season as indicated by the individual trends for the different days. Ordering all electricity price provides a sense of the frequency that each price point arises. A three-period approximation of the ordered electricity price profile can be formulated (Figure 3B). There is a trade-off between the accuracy of the approximation (i.e. more periods) and the computational burden and is an area that may attract attention in future work.

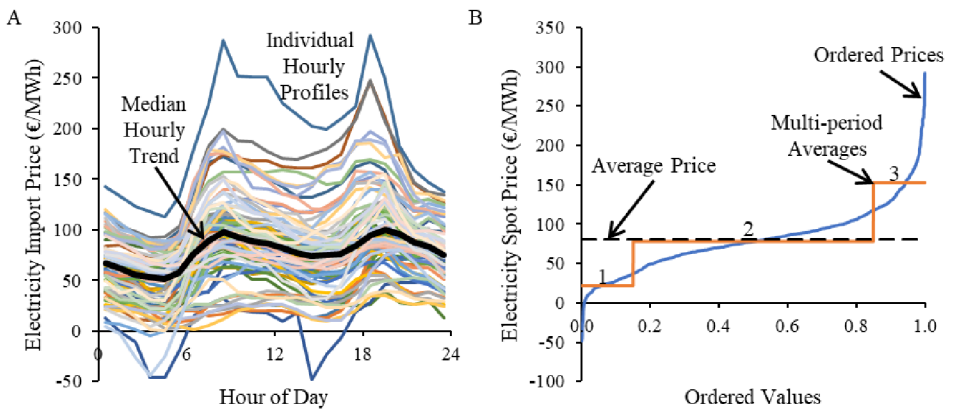


Figure 3: Electricity price with daily profiles for a sample of 84 d in 2017 (A) and by ordered values with single average and multi-period averages (B).

Using the average prices of the three periods from Figure 3B, the USP superstructure may be converted into a multi-period problem. This means the superstructure is replicated

three times. Nodes that represent the physical unit operation and its capital cost connect to each of the three superstructures. This is necessary to ensure that when a unit operation is required in any period, the capital cost is precisely accounted.

Solving the new multi-period superstructure to minimise TAC leads to the structural design as shown in Figure 4. Compared to the base case without multi-period analysis, the TAC savings is 7.5 % but a 90 % increase in GHG emissions due to the export of electricity, 29.3 MW_e. The final design includes an electric boiler, a gas boiler and a gas engine. When the electricity price is low, the electric boiler is switched on and the gas engine off. As the electricity price rises from period 1 to 2, it becomes uneconomic to use electricity as the heat source rather the site becomes a net electricity exporter. During the two higher electricity price periods, the gas engine runs at peak capacity to maximise electricity exports. The gas boiler has a steady load through all three periods of about 43 MW_f. The optimal solution suggests that an intensified and integrated gas boiler and engine with electric water heating as a single unit is a key technological solution.

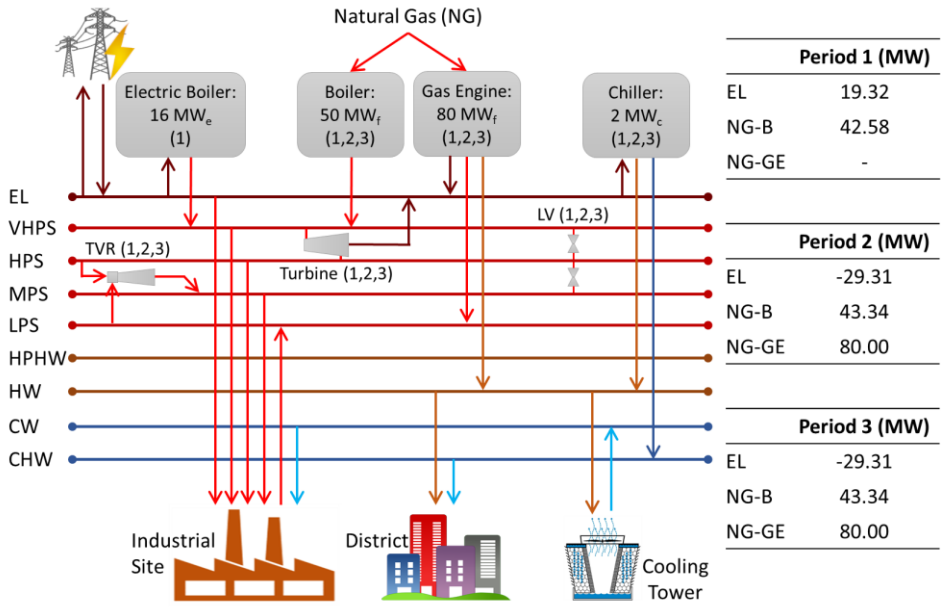


Figure 4: Utility systems structural design with key fuel and equipment for minimum TAC.

6. Conclusions

The Utility Systems Planner (USP) model, built using P-graph, incorporated a comprehensive range of possible fuels, resources, energy conversion operations, and auxiliary equipment. This superstructure has been extended in this study to encompass a multi-period analysis. This step forward enabled consideration for considering peak and off-peak electricity prices as part of the utility systems design phase. A comparison of the best utility systems structural designs showed the addition of an Electric Boiler (EB). The installation of an EB unlocked a Total Annual Cost savings of 7.5 % but a 90 % increase in GHG emissions due to an electricity export of 29.3 MW. Future work will focus on applying the USP to additional case studies as well as greater emphasis on environmental concerns beyond GHG emissions.

Acknowledgement

This research has been supported by the EU project "Sustainable Process Integration Laboratory – SPIL", project No. CZ.02.1.01/0.0/0.0/15_003/0000456 funded by EU "CZ Operational Programme Research, Development and Education", Priority 1: Strengthening capacity for quality research, in a collaboration with the Technische Hochschule Ingolstadt, Germany.

References

S. Aflaki, S. Netessine, 2017, Strategic Investment in Renewable Energy Sources: The Effect of Supply Intermittency, Manufacturing & Service Operations Management, 19(3), 489–507.

K.B. Aviso, J.-Y. Lee, J.C. Dulatre, V.R. Madria, J. Okusa, R.R. Tan, 2017, A P-graph model for multi-period optimization of sustainable energy systems, Journal of Cleaner Production, 161(Supplement C), 1338–1351.

F. Friedler, J.B. Varga, E. Fehér, L.T. Fan, 1996, Combinatorically Accelerated Branch-and-Bound method for solving the MIP model of Process Network Synthesis, Chapter In: State of the Art in Global Optimization, Springer, Boston, USA, 609–626.

I. Heckl, L. Halász, A. Szlama, H. Cabezas, F. Friedler, 2015, Process synthesis involving multi-period operations by the P-graph framework, Computers & Chemical Engineering, 83(Supplement C), 157–164.

J.J. Klemeš, V.R. Dhole, K. Raissi, S.J. Perry, L. Puigjaner, 1997, Targeting and design methodology for reduction of fuel, power and CO_2 on Total Sites, Applied Thermal Engineering, 17(8–10), 993–1003.

X. Luo, J. Hu, J. Zhao, B. Zhang, Y. Chen, S. Mo, 2014, Multi-objective optimization for the design and synthesis of utility systems with emission abatement technology concerns, Applied Energy, 136(Supplement C), 1110–1131.

A. Nemet, J.J. Klemeš, Z. Kravanja, 2015, Designing a Total Site for an entire lifetime under fluctuating utility prices, Computers & Chemical Engineering, 72(Supplement C), 159–182.

B.H.Y. Ong, T.G. Walmsley, M.J. Atkins, M.R.W. Walmsley, 2017, Total site mass, heat and power integration using process integration and process graph, Journal of Cleaner Production, 167(Supplement C), 32–43.

B.H.Y. Ong, T.G. Walmsley, M.J. Atkins, M.R.W. Walmsley, J.R. Neale, 2016, Approximation of non-linear cost functions in P-graph structures, Chemical Engineering Transactions, 52, 1093–1098.

S. Perry, J.J. Klemeš, I. Bulatov, 2008, Integrating waste and renewable energy to reduce the carbon footprint of locally integrated energy sectors, Energy, 33(10), 1489–1497.

P-graph Studio, 2017, P-graph Studio, P-graph <p-graph.com> accessed 07.12.2017.

L. Sun, L. Gai, R. Smith, 2017, Site utility system optimization with operation adjustment under uncertainty, Applied Energy, 186(Part 3), 450–456.

T.G. Walmsley, X. Jia, M. Philipp, A. Nemet, P.Y. Liew, J.J. Klemeš, P.S. Varbanov, 2018, Total Site utility system structural design using P-graph, Chemical Engineering Transactions, 63, 31–36.

Q. Zhu, X. Luo, B. Zhang, Y. Chen, S. Mo, 2016, Mathematical modeling, validation, and operation optimization of an industrial complex steam turbine network-methodology and application, Energy, 97(Supplement C), 191–213.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50189-0

Sustainable and Innovative Solutions through an Integrated Systematic Framework

Nipun Garg[a*], Georgios M. Kontogeorgis[a], John M. Woodley[a], Rafiqul Gani[a,b]

[a]*Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads, DK-2800 Kgs. Lyngby, Denmark*
[b]*PSE for SPEED, Skyttemosen 6, DK-3450 Allerod, Denmark*
nipgar@kt.dtu.dk

Abstract

Chemical and related industries need to find more sustainable and innovative processes, as due to rising environmental concerns and strict governmental policies it is no longer sufficient for the synthesized-design processes to be just profitable and operable. Process intensification as an integrated part of process synthesis design methods has the potential to determine more sustainable and innovative solutions. In this paper, an integrated and systematic framework based on a generic 3-stage approach to perform sustainable process synthesis, design and intensification is presented. The database and algorithms in the framework have been extended to be applicable over a wide range of applications and have been validated through a case study for the production of bio succinic acid. In this case study, a novel superstructure of alternatives has been generated to identify the optimal processing route followed by the application of an extended phenomena based synthesis-intensification step to generate non-trade off, more sustainable and intensified solutions for the production of bio succinic acid.

Keywords: Generic methodology, Process Intensification, Sustainable solutions.

1. Introduction

Process Intensification (PI) can be defined as improvements of a process at unit operational, functional and/or phenomena levels that can be obtained by integration of unit operations, integration of functions and phenomena's or targeted enhancement of the phenomena for a target operation. Earlier process synthesis methodologies operated at the unit operations (Unit-Ops) level (Jaksland et al. 1995) and functional (task) level (Jaksland et al. 1995, Siirola 1996) which has further been extended to incorporate intensification at phenomena level (Babi et al., 2015). Generation of novel and more sustainable designs, i.e., going beyond the existing intensified unit-ops one must target lower level of aggregation, i.e. phenomenon level (Papalexandri et al. 1996, Lutze et al., 2013). At this level one can investigate underlying driving forces associated with unit-ops to create new versions of them.

In this framework, the process flowsheets options are synthesized and/or evaluated that corresponds to acceptable design based on a given set of process specifications. This defined set of performance parameters is then referred to as the base case for intensification and through evaluation of the base case design, process "hot-spots" are identified and transformed into targets for improvement. Then at phenomena level, the phenomena involved in performing specific tasks are identified and while increasing search space these phenomena are recombined using a systematic and generic approach to find novel and innovative intensified alternatives that match the improvement targets.

These alternatives match the same specifications as the base case, but because they also satisfy the new targets for improvement, are by definition, non-trade off and more sustainable than the base case. That is, more flowsheet options that improve the base case and correspond to even better values of performance parameters are generated.

In this work, the phenomena based synthesis approach is extended by applying it to bio processes. It includes the expansion of database and algorithms to identify phenomena building blocks (PBB's) and generation of basic structures translating to unit-ops. In its application, new synthesis tasks have also been introduced, which is a novel approach for selection of different fermentation cell strains that are formulated in the form of a superstructure to identify the optimal along with downstream separation process.

2. Integrated framework for Process Synthesis, Design and Intensification

An integrated, systematic and computer aided framework for process synthesis, design and intensification is shown in figure 1. This computer aided framework is based upon a generic 3 stage approach to sustainable process design (Babi et al., 2015).

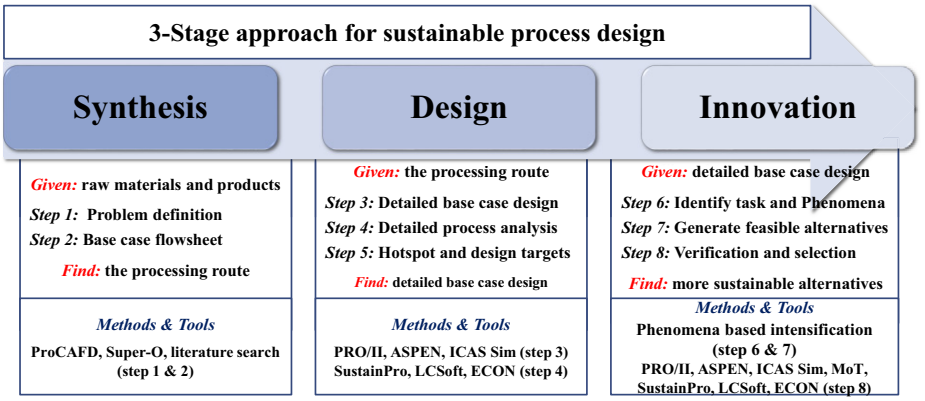


Figure 1. Overview of integrated framework for Process Synthesis, Design and Intensification

As shown in figure 1, the first stage of the framework involves process synthesis where an optimal process flowsheet is identified from numerous alternatives. In step 1 the synthesis problem is defined while in step 2, the optimal process flowsheet is identified. This can be done by using computer aided flowsheet design (ProCAFD) (Tula et al., 2015), performing the superstructure optimization using Super-O tool (Bertran et al., 2017) or by doing literature search. In stage 2, detailed simulation is performed to extract the detailed mass and energy balance data (step 3) and perform process analysis (step 4) in terms of sustainability, life cycle assessment and economics to identify the process "hot-spots" and based on these, set the targets for improvement (step 5) to be achieved in stage 3. In stage 3, a phenomenon based process synthesis intensification method (Babi et al., 2015) is applied to generate flowsheet alternatives that match the specified set of design targets and therefore determines more sustainable alternatives compared to the base case. In this approach the tasks performed by the unit operations involved in the optimal route are identified and from which further associated phenomena are also identified (step 6). These phenomena are combined using combination rules to generate new and/or innovative, intensified and more sustainable flowsheet alternatives (step 7). The final designs are then verified and compared with base case (step 8) through a set of pre-defined performance parameters or criteria.

3. Case study

The generic framework is applied to the production of bio succinic acid. Succinic acid is one of the most widely used platform chemical with broad range of industrial applications in food, pharma and chemical industries (Cok et al., 2014). Thus, the main objective of this case study is to identify more sustainable, intensified and non-trade off solutions along with testing of current methods and identify limitations for improvement in production of succinic acid.

3.1. Stage 1: Synthesis

Step 1 Synthesis problem definition: The synthesis problem as defined is to find the optimal route for production of bio succinic acid with a purity of at least 99.9%. The starting raw materials considered are Glucose, Glycerol, Maltose and Sucrose along with CO_2. Using these raw materials different reaction paths with different bacterial and yeast strains are identified from literature along with required reaction data.

Step 2 Generation of base case flowsheet: The base case flowsheet is generated using the superstructure optimization approach. Here, a novel superstructure with 8 processing steps and 33 processing intervals excluding raw material and product steps is generated that includes new synthesis tasks such as inclusion of different cell strains to produce bio succinic acid. Overall a superstructure describing the network of configurations for different processing routes, consists of 10 different fermentation models followed by 7 recovery and purification processing steps, where data was collected through available information on reaction as well as succinic acid production related technologies reported in literature. Then the superstructure based optimization problem is set up in Super-O, where the MILP problem is solved to determine the optimal processing route.

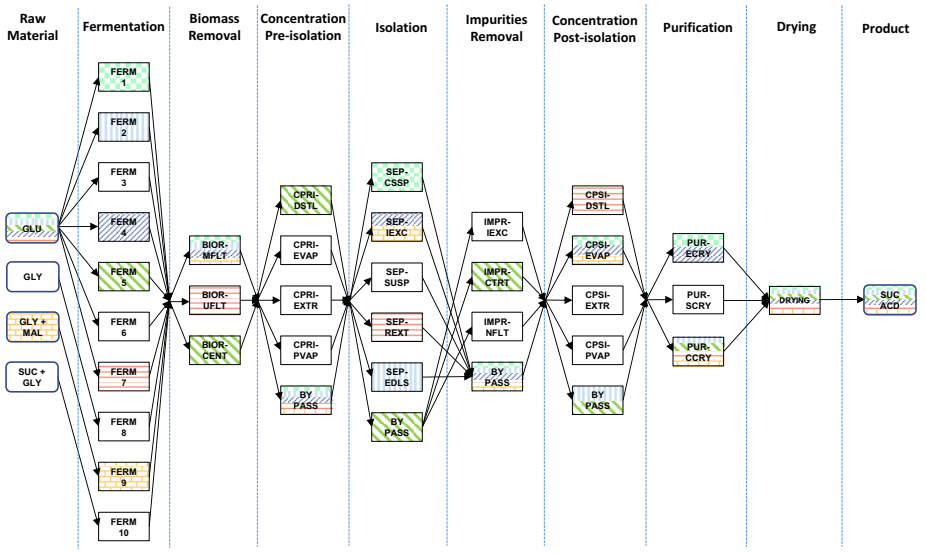


Figure 2. Superstructure (PSIN) of alternatives for succinic acid case study

The optimal process flowsheet alternative obtained is a novel processing route (intervals with green slanted lines) while five other existing processing routes (rest of the notations) reported in literature for example by Schroder et al., 2014 are also identified in the superstructure as shown in figure 2. Thus, potentially from this superstructure

2604 novel and existing feasible process flowsheets alternatives can be generated. The optimal process synthesis flowsheet (figure 3) consists of a fermenter, centrifuge for broth clarification, distillation to concentrate the broth and remove unwanted organic acids and alcohols, followed by carbon treatment for removal of soluble solids causing colour and then finally crystallization and drying to get pure crystals of succinic acid.

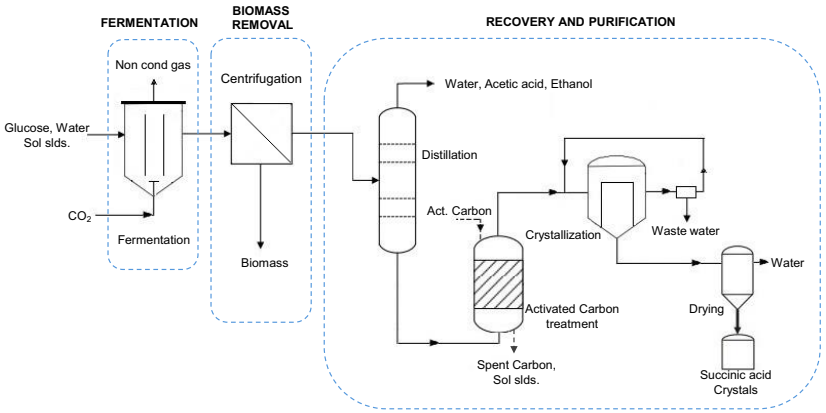


Figure 3. Novel and optimal processing route (Base case flowsheet)

3.2 Stage 2: Detailed design and analysis

Step 3 Detailed base case design: The detailed design for the optimal process flowsheet is performed using PRO/II simulator to extract the mass and energy balance data.

Step 4 Process analysis: The process analysis is performed using in-house tools SustainPro (Carvalho *et al.* 2013), LCSoft (Kalakul et al., 2014) and ECON (Saengwirum 2011). The sustainability analysis indicates loss of product and raw material in open path (in-out streams) containing crystallizer. The economic analysis shows high utility costs related to unit-op in open path i.e. reboiler of distillation column. As expected, the LCA analysis confirms with the high carbon footprint for reboiler of the distillation column.

Step 5 Identify process hotspots and design targets: Based on above analysis, process hot-spots are identified that are translated into design targets. The identified process hot-spots are high energy consumption/demand for succinic acid recovery and loss of product in above mentioned open paths. The corresponding translated design targets are reduction of energy demand, increase in product recovery, reduction in number of unit operations and improvement in sustainability and LCA factors.

3.3 Stage 3: Innovation (Phenomena based intensification)

Step 6 Identify desirable tasks and phenomena: The base case flowsheet is represented in terms of tasks and then in terms of phenomena. The identified PBBs that make up the initial search space of PBBs are R (reaction), M (mixing), 2phM (two phase mixing), PC(VL), PT(VL), PS(VL), PC(LS), PT(LS), PS(LS), H (heating) and C (cooling) where, PC is phase contact; PT is phase transfer and PS is phase separation; while V, L and S represents vapor, liquid and solid phases. Here, the tool library to identify the PBB's has been extended in order to incorporate wide range of unit-ops. The pure component and mixture property analysis is performed using ICAS (Gani et al., 1997), which identified one azeotrope between water and ethanol. As in this case study, the

main objective is to get pure succinic acid, the separation of water and ethanol is not considered in further steps. Additional phenomena are identified from the database based on process hotspots to increase the search space. The phenomena's identified are PT(PVL), PT(VV), PS(VV), PC(LL) and PS(LL). Thus, the total list contain 20 PBB's. The total number of possible combinations of these PBB's to form simultaneous phenomena building blocks (SPB's) are 519252. However, all combinations are not possible. Thus, using connectivity rules, 244 SPB's are found to be feasible.

Step 7 Generation of feasible flowsheet alternatives: Intensified flowsheet alternatives are generated by combining the SPB's to form basic structures. These basic structures are formed in a way that they satisfy the identified reaction and separation tasks. In alternative 1, starting with the first task of reaction, the second task, cell removal or clarification of broth could be combined to obtain a new basic structure of phenomena's that translates into a membrane bioreactor (figure 4). In this unit operation, the fermentation broth is clarified i.e. the reaction product is removed continuously and the cell culture remains in the membrane bioreactor leading to increase in cell concentration and high product yield, also observed by Wang et al. (2014).

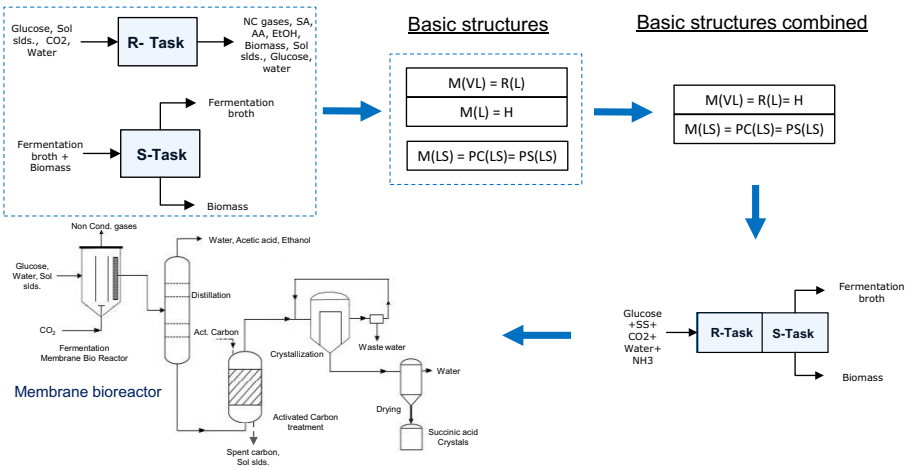


Figure 4. Intensified flowsheet alternative 1 (Identified task, combination of basic structures)

In alternative 2, the integration of basic structures for the last two separation tasks is considered. This is possible as both the separation tasks involve SPB involving separation of crystalline product from the liquid phase. Thus, a basic structure having phenomena PT(LS) and PS(LL) is selected. Additional PBB like cooling, heating and mixing can also be added to the selected SPB's. Thus, in flowsheet alternative 2 the combined basic structure is translated to membrane crystallizer to obtain desired purity crystals of succinic acid. Alternative 3 is a combination of alternative 1 and 2 consisting both membrane bio reactor and membrane crystallizer.

Step 8 Verification and selection: The generation of more sustainable and non-trade off process designs by matching a set of design targets is illustrated via radar plot in figure 5, where, the ratios of different performance criteria with respect to the base-case multiplied by 100 have been plotted (for profit, the inverse has been taken). The base-case design is at the boundary while the more sustainable alternatives are all within the boundary, indicating that these intensified alternatives are non-trade off alternatives. Note that for each alternative the required purity of product has been achieved, no

solvents have been used and the number of unit operations have been reduced from 6 (the base case) to 4 (alternative 3). Out of the intensified alternatives generated, alternative 3 give the best results.

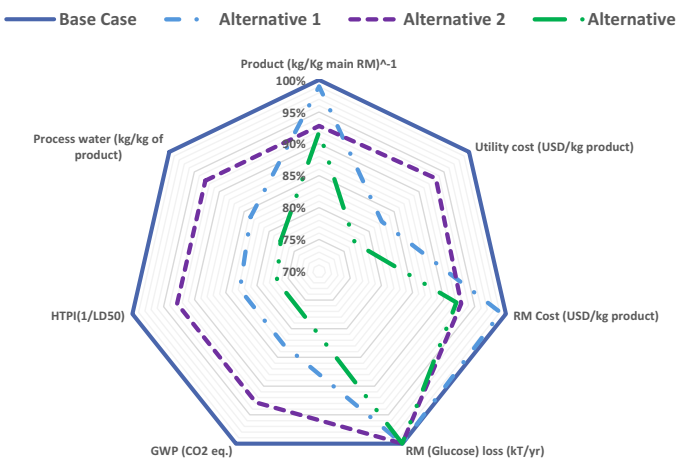


Figure 5. Economic, sustainability and LCA improvements relative to the base case design (HTPI: Human toxicity potential by ingestion, GWP: Global warming potential)

4. Conclusions

An integrated and systematic framework for process synthesis intensification has been extended and further applied to bio processes (bio succinic acid). The application of the extended framework gave a novel optimal process flowsheet along with more than 2600 other alternatives at the synthesis stage while 3 intensified and more sustainable alternatives were generated at the innovation stage for production of bio succinic acid. Current and future work will further extend the phenomena based approach and its implementation in a computer aided tool.

References

C.A. Jaksland, R. Gani, K.M. Lien, 1995, Chem Eng Sci., 50, 511-530.
J. J. Siirola, 1996. Comput. Chem. Eng., Supplement 2, 20, S1637-S1643.
D. K. Babi, J. Holtbruegge, P. Lutze, A. Gorak, J. M. Woodley, R. Gani, 2015. Comput. Chem. Eng., 81, 218-244.
K.P. Papalexandri, E.N. Pistikopoulos, 1996, AIChE J, 42, 4, 1010-1032.
P. Lutze, D. K. Babi, J. M. Woodley, R. Gani, 2013, Ind. Eng. Chem. Res., 52, 7127-7144.
A. K. Tula, M. R. Eden & R. Gani, Comp. & Chem. Eng., 81(2015), 245-259.
M. O. Bertran, R. Frauzem, A. S. Arcilla, L. Zhang, J. M. Woodley, R. Gani, 2017, Comput. Chem. Eng., 106, 892-910.
H. Schroder, S. Haefner, G. Abendroth, R. Hollmann, A. Raddatz, H. Ernst, H. Gurski, 2014, U.S. Patent 8,673,598 B2.
B. Cok, I. Tsiropoulos, A. L.Roes, M. K. Patel, 2014, Biofuel Bioprod. Bioref., 8, 16–29.
A. Carvalho, H. A. Matos, R. Gani, 2013, Comput. Chem. Eng., 50, 8-27.
S. Kalakul, P. Malakul, K. Siemanond, R. Gani, 2014, Journal of Cleaner Production, 71, 98-109.
P. Saengwirun, 2011, MSc thesis, Chulalongkorn University.
R. Gani, G. Hytoft, C. Jaksland, A. K. Jensen, 1997, Comput. Chem. Eng., 21, 1135-1146.
C. Wang, W. Ming, D. Yan, C. Zhang, M. Yang, Y. Liu,Y. Zhang, B. Guo, Y. Wan, J. Xing, 2014, Bioresource technology, 156, 6-13.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50190-7

Uncertainty aware integration of planning, scheduling and multi-parametric control

Vassilis M. Charitopoulos, Adrian M. Aguirre, Lazaros G. Papageorgiou, Vivek Dua
Department of Chemical Engineering, Centre for Process Systems Engineering, University College London, Torrington Place, London WC1E 7JE, United Kingdom.
v.dua@ucl.ac.uk

Abstract
In this work we investigate the integrated planning, scheduling and control (iPSC) of process systems under uncertain conditions throughout the three levels of decision making. The planning problem is explored in a rolling horizon fashion coupled with demand forecasts. Proactive and reactive approaches are employed to handle the effect of stochastic variations. Depending on the nature of the uncertain parameters robust optimisation and chance constrained programming are employed. For the closed-loop implementation of the control, novel multi-parametric controllers are designed. The proposed framework is tested on the iPSC of a polymerisation process. Finally, Monte Carlo simulations are conducted to highlight the benefits of the "uncertainty-aware" solutions when compared to the deterministic ones.

Keywords: enterprise wide optimisation, uncertainty, robust optimisation, multi-parametric programming, chance constrained programming.

1. Introduction
Contemporary process industries are part of a progressively complex global market network and the need to account for efficient and integrated solutions has become increasingly mandatory. Integrating control with operations has gained considerable amount of research interest (Chu and You, 2013; Zhuge and Ierapetritou, 2012). So far, most of the research works have dealt with integrating cyclic scheduling and control with little work done on the iPSC (Dias and Ierapetritou, 2016). Guittierez-Limon et al. (2014) studied the iPSC using a time-slot based formulation and solved a monolithic MINLP problem, while Shi et al. (2015) studied the iPSC based on a flexible recipe method and formulated the problem as an MILP for which bilevel decomposition was employed. Charitopoulos et al. (2017a) used a TSP-based model for the integration of planning and scheduling, with the control decisions modelled through the use of offline calculations and linear metamodels. Later, Charitopoulos et al. (2017b) investigated the use of explicit controllers (mp-MPC) for the efficient closed-loop iPSC. Even though the integrated problem can lead to improved operations, because of the synergies between the different levels of decision

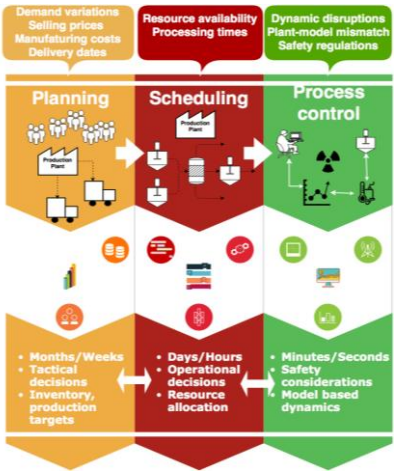


Figure 1. Integrated planning, scheduling and control problem under uncertainty

making, uncertainty is ubiquitous and its impact on optimal operations can be tremendous. To the best of our knowledge, no previous research work has addressed the iPSC under uncertainty in all three levels of decision making (Dias and Ierapetritou, 2016). As shown in Figure 1, the integrated nature of the problem results in diverse kinds of uncertainties that call for different policies. To this end, in the present work we employ reactive and proactive approaches by means of multi-parametric programming (mp-P) at the level of control (Charitopoulos et al., 2017b), robust optimisation, chance constrained programming and rolling horizon with demand forecasts for the planning and scheduling (Li and Ierapetritou, 2009). The uncertainty-aware integrated framework is applied to the iPSC of a polymerisation process and its benefits are highlighted via Monte-Carlo analysis.

2. Mathematical modelling of iPSC

In this section a brief overview of the mathematical modelling of the iPSC with emphasis on the uncertain entities under consideration is given; for a detailed exposition the reader is referred to Charitopoulos et al. (2017a). Planning periods (p) are modelled as discrete time points and the scheduling decisions use continuous-time formulation. The overall scope of the integration has as follows:

$$\text{iPSC Problem: } \begin{cases} \max\limits_{\mathbf{x_{pl}},\mathbf{x_{sc}},\mathbf{x_c}} \ \text{Profit} \\ \text{Subject to: } \mathbf{x}_{\text{pl}}, \mathbf{x}_{\text{sc}}, \mathbf{x}_{\text{c}} \in \Omega^{\mathbf{p}} \times \Omega^{\mathbf{s}} \times \Omega^{\mathbf{c}} \end{cases}$$

where $\mathbf{x_{pl}}, \mathbf{x_{sc}}, \mathbf{x_c}$ stand for the planning, scheduling and control decisions respectively while $\Omega^{\mathbf{p}}, \Omega^{\mathbf{s}}, \Omega^{\mathbf{c}}$ their related feasible sets. The production output (Q_{ip}) of each product (i) per planning period is calculated by eq. (1) and is a function of the production rate (r_{ip}), which is considered to be fluctuating, and the production time (T_{ip}). The sales (S_{ip}) that satisfy the demand for each product (D_{ip}) are calculated by eq. (2) along with the level of backlogged demand in each period (B_{ip}).

$$Q_{ip} = r_{ip} T_{ip} \ \forall i, p \qquad \text{and} \qquad S_{ip} = D_{ip} + B_{i,p-1} - B_{ip} \ \ \forall i, p \qquad (1)\text{-}(2)$$

For the closed loop case a rescheduling framework has been proposed in our previous work (Charitopoulos et al., 2017b) where variations in transition times were considered from a systems dynamics' perspective. An open question in the literature, that has yet to be addressed, is the systematic treatment of uncertainty that is present simultaneously across the three different levels of decision making and this constitutes the main contribution of the present work.

3. A hybrid framework for uncertainty-aware iPSC

3.1. Rolling horizon for planning decisions

A rolling horizon (RH) approach is proposed to deal with planning decisions under uncertainty. Firstly, an initial demand forecast (d_{ip}) which is known in advance for the whole planning horizon is considered. The uncertainty on the demand forecast is modelled as a uniform distribution between minimum and maximum values ($\pm \delta_{ip}$) of the demand forecast for the current time period (Liu et al, 2009). This demand uncertainty is revealed only at the beginning of the current time period p^* allowing for rescheduling in order to revise the schedule based on the new state of the system. Then, the new demand of the current time period is calculated by the eq. (3) while for the subsequent time periods the demand forecast is considered.

$$d_{ip} = \text{Uniform}\left[\left(1-\delta_{ip}\right)\cdot d_{ip}, \left(1+\delta_{ip}\right)\cdot d_{ip}\right] \quad \forall i, p = p^* \tag{3}$$

For its implementation, a control horizon (L^{CH}) should be defined. The control horizon determines implicitly the scheduling horizon (H) to be solved. Thus, starting from the first time period $p^* = 1$ our control horizon is solved iteratively considering a fixed step size until the end of the planning horizon $p^* = PH$. At each step, an MILP is solved with relaxed integer variables for the time periods outside of the control horizon. At each iteration, all the information and decisions taken of the current time period are fixed and the scheduling horizon is updated as $H = [p^*, p^* + L^{CH} - 1]$. At each run, new information about the future demand is revealed and the optimal solution of the planning problem is re-computed.

3.2. Proactive scheduling formulations

In order to secure smooth operations on the level of scheduling, a hybrid chance constrained (CCP)/robust optimisation (RO) approach is employed. Within the context of iPSC the production rates are related to the performance of the dynamic system and in order to avoid excessive need for rescheduling, chance constrained programming is employed (Li et al., 2006). Assuming normal distribution of $\sigma_{r_{ip}}$ standard deviation, a mean value of the production rate $\hat{r}_{ip}$ and a desired reliability level α, the equivalent deterministic chance constraint of eq. (1) is derived as follows:

$$\Pr\left(Q_{ip} \le r_{ip} T_{ip}\right) \ge \alpha \leftrightarrow Q_{ip} \le \left(\hat{r}_{ip} - \Phi^{-1}(\alpha)\sigma_{r_{ip}}\right) T_{ip} \tag{4)-(5}$$

where $\Phi^{-1}(\alpha)$ stands for the inverse cumulative distribution for reliability α. Next, the budget parameter approach of Bertsimas and Thiele (2006) was chosen in order to account for deviation in products demand. The nominal demand ($\overline{d_{ip}}$) is considered to vary as $[\overline{d_{ip}} - \check{d}_{ip}, \overline{d_{ip}} + \check{d}_{ip}]$ and its scaled deviation is defined as follows: $z_{ip} = \frac{d_{ip} - \overline{d_{ip}}}{\check{d}_{ip}}$

In brief, eq. (2) is rewritten as eq. (6).

$$S_{ip} \le \max_{w \in U}\{d_{ip}\} \quad + B_{i,p-1} - B_{ip} \tag{6}$$

The uncertainty set (U) is defined by eq. (7)-(9) so as to avoid a cumulative overestimation of the uncertain demand using the budget parameter (Γ_p).

$$D_{ip} = \check{d}_{ip} z_{ip}, \forall i, p \text{ and } 0 \le z_{ip} \le 1 \ \forall i, p \text{ and } \sum_{k=1}^{p} z_{ip} \le \Gamma_p \quad \forall, i, p \tag{7)-(9}$$

Next, the robust counterpart (RC) is derived based on the related dual problem and eq. (6) is substituted by eq. (10)-(11) using the related dual variables (q_{ip},y_{ipk}).

$$S_{ip} \le \overline{d_{ip}} + \Gamma_p q_{ip} + \sum_{k=1}^{p} y_{ipk} + B_{ci,p-1} - B_{cip} \tag{10}$$

$$q_{ip} + y_{ipk} \ge \check{d}_{ip} \ \forall i, p, k \le p \tag{11}$$

3.3. The hybrid framework for uncertainty-aware iPSC

Overall, the hybrid framework consists of proactive policy at the scheduling level while the reactive policy is considered at the levels of planning and control and is shown in Figure 2. The rolling horizon solution of the problem allows for consideration of updated demand forecasts and in order to prevent low levels of customer satisfaction, the variations in the demand are addressed proactively through the RO approach. Uncertain production rates that affect the scheduling relevant decisions are handled through individual chance constraints and based on the profitability of each product, in order to avoid over-conservative solutions, a different level of reliability is specified. For the control layer novel mp-MPC controllers are employed. The novelty is the

consideration of kinetic/quality data as uncertain parameters that results in left-handside uncertainty along with the different set-points that are provided by the level of scheduling (Charitopoulos et al., 2017b). Overall the integrated hybrid framework is an MILP problem with the control decisions made in an outer loop resulting in an additional rescheduling mechanism that reactively handles variations in transition times as shown in Figure 2.

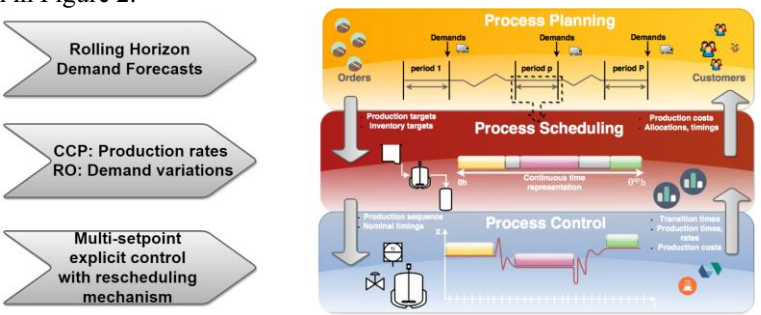


Figure 2. Conceptual graph of the hybrid framework for the uncertain iPSC

4. Case study: MMA polymerisation process

The proposed approach was tested on the iPSC of a continuous MMA polymerisation process. Related data about the economics and dynamics of this case study can be found in Chu and You (2013). In this instance, 16 products and a planning horizon of 6 weeks are examined.

4.1. Design of the hybrid framework for uncertainty-aware iPSC

The tuning parameters of the integrated framework are as follows: $L_{CH} = 2, a = 70 - 95\%, |\sigma_{r_i}| = 0.2 r_i, \check{d}_{ip} = [0.05\ 0.1\ 0.2\ 0.15\ 0.15\ 0.2]\overline{d_{ip}}, \Gamma_p = [0\ 1\ 2\ 2\ 3\ 3\ ]$. At the level of control the uncertain parameters are: (i) the states of the system at each sampling instance, (ii) the desired setpoints and (iii) the inlet initiator concentration ($C_{l_{in}}$). The model of the system is given by eq. (12)-(15) while a total of 9 uncertain parameters were considered during the design of the mp-MPC.

$$\dot{C_m} = -(k_p + k_{fm})C_m\Psi + \frac{F(C_{m_{in}} - C_m)}{V} \tag{12}$$

$$\dot{C_l} = -k_l C_l + \frac{F_l C_{l_{in}} - F C_l}{V} \tag{13}$$

$$\dot{D_0} = (0.5k_{Tc} + k_{Td})C_l\Psi^2 + k_{fm}C_m\Psi - \frac{FD_0}{V} \tag{14}$$

$$\dot{D_I} = M_m(k_p + k_{fm})C_m\Psi - \frac{FD_I}{V} \tag{15}$$

First the mp-MPC is designed (Charitopoulos et al., 2017b). The solution of the related mp-P is computed using Mathematica 11 and it involves 5 explicit expressions of the control law, with an example given by eq. (16) while the detailed results are omitted for the sake of brevity.

$$\text{if } \boldsymbol{\theta} \in CR_1 \ \rightarrow F_l(t_0) = \frac{988.98(1.0114\theta_6 - \theta_2)}{\theta_9} \tag{16}$$

where $\theta_6, \theta_2, \theta_9$ represent the set-point for C_l, the measurement of C_l at each instance and the $C_{l_{in}}$ respectively. Note that for this case study, a prediction horizon of 1 was used for simplicity.

4.2. Monte Carlo simulation results

Once the solution from the hybrid framework is computed the benefits of the proactive mechanisms and the RH- planning are evaluated through Monte Carlo (MC) simulation

runs. For the MC runs all the binary variables are fixed and 1,000 iterations were conducted. At each iteration, a different value of the uncertain parameters is realised, i.e. r_{ip}, and d_{ip}. Within the MC runs the problem is solved with the addition of eq. (17)-(18) that represent a nervousness cost for the deviation of backlog and inventory levels, respectively, to be penalised during the MC runs.

$$DV_{ip} \geq V_{ip} - V_{ip}^{MC} \quad \forall i, p \tag{17}$$

$$DB_{ip} \geq B_{ip} - B_{ip}^{MC} \quad \forall i, p \tag{18}$$

Notice that the backlog (B_{ip}), acts as a slack variable and prevents infeasible runs. The modified objective of the MC simulations is given by eq. (19).

$$Prof_{MC} = Prof - \sum_{ip}[\delta_{ip} DV_{ip} + \beta_{ip} DB_{ip}] \tag{19}$$

The expected value (EV) of the MC runs is computed as the average objective over all the runs, i.e. $EV = \sum_{n}^{N_{MC}} \frac{Prof_{MC}}{n}$, where N_{MC} is the total number of runs. In Figure 3, the comparative results are illustrated.

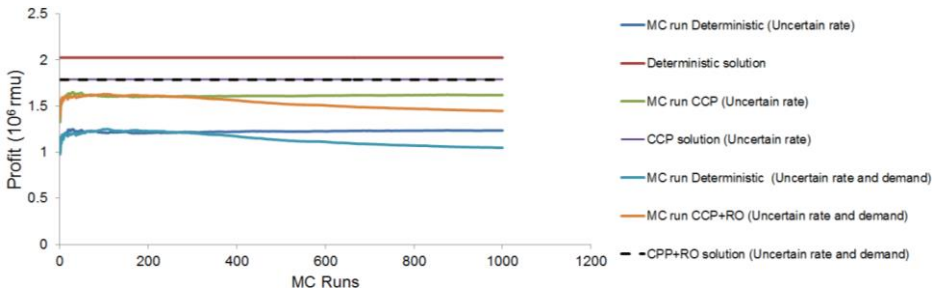


Figure 3. Comparative plot of results from MC simulations.

As indicated in Figure 3 the consideration of uncertain variations in the production rate has a significant impact on the profitability of the process, something that is exacerbated when the demand variations are also taken into account. Following the proposed framework, even though less profit is achieved, the deviation from its EV as computed by the MC simulation is significantly less when compared to the deterministic one. More specifically, the reduction on the profit for the deterministic case is 48.22% while for the one computed by the hybrid framework is less than 18.5%. With regards to the computational statistics an overview is given in Table 1.

Table 1. Overview of computational statistics using CPLEX 12.7 / GAMS 24.8.3

	Deterministic	**Hybrid framework**
Continuous variables	7,835	8,267
Equations	9,656	10,088
Integer variables	3,104	3,104
CPU (s)	24.23	31.42

As shown in Table 1, the time needed for the solution by the hybrid framework allows for the online implementation of the iPSC-relevant decisions, despite the increased number of variables and equations. The reactive components of the proposed framework are next tested by imposing a number of different dynamic disruptions that lead to rescheduled solutions. More specifically, dynamic disruptions during both transition and production periods were successfully handled due to the computational merits of the

proposed novel mp-P controllers and the rescheduling mechanism. In Figure 4 different instances of the uncertainty-aware iPSC solution for the planning horizon are shown.

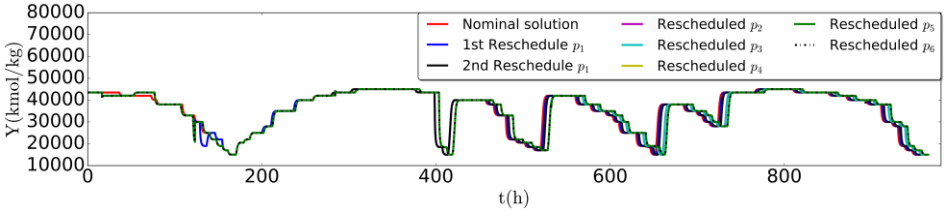


Figure 4. Y=f(t) plot for the uncertainty-aware iPSC solution computed by the hybrid framework

5. Concluding Remarks

Uncertainty is ubiquitous within the integrated problems considered by enterprise wide optimisation and can lead to significant loss of profitability, if left untreated. In the present work, we solve the iPSC problem under multiscale uncertainty using a hybrid proactive/reactive framework based on the nature of the stochastic variations. The results underline the benefits of the proposed approach that through the rolling horizon solution and the outer multi-parametric control loop allow for online implementation of the integrated planning, scheduling and control decisions, thus safeguarding optimal operations. Ongoing research in our group involves the development of decomposition techniques that will enable further improvements in terms of computational times and the solution of industrial-scale case studies.

Acknowledgements

The authors gratefully acknowledge financial support from EPSRC grants EP/M027856/1 and EP/M028240/1.

References

P. Li, H. Arellano-Garcia, G. Wozny (2006). Chance constrained programming approach to process optimization under uncertainty, Comput. Aid. Chem. Eng. 21, 1245-1250

Y. Chu, F.You (2013). Integrated scheduling and control of a polymerization reactor with online closed-loop implementation, Comput. Aid. Chem. Eng. 32, 481-486.

Z. Li, M. Ierapetritou (2009). Integration of Planning and Scheduling and Consideration of Uncertainty in Process Operations. Comput. Aid. Chem. Eng. 27, 87-94

S. Liu, J.M. Pinto, L.G. Papageorgiou (2009) Medium-term planning of multistage multiproduct continuous plants using mixed integer optimisation. Comput. Aid. Chem. Eng. 26, 393-398

V. M. Charitopoulos, V. Dua, L.G. Papageorgiou (2017a). Traveling Salesman Problem-Based Integration of Planning, Scheduling, and Optimal Control for Continuous Processes. Ind. Eng. Chem. Res., 56(39), 11186-11205.

V.M. Charitopoulos, V. Dua, L.G. Papageorgiou (2017b). Closed loop integration of planning, scheduling and control via exact multi-parametric nonlinear programming. Comput. Aid. Chem. Eng., 40, 1273-1278

L.S. Dias, M.G. Ierapetritou (2016). Integration of scheduling and control under uncertainties: Review and challenges. Chem. Eng. Res. Des., 116, 98-113.

H. Shi, Y. Chu, F. You (2015). Novel optimization model and efficient solution method for integrating dynamic optimization with process operations of continuous manufacturing processes. Ind. Eng. Chem. Res., 54(7), 2167-2187.

D. Bertsimas, A. Thiele (2006). A robust optimisation approach to inventory theory. Oper. Res., 54(1), 150-168.

M.A. Gutiérrez-Limón, A. Flores-Tlacuahuac, I.E. Grossmann (2014). MINLP formulation for simultaneous planning, scheduling, and control of short-period single-unit processing systems. Ind. Eng. Chem. Res., 53(38), 14679-14694.

J. Zhuge, M.G. Ierapetritou (2012). Integration of scheduling and control with closed loop implementation. Ind. Eng. Chem. Res., 51(25), 8550-8565.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50191-9

Synthesis of Refinery Hydrogen Networks with Parametric Uncertainties

Zhi-Ting Chen[a], Jui-Yuan Lee[a*], Viknesh Andiappan[b], Chun Deng[c], Denny K.S. Ng[d]

[a]*Department of Chemical Engineering and Biotechnology, National Taipei University of Technology, 1, Sec. 3, Zhongxiao E. Rd., Taipei 10608, Taiwan, ROC*
[b]*Department of Chemical Engineering, University of Bath, Bath BA2 7AY, UK*
[c]*State Key Laboratory of Heavy Oil Processing, College of Chemical Engineering, China University of Petroleum, 18 Fuxue Rd., Changping, Beijing 102249, China*
[d]*Department of Chemical and Environmental Engineering/Centre of Sustainable Palm Oil Research, The University of Nottingham Malaysia Campus, Broga Road, Semenyih, 43500, Malaysia*
juiyuan@ntut.edu.tw

Abstract

This work presents a fuzzy optimisation model for the synthesis of refinery hydrogen networks under uncertainty. Based on a generic superstructure that incorporates all the possible connections between the hydrogen system elements (sources, sinks, compressors etc.), the mathematical formulation consists mainly of flowrate and impurity load balances, where fuzzy intervals are introduced to account for parametric uncertainties. Pressure constraints and the optimal placement of compressors are also taken into account. A case study is presented to illustrate the proposed approach.

Keywords: hydrogen management, process integration, mathematical programming, fuzzy optimisation.

1. Introduction

Hydrogen is an important and expensive utility in oil refining and petrochemicals processing, required in operations such as hydrotreating and hydrocracking. Demand for hydrogen in refineries and petrochemical plants has increased with stricter regulations on sulphur content in fuels and gasoline aromatic composition, as well as increased processing of heavier crude oils and the shrinking market for heavy fuel oil. Increased hydrogen requirements can lead to a deficit of hydrogen in a refinery, resulting in the need for additional hydrogen production or hydrogen import. Therefore, it is desirable for refineries to make efficient use of hydrogen, in order to reduce the capital investment of a new hydrogen plant or the operating costs of existing hydrogen plants or imports.

Synthesis of refinery hydrogen networks is widely accepted as a useful means of improving the efficiency of hydrogen systems. The methodologies can be classified into pinch analysis and mathematical programming approaches. Alves and Towler (2002) proposed the hydrogen surplus diagram to set targets for the minimum flowrate of fresh hydrogen before detailed network design. Various pinch-based targeting techniques were later developed, such as the material recovery pinch diagram (El-Halwagi et al., 2003), the source composite curve (Bandyopadhyay, 2006) and gas cascade analysis

(Foo and Manan, 2006). However, most of the pinch methods cannot handle cost considerations, multiple impurities or pressure/connection constraints effectively. Hallale and Liu (2001) presented a superstructure-based mathematical approach to hydrogen network retrofitting. Their method accounts for pressure constraints, existing equipment and capital costs. Many other mathematical techniques were later developed, including those for the design of hydrogen networks under uncertainty using chance constrained programming (Jiao et al., 2012), robust optimisation (Lou et al., 2014) and stochastic programming approaches (Jagannath and Almansoori, 2014).

It is normally assumed in the synthesis problems that the stream flowrates and qualities are known precisely. In practice, however, process data may be subject to uncertainty or variability. In this work, a fuzzy optimisation model is developed for refinery hydrogen network synthesis with parametric uncertainties. Based on a superstructure incorporating all possible options for hydrogen recovery, the mathematical formulation consists mainly of mass balance equations and pressure constraints, with the objective function considering both the hydrogen requirement and the total annualised cost. The optimal placement of compressors is also taken into account. A case study is solved to illustrate the proposed approach.

2. Problem statement

The problem addressed in this paper may be briefly stated as follows. Given a system with a set of internal hydrogen sources $i \in \mathbf{I}$ and a set of hydrogen sinks $j \in \mathbf{J}$. These sources and sinks may be matched to meet the objective of minimising the requirement for additional hydrogen from external sources, as well as the total annualised cost. A fuzzy resource consumption/cost goal is specified in terms of upper and lower bounds, corresponding to the least and most desirable fresh hydrogen usage/total cost levels. Each internal source is characterised by a flowrate and one or more quality indices, which are defined in terms of fuzzy intervals. Similarly, each sink is defined in terms of fuzzy flowrate and fuzzy quality limits. The model covers a range of solutions ranging from optimistic (i.e. high-risk) and conservative (i.e. low-risk) extremes. Solving the model identifies an optimal solution that achieves the optimal compromise between the need to conserve fresh hydrogen/reduce cost on one hand, and the need to avoid adverse process effects on the other hand. The final source-sink network has to satisfy all the flowrate/component balances, quality constraints and resource conservation/cost saving goals defined by the fuzzy intervals.

3. Model formulation

The fuzzy optimisation model is presented below. The objective function is to maximise the overall degree of fuzzy constraint satisfaction:

$$\max \lambda \tag{1}$$

The fuzzy degree of satisfaction lies between 0 and 1; 0 corresponds to unsatisfactory, 1 is completely satisfactory, and fractional values indicate partial satisfaction.

$$0 \le \lambda \le 1 \tag{2}$$

In fuzzy optimisation using "max-min" aggregation, every fuzzy constraint should be satisfied partially at least to the degree λ. Fuzzy goals for hydrogen consumption (ϕ_{HC}) and total cost (ϕ_{TC}) are specified for the network using the linear membership function, as follows:

$$\phi_{\mathrm{HC}} - \phi_{\mathrm{HC}}^{\mathrm{U}} \le \lambda\left(\phi_{\mathrm{HC}}^{\mathrm{L}} - \phi_{\mathrm{HC}}^{\mathrm{U}}\right) \tag{3}$$

$$\phi_{\mathrm{TC}} - \phi_{\mathrm{TC}}^{\mathrm{U}} \le \lambda\left(\phi_{\mathrm{TC}}^{\mathrm{L}} - \phi_{\mathrm{TC}}^{\mathrm{U}}\right) \tag{4}$$

Note that the hydrogen consumption or total cost is at the upper bound when $\lambda = 0$, and at the lower bound when $\lambda = 1$.

Eq. (5) describes the flowrate balance for each hydrogen source. The hydrogen from source i may be sent to hydrogen sinks j (f_{ij}), compressors m (f_{im}) and fuel systems w (f_{iw}). It can be seen that when $\lambda = 0$, the amount of hydrogen available from source i is F_i^{U}, while when $\lambda = 1$, the amount available is F_i^{L}. This relationship means that lower values of λ are inherently more optimistic (and thus riskier), while higher values of λ are inherently more conservative. The objective function thus forces the model towards a more conservative solution with a robust network. If the hydrogen available from source i is known precisely, then $F_i^{\mathrm{L}} = F_i^{\mathrm{U}}$ and the fuzzy interval has zero width.

$$F_i^{\mathrm{U}} + \lambda\left(F_i^{\mathrm{L}} - F_i^{\mathrm{U}}\right) = \sum_{j\in\mathbf{J}} f_{ij} + \sum_{m\in\mathbf{M}} f_{im} + \sum_{w\in\mathbf{W}} f_{iw} \quad \forall i \in \mathbf{I} \tag{5}$$

Eq. (6) describes the flowrate balance for each hydrogen sink. The hydrogen to sink j may come from hydrogen sources i, compressors m (f_{mj}) and external fresh hydrogen sources r (f_{rj}). It can also be seen that the hydrogen demand of sink j is F_j^{L} when $\lambda = 0$, and is F_j^{U} when $\lambda = 1$; the logic used is similar, with higher values of λ denoting more conservative (or less risky) conditions. In addition, if the hydrogen demand of sink j is known precisely, the fuzzy interval will also have zero width.

$$F_j^{\mathrm{L}} + \lambda\left(F_j^{\mathrm{U}} - F_j^{\mathrm{L}}\right) = \sum_{i\in\mathbf{I}} f_{ij} + \sum_{m\in\mathbf{M}} f_{mj} + \sum_{r\in\mathbf{R}} f_{rj} \quad \forall j \in \mathbf{J} \tag{6}$$

Eq. (7) describes the impurity load balance for each hydrogen sink. It is stated that the load of component c for the mixed stream entering a given sink cannot exceed its load tolerance. Note that when $\lambda = 0$, the streams entering the sink are of the best quality (Y_{ic}^{L} and Y_{rc}^{L}), leading to an optimistic solution with maximised opportunities for hydrogen reuse/recycling. However, the hydrogen recovery potential has to be counterbalanced by considerations of data uncertainty. Thus, $\lambda = 1$ corresponds to the most conservative solution, which in turn assumes that the available streams are of the poorest quality (Y_{ic}^{U} and Y_{rc}^{U}). Likewise, it can be observed from the right-hand side of Eq. (7) that the least conservative solution ($\lambda = 0$) assumes maximum sink tolerance to the quality load (M_{jc}^{U}), while the most conservative solution ($\lambda = 1$) assumes minimum tolerance (M_{jc}^{L}).

$$\sum_{i\in\mathbf{I}} f_{ij}\left[Y_{ic}^{\mathrm{L}} + \lambda\left(Y_{ic}^{\mathrm{U}} - Y_{ic}^{\mathrm{L}}\right)\right] + \sum_{m\in\mathbf{M}} f_{mj} y_{mc}^{\mathrm{out}} + \sum_{r\in\mathbf{R}} f_{rj}\left[Y_{rc}^{\mathrm{L}} + \lambda\left(Y_{rc}^{\mathrm{U}} - Y_{rc}^{\mathrm{L}}\right)\right]$$
$$\le M_{jc}^{\mathrm{U}} + \lambda\left(M_{jc}^{\mathrm{L}} - M_{jc}^{\mathrm{U}}\right) \quad \forall c \in \mathbf{C}, j \in \mathbf{J} \tag{7}$$

Eqs. (8)-(10) are used to determine the compressor inlet and outlet pressures. The inlet pressure of compressor m (p_m^{in}) should be lower than or equal to the pressures of all inlet streams from internal (P_i) and external hydrogen sources (P_r), whilst the outlet pressure of compressor m (p_m^{out}) should be higher than or equal to the operating pressures of all hydrogen sinks connected to it.

$$p_m^{\text{in}} \le P_i + \Gamma\left(1 - z_{im}\right) \quad \forall i \in \mathbf{I}, m \in \mathbf{M} \tag{8}$$

$$p_m^{\text{in}} \le P_r + \Gamma\left(1 - z_{rm}\right) \quad \forall m \in \mathbf{M}, r \in \mathbf{R} \tag{9}$$

$$p_m^{\text{out}} \ge P_j z_{mj} \quad \forall j \in \mathbf{J}, m \in \mathbf{M} \tag{10}$$

where z_{im}, z_{rm} and z_{mj} are binary variables indicating if the connection exists; Γ is a large positive value. Note that the inlet compressor pressure will be forced to the lowest of the inlet streams, and the outlet pressure to the highest of the downstream hydrogen sinks, with the inclusion of the compressor cost in the objective function to be minimised.

The pressure and connection constraints are given in Eqs. (11)-(13). In particular, Eq. (13) forbids the connections from lower pressure to higher pressure.

$$\varepsilon - \Gamma\left(1 - z_{ab}^{\text{P}}\right) \le P_b - P_a \le \Gamma z_{ab}^{\text{P}} \tag{11}$$

$$F_{ab}^{\text{L}} z_{ab} \le f_{ab} \le F_{ab}^{\text{U}} z_{ab} \tag{12}$$

$$z_{ab}^{\text{P}} + z_{ab} \le 1 \tag{13}$$

where ε is a small positive value; z_{ab}^{P} is a binary variable indicating if P_b ($\in \left\{P_j, p_m^{\text{in}}\right\}$) is higher than P_a ($\in \left\{P_i, p_m^{\text{out}}, P_r\right\}$), with $ab \in \{ij, im, mj, rj, rm\}$.

The hydrogen consumption and the total cost of the network are given by Eqs. (14) and (15) respectively. Detailed cost functions for hydrogen utility (φ_{H_2}), compressor power (φ_{power}), fuel (φ_{fuel}) and piping (φ_{pipe}) as well as the relevant parameters can be found in the previous works (Hallale and Liu, 2001; Deng et al., 2014).

$$\phi_{\text{HC}} = \sum_{r \in \mathbf{R}} \sum_{j \in \mathbf{J}} f_{rj} + \sum_{r \in \mathbf{R}} \sum_{m \in \mathbf{M}} f_{rm} \tag{14}$$

$$\phi_{\text{TC}} = \varphi_{\text{H}_2} + \varphi_{\text{power}} - \varphi_{\text{fuel}} + \varphi_{\text{pipe}} \tag{15}$$

Other mass balance equations for compressors and the fuel system are omitted here due to space limitations.

4. Case study

This case study is representative of a real refinery system. The data used is adapted from Hallale and Liu (2001), but fictitious uncertainty margins are added to the flowrates and concentration levels to demonstrate the developed fuzzy model. There are six hydrogen consumers using hydrogen from a catalytic reformer and a hydrogen plant. There are two make-up compressors in the system and all the consumers except the isomerisation plant have an internal recycle compressor. The process flowrate and pressure data are given in Table 1. The lower and upper limits of the fuzzy goal for fresh hydrogen consumption and total annual cost are based on the results reported by Hallale and Liu (2001) of 45 MMscfd and \$22.5 million/y, and the total fresh hydrogen demand and system cost of 55.57 MMscfd and \$27.55 million/y without hydrogen management, respectively.

Solving the model yields the network shown in Figure 1, where it can be seen that the fresh hydrogen requirement is 48.96 MMscfd. The corresponding level of satisfaction of the fuzzy limits is $\lambda = 0.625$. At the optimal value of λ, the actual allocation of hydrogen to sink HCU-K is 38.94 MMscfd (= 38 MMscfd + 0.625 × (39.5 − 38 MMscfd)), and the amount of hydrogen from source HCU is 11.13 MMscfd (= 13 MMscfd + 0.625 ×

(10 − 13 MMscfd)). As a result, the corresponding hydrogen flowrate to the fuel system is 26.75 MMscfd. By comparison, the solution to the crisp (non-fuzzy) problem is 45 MMscfd of fresh hydrogen from the hydrogen plant and 23.67 MMscfd of low-purity hydrogen to the fuel system (Hallale and Liu, 2001).

The optimal value of λ can be interpreted as follows. For any given process, the parameters may be defined in terms of a fuzzy interval, which ranges from an optimistic (or risky) value to a conservative (or low-risk) value. For example, the concentration limit for JHT-K ranges from 30 to 40 %. The lower limit of this range corresponds to a conservative or safe assumption, whilst the upper limit entails more risk by assuming that the process is able to physically tolerate more impurity. In other words, using 30 % as the limit for this sink is equivalent to setting $\lambda = 1$, whilst using 40 % assumes that $\lambda = 0$. As the optimisation model allows λ to be adjusted until the optimal value of 0.625 is found, it follows that the corresponding concentration limit for this solution is 33.75 % (= 40 % − 0.625 × (40 − 30 %)). Note that this result is closer to the conservative assumption than to the risky one. A similar argument applies for all the fuzzy parameters in the model.

Table 1. Process Data for the Refinery Case Study

Source	Flowrate (MMscfd)	Impurity concentration (%)	Pressure (psia)	Sink	Flowrate (MMscfd)	Impurity concentration (%)	Pressure (psia)
CRU	(23,25)	(25,30)	300	DHT-K	11.31	(19.03,30.03)	600
DHT	(8,9)	(30,40)	400	CNHT-K	8.21	(15.47,20,47)	500
CNHT	3.47	(37,40)	350	JHT-K	(7,9.5)	(30,40)	500
JHT	4.32	(30,40)	350	NHT-K	(12,13.5)	(27.56,33.56)	300
NHT	(6,8)	(15,45)	200	IS4-K	0.04	(25,30)	300
IS4	0	-	-	HCU-K	(38,39.5)	(7,13)	2000
HCU	(10,13)	(20,50)	1200				
H_2 plant	(45,55.57)	(8,10)	300				

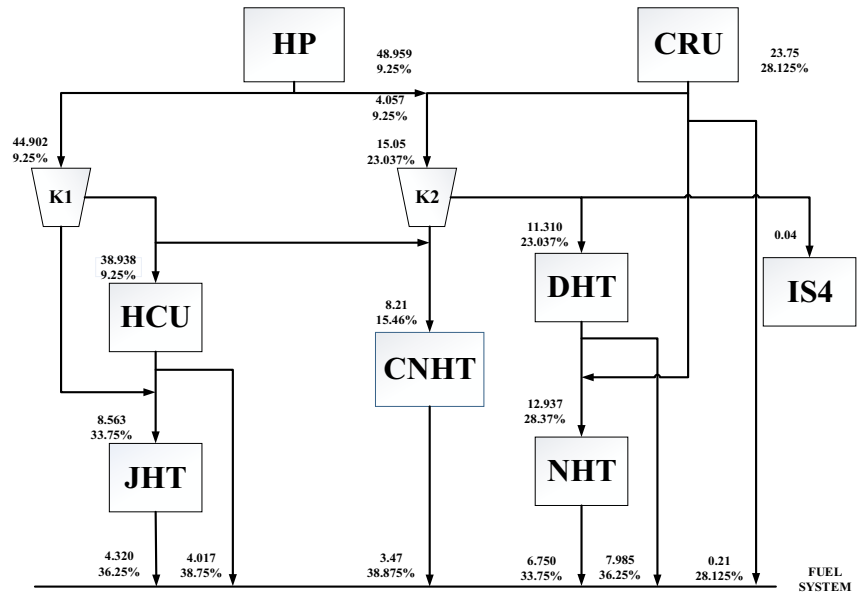


Figure 1. Optimised network considering uncertainties.

Table 2. Comparison of Results

Results	Hallale and Liu (2001)	This work
Hydrogen cost	$32,900,000/y	$35,740,310/y
Electricity cost	$1,770,000/y	$2,209,399/y
Fuel cost	$12,200,000/y	$14,691,790/y
Piping cost	-	$693,491/y
Total cost	$22,470,000/y	$23,951,410/y
Degree of satisfaction	-	0.625

Table 2 shows a result comparison. Though not reported, the piping cost of the existing hydrogen system should account for only a small portion of the total, as is presented in this work. Therefore, the slight increase in the total annual cost (as well as the hydrogen utility requirement/cost) is the penalty incurred to provide a margin of safety based on process mass balances, particularly with regard to the impurity load.

5. Concluding remarks

A fuzzy optimisation model for refinery hydrogen network synthesis with parametric uncertainties has been developed in this work. The formulation consists mainly of mass balance equations and pressure constraints, and accounts for the compromise between resource conservation and cost saving. A case study was solved to illustrate the proposed approach. Future work will consider the placement of purification units and carry out more detailed cost analysis.

References

J.J. Alves, G.P. Towler, 2002, Analysis of Refinery Hydrogen Distribution Systems, Ind. Eng. Chem. Res., 41, 23, 5759-5769.

S. Bandyopadhyay, 2006, Source Composite Curve for Waste Reduction, Chem. Eng. J., 125, 2, 99-110.

C. Deng, H. Pan, Y. Li, Y. Zhou, X. Feng, 2014, Comparative Analysis of Different Scenarios for the Synthesis of Refinery Hydrogen Network, Appl. Therm. Eng., 70, 2, 1162-1179.

M.M. El-Halwagi, F. Gabriel, D. Harell, 2003, Rigorous Graphical Targeting for Resource Conservation via Material Recycle/Reuse Networks, Ind. Eng. Chem. Res., 42, 19, 4319-4328.

D.C.Y. Foo, Z.A. Manan, 2006, Setting the Minimum Utility Gas Flowrate Targets Using Cascade Analysis Technique, Ind. Eng. Chem. Res., 45, 17, 5986-5995.

N. Hallale, F. Liu, 2001, Refinery Hydrogen Management for Clean Fuels Production, Adv. Environ. Res., 6, 1, 81-98.

A. Jagannath, A. Almansoori, 2014, Modeling of Hydrogen Networks in a Refinery Using a Stochastic Programming Appraoch, Ind. Eng. Chem. Res., 53, 51, 19715-19735.

Y. Jiao, H. Su, W. Hou, Z. Liao, 2012, Optimization of Refinery Hydrogen Network Based on Chance Constrained Programming, Chem. Eng. Res. Des., 90, 10, 1553-1567.

J. Lou, Z. Liao, B. Jiang, J. Wang, Y. Yang, 2014, Robust Optimization of Hydrogen Network, Int. J. Hydrogen Energy, 39, 3, 1210-1219.

R.R. Tan, 2011, Fuzzy Optimization Model for Source-Sink Water Network Synthesis with Parametric Uncertainties, Ind. Eng. Chem. Res., 50, 7, 3686-3694.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50192-0

Kaibel Column: Modeling and Optimization

E. Soraya Lopez-Saucedo[a*], Qi Chen[b], Ignacio E. Grossmann[b], Jose A. Caballero[c]

[a]*Departamento de Ingenieria Quimica, Instituto Tecnologico de Celaya, 600 Antonio Garcia Cubas Esq. Av Tecnologico, Celaya 38010, Mexico*
[b]*Department of Chemical Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh 15213, USA*
[c]*Departamento de Ingenieria Quimica, Universidad de Alicante, Carretera San Vicente S/N, Alicante 03690, España*
sorayadex@gmail.com

Abstract

In this work, we propose a rigorous model for the simulation and optimization of a dividing-wall configuration known as Kaibel column (KC). The rigorous model, based on the well-known MESH equations of a conventional Continuous Distillation Column (CDC), is separated into four sets of equations that represent the main sections of the column: above, below, left side, and right side sections of the dividing wall, including secondary sections formed by the trays in between the side feed and outlets. The non-linear programming optimization of this rigorous model is performed in order to determine the Kaibel Column potential benefits, such as energy savings and greater purities for the middle products within a single column. The proposed steady-state Kaibel Column performs the separation of a quaternary methanol, ethanol, propanol, and butanol mixture while minimizing a total annual cost function using the reflux ratio, vapor and liquid rates, the product rates, and the heat duties as the manipulated variables of the system. The results show that the Kaibel Column is able to reduce the energy consumption in the reboiler and condenser by more than 40 % compared to the conventional sidedraw continuous counterpart. Also, the reduction in the number of trays of the dividing wall proves to be an important factor since small reductions in energy consumption were observed. The model was coded in Pyomo and solved using the NLP solver IPOPT.

Keywords: Dividing Wall Columns, Kaibel Columns, NLP Optimization.

1. Introduction

Over the last decades, the study of new dividing wall configurations has been addressed with the purpose of solving the inherent problem of the traditional distillation processes: their high energy consumption. One of the innovative solutions to overcome this energy problem is the use of intensive configurations, such as Dividing Wall Distillation Columns (DWDC), created by the addition of a wall that splits the column in two sections. While its construction and control might still represent a challenge, this configuration has proven to generate savings up to 30 % in energy consumption and, in some cases, savings in the capital and investment costs by reducing the number of distillation columns needed to perform a multicomponent separation (Dejanovic et al., 2010). Among these divided columns different configurations have been proposed where the Kaibel Column (KC) (Kaibel, 1987) is considered a promising option since it

can replace a sequence of two, three or more distillation columns needed to purify a multicomponent mixture (Yildrim et al., 2011) since it is able to separate more than three products within a single column (Kiss et al., 2012). The simulation and optimization of this multi-product configuration has been performed first by using commercial software to simulate the column then by the use of an external algorithm which performs the optimization. These results have shown for KC good controllability properties and energy savings (Qian et al., 2016, Tututi-Avila et al., 2017) but no rigorous open-source model has been reported for its study.

In this paper, we propose a model to simulate and optimize a Kaibel Column to obtain the system variable profiles that minimize a total annual cost while producing four high purity products: methanol, ethanol, n-propanol, and butanol. The proposed equations are presented in Section 3 while the case study and operational conditions are given in Section 2. Finally, the results are presented in Section 4 followed by the Conclusions in Section 5.

2. NLP Problem Statement

In a general form, the problem is stated as follows. Given a Kaibel Column configuration and a desired final product purity, a feed mixture of NC components is to be separated into NC high-purity products while minimizing the total annual cost function. To accomplish this goal the reflux ratio, the vapor flowrate, the liquid distributors, and the heat duties are used as the manipulated variables in the NLP optimization problem, allowing the purification (to some pre-specified tolerance) of the NC components in the feed.

3. Model Equations

The proposed model for the KC was obtained by modifying the mass and energy balances on the internal trays of a conventional Continuous Distillation Column (CDC) with two side outlets, S_1 and S_2. This model is given in Eq.(1) to Eq.(13) with the objective function (TAC) in Eq.(14). The proposed set of MESH equations is split into four main sections: above and below the dividing wall and the left and right sections of the dividing wall (section 1 and 2, respectively). Some secondary sections, such as the trays in between the side feed F and the side outlets S_1 and S_2, are also considered in the model. All these sections can be seen in the KC scheme in Figure 1. Due to space limitations, only the mass balances per component are given here and will have to be further modified by the reader to complete the MESH equations of the model. The mass balances equations are given in the next order: reboiler and vapor distributor in Eq.(1) and Eq.(2), starting tray of the dividing wall in section 1, side feed tray, and ending tray of the dividing wall in section 1 in Eq.(3) to Eq.(5), starting tray of the dividing wall in section 2, side outlet 1, side outlet 2, and ending tray of the dividing wall in section 2 in Eq.(6) to Eq.(9), and liquid distributor and condenser in

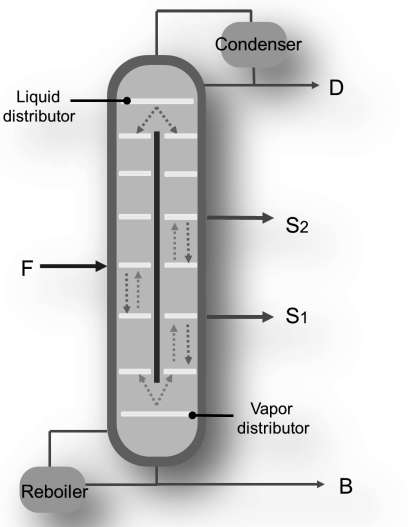


Figure 1. Kaibel Column.

Eq.(10) and Eq.(11). The total mass balances are obtained when the liquid and vapor compositions terms are removed from Eq.(1) to Eq.(13) while the energy balances are obtained when the liquid and vapor composition terms are replaced by the liquid and vapor enthalpy terms in Eq.(1) to (13). The vapor compositions, composition summations, vapor-liquid equilibrium constants, activity coefficients (UNIQUAC), and vapor and liquid enthalpy equations are also included and solved in the model. The Total Annual Cost (TAC) is given in Eq.(14). The trays are numbered from the bottom to the top of the column, with tray *1* the reboiler and tray *NT* the condenser.

$$0 = \left(L_{j+1} + F - S_1 - S_2\right)x^1_{j+1,i} - V_j y_{j,i} - B x_{j,i} \tag{1}$$

$$0 = \left(L^1_{j+1} + F\right)x^1_{j+1,i} + \left(L^2_{j+1} - S_1 - S_2\right)x^2_{j+1,i} + V_{j-1}y_{j-1,i} - \left(L_j + F - S_1 - S_2\right)x_{j,i} - V^1_j y_{j,i} - V^2_j y_{j,i} \tag{2}$$

$$0 = \left(L^1_{j+1} + F\right)x^1_{j+1,i} + V^1_{j-1}y_{j-1,i} - \left(L^1_j + F\right)x^1_{j,i} - V^1_j y^1_{j,i} \tag{3}$$

$$0 = \left(L^1_{j+1} + F\right)x^1_{j+1,i} + \left(L^2_{j+1} - S_1 - S_2\right)x^2_{j+1,i} + V_{j-1}y_{j-1,i} - \left(L_j + F - S_1 - S_2\right)x_{j,i} - V^1_j y_{j,i} - V^2_j y_{j,i} \tag{4}$$

$$0 = \left(L^1_{j+1} + F\right)x^1_{j+1,i} + V^1_{j-1,i}y_{j-1,i} - \left(L^1_j + F\right)x^1_{j,i} - V^1_j y^1_{j,i} \tag{5}$$

$$0 = L^1_{j+1}x^1_{j+1,i} + V^1_{j-1}y^1_{j-1,i} - \left(L^1_j + F\right)x^1_{j,i} - V^1_j y^1_{j,i} + F x_{F,i} \tag{6}$$

$$0 = L^1_{j+1}x_{j+1,i} + V^1_{j-1}y^1_{j-1} - L^1_j x^1_{j,i} - V^1_j y^1_{j,i} \tag{7}$$

$$0 = (L^2_{j+1} - S_1 - S_2)x^2_{j+1,i} + V^2_{j-1}y_{j-1,i} - (L^2_j - S_1 - S_2)x^2_{j,i} - V^2_j y^2_{j,i} \tag{8}$$

$$0 = (L^2_{j+1} - S_2)x^2_{j+1,i} + V^2_{j-1}y^2_{j-1,i} - (L^2_j - S_1 - S_2)x^2_{j,i} - V^2_j y^2_{j,i} - S_1 x^2_{j,i} \tag{9}$$

$$0 = L^2_{j+1}x^2_{j+1,i} + V^2_{j-1}y^2_{j-1,i} - \left(L^2_j - S_2\right)x^2_{j,i} - V^2_j y^2_{j,i} - S_2 x^2_{j,i} \tag{10}$$

$$0 = L^2_{j+1}x_{j+1,i} + V^2_{j-1}y^2_{j-1,i} - L^2_j x^2_{j,i} - V^2_j y^2_{j,i} \tag{11}$$

$$0 = L_{j+1}x_{j+1,i} + V^1_{j-1}y_{j-1,i} + V^2_{j-1}y^2_{j-1,i} - L^1_j x_{j,i} - L^2_j x_{j,i} - V_j y_{j,i} \tag{12}$$

$$0 = V_{j-1}y_{j-1,i} - L_j x_{j,i} - D x_{j,i} \tag{13}$$

$$TAC = C_H Q_{Reb} + C_c Q_{Con} + \frac{I_R(I_R+1)^{PL}}{(I_R+1)^{PL}}(C_{shell} + C_{int} + C_{HE}) \tag{14}$$

The notation is: j are the trays, i are the components, superscripts 1 and 2 represent the section 1 (left side) and 2 (right side) of the dividing wall, x and y are the liquid and vapor composition, S_1 and S_2 are the side outlets flowrate, D and B are the distillate and bottoms flowrate in kg-mol/h, F is the side feed flowrate in kg-mol/h, L and V are the liquid and vapor flowrate in kg-mol/h, Q_{Reb} and Q_{Con} are the energy consumption in the reboiler and condenser in kW, rr is the internal reflux ratio, I_R and PL are the interest rate and plant life, C_H, C_C, and C_{HE} are the costs of heat and cold utilities and the heat exchangers, while C_{shell} and C_{int} are the costs of the column shell and internal with installation. The costs for the TAC are: I_R = 0.1, PL = 10 y, C_H = 0.0091 \$/kWh, C_C = 0.0023 \$/kWh, C_{HE} = 4,500 \$, C_{shell} = 70,000 \$, and C_{int} = 132,600 \$.

3.1. Case Study

The study of the KC is performed by separating an equimolar quaternary mixture of methanol (1), ethanol (2), n-propanol (3), and butanol (4). The separation is carried out with the objective of obtaining four high-purity products with the following specifications: product flowrates higher than 0.2 kg-mol/h and final compositions higher than 0.99 for the four components. The proposed model was solved under steady-state conditions with the operational conditions given in Table 1 and considering the following degrees of freedom: the condenser and reboiler duties, the product flowrates, the reflux ratio, the liquid and vapor rates, and the liquid distributor. The vapor distributor is assumed to be constant since it is known to be a very difficult variable to control during the column operation (Yildrim et al., 2011). The problem is initialized by performing the simulation of the KC column (*obj=1*) while using the warm start option in IPOPT. This simulation is followed by the solution of the NLP optimization problem using the TAC as the objective function. In order to compare the energy consumption in the reboiler and condenser, the KC results are compared to the results obtained from the optimization of a conventional CDC under the same operational conditions. The proposed KC and the CDC NLP models were written in Pyomo and solved using IPOPT.

4. Results

Table 1. Operating conditions for the KC.

Number of trays	70
Side outlets trays, S_1-S_2	20-40
Side feed tray, F	35
Dividing wall starting tray,	10
Dividing wall ending tray,	50
Vapor distributor, dv_1	0.60
Initial vapor flowrate, V	5
Initial liquid composition	$x_{i\,i}^{0}$
(1) Methanol	0.25
(2) Ethanol	0.25
(3) n-Propanol	0.25
(4) Butanol	0.25
Side feed liquid composition	$x_{i\,i}^{0}$

The optimization results for the KC and the conventional CDC for the quaternary methanol-ethanol-propanol-butanol mixture separation are presented in Table 2. The four high-purity products specification of 0.99 was achieved by the KC while the CDC achieved it only in the distillate and bottom products with middle products a purity of 0.93. The KC comprises 8,226 equations solved in 481 seconds while the CDC comprises 5,258 equations solved in 190 seconds. Even though the middle products were not highly purified by the CDC, the conventional column energy consumption is higher than the heat duty consumed by the KC column. This is consistent with the behavior observed for other mixtures, where savings around 14 % are observed with respect to the conventional CDC separation sequence (Tututi-Avila et al., 2017). For this separation case, a comparison between the results in Table 2 shows that KC reduces the energy consumption in the reboiler by 42.14 %, followed by a reduction of 45.6 % in the condenser. The change in the position of the dividing wall slightly affects the energy consumption when the number of dividing-wall trays is reduced (34 dividing-wall trays, from tray 13 to tray 47 of the main column), slightly higher energy savings of 42.6 % in the reboiler and 46 % for the

condenser are observed when compared to the base case (40 dividing-wall trays, from tray 10 to 50 of the main column). On the other hand, the addition of trays on the dividing-wall will lead to a reduction on the energy efficiency by moving from energy savings from the base case of 42.14 % for the reboiler to 41.73 while the condenser changes from 45.6 % to 45.13 %. These changes might not seem significant but it is important to point out the importance on the selection of the dividing wall trays and their position. The side feed flowrate also has a direct effect on the energy savings. On the KC we observed that for flowrates between 0.80 and 1 kg-mol/h, the KC still consumes less energy when compared to the CDC. We can therefore say that the proposed KC model gives a consistent solution that can predict the behavior of the different non-linear variables involved in this intensified distillation configuration.

To study the behavior of the variables in the system we present the KC liquid composition profiles in Figure 2. The composition profiles above and below the dividing wall with section 1 and section 2 of the dividing column are given in Figures 2(a) and 2(b), respectively. In Figure 2(a) we observe how the propanol composition is diluted by the side feed *F* on tray 35, with smaller composition values on the trays below it. On the other hand, section 2 profiles in Figure 2(b) show how the four products achieve the desired purity by keeping a reflux ratio value close to 1 and a temperature profile on the internal trays of the column close to the pure component boiling points, allowing for a better separation compared to section 1. Even though a remixing effect on the middle products after their removal from tray 20 and 40 (shaded area in Figure 2(b)) is observed, it is not significant for this particular case since we are already achieving energy savings. This dilution could lead to small thermal inefficiencies in the KC, reducing the energy savings. However, in order to reduce these inefficiencies in further examples we would have to perform a more in depth study on the optimal position of the side outlets.

Table 2. Results for the CDC and the KC.

	CDC	KC		
Dividing Wall Trays		10-50*	13-47	9-51
Q_{Reb}	61.628	35.652	35.405	35.911
Q_{Con}	56.950	30.997	30.764	31.249
Reflux ratio	0.963	0.937	0.937	0.937
Side feed	0.855	0.802	0.801	0.803
Distillate/Bottoms	0.203[1]/0.199[1]	0.200/0.202	0.200/0.201	0.200/0.202
Side outlet 1	0.227[2]	0.200	0.200	0.200
Side outlet 2	0.224[2]	0.199	0.200	0.199
Liquid distributor, dl_1	-	0.50	0.50	0.50

* Base case. [1] Methanol and butanol composition in *D* and *B* is 0.99.

[2] Ethanol and propanol composition in S_1 and S_2 is 0.93.

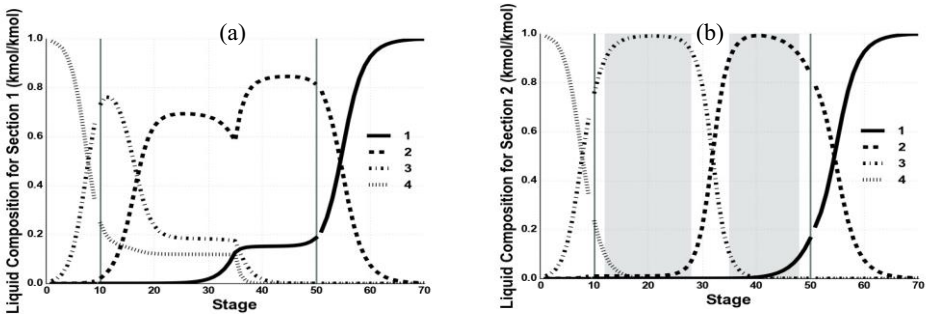


Figure 2. Liquid composition profiles of the Kaibel Column. (a) Liquid composition profiles of the main column with section 1 of the dividing wall. (b) Liquid composition profiles of the main column with section 2 of the dividing wall.

5. Conclusions

In this work, we proposed a model for the simultaneous simulation and optimization of a Kaibel Column for the separation of high-purity products of a methanol-ethanol-propanol-butanol mixture. In order to study the behavior of all the variables involved in a distillation process and the energy efficiency of a KC, a simulation was performed first, followed by the NLP optimization problem that minimizes a total annual cost function under fixed product specifications. To be able to compare and define possible energy savings, a conventional Continuous Distillation Column was also solved. By comparing these two columns we observe that the KC proposed model shows accurate profiles by achieving energy savings of around 42 % in the reboiler and 45 % in the condenser for the base case, while the reduction of 2 trays in the dividing wall increased this savings by an extra 0.50 %, proving how important it is to decide the number of trays and their position in a distillation configuration. These results show that the proposed model can be used as an accurate option for the solution of Kaibel Column configurations.

References

Dejanovic, I., Matijasevic, L., and Olujic, Z., 2010, Dividing Wall Column – A Breakthrough Towards Sustainable Distilling, Chemical Engineering and Processing, 49, 559-580

Kaibel, G., 1987, Distillation Columns with Vertical Partitions, Chemical Engineering & Technology, 10

Kiss, A. A., Landaeta, S. J., and Ferreira, C. A. I., 2012, Towards Energy Efficient Distillation Technologies – Making the Right Choice, Energy, 47, 531-542

Qian, X., Jia, S., Skogestad, S., and Yuan, X., 2016, Control Structure Selection for Four-Product Kaibel Column, Computers and Chemical Engineering, 93, 372-381

Tututi-Avila, S., Domiguez-Diaz, L. A., Medina-Herrera, N., Jimenez-Gutierrez, A., and Hahn, J., 2017, Dividing-Wall Columns: Design and Control of a Kaibel and a Satellite Distillation Column for a BTX Separation, Chemical Engineering and Processing, 114, 1-15

Yildrim, O., Kiss, A. A., and Kenig, E. Y., 2011, Dividing Wall Columns in Chemical Process Industry: A Review on Current Activities, Separation and Purification Technology, 80, 408-417

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50193-2

Systematic Procedure for the Removal of Violations in Water Sources Diagrams

Flávio da S. Francisco[a*], Ewerton E. S. Calixto[a,b], Fernando L. P. Pessoa[a,b], Eduardo M. Queiroz[a]

[a]*DEQ/EQ/UFRJ – Sala 209, Bloco E, Escola de Química, Centro de Tecnologia, Cidade Universitária, Ilha do Fundão, Universidade Federal do Rio de Janeiro, Rio de Janeiro – RJ, Brazil.*
[b]*Centro Universitário SENAI CIMATEC, Av. Orlando Gomes, 1845, Piatã. Cep. 41.650-010. Salvador. Bahia, Brazil*
**flaviosfrancisco@eq.ufrj.br*

Abstract

Over the years, the Water Sources Diagram (WSD) has become a consolidated approach to determine the minimum freshwater consumption and simultaneously generate the system structure in water networks in many kinds of industrial processes. Some efforts have been employed to, for example, automatically determine the reference contaminant and reference operation and to predict in which operation(s) violations will occur, before the beginning of the calculations. Those violations are then removed as part of the last step of the method and it is currently performed by hand. Seeking to improve this step and to create an automated procedure that can be linked with the MINEA software was the motivation of this research.

Keywords: Concentration Violation, Water Networks, Automated Procedure.

1. Introduction

The WSD is an algorithmic-heuristic technique (Gomes et al., 2007) that has been employed in many water allocation problems in industry over the years. Notwithstanding, some authors focused their attention on innovative methods to improve both the performance and quality of the WSD results. For example, one of the WSD drawbacks reported by many users in the past, was the lack of an automated way to generate scenarios (water networks (WN)) without made-by-hand. This cumbersome procedure was solved by the development of a software called MINEA (Mirre et al., 2013). The user was then able to solve many problems in less time with more confidence in the results. Since then, MINEA has been updated to add the most recent algorithms, improve performance and to become more user-friendly.

The algorithms to determine the reference contaminant and reference operation were proposed by Calixto et al. (2015). It's now possible to skip these choices right to the WSD generation. Francisco et al. (2015) extended the WSD method to fixed flowrate operations with one contaminant and (Francisco et al., 2015). Francisco et al. (2017) added new characteristics to determine the reference operation and reference contaminant in fixed flowrate operations involving multiple contaminants. These works filled a gap in the WSD method. However, even those efforts are not enough to avoid the presence of violations in the final water network when dealing with multiple contaminants. With this in mind, Calixto et al. (2015) provided an approach to identify the violated contaminants and the operations in which that occurs. It helps to previously know the violations and

take the proper action while the flowsheet is elaborated. Removing violations in the mass balance flowsheet is one of the last step of the WSD procedure and it is currently made almost by hand. This is one of the reasons that affects the time to get the final results.

This last step is responsible for the removal violation process and it is based on the increasing of the external fresh water consumption or by redirecting the outlet streams of some splitters that are upstream of the operation where the violation is observed. The second option consists of removing the violation without increasing the external fresh water consumption when the operation does not require external fresh water.

Currently, MINEA software provides only the WSD and the flowsheet generation is made separately in conjunction with the violations removal and evolution of the WN. In future version of the software, it will be possible to get the final flowsheet without the presence of violations and evolve the WN. Note that in this work the attention is only on the removal of the violations without observe other options to reduce the freshwater consumption like discussed by Gomes et al. (2013) and Francisco et al. (2015,2017), which means that the evolution of the WN is not performed.

We propose a mathematical model to remove violations with (or without) the minor increasing in the fresh water consumption in water allocation problems with multiple contaminants for the respective WN. This novel approach will be implemented in the future version of the software MINEA. For brevity reasons the WSD will not be explained in detail here, for more details see Gomes et al. (2007, 2013) or Francisco et al. (2015, 2017).

2. Mathematical model

The violations are present in each inlet or outlet of an operation *u*. Figure 1 illustrates the mass flowrate in t/h and concentrations in ppm for the inlet and outlet of an operation *u*.

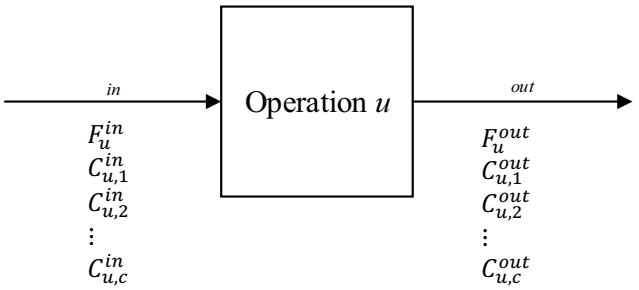


Figure 1. Representation of an operation *u* and its inlet and outlet contaminants concentrations and flowrates.

Where F_u^{in} and F_u^{out} are the inlet and outlet flowrates of an operation *u*, $C_{u,c}^{in}$ and $C_{u,c}^{out}$ are the inlet and outlet concentrations for the operation *u* for a certain contaminant *c*. $C_{u,c}^{out}$ can be determined by the following linear mass transfer equation:

$$C_{u,c}^{out} = \frac{\Delta m_{u,c}}{F_u^{out}} + C_{u,c}^{in} \tag{1}$$

Where $\Delta m_{u,c}$ is the mass load of the contaminant *c* in the operation *u* in kg/h.

A violation usually occurs when a mixer is present upstream the operation u, due to the additional contribution of a poor-quality water stream. Figure 2 shows the representation of a mixer m with an operation u in a flowsheet.

If an amount of s mixed streams is not enough to provide an operation without violations, another water stream supply must be added, from internal or external sources. Eq. (2) shows the mass balance in this system.

$$\sum_{s} C_{m.c}^{s} F_{m}^{s} + C_{m.c}^{a} F_{m}^{a} = F_{u}^{in} C_{u,c}^{in} \tag{2}$$

Where $C_{m.c}^{s}$ is the inlet concentration of contaminant c for a mixer m in a stream s, F_{m}^{s} is the inlet flowrate of the stream s for a mixer m, $C_{m.c}^{a}$ is the inlet concentration of contaminant c in the supplied stream a, that can be internal or external. F_{m}^{a} is the inlet flowrate of the supplied stream a for a mixer m.

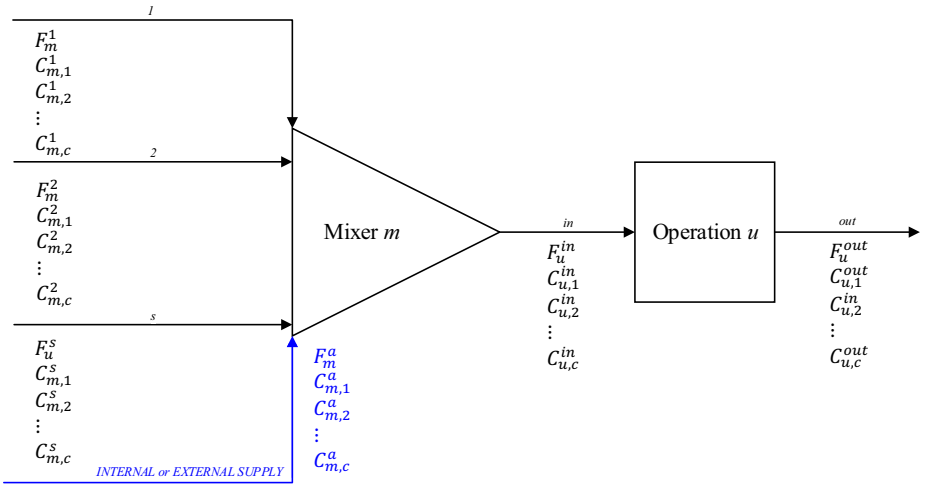


Figure 2. Scheme for the addition of water supply.

We need to know the minimum amount of external or internal water that must be added or allocated to meet the restrictions imposed by operation u. If, $C_{m.c}^{s}$and F_{m}^{s} are known, then it becomes a nonlinear optimization problem.

$$\begin{aligned} &Min \sum_{a} F_{m}^{a} \\ &\text{s. t.} \\ &C_{u,c}^{in} \leq C_{u,c}^{in,max} \\ &C_{u,c}^{out} \leq C_{u,c}^{out,max} \\ &F_{u}^{in} \leq F_{u}^{max} \\ &C_{u,c}^{in}, C_{u,c}^{out} \geq 0 \end{aligned} \tag{3}$$

Where $C_{u,c}^{in,max}$ and $C_{u,c}^{out,max}$ are the maximum inlet and outlet concentrations in the operation u for a contaminant c, respectively and F_{u}^{max} is the maximum allowable flowrate in the operation u.

The model is nonlinear because it is not known the value of $C_{m.c}^{s}$and F_{m}^{s}in Eq. (3). The term $C_{m.c}^{s}F_{m}^{s}$ brings the nonlinearity. If we know some of these values the model can be modeled as linear. Another way to linearize the model is by changing the variables, which can be done by replacing the nonlinear term.

Besides the mathematical model, some basic rules are necessary to be performed in the last step when the water network is obtained by WSD, as follows: a) all the operations are "checked" to verify if there are some violation; b) If any violation is observed, the method starts by considering the addition or removal of internal water streams that are used in the operations which present violation. This is done to minimize the total value of internal streams that are possible to remove the violation in each operation subjected to the respective restrictions. c) If it is not possible to remove the violations in each operation by splitting or redirecting internal water streams, the minimization of external water source with the internal water sources consumptions is carried out. The optimization model was implemented in Microsoft Excel and solved using the GRG nonlinear solution method presented in the *Solver* package.

To better explain how the approach works, a classical case study introduced by Doyle and Smith (1997) is presented. This case study is composed by four water-using processes and three contaminants.

3. Case Studies

3.1. Example 1

This case study was introduced by Doyle and Smith.(1997). Table 1 shows the limiting data. Figure 3 shows the WSD generated.

Table 1. Limiting data for Example 1.

Op.	f_u (t/h)	Cont.	$C_{u,c}^{max}$(ppm)	$C_{u,c}^{out,max}$(ppm)	$\Delta m_{u,c}(kg/h)$
1	34	A	0	160	5,44
		B	0	450	15,3
		C	0	30	1,02
2	75	A	200	300	7,5
		B	100	270	12,75
		C	500	740	18
3	20	A	600	1240	12,8
		B	850	1400	11
		C	390	1580	23,8
4	80	A	300	800	40
		B	460	930	37,6
		C	400	900	40

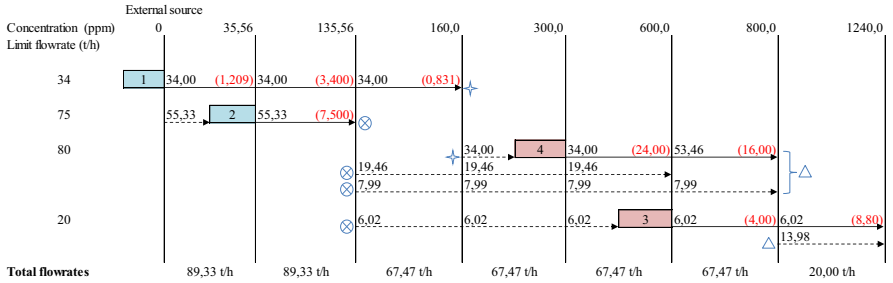


Figure 3. WSD generated for example 1.

Operation 1 and 2 use 34 t/h and 55,33 t/h from the external fresh water source. Part of this amount is distributed to operation 3 and 4. Operation 4 uses all 34 t/h from operation 1 and 27,45 t/h from operation 2. Therefore, operation 3 uses 6,02 t/h from operation 2 and in the last interval of the diagram, a flowrate of 13,98 t/h is calculated to operation 3 that comes from operation 4. The total fresh water consumption is 89,33 t/h. Figure 4 shows the representation of the diagram in a form of a flowsheet.

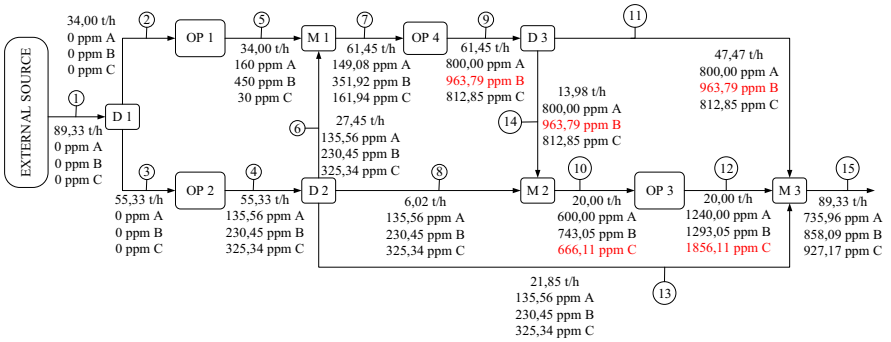


Figure 4. WN for maximum reuse with violations.

The highlighted concentrations in operations 3 and 4 are violated for the contaminants B and C. To remove them, the rules previously described must be performed. Initially, the available internal water supply flowrates need be identified and minimized using Eq. (3). In this case, the flowrates for the streams 6, 8 and 14 are adjusted to meet the limiting concentration data in the operations 3 and 4. Figure 5 shows the new flowsheet with the new minimum internal flowrate supply.

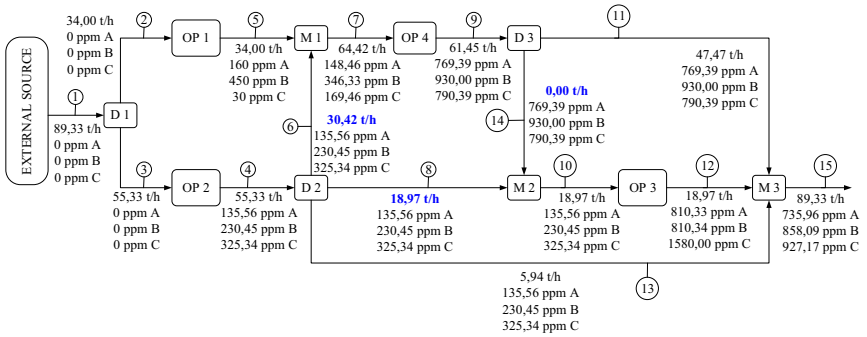


Figure 5. New WN for maximum reuse without violation and new internal flowrate.

The internal flowrates were successfully adjusted, and no violation is present in operations 3 and 4. In this case, the use of external water was not necessary, but only a rearrange of the internal water streams. It is important to note that the WN is not optimal, because, in this step, only the violations are removed, and the evolution step was not implemented. With the evolutionary step the WN will be improved to achieve its optimal freshwater value, modifying or not the WN.

4. Conclusions

A systematic way to remove violations in WNs that were generated by the WSD is presented. For the analyzed case study, the use of internal water supply was enough to eliminate the violations in the flowsheet. In the cases where internal sources are not available, the use of external fresh water supply is possible. For future implementations, this method will be integrated with MINEA WSD algorithm, Microsoft Excel *Solver* and GAMS *Solvers* in only one software package. Research directions are also made to improve this algorithm and to include other options for reducing the freshwater consumption in the WN obtained by WSD, as part of the evolutionary step.

References

E.E.S. Calixto, E.E.S., Francisco, F.S., Pessoa, F.L.P., Queiroz, E.M., 2015. A Novel Approach to Predict Violations and to Define the Reference Contaminant and Operation in Water Using Networks, in: Computer Aided Chemical Engineering. Elsevier, pp. 1901–1906. https://doi.org/10.1016/B978-0-444-63576-1.50011-X

Francisco, F., Bagajewicz, M.J., Pessoa, F.L.P., Queiroz, E.M., 2015. Extension of the water sources diagram method to systems with simultaneous fixed flowrate and fixed load processes. Chem. Eng. Res. Des. 104, 752–772. https://doi.org/10.1016/j.cherd.2015.10.024

Doyle, S.J., Smith, R., 1997. Targeting Water Reuse with Multiple Contaminants. Process Saf. Environ. Prot. 75, 181–189. https://doi.org/10.1205/095758297529020

Francisco, F.S., Mirre, R.C., Calixto, E.E.S., Pessoa, F.L.P., Queiroz, E.M., 2017. Water sources diagram method in systems with multiple contaminants in fixed flowrate and fixed load processes. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2017.11.043

Gomes, J.F.S., Mirre, R.C., Delgado, B.E.P.C., Queiroz, E.M., Pessoa, F.L.P., 2013. Water Sources Diagram in Multiple Contaminant Processes: Maximum Reuse. Ind. Eng. Chem. Res. 52, 1667–1677. https://doi.org/10.1021/ie301537c

Gomes, J.F.S., Queiroz, E.M., Pessoa, F.L.P., 2007. Design procedure for water/wastewater minimization: single contaminant. J. Clean. Prod. 15, 474–485. https://doi.org/10.1016/j.jclepro.2005.11.018

Mirre, R.C., Francisco, F. da S., Calixto, E.E. da S., Pessoa, F.L.P., Queiroz, E.M., 2013. MINEA, Software para a Minimização do Consumo de Água em Processos Químicos com Base no Diagrama de Fontes de Água (DFA), in: 11º Congreso Interamericano de Computación Aplicada a La Industria de Procesos - CAIP. Pontificia Universidad Católica del Perú, Lima, Peru.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50194-4

A heat integration method with location-dependent heat distribution losses

Hür Bütün[a*], Ivan Kantor[a], François Maréchal[a]

[a]*École Polytechnique Fédérale de Lausanne, Rue de l'industrie 17, Case Postale 440, Sion, CH-1951, Switzerland*

hur.butun@epfl.ch

Abstract

Energy consumption in industrial processes is mainly in the form of heat. Thus, heat recovery is one of the main focuses in industrial energy efficiency problems. Heat integration (HI) techniques have been studied extensively to solve such problems. One of the main drawbacks of the classical heat integration approaches is that heat can be transferred from any stream to another as long as it flows from higher temperature intervals to lower ones, which results in impractical scenarios, in which heat is recovered over long distances. This work proposes a heat integration method which takes into account heat distribution losses. The heat losses are calculated as a function of the distance between the original location of the stream and the location it is used and the supply and return temperatures. The heat cascade is written so that the energy balance is closed for each location. This way, while heat recovery within the same or close location is promoted, heat transfer over long distances is discouraged. Using the proposed method, practically infeasible solutions are eliminated at the level of optimisation. At the same time, the temperature drop and the heat losses resulting from heat exchange over long distances are calculated. The method is applied to a case study with two plants. While the total operating cost can be reduced by 25% by heat integration within and between the sites, not exchanging heat between the two sites is found to be more beneficial when heat losses are taken into account.

Keywords: hear integration, heat distribution losses, multiple locations

1. Introduction

Energy efficiency has been a research-intensive field for more than forty years, initially motivated by fossil fuel prices and afterwards by environmental concerns. As one of the main energy consumers, the industrial sector, and more specifically process industry, has become the target of energy efficiency research. Energy consumption in industry is mainly in the form of heat. Hence, the research in the field focuses primarily on heat recovery and waste heat valorisation. Pinch analysis (PA) is a technique proposed by Linnhoff and Hindmarsh [1] which uses thermodynamics and a graphical representation of the streams to obtain the maximum energy recovery (MER) between the processes. Although PA has proven to be effective in setting targets for energy consumption, it can yield scenarios that are practically inefficient because of direct heat exchange between the processes. Dhole and Linnhoff [2] developed total site analysis (TSA), a method based on PA, to overcome this drawback. In TSA, heat is recovered from processes by means of utilities (e.g. steam, hot water, hot oil) and transferred to other processes that require heating. HI methods using mathematical programming (MP) have emerged to find the optimal utility configuration satisfying the MER. Typically, the problem is formulated

with mixed integer linear programming (MILP), where the selection of the utilities is decided using binary variables and the size of each utility using continuous variables [3].

HI has been used for problems at different scales, including the optimization of a single process unit, a single plant and multiple plants. HI across industrial plants is typically achieved using intermediate fluids. Ahmad and Hui [4] used steam at different pressure levels to transfer heat between plants. Hackl et al. [5] proposed using a hot water loop to recover heat at low temperature in a chemical cluster. Rodera and Bagajewicz [6] studied both direct and indirect heat transfer across plants and concluded that more energy savings can be achieved by direct heat transfer and that steam is not always effective in indirect heat transfer.

Chew et al. [7] listed the layout of the plants as one of the crucial issues to be considered in heat integration. Several authors addressed this issue using different techniques. Wang et al [8] proposed a graphical method to consider interplant heat integration in parallel, split and series connection patterns and compared the energy savings and pipe length under different configurations. Song et al. [9] developed a graphical technique called interplant shifted composite curve to determine the maximum feasible heat recovery by indirect heat integration between plants. Chang et al. [10] proposed a mixed integer non-linear programming (MINLP) method to optimize waste heat integration between plants, including piping cost in the objective function. Stijepovic and Linke [11] presented a two-step optimization framework: in the first step, a linear programming (LP) problem is solved to obtain the maximum possible heat recovery; in the second step, a MINLP is used to achieve waste heat recovery with optimal design, accounting for piping cost.

A gap is identified when considering interplant heat integration since most of the existing methods do not consider the distance between plants in different locations. Some authors addressed the issue from an economic perspective, by including the cost of piping. However, the heat losses and the temperature drop due to the heat transfer over long distances have not been considered. This work focuses on heat losses resulting from interplant heat integration and proposes a MILP method based on [3] to obtain optimal heat integration scenarios for problems with multiple locations.

2. Methodology

2.1. Estimation of heat losses

The heat losses due to interplant heat transfer depend both on the pipe geometry and on the temperature of the heat transfer fluid. This work considers heat losses from pipes buried under ground as well as pipes above ground.

Heat losses from underground pipes

Heat distribution using pipes buried underground is common in urban district energy networks since it would be impossible to have the pipes above the surface in urban centres. This application is not very common in industry since urban planning rules do not apply to the industrial zones and installing pipes underground is costly. However, it is considered as an option in this work, since the ground provides insulation for the pipes, resulting in lower heat losses.

The formulas used in this work are adapted from the calculation of steady-state heat losses from buried pre-insulated district heating pipes (Bohm [12]) (Eqns. 1-5)

$$H = H \ + \frac{\lambda_g}{\hbar_{air}} \tag{1}$$

where H represents the depth at which the pipes are buried, λ_g the thermal conductivity of the ground and $\hbar_{air}$ the convective heat transfer coefficient of air.

$$R_m = \frac{1}{4\cdot\pi\cdot\lambda_g}\ln\left(1+\left(\frac{2\cdot H}{E}\right)^2\right),\quad R_g = \frac{1}{2\cdot\pi\cdot\lambda_g}\ln\left(\frac{4\cdot H}{D_i}\right),\quad R_i = \frac{1}{2\cdot\pi\cdot\lambda_i}\ln\left(\frac{D_i}{D_p}\right) \tag{2}$$

where R_m, R_g, and R_i are the thermal resistance of the mutual action of the two pipes (i.e. supply and return), the ground and the insulation material, respectively. E is the distance between the pipes, λ_i is the thermal conductivity of the insulation material, and D_i and D_p are the diameter of the insulated pipe and of the pipe itself, respectively.

$$U_1 = \frac{R_g+R_i}{\left(R_g+R_i\right)^2-R_m^2},\quad U_2 = \frac{R_m}{\left(R_g+R_i\right)^2-R_m^2} \tag{3}$$

$$\dot{Q}_{sup} = ((U_1 - U_2)\left(T_s - T_g\right) + \ U_2(T_s - T_r)) \cdot L_p \tag{4}$$

$$\dot{Q}_{ret} = ((U_1 - U_2)\left(T_r - T_g\right) - \ U_2(T_s - T_r)) \cdot L_p \tag{5}$$

where U_1 and U_2 are heat loss coefficients, L_p is the pipe length, T_s, T_r, and T_g are the temperatures of supply, return and ground, and $\dot{Q}_{sup}$ and $\dot{Q}_{ret}$ are the specific heat losses from supply and return.

Heat losses from above surface pipes

A simplified formulation is used to calculate the heat losses from above-ground pipes (Eq. 6-9). The temperature of the pipe is assumed to be equal to the temperature of the fluid flowing inside the pipe. This way, the convective heat transfer inside the pipe can be neglected.

$$\frac{1}{U} = \frac{1}{\hbar_{air}} + \frac{t_p}{\lambda_p} + \frac{t_i}{\lambda_i} \tag{6}$$

$$A = 2 \cdot \pi \cdot D_i \cdot L_p \tag{7}$$

$$\dot{Q}_{sup} = U \cdot A \cdot (T_s - T_{amb}) \tag{8}$$

$$\dot{Q}_{ret} = U \cdot A \cdot (T_r - T_{amb}) \tag{9}$$

where t_p and t_i are the thickness of the pipe wall and of the insulation material, respectively, U is the overall heat transfer coefficient, A is the surface area of the insulated pipe and T_{amb} is the temperature of the ambient air.

Heat losses in the supply and return pipes are subtracted from the total heat load of the stream (Eq. 10).

$$\dot{Q}_s = \dot{Q}'_s - \dot{Q}_{sup} - \dot{Q}_{ret} \tag{10}$$

where $\dot{Q}'_s$ and $\dot{Q}_s$ are the heat loads of the stream prior to and after heat losses.

2.2. Modified MILP formulation

Maréchal and Kalitventzeff's MILP formulation [3] is modified to take into account the location of the streams and the heat losses. The set of locations (***L***) is introduced and all units (***U***) and streams (***S***) are assigned to their corresponding locations by units of location

(**UL**) and streams of location (**SL**) sets. The streams that can exchange heat over distances are defined in a new set of interlocation streams ($\boldsymbol{IS} \in \boldsymbol{S}$) and added in **SL**. The temperature and heat load ($\dot{Q}_{s,t,k}$) of the interlocation streams are modified accounting for heat losses using Eq. 10. A stream can be split and used in several locations and to enforce the mass balance, a set of parents (**P**) and streams of parents in locations (**SP**) are defined to assign the usage of the same stream in different locations to the same parent.

The objective function (Eq. 11) is selected as the minimisation of the operating cost. The heat cascade equations (Eq. 12-14) ensure that heat flows from higher temperature intervals to lower ones. The objective function and the heat cascade are adapted from [3]. In the classical formulation, the heat cascade equations are written for each time step ($t \in \boldsymbol{T}$) and temperature interval ($k \in \boldsymbol{K}$). In the proposed formulation, they are also indexed over locations ($l \in \boldsymbol{L}$), to close the heat balance for each location.

$$\min_{y,f} \sum_{u=1}^{\boldsymbol{U}} \sum_{t=1}^{\boldsymbol{T}} (C_{u,t}^{op1} \cdot y_{u,t} + C_{u,t}^{op2} \cdot f_{u,t}) \cdot t^{op} \tag{11}$$

$$\sum_{u=1}^{\boldsymbol{U}} \sum_{s=1}^{\boldsymbol{SL}} \dot{Q}_{s,t,k} \cdot f_{s,t} + \dot{R}_{t,k+1,l} - \dot{R}_{t,k,l} = 0 \quad \forall\, t \in \boldsymbol{T},\ k \in \boldsymbol{K},\ l \in \boldsymbol{L} \tag{12}$$

$$\dot{R}_{t,k,l} \geq 0 \quad \forall\, t \in \boldsymbol{T},\ k \in \boldsymbol{K},\ l \in \boldsymbol{L} \tag{13}$$

$$\dot{R}_{t,1,l} = 0,\ \dot{R}_{t,k+1,l} = 0 \quad \forall\, t \in \boldsymbol{T},\ l \in \boldsymbol{L} \tag{14}$$

The contribution of this work is on determining the sizing factor of the streams ($f_{s,t}$). In the classical formulation, it is equal to the sizing factor of the unit ($f_{u,t}$) to which the streams belong. However, this must be modified as several streams are created to represent a stream in different locations. Eq. 15 is used instead, to set the sum of the sizing factor of a parent in different locations ($f_{p,l,t}$) to the sizing factor of the unit. Afterward, the stream sizing factors are fixed to the sizing factor of their parents (Eqn. 16).

$$\sum_{l=1}^{\boldsymbol{L}} f_{p,l,t} = f_{u,t} \quad \forall\, p \in P,\ u \in \boldsymbol{U}, t \in \boldsymbol{T} \tag{15}$$

$$f_{p,l,t} = f_{s,t} \quad \forall\, p \in P,\ l \in \boldsymbol{L}, s \in \boldsymbol{SP}, t \in \boldsymbol{T} \tag{16}$$

The descriptions of the parameters and variables used in the MILP formulation are given in Table 1.

Table 1. Description of the parameters and variables (bold) in the MILP formulation

Symbol	Description
$C_{u,t}^{op1}$	Fixed operating cost [€/h]
$C_{u,t}^{op2}$	Variable operating cost [€/h]
t^{op}	Operating time [h]
$\dot{Q}_{s,t,k}$	Heat from/to streams [kW]
$\boldsymbol{y_{u,t}}$	Binary variable to use a unit or not [-]
$\boldsymbol{f_{u,t}}$	Sizing factor of a unit [-]
$\boldsymbol{f_{p,l,t}}$	Sizing factor of a parent in a location [-]
$\boldsymbol{f_{s,t}}$	Sizing factor of a stream [-]
$\boldsymbol{\dot{R}_{t,k}}$	Residual heat in a temperature interval [-]

3. Case study and results

A case study with a single time step and two industrial plants that are 500 m apart from each other in both coordinates is considered (see Figure 1). Currently, both sites operate independently, using their own boilers and steam networks. However, the capacity of the boiler and of the steam network on Site2 can be extended in case the sites share their utility network.

Several scenarios are considered and compared with each other:

- Scenario 0 (s0): Both sites are operated at business as usual state. This scenario forms a basis for comparison with the energy saving solutions;
- Scenario 1 (s1): Heat integration is allowed within the boundaries of each site;
- Scenario 2 (s2): Heat integration is allowed within and between locations, without considering losses;
- Scenario 3 (s3): Heat integration is allowed within and between locations, considering losses for underground pipes. The steam network of Site2 is shared, Site1 does not have additional heating utilities;
- Scenario 4 (s4): Heat integration is allowed within and between locations, considering losses for the pipes above the ground. The steam network of Site2 is shared, Site1 does not have additional heating utilities.

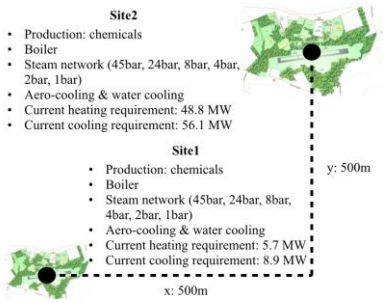


Figure 1. Case study layout

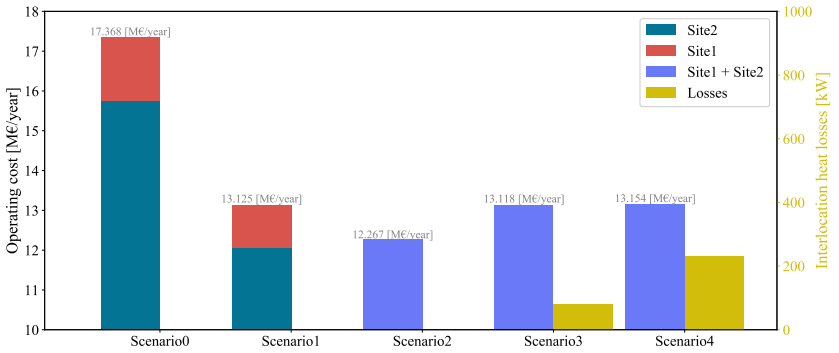


Figure 2. Operating cost and heat losses comparison of scenarios

The introduction of locations results in 10% increase in the number of variables, while a significant change in the solution time is not observed. The results of the case study are depicted in Figure 2. Comparing s0 and s1, when the two sites are optimised internally (i.e. without interplant heat transfer), the total operating cost reduces by 25% due to internal heat recovery on both sites and more efficient use of the steam network. In s2, the overall operating cost reduces by 30% compared to s0, due to heat recovery within and between sites. This scenario, representing the current state of the art in the literature, assumes no heat losses for interplant heat transfer. Hence, it sets the theoretical maximum heat recovery but it is likely to be impractical. In s3 the operating cost is 25% less compared to s0 and slightly less compared to s1. Thus, it is economically beneficial for the sites to use a common steam network with underground pipes. Heat is transferred

using 1 bar and 2 bar steam between sites, which results in 82 kW of heat losses. In s4, the heat losses increase to 232 kW, since the pipes are considered to be above-ground. This scenario yields in an operating cost lower than the business as usual case, but slightly higher than s1. Therefore, it is better to optimise both sites internally, instead of using a common steam network.

4. Conclusions

In this work, a mathematical programming method is proposed to solve heat integration problems with several locations in which interplant heat exchange occurs. While the state of the art either neglects layout issues or considers them only from an economic point of view by accounting for piping cost, the proposed method gives additional insights by considering heat losses due to heat transfer over long distances. The method is applied to a case study with two sites in different locations. The total operating cost of the sites reduces by 30% applying the state of the art methods however the reduction becomes 25% when interplant heat losses are taken into account. While interplant heat transfer is still beneficial when heat distribution pipes are underground, it is not economically attractive when the piping is above the ground. In order to have a better economic analysis, the cost of piping should be included as a post calculation or in the objective function. This will be addressed in future work. The method can be used in heat integration studies to properly consider heat transfer over long distances and therefore obtain more realistic results.

Acknowledgements

The authors acknowledge the support of EU H2020 EPOS project (Grant 679386) and of the Swiss Competence Center in Energy Research (SCCER-EIP).

References

[1] Linnhoff B, Hindmarsh E. The pinch design method for heat exchanger networks. Chem Eng Sci 1983;38:745–63.
[2] Dhole VR, Linnhoff B. Total site targets for fuel, co-generation, emissions, and cooling. Comput Chem Eng 1993;17:S101–9.
[3] Marechal F, Kalitventzeff B. Targeting the integration of multi-period utility systems for site scale process integration. Appl Therm Eng 2003;23:1763–84.
[4] Ahmad S, Hui DCW. Heat recovery between areas of integrity. Comput Chem Eng 1991;15:809–32.
[5] Hackl R, Andersson E, Harvey S. Targeting for energy efficiency and improved energy collaboration between different companies using total site analysis (TSA). Energy 2011.
[6] Rodera H, Bagajewicz MJ. Multipurpose Heat-Exchanger Networks for Heat Integration Across Plants. Ind Eng Chem Res 2001;40:5585–603.
[7] Chew KH, Klemeš JJ, Wan Alwi SR, Abdul Manan Z. Industrial implementation issues of Total Site Heat Integration. Appl Therm Eng 2013;61:17–25.
[8] Wang Y, Feng X, Chu KH. Trade-off between energy and distance related costs for different connection patterns in heat integration across plants. Appl Therm Eng 2014;70:857–66.
[9] Song R, Feng X, Wang Y. Feasible heat recovery of interplant heat integration between two plants via an intermediate medium analyzed by Interplant Shifted Composite Curves. Appl Therm Eng 2016;94:90–8.
[10] Chang C, Chen X, Wang Y, Feng X. An efficient optimization algorithm for waste Heat Integration using a heat recovery loop between two plants. Appl Therm Eng 2016.
[11] Stijepovic MZ, Linke P. Optimal waste heat recovery and reuse in industrial zones. Energy 2011.
[12] Bohm B. On transient heat losses from buried district heating pipes. Int J Energy Res 2000;24:1311–34.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50195-6

Techno-economic evaluation of multistage membrane combinations using three different materials to recover helium from natural gas

M. Abdul Quader[a,b], Simon Smart[a,c], Thomas E. Rufford[a,b*]

[a]*School of Chemical Engineering, The University of Queensland, St Lucia 4072, Australia*
[a]*Australian Centre for LNG Futures (ACLNGF), School of Chemical Engineering, The University of Queensland, St Lucia 4072, Australia*
[c]*Dow Centre for Sustainable Engineering Innovation, School of Chemical Engineering, The University of Queensland, Brisbane, QLD, 4072, Australia*
t.rufford@uq.edu.au

Abstract

Our study evaluates membrane processes to recover helium from natural gas as an alternative to the current cryogenic distillation technologies used in liquefied natural gas (LNG) plants. The process simulations were implemented in Aspen HYSYS v.10 to separate helium >90% purity at a recovery 99% of helium from a feed gas containing 1% He + 98.3% N_2 + 0.7% CH_4, which is a typical gas coming from a N_2 rejection unit (NRU) in an LNG plant. In this study, we designed six four-stage cascade membrane process configurations (cross-flow and counter-current flow pattern used) based on the glassy membranes such as Hyflon AD60X, Polyimide and Polysulfone: Concept-A-1, A-2, B, C-1, C-2 and D. A techno-economic analysis for the process configuration of >90% He purity was conducted using a net present value (*NPV*) method and internal rate of return (*IRR*). Advanced cascade membrane configuration concepts-C-2, C-1 and A-1 show the best performance in terms of upgraded-He production price respectively.

Keywords: Helium recovery, Multi-stage membrane process, Techno-economic analysis

1. Introduction

Membrane-based gas separation process could be complicated when the concentration of the target component in the inlet feed gas is comparatively low such as 0.5-4% helium concentration in the natural gas source (Xiong et al. 2017; Scholes and Ghosh 2017). In that case the membrane can neither achieve high purity permeate nor low-concentrate retentate. At such low feed-concentration, the selectivity and permeability of membranes available in market determine that a single-stage or two-stage membrane cannot produce high helium purity and recovery from natural gas even at a very high inlet pressure and/or over large membrane areas (Alders et al., 2017). The solution is a combination of few-stage membranes, in cascade, parallel or in series to reach higher qualities of permeate and retentate (Scholes and Ghosh 2016). Such arrangements result in higher capital costs (CAPEX) (due to high membrane area) and operation costs (OPEX) (due to high compression costs) making the appropriacy of membrane debatable for such conditions of low concentration feeds. Therefore, selection of the right option will require detailed techno-economical investigations. The success of membrane systems will be very much

dependent on process synthesis, advanced configuration, and design (Scholes et al., 2017).

This paper evaluates cascade membrane processes integrating two different materials in each design concept for helium recovery from natural gas. We introduce six multi-stage cascade membrane process design concepts using combinations of Hyflon AD60X, Polyimide and Polysulfone glassy polymeric membranes that are commercially available. An economic evaluation was completed for each design concept.

2. Approach to performance and cost estimation

2.1. Membrane materials and conditions for the simulation

For the multi-stage hollow-fibre membrane processes simulation, three commercially available glassy polymeric membranes–Hyflon AD60X, Polyimide and Polysulfone, used to produce a He stream with at least 90% purity at a recovery 99 % of helium from natural gas (Table 1). The same N_2 rejection unit (NRU) feed gas composition and parameters were used in all multiple-stage hollow fiber cascade membrane processes. The NRU exit gas conditions are shown in Table 2. NRU off-gas temperature, pressure and flow rate took from BOC Helium Plant, Darwin Australia (Lindemann et al. 2010).

Table 1. Permeabilities of selected glassy membrane materials and He-selectivities over N_2 and CH_4.

Materials	He (Barrer)	CH_4 (Barrer)	N_2 (Barrer)	He/CH_4 selectivity	He/N_2 selectivity	Ref.
Hyflon AD60X	476	3.03	9.46	157.1	50.30	(Macchione et al. 2007)
Polyimide	98.5	0.63	1.27	156	77.5	(Walker and Koros 1991)
Polysulfone	12	0.07	0.09	171	133	(McHattie et al., 1991)

1 barrer = 1×10^{-10} cm^3 (STP) cm cm^{-2} s^{-1} $cmHg^{-1}$

Table 2. Feed gas compositions and conditions of NRU off gas for all multi-stage membrane processes.

NRU-stream Feed Conditions		*Composition*	
Temperature	286 k (12. 85^{0}C)	He	1%
Pressure	120 kPa	N_2	98.3%
Flow arte	876.6 kmol/h	CH_4	0.7%

2.2. Membrane material integration and process combination

Six different membrane processes combination and material integration such as concept-A-1 & A-2, concept-B, concept-C-1 & C-2, and concept-D were designed and simulated. All processes were simulated with a target of 90% helium purity in product stream (peremate4) and fixed 99% recovery from the feed. Another stream output of the whole is retentate1 which is nitrogen-rich stream with small methane composition. All membrane simulations were undertaken in Aspen HYSYS simulation software package (AspenTech, Bedford, MA, USA) V.10 using the Peng-Robinson fluid package. Aspen HYSYS does not have a gas separation membrane process unit, so an in-house programmed module was incorporated into HYSYS. Crossflow, co-current and counter-current flow module pattern with or without sweep gas on the permeate side are

considered in this program in hollow fiber membrane module. Advanced process configurations are summarised in Figure 1 and description in Table 3.

Table 3. Descriptions of advanced multi-stage cascade membrane process design concepts

Design concepts	Descriptions of advanced multi-stage cascade membrane configurations
A-1	Cross-flow: High permeability (P) Polyimide material in module Mem1 & Mem2 + high selectivity (α) Polysulfone material in module Mem3 & Mem4.
A-2	Cross-flow: High permeability (P) Hyflon AD60X material in module Mem1 & Mem2 + high selectivity (α) Polysulfone material in module Mem3 & Mem4.
B	Cross-flow: High selectivity (α) Polysulfone material in module Mem1 & Mem2 + high permeability (P) Polyimide material in module Mem3 & Mem4.
C-1	Counter-current flow: High permeability (P) Polyimide material in module Mem1 & Mem2 + high selectivity (α) Polysulfone material in module Mem3 & Mem4.
C-2	Counter-current flow: High permeability (P) Hyflon AD60X material in module Mem1 & Mem2 + high selectivity (α) Polysulfone material in module Mem3 & Mem4.
D	Counter-current flow: High selectivity (α) Polysulfone material in module Mem1 & Mem2 + high permeability (P) Polyimide material in module Mem3 & Mem4.

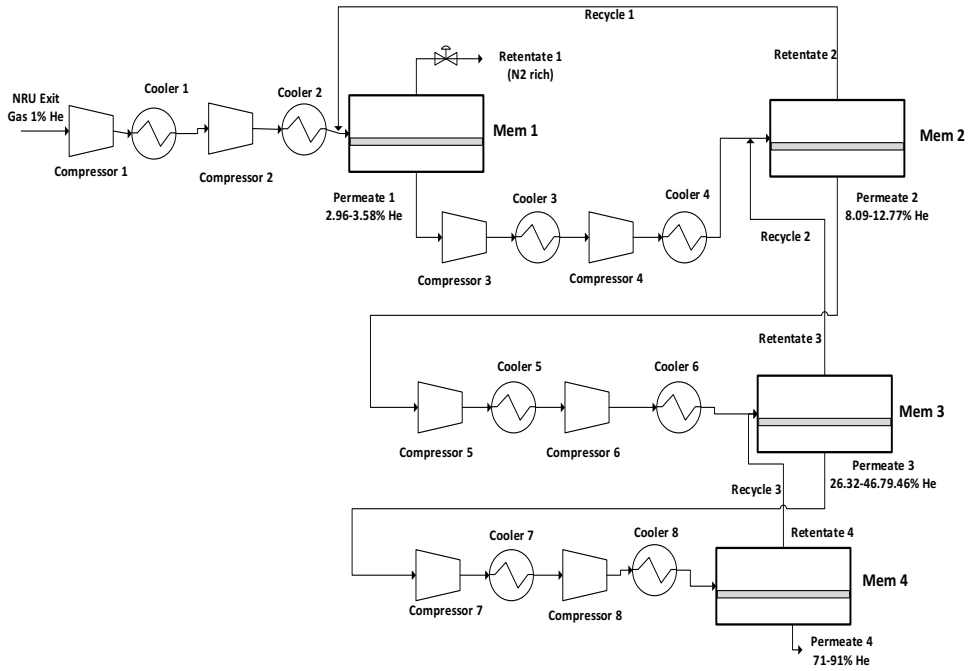


Figure 1. Illustration of four-stage cascade membrane processes with cross-flow and counter-current flow patterns used for Concept-A-1 & A-1, Concept-B, Concept-C-1& C-2 and Concept-D. All simulations are performed to produce >90% He with fixed 99% recovery from 1% He feed concentration at Nitrogen Rejection Unit off-gas.

3. Results and discussions

3.1. Simulation of multi-stage cascade membrane process concepts

Figure 2 (a-c) show the He purity, total membrane area, and compressor power required, in four-stages/cross-flow cascade membrane processes A-1 (Polyimide+ Polysulfone), A-2 (Hyflon AD60X+Polysulfone), and B (Polysulfone+Polyimide). Simulations were performed with a pressure range 19-25 bar to meet the target purity >90% He by fixing 99% recovery from the feed. In terms of permeate He mole fraction concept-A-1 shows better performance rather than concept-A-2 and B. Both design concept-A1 and A-2 ensure minimum required membrane area while concept-B shows less compressor energy conception.

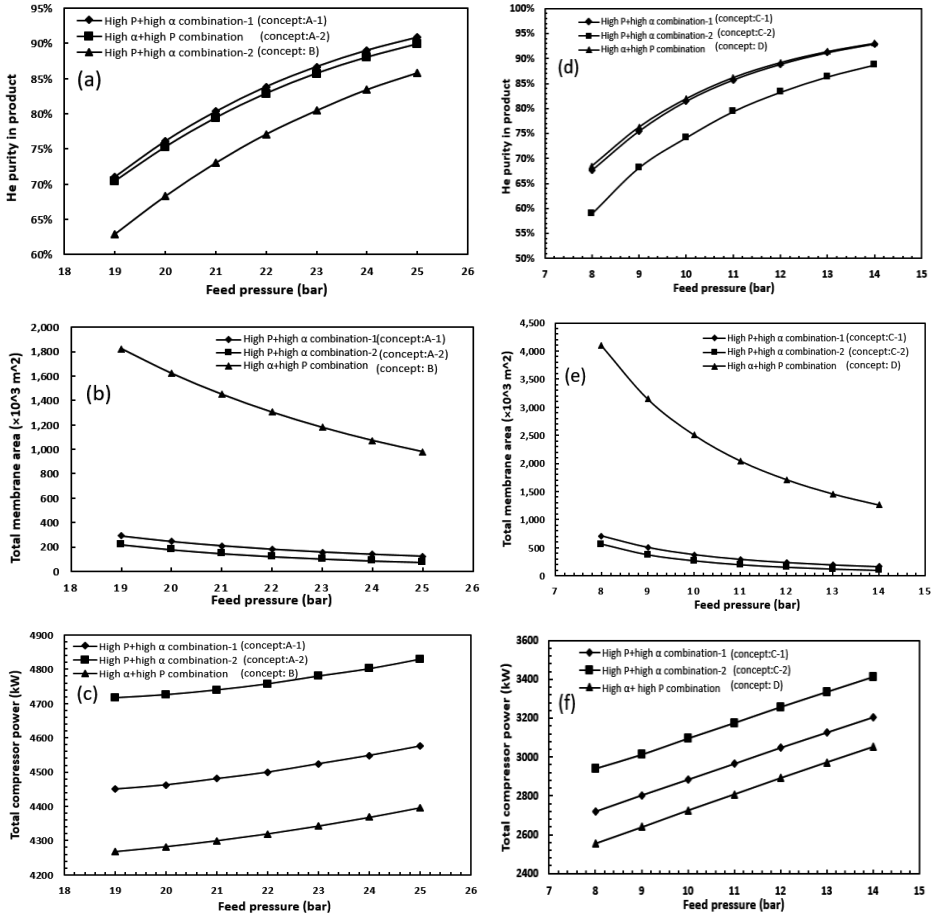


Figure 2. Helium purity in product, total membrane area, and required compressor power consumption in 4-stages/cross flow cascade membrane design concepts-A1, A-2 &-B (a-c),and 4-stages/counter-current flow cascade membrane design concepts-C-1, C-2 &-D (d-f).

On the other hand, Figure 2 (d-f) show results of four-stages/counter-current flow pattern cascade membrane processes C-1 (Polyimide+Polysulfone), C-2 (Hyflon AD60X+Polysulfone), and D (Polysulfone+Polyimide). These simulations were performed with pressure range 8-14 bar to meet the target purity >90% He with 99%

recovery from the feed stream. Cascade membrane design concept-C-1 and concept-D show same purity level against same feed pressure. Concept-D needs much more membrane area then C-1 and C-2 while compressor energy conception for concept-D is less. It was expected that high permeability membrane needs less membrane area and high selective membrane consume less compressor power and give higher purity.

3.2. Economic evaluation of multi-stage membrane processes

Economic analysis of the flowsheet was carried out only for the case that produced >90% He purity with 99% recovery. Techno-economic analysis (TEA) reported here incorporates capital cost estimates with relatively high contingency factors, reflecting a higher risk investment of first-of-a-kind (FOAK) facility. Aspen Process Economic Analyzer (APEA) v.10 was used to develop a conceptual process model for calculating the purchased cost of equipment (PCE) from equipment sizing and costing data. The economic model considered 35% company tax rate, 10% internal rate of return (*IRR*), and 15% project contingency. The total project capital cost was considered to depreciate, via straight-line depreciation method, over the 10 years of the economic life of the facilities. Membrane module cost considered \$50/m^2 for all stages during capital investment and replacement cost considered \$10/m^2 with 5 years membrane lifetime. This skid cost includes the cost of membrane modules, module housings, connecting valves, supporting structure, and instrumentation. The economic model was used to calculate minimum helium production price to reach a breakeven point for a 10% internal rate of return (*IRR*) for all selected design concepts of helium upgrading plant (i.e. >90% purity). The Economic evaluation was undertaken using an *NPV* (Net Present Value) method based on the amount of helium produced (thousand cubic feet, MSCF). The *NPV* calculation does not include high-pressure final compression of helium product to meet transport pressure specification.

Table 4. Helium production price (\$/thousand cubic feet) from natural gas with IRR, NPV value for different design concepts.

Membrane process concepts	**Feed pressure (bar)**	**Purity**	**Recovery**	**Total Power (kW)**	**Total Area (m^2)**	**US. \$/MSCF**	**IRR**	**NPV (US.\$ billions)**
Concept-A-1	25	>90%	99%	4575.8	1.23E+05	9.04	118.89%	2.107
Concept-B	25	>90%	99%	4396.6	9.81E+05	20.16	64.55%	1.984
Concept-C-1	13	>90%	99%	3127.7	1.93E+05	8.89	115.87%	2.109
Concept-C-2	14	>90%	99%	3413.5	9.54E+04	7.72	130.00%	2.122
Concept-D	13	>90%	99%	2972.4	1.46E+06	25.4	53.94%	1.927

Table 4 shows the results of *NPV* value, *IRR* and helium production price (\$/ thousand cubic feet) for selected flowsheets of upgraded helium (>90% purity). It is seen that concept-C-2 have maximum *NPV* of \$ 2.122 billion at 14 bar feed pressure with minimum helium production price \$ 7.72/MSCF. Membrane process design concept-C-1 and concept-A-1 can produce helium with the second (\$8.89/MSCF) and third (\$9.04/MSCF) most low price from natural gas where 1% He feed concentration respectively, which is much lower than current US crude helium (50-70%) auction price \$119.31/MSCF (2018).

Overall, a combination of high permeability membrane in the initial stage and high selective membrane in the last stage shows low cost-optimal and efficient process.

4. Conclusions and Outlook

In this paper, the effect of combined membrane process was evaluated in four-stage cascade hollow fiber membrane process to recover helium from natural gas where 1% helium concertation considered. Hollow fiber membrane module pattern such as cross-flow and counter-current were considered. To separate helium from a mixture of $He+N_2+CH_4$ with minimum 90% purity and 99% recovery, total six different process concepts were designed and simulated by using there different glassy polymeric membranes such as Hyflon AD60X, Polyimide, and Polysulfone. Design concept-C-2, C-1 and A-1 show the best performance for He upgrading respectively. The three-stage membrane process will be evaluated by integrating membrane materials to produce helium product such as crude-He, upgraded-He and A grade-He from feed 1-4% He with different recovery rate in the future.

Acknowledgements

This research was funded by The Australian Centre for LNG Futures (ACLNG) with additional scholarship support for Mr M. Abdul Quader provided from the Australian Government Research Training Program (RTP). We thank Dr Colin Scholes at the Cooperative Research Centre for Greenhouse Gas Technologies (CO2CRC) for sharing the membrane module code.

References

Alders, Michael, Dominik Winterhalder, and Matthias Wessling. 2017. 'Helium recovery using membrane processes', *Separation and Purification Technology*, 189:433-440

Lindemann, U, S Boeck, L Blum, and K Kurtcuoglu. 2010. "Turnkey helium purification and liquefaction plant for Darwin, Australia." In *AIP Conference Proceedings*, 271-74. AIP.

Macchione, Marialuigia, Johannes Carolus Jansen, Giuseppina De Luca, Elena Tocci, Marcello Longeri, and Enrico Drioli. 2007. 'Experimental analysis and simulation of the gas transport in dense Hyflon® AD60X membranes: influence of residual solvent', *Polymer*, 48: 2619-35.

McHattie, JS, William J Koros, and Donald R Paul. 1991. 'Gas transport properties of polysulphones: 1. Role of symmetry of methyl group placement on bisphenol rings', *Polymer*, 32: 840-50.

Scholes, Colin A, and Ujjal K Ghosh. 2017. 'Review of Membranes for Helium Separation and Purification', *Membranes*, 7: 9.

Scholes, Colin A, Ujjal Kumar Gosh, and Minh T Ho. 2017. 'The Economics of Helium Separation and Purification by Gas Separation Membranes', *Industrial & Engineering Chemistry Research*, 56: 5014-20.

Scholes, Colin A., and Ujjal Ghosh. 2016. 'Helium separation through polymeric membranes: selectivity targets', *Journal of Membrane Science*, 520: 221-30.

Walker, David RB, and William J Koros. 1991. 'Transport characterization of a polypyrrolone for gas separations', *Journal of Membrane Science*, 55: 99-117.

Xiong, L, N Peng, L Liu, and L Gong. 2017. "Helium extraction and nitrogen removal from LNG boil-off gas." In *IOP Conference Series: Materials Science and Engineering*, 012003. IOP Publishing.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50196-8

Design of Flexible Multiperiod Heat Exchanger Networks with Reducing Redundancy in Heat Transfer Areas

Lixia Kang[a], Yongzhong Liu[a*]

[a]*Department of Chemical Engineering, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, China*
yzliu@mail.xjtu.edu.cn

Abstract

This paper addresses a design procedure for flexible multiperiod heat exchanger network (HEN) with consideration of bounded disturbances in subperiods, aiming to minimize the total annual cost and reducing redundancy of heat transfer areas in the multiperiod HEN. In the proposed procedure, an initial multiperiod HEN is determined by the optimal stream matches and heat transfer area assignment of the HEN in each subperiod, which is then used as the basis for the design of flexible multiperiod HEN after identification of the critical points. The flexibility of the resulting multiperiod HEN is tested by calculating the flexibility index of each subperiod. The application of the proposed procedure is finally verified through a case study of a three-subperiod HEN. The results of this work are compared with those obtained by using the previous methods.

Keywords: multiperiod operation; heat exchanger network; flexible analysis.

1. Introduction

In process industries, stochastic shocks in both external environment and internal operating parameters may cause operational fluctuations in heat exchanger networks (HENs). Under these circumstances, the design of HEN can be achieved by partitioning the recurrent operational conditions into a finite number of subperiods with fixed operational parameters, which is referred as multiperiod HEN optimization problems.

Theoretically, the multiperiod HEN synthesis problem is formulated as a mixed-integer nonlinear programming (MINLP) model that can be solved by either simultaneous methods or sequential ones. Studies on simultaneous methods are mainly focused on the development of the decomposition strategies, such as the Lagrangean approach(Escobar et al., 2014), interval-based frameworks(Isafiade and Fraser, 2010) and the advanced algorithms(Ahmad et al., 2012). However, when the number of subperiods and/or streams in multiperiod HEN increases, these methods become time-consuming due to the complexity of the models and the limitation of the algorithms. Therefore, the MINLP models of multiperiod HEN are usually solved in a stepwise manner, where a promising structure of multiperiod HEN is determined in the synthesis step, and the operational parameters are optimized in the optimization step. Examples of this kind of methods are the simplified model methods where a simplified multiperiod model is built by fixing the promising stream matches that obtained through single period optimization(Kang et al., 2016) or multiperiod optimization(Isafiade et al., 2015), the representative subperiod methods where a representative subperiod is selected

according to the durations of subperiods(Kang et al., 2015) or the temperature profile of streams(Isafiade and Short, 2016), and the time sharing methods where the time sharing mechanism is introduced to assign the heat transfer areas(Jiang and Chang, 2013).

However, the abovementioned methods are mainly developed to meet the operational requirement of multiperiod HEN in nominal conditions. The fluctuations of operational parameters in subperiods are usually ignored. Moreover, the heat transfer areas are conventionally assigned to be the maximum ones required in all subperiods, which causes a remarkable redundancy in heat transfer areas of the multiperiod HENs. To some extent, this redundancy may render the HEN deviating from their optimal operational conditions, leading to unnecessary increases in both the capital cost and operational cost.

To address these problems, a method for design of flexible multiperiod HENs with reducing redundancy in heat transfer areas is proposed, taking the disturbances of parameters in subperiods into consideration. The rest of this paper is organized as follows. In section 2, the problem of the design of flexible multiperiod HEN is stated, followed by the proposed procedure in section 3. A case study for a three-subperiod HEN is adopted to illustrate the application of the proposed method in section 4, and the conclusions are drawn in section 5.

2. Problem statement

There are N_h hot streams and N_c cold streams, spanning N_s subperiods. Given are (1) nominal multiperiod conditions, including the inlet and outlet temperatures, heat capacity flow rates, heat transfer coefficients of all streams in each subperiod; (2) the parameters related to the utilities in each subperiod, such as the types, inlet temperatures, outlet temperatures and annual costs; (3) cost coefficients of heat transfer units, including the fixed cost coefficients, area cost coefficients and area cost exponents etc.; (4) the expected variations of uncertain parameters, mainly the inlet conditions. The main aim of the proposed method is to determine a cost-effective multiperiod HEN with less redundancy of heat transfer area and enough flexibility to accommodate the disturbances during operations.

3. Method for design of flexible multiperiod HENs

Figure 1 shows the proposed procedure for the design of flexible multiperiod HEN. In this procedure, an initial multiperiod HEN is first synthesized and taken as the basis to design the flexible multiperiod HEN. The flexibility of the multiperiod HEN is then tested through the calculation of the flexibility index of each subperiod.

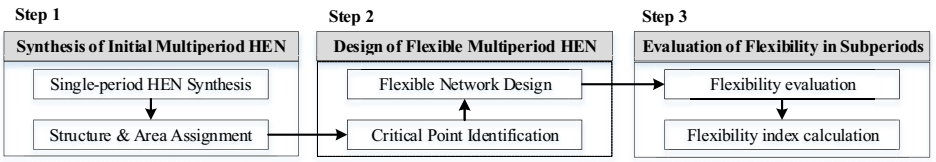


Figure 1 Procedure for design of flexible multiperiod HEN

3.1. Step 1: Synthesis of initial multiperiod HEN

In this step, an initial multiperiod HEN is first determined based on single period HEN synthesis, where the nominal multiperiod parameters are taken as the inputs. For the $p^{\text{-th}}$ subperiod, the outcome of the single period synthesis model is the set of the optimal stream matches $\mathbf{M}_p$, and the corresponding area set $\mathbf{A}_p$. The initial multiperiod HEN is then constructed according to the following rules.

$$\begin{cases} \mathbf{M^I} = \bigcup_{p=1}^{p=N_s} \mathbf{M}_p \\ \mathbf{A^I} = \{a^I \mid a^I = \min(a_p),\ a_p \in \mathbf{A}_p\} \end{cases} \tag{1}$$

It implies that the stream matches in the initial multiperiod HEN is a union of the optimal stream matches in all subperiods, and the minimum areas of heat exchangers in different subperiods are chosen as the initial areas. Note that although the structure determined by eq. (1) ensures the optimal operation of multiperiod HEN, the areas are insufficient to meet the requirement of the multiperiod operation.

3.2. Step 2: Design of flexible multiperiod HEN

Based on the initial stream matches and area assignment, the set of critical points **S**, which restrict the flexibility of the initial multiperiod HEN are identified by solving the model proposed by Novak Pintarič and Kravanja (2015). The Initial areas are then increased to accommodate the disturbances in subperiods by solving the following multi-scenario model.

$$\begin{cases} \min\limits_{d,y,x,\theta} TAC = c^{inv} + \sum_{s\in\mathbf{S}} c_s^{\text{oper}} \\ s.t. h(d_s, y_s, x_s, \theta_s \mid \mathbf{M}^I) = 0 \\ \quad g(d_s, y_s, x_s, \theta_s \mid \mathbf{M}^I) \le 0 \\ \quad g_{\text{A}}(d_s, y_s, x_s, \theta_s \mid \mathbf{M}^I) \le A^I + \Delta A \\ \quad \theta_s = \theta_s^N + \text{sgn}_s\, \delta_s \Delta\theta_s \\ \quad \delta_s = 1, \text{sgn} = \{-1, 1\}, s \in \mathbf{S} \end{cases} \tag{2}$$

where TAC is the total annual cost, consisting of the investment cost, c^{inv} and operating cost, c^{oper}. h and g are equality and inequality constraints. d, x, y and θ stand for design variables, state variables, control variables and uncertain variables, respectively. A^I and ΔA are the initial area and increased area of a stream match included in $\mathbf{M}^{\text{I}}$. θ^N are the nominal values of uncertain variables; sgn represents the direction of deviation from the nominal point towards the critical point either in the negative or positive sides. $\Delta\theta$ are the expected variations of uncertain variables. δ is the flexibility index, where the network is sufficiently flexible to meet the requirements of the expected change range of uncertain parameters when $\delta \geqslant 1$. Note that the flexibility index, in this case, is set to be one, aiming to obtain the least area increase that is just sufficient to meet the requirement of flexibility.

3.3. Step 3: Evaluation of flexibility in subperiods

In this step, the flexibility of the resulting flexible multiperiod HEN is evaluated by solving a flexibility index problem(Swaney and Grossmann 1985).

$$\begin{cases} F = \min\limits_{s} \max\limits_{y,x,\theta} \delta_s \\ s.t. \mathrm{h}(y_s, x_s, \theta_s \mid d) = 0 \\ \mathrm{g}(y_s, x_s, \theta_s \mid d) \le 0 \\ g_A(y_s, x_s, \theta_s \mid d) \le A^I + \Delta A \\ \theta_s = \theta_s^N + \mathrm{sgn}_s \cdot \delta_s \cdot \Delta\theta_s \end{cases} \tag{3}$$

The resulting multiperiod HEN is considered flexibility-qualified when the flexibility indexes in all subperiods are greater than one.

4. Case study

In this section, a three-subperiod HEN, including two hot streams and two cold ones, is employed to illustrate the proposed procedure. To consider the disturbances in subperiods, we assume ±5% disturbances of inlet temperatures and heat capacity flowrates of all streams in this work. In addition, the cost of the heat exchanger is calculated by $5500+4333A^{0.6}$, where a fixed cost efficient of 5500 is added to take the effect of the number of heat exchangers on TAC into account. The rest of the parameters are consistent with those in the literature(Jiang and Chang, 2013).

All the calculations in this work were carried out on GAMS 24.3. ALPHAECP was used as the global solver, and the CPLEX and IPOPTH were used as the MIP and NLP solver, respectively.

4.1. Design of the flexible multiperiod HEN with redundancy control

Table 1 Initial and final assigned heat transfer areas of the multiperiod HEN

Stream matches	Heat transfer areas/m^2			Initial areas/m^2	Additional areas/m^2	Final areas//m^2
	Subperiod 1	Subperiod 2	Subperiod 3			
(1,1,1)	80.9	66.8	3.9	3.9	77.1	81.0
(1,2,2)	102.4	83.2	72.2	72.2	46.0	118.1
(2,1,2)	85.3	264.3	188.1	85.3	171.7	257.0
(2,2,2)	/	14.6	/	0	23.7	23.7
(2,CU,3)	42.2	52.7	59.7	42.2	26.2	68.4
(HU,1,0)	9.6	8.1	14.3	8.1	22.3	30.4
Total	320.5	489.7	338.2	211.8	366.9	578.7

According to the proposed procedure in Figure 1, the single period HEN synthesis model is first solved to determine the optimal stream matches and area assignment of the HEN in each subperiod, as listed in the first four columns in Table 1. Then, the fifth column gives the stream matches and the corresponding areas of the initial multiperiod HEN according to the rules presented in eq. (1). After identifying the critical points, the increased areas and final areas of the multiperiod HEN are obtained by solving the flexible design model in eq. (2), as is given in the last two columns.

Results in Table 1 show that the initial areas of all six heat exchangers need to be increased to accommodate the disturbances in subperiods. Note that, for the stream match (2, 1, 2), the area after expansion (257 m^2) is smaller than the maximum one (264 m^2) obtained by single period synthesis in different subperiods. This indicates that it is necessary to adopt the minimum area of a heat exchanger in different subperiods as the initial area of the multiperiod HEN.

4.2. Discussion on the economic efficiency of flexible multiperiod HENs

To verify the economic-effectiveness of the resulting flexible multiperiod HEN, the flexibility of the multiperiod HENs obtained by the methods in literature are checked and improved to meet the requirement of flexibility in all subperiods by solving the proposed flexible design model presented in eq. (2).

Table 2 Comparison of the results obtained in this work and those obtained by previous methods

Methods	No. of units	Total areas/m^2	Capital cost/$	Operating cost /$ · y^{-1}	Total annual cost /$ · y^{-1}
Representative subperiod method	7	486	380,967	225,253	263,349
Time sharing method	6	602	416,567	176,077	217,733
The proposed method	6	579	407,641	175,865	216,629

Results in Table 2 show that, when the disturbances in subperiods are considered, the multiperiod HEN obtained by the proposed method has the lowest TAC. Although the total areas and capital cost of the multiperiod HEN obtained by the representative subperiod method are less than those obtained by the proposed method, the TAC is 22% higher than the one obtained in this work. Although the time-sharing method is effective for reducing area redundancy and lowering both capital cost and operating cost in multiperiod HEN, the proposed method does get a better result.

4.3. Discussion on the flexibility of multiperiod HEN

Figure 2 shows a comparison of flexibility indexes and area redundancy of the multiperiod HENs obtained in this work and those obtained on the basis of previous methods. Results in Figure 2(a) show that the multiperiod HEN obtained by the proposed method has sufficient flexibility to accommodate the disturbances in subperiods. It has the greatest flexibility in the subperiod 1, followed by the subperiod 3, and features the least flexibility in subperiod 2. Note that this order is exactly the same as that of area redundancy in Figure 2(b). It indicates that the more the area redundancy of the subperiod is, the greater the flexibility of the subperiod will be. Although the flexibility indexes of three subperiods obtained by the time-sharing method are similar to those obtained by the proposed method, the proposed method obtains less area redundancy. In contrast to the results obtained by these two methods, the flexibility indexes of three subperiods obtained by the representative subperiod method are much closer to each other, as well as the redundancy of their heat transfer areas.

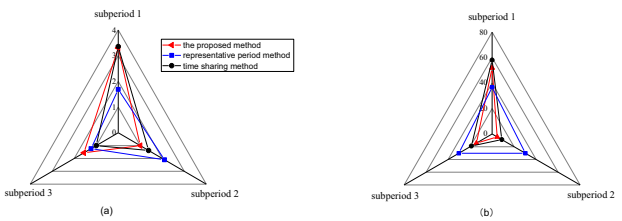


Figure 2 Comparison of results obtained by the proposed method and those in literature
(a) Flexibility index and (b) area redundancy in each subperiod

5. Conclusions

In this work, a stepwise method for the design of flexible multiperiod HEN is proposed, where the disturbances of the parameters in subperiods are considered. In the proposed method, an initial multiperiod HEN is determined on the basis of the single period HEN synthesis. Then, the critical points that restrict the flexibility of the initial multiperiod HEN are identified to finalize the design of the flexible multiperiod HEN by solving a multi-scenario model. The flexibility index of the resulting multiperiod HEN in each subperiod is finally calculated to verify the effectiveness of the proposed method. The results of the case study show that the proposed method can reach a cost-effective multiperiod HEN with sufficient flexibility and less area redundancy.

Acknowledgments

The authors gratefully acknowledge funding by the projects (No. 21676211 and No.21376188) sponsored by the National Natural Science Foundation of China (NSFC).

References

A. J. Isafiade, D. M. Fraser, 2010, Interval based MINLP superstructure synthesis of heat exchanger networks for multi-period operations, Chemical Engineering Research and Design, 88, 10, 1329-1341.

A. Isafiade, M. Bogataj, D. Fraser, Z. Kravanja, 2015, Optimal synthesis of heat exchanger networks for multi-period operations involving single and multiple utilities, Chemical Engineering Science, 127, 175-188.

A. J. Isafiade, M. Short, 2016, Simultaneous synthesis of flexible heat exchanger networks for unequal multi-period operations, Process Safety and Environmental Protection, 103, Part B, 377-390.

D. Jiang, C.-T. Chang, 2013, A New Approach to Generate Flexible Multiperiod Heat Exchanger Network Designs with Timesharing Mechanisms, Industrial & Engineering Chemistry Research, 52, 10, 3794-3804.

M. Escobar, J. O. Trierweiler, I. E. Grossmann, 2014, A heuristic Lagrangean approach for the synthesis of multiperiod heat exchanger networks, Applied Thermal Engineering, 63, 1, 177-191.

M. I. Ahmad, N. Zhang, M. Jobson, L. Chen, 2012, Multi-period design of heat exchanger networks, Chemical Engineering Research and Design, 90, 11, 1883-1895.

L. Kang, Y. Liu, J. Hou, 2015, Synthesis of multi-period heat exchanger network considering characteristics of sub-periods, Chemical Engineering Transactions, 45, 49-54.

L. Kang, Y. Liu, L. Wu, 2016, Synthesis of multi-period heat exchanger networks based on features of sub-period durations, Energy, 116, Part 2, 1302-1311.

R. E. Swaney, I. E. Grossmann, 1985, An index for operational flexibility in chemical process design. Part II: Computational algorithms, AIChE Journal, 31, 4, 631-641.

Z. Novak Pintarič, Z. Kravanja, 2015, A methodology for the synthesis of heat exchanger networks having large numbers of uncertain parameters, Energy, 92, Part 3, 373-382.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50197-X

Integrated scheduling of on-line blending and distribution of oil products in refinery operation

Lijie Su[a], Lixin Tang[a], David E. Bernal[b], Ignacio E. Grossmann[b], Bowen Wang[a]

[a] *Institute of Industrial and Systems Engineering, Northeastern University, Shenyang 110819, P. R. China*
[b] *Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh 15213, PA, USA*
sulijie@ise.neu.edu.cn

Abstract

This paper addresses the integration of scheduling of on-line blending and distribution for refinery oil products, which needs to determine simultaneously the blending recipes and the distribution times of oil products in every oil pipeline, satisfying the time-varying demand and product properties specifications for each oil product. We formulate a novel hybrid Mixed-Integer Nonlinear Programming (MINLP) and Generalized Disjunctive Programming (GDP) model for the integrated scheduling problem. Instead of using a general reformulation, we design an improved logic-based Outer-Approximation to solve the hybrid model directly. Numerical experiments demonstrate the usefulness and efficiency of the proposed model and the solution method.

Keywords: On-line blending, scheduling, Mixed Integer Nonlinear Programming, Generalized Disjunctive Programming, Outer Approximation

1. Introduction

The blending and distribution operations are placed at the end of refinery operations, which determine the qualities of oil products and part of production and transportation cost. On-line blending and distribution operation in a refinery is an integrated process, which blends the oil products in pipelines and directly distributes it to customers. Different from traditional off-line blending and distribution operations, on-line blending and distribution needs fewer blending units and storage tanks for different kinds of oil products. Research on the on-line blending and distribution scheduling can optimize the blending recipes, control the qualities of oil products, and reduce the product inventory and market response time.

Pinto et al. (2000) addressed the on-line blending scheduling as a subproblem of the planning in a refinery production. The authors proposed a mathematical model that represents this scheduling problem without consideration of nonlinear blending properties. Jia and Ierapetritou (2003) considered the gasoline blending and distribution scheduling problem, which is off-line blending operation. Xue et al. (2005) proposed a MINLP model for off-line blending problem, which determined the blending recipe through a predicting function for oil product qualities. Li et al. (2005) presented a nonlinear blending functions for components and oil products. Méndez et al. (2006) considered the off-line batch blending scheduling problem with varying blending recipes. Kelly et al. (2014) proposed a new method to optimize both the component

distillation curve and the blending recipes using monotonic interpolation. We can notice that most of the previous work focused on the blending operations including nonlinear blending functions, but few aimed at the integration of on-line blending and distribution.

In this work, we define the on-line blending and scheduling problem with consideration of nonlinear blending operation and distribution in oil product pipelines described in Section 2. We formulate a hybrid MINLP and GDP model to descripe the problem in Section 3. In order to solve the proposed model, we design a logic-based Outer Approximation (OA) algorithm in Section 4. Section 5 presents some numerical experiments, to show the validity of the model and efficiency of the solution method. Finally we give a conclusion for this work.

2. Problem description

The refinery on-line blending and distribution process is shown in Fig. 1.

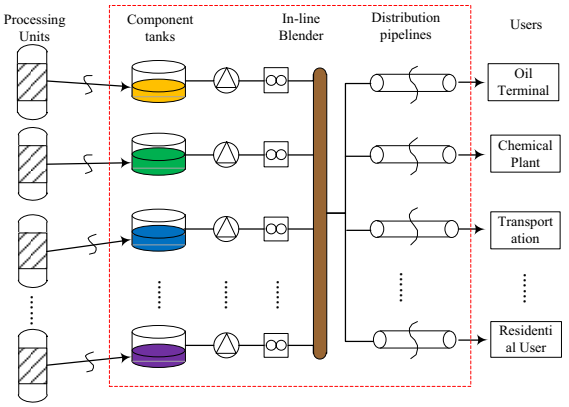


Fig. 1 Illustration of on-line blending and distribution process in a refinery

All kinds of component oils are produced from refinery units and stored respectively in the component tanks. Through pumps and valves, different kinds of component oils are injected into the in-line blender. After the blending operation, the oil products are transported into the long-distance oil product pipelines, which are connected with terminal users of the refinery. We consider the standing times in in-line blender negligible. The consistency for the pipelines is considered, which is once the blending operations happen, that they continue until the end of scheduling horizon.

The model presented here receives as parameters:

1) Scheduling horizon H; 2) Set of component oils, supplying capacity, initial inventory quantity, properties; 3) Set of pipelines, transportation capacity; 4) Set of oil product, demand quantity in each pipeline, properties standard; 5) Parameters of blending operations, bounds of blending ratios for components to each oil products; 6) Operations constraints; and 7) Set of cost, component oil cost, pumping cost, changeover of different oil products in pipelines.

Using these parameters, it determines:

1) Blending recipes for each oil product in every pipeline; 2) Properties of oil product in every pipeline; 3) Inventory quantity of each component oil during each time periods; and 4) Distribution time of each oil product in pipelines. The objective function is to minimize the total cost, which includes the cost of blended components, blending operations, changeover of oil products in pipelines and penalties for extra qualities of oil products.

3. Mathematical Formulation

The discrete-time representation is adopted in the following mathematical formulation. We use the following indices: i to refer to a blending component, j as an oil product pipeline, p as oil product and t as discrete time point. The objective function of the on-line blending and scheduling problem is to minimize the total cost, as Eq. (1).

$$\min Z = \sum_{i,j,p,t} \left(\left(C_i + CP_i\right) \cdot FI_{ijpt} + CN_i \cdot V_{it} + CL_p^{ON} \cdot (OL_{jpt} - ON_p^L) + \sum_{p'} CH_{pp'} \cdot w_{jpp'} \right) \tag{1}$$

Here, C_i is the unit cost of component i, CP_i is the cost of pumping operation for component i, FI_{ijpt} is the flow rate of component i to blend oil product p in pipeline j at time t, CN_i is the unit inventory cost of component i, V_{it} is inventory quantity of component oil i in time t, CL_p^{ON} is the penalty for the extra Octane number of oil product p, OL_{jpt} is the variable for Octane number of oil product p in pipeline j at time t and ON_p^L is the lower bounds of the Octane number for oil product p, $CH_{pp'}$ is the changeover cost between oil products in pipelines, $w_{jpp'}$ is the binary variable whether exits a changeover from p to p' in pipeline j.

The mass balance equations of component tanks are considered with general inventory constraints. The inventory quantity must be within the limiting ranges. The upper bound of component inventory is storage capacity. The lower bound is the requirement of blending operation, which is used to ensure the constant component properties.

For oil pipelines, the flow rate of one oil product is the total of the injected components. The ratio relationship between the amount of components and oil product is the blending recipe. The properties of oil products, like Octane number, Sulphur content and RVP, are calculated based on nonlinear blending functions (Alhajri and Elkamel, 2008), which should satisfy the national standard criterions for each kind of oil product. The demand quantity of all kinds of oil product in each pipeline must be satisfied.

The time constraints include the starting and ending times of blending oil product p in pipeline j. The starting blending time must be prior to the ending time. Once an oil product is beginning to blended, which must be ended on the horizon.

The disjunctive constraint is introduced as Eq. (2). X_{jpt} is the boolean variable, which is true when blended oil product p in pipeline j at time point t. The blending ratio R_{ijpt}, the component flow FI_{ijpt} and the flow of oil product FL_{jpt} should be within proper ranges, given by their bounds. And the properties of blended oil product, like Octane number, sulfur content and RVP for gasoline, must satisfy the quality standards. If X_{jpt} is false, all the left variables are equal to zero.

$$\begin{bmatrix} X_{jpt} \\ \sum_i R_{ijpt} = 1, R_{ip}^L \le R_{ijpt} \le R_{ip}^U \\ L_j^L \le FL_{jpt} \le L_j^U \\ R_{ip}^L \cdot L_j^L \le FI_{ijpt} \le R_{ip}^U \cdot L_j^U \\ ON_p^L \le ON_{jpt}, SL_{jpt} \le S_p^U, RL_{jpt} \le RVP_p^U \end{bmatrix} \vee \begin{bmatrix} \neg X_{jpt} \\ R_{ijpt} = 0, FI_{ijpt} = 0 \\ FL_{jpt} = 0 \\ ON_{jpt} = 0 \\ SL_{jpt} = 0 \\ RL_{jpt} = 0 \end{bmatrix} \quad \forall j, p, t \tag{2}$$

The Eq. (3) is the logic proposition, which means there is only one oil product being blended in one time period, or the pipeline is idle.

$$\left(\vee_p X_{jpt} \right) \vee X_{jt}^{Free} \quad \forall j, t \tag{3}$$

Here the boolean variable X_{jt}^{Free} is true if the pipeline j at time t is idle.The constraints of operation rule include that there is only one blending operation per oil product in a pipeline if there is demand for it in this pipeline.

For the logic variable X_{jpt}, we also define responding binary variable x_{jpt} to formulate the assignment constraints. The relationships between the assignment variables and changeover variables are presented with algebraic equations. The unique sequence constraints and the number of changeovers are also included in the presented formulation.

The proposed mathematical model is a hybrid MINLP and GDP model with nonconvex nonlinear functions, such as the blending equations and the octane number functions.

4. Solution method of Logic-based OA

Based on the structure of the proposed model for the on-line blending scheduling of oil products, we design a logic-based OA solution method, which solves the proposed model directly without transforming it into an MINLP problem. Making use of the logic relationships behind the formulation can derive tighter bounds (Turkay&Grossmann, 1996).

The presented logic-based OA is also decomposition methods, which divides the hybrid model into a MILP master problem and a NLP subproblem. The NLP subproblems are obtained by fixing integer and logic variables. When the NLP subproblem is feasible, we get an upper bound of the original minimization problem. The master problems are constructed with gradient based linearization generated at the NLP solutions for the nonlinear constraints (OA cuts) and the convex hull reformulation for the constraints in disjunctive terms. The solution of the master problem is a local lower bound of the original problem, given that it is a relaxation of the original problem. Iterating the NLP subproblem and MILP master problem, we can continuously decrease the gap between lower and upper bounds given by their solutions and terminate the solution process when the gap is within a tolerance, or the MILP problem becomes infeasible. The procedure of the logic-based OA methods is as follows.

Step1. Design a heuristic to obtain few initial scheduling solutions of oil-product in the pipelines, which satisfies the end demands for oil products.

Repeat.

Step2. Construct the NLP subproblems by fixing the assignment and logic variables in the scheduling solutions.

Step3. Solve the NLPs, where feasible or unfeasible solutions of the original problem can be obtained. The upper bound of the original problem is updated with the minimal objective value among all feasible NLP subproblems.

Step4. Based on the solutions of NLPs, we construct the master problem with OA linearization cuts for nonlinear equations and convex hull reformulation constraints for disjunctive terms.

Step5. Solve the constructed master MILP problem, which provides a lower bound, and simultaneously obtaine a new solution for the integer, logic variables.

Until the termination conditions are satisfied.

The heuristic for generating the initial distribution scheduling is sequencing the demanding oil products in each pipeline at once. Here the changeovers of oil products are not considered in the initial step.

5. Numerical experiments

Our proposed solution method is implemented and used to solve three study cases. The network includes 5 components, 2-4 pipelines, 2 kinds of oil products. The range of scheduling horizon varies from 5 to 10 hours, and 3 kind of properties of oil product are considerated (Pinto et al., 2000).

The implementation is made in GAMS 23.7 (Bussieck&Meeraus, 2004) and solved using a desktop running Windows 10, with Intel Core 2 Duo processor, CPU 3.3 GHz and RAM 4.00 GB. The NLP solver is CONOPT3.14, and MILP solver is CPLEX12.3.

The computational results of the three cases are compared against LOGMIP2.0 solver (Vecchietti&Grossmann, 1999). Here the big-M reformulation is used in LOGMIP. For the MINLP model of LOGMIP, the numbers of binary variables in three cases are 88, 148, 176, the numbers of continuous variables are 618, 1108, 1170, the number of constraints are 1106, 1981, 2060. The solution iterations of LOGMIP are 4, 5, 4, and CPU solution times are 3.9, 8.9, 8.6 seconds. The problem size of our proposed hybrid model includes the numbers of boolean variables 24, 44, 48, integer variables 64, 104, 128, disjuntive term constrains 24, 44, 48. We construct 2 initial boolean and integer solutions to generate 2 NLP sublproblems. The solution iterations of our logic-based OA is 1, 1, 3, and CPU solution times changes from 1.5, 7.7, 4.5 seconds. For the same case, two solution methods obtain the same optimal solution. Our proposed method needs fewer iterations and shorter CPU time than LOGMIP solver. The reason is that we directly solve the proposed model otherwise solve the big-M reformulation as LOGMIP solver.

The optimal scheduling solution for case 3 and the optimization blending recipes for $97^{\#}$ gasoline is shown in Fig. 2. The Octane Numbers of the blended gasoline $97^{\#}$ are 97.207, which is close to the criteria value and the sulphur content of gasoline $97^{\#}$ is 28.383 ppm, which is near half of the criteria value 50. Our proposed method provides a optimization scheduling and blending recipes simultaneously for on-line blending and distribution process.

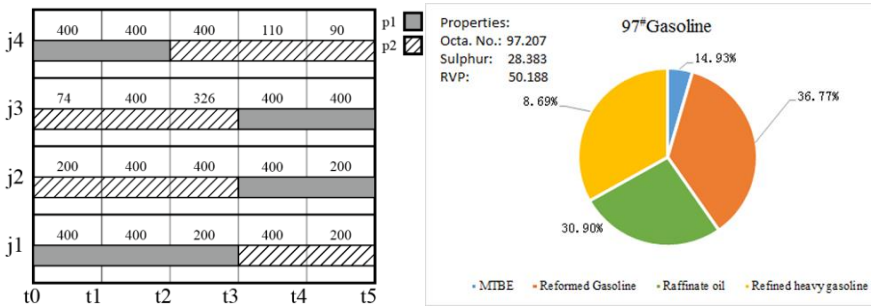


Fig. 2. Scheduling results of case 3 and blending recipe for one oil product

6. Conclusions

In this paper, we discussed an integrated scheduling for on-line blending and distribution problem with simultaneous consideration of blending recipes and distribution schedule. Based on discrete-time representation, a hybrid MINLP and GDP model is formulated. We also designed a logic-based OA method to solve the proposed model. Using some numerical experiments, a comparison with the LOGMIP solver demonstrates the usefulness and efficiency of the proposed model and solution method.

Acknowledgment

This work was supported by Major International Joint Research Project of National Natural Science Foundation (71520107004), National Natural Science Foundation (61374203), 111 Project (B16009).

References

I. Alhajri, A. Elkamel, 2008, A nonlinear programming model for refinery planning and optimisation with rigorous process models and product quality specifications. Inter J Oil, Gas and Coal Tech, 1(3), 283-307.

M.R. Bussieck, A. Meeraus, 2004, General algebraic modeling system (GAMS). In: Optimization, Modeling Languages in Mathematical, Kallrath, J. (Eds.), Applied Optimization. Springer, US, pp. 137-157.

Z.Y. Jia, M. Ierapetritou, 2003, Mixed-Integer Linear Programming Model for Gasoline Blending and Distribution Scheduling, Ind. Eng. Chem. Res. 42, 825-835

J.D. Kelly, B.C. Menezes, I.E. Grossmann, 2014, Distillation Blending and Cutpoint Temperature Optimization Using Monotonic Interpolation. Ind. Eng. Chem. Res. 53, 15146-15156.

C.A. Méndez, I.E. Grossmann, I. Harjunkoski, P.Kaboré, 2006, A simultaneous optimization approach for off-line blending and scheduling of oil-refinery operations, Comput Chem Eng, 30, 614-634.

J. Li, L.C. Liao, Y.J. Tan, 2005,The model of gasoline blending optimization. J. Natl. Uni. Deff. Tech. 27(3), 125-128. (Chinese)

J.M. Pinto, M. Joly, L.F.L. Moro, 2000, Planning and scheduling models for refinery operations, Comput Chem Eng, 24, 2259-2276.

M.Turkay, I.E. Grossmann, 1996, Logic-based MINLP algorithms for the optimal synthesis of process networks, Comput Chem Eng, 20(8), 959-978.

A.Vecchietti, I.E. Grossmann, 1999, LOGMIP: A disjunctive 0-1 nonlinear optimizer for process system models. Comput Chem Eng, 23, 555-565.

M.S. Xue, Z.K. Li, G. Wu, D.M. Sun, 2005, A scheduling model for oil products blending and its application. Petro Proc Petroch, 36(3), 64-68. (Chinese)

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50198-1

Emission Conscious Scheduling of Crude Unloading, Transferring, and Processing for Petroleum Refineries

Jialin Xu[a], Honglin Qu[a], Jian Zhang[a], Sujing Wang[b], Qiang Xu[a*]

[a]*Dan F. Smith Department of Chemical Engineering, Lamar University, Beaumont, TX 77710, USA*
[b]*Department of Computer Science, Lamar University, Beaumont, TX 77710, USA*
Qiang.xu@lamar.edu

Abstract

Emission reduction becomes increasingly important for petroleum refineries nowadays as more environmental concerns and stringent regulations have been posted. Cost-effective solution strategies require emission reductions to be addressed from the entire plant point of view, where emission source generation and utilization should be well balanced. In this paper, a new methodology framework and a general plant-wide scheduling model have been developed for emission-conscious crude unloading, transferring, and processing (ECUTP) operations, which couple generic production activities and characterizations of major air emissions from refineries, such as CO_2, VOCs, NO_X, and PM. The efficacy of the proposed ECUTP scheduling model has been demonstrated by a large-scale case study.

Keywords: Refinery processing, production scheduling, MINLP, optimization.

1. Introduction

Scheduling of front-end crude supply and refinery operations is an important part of petroleum supply-chain management. Emissions from refineries are significant and the emission reduction is a very challenging task. It involves a large-scale complex manufacturing system, which consists of over 17 major facilities with more than 85 main operation units, and thousands of streams. Due to the increasingly strict economic competitions and environmental regulations, refineries in one hand are eager to improve production solutions to leverage profitability margins in nowadays volatile market; on the other hand, they have to pursue cost-effective pollution prevention technologies to comply with more stringent environmental mandates, such as air emission requirements.

Cost-effective solution strategies require emission reductions to be addressed from the entire plant point of view. Mendez et al. presented a simultaneous optimization approach for blending and scheduling of refinery plant (Méndez et al. 2006). Meanwhile, Pinto et al. presented a general modelling framework for petroleum supply chain optimization (Neiro and Pinto, 2004). Very recently, Xu et al. simultaneously optimized crude unloading, transferring, and processing (CUTP) operations (Xu et al. 2017a). It should be highlighted that plant-wide scheduling model with emission characterization is still lacking based on our conducted literature survey.

In this paper, a new methodology framework and a general plant-wide scheduling model have been developed for emission-conscious CUTP (ECUTP) operations. Major

emissions from refineries: such as CO_2, volatile organic compounds (VOCs), nitrogen oxides (NO_X), and particulate matters (PM) are characterized and quantified. The scheduling model is a large-scale mixed integer nonlinear programming (MINLP) problem. The efficacy of the proposed model has been demonstrated by a case study.

2. Methodology framework

The methodology starts with collecting economic, process, and initial operating data. Since this paper addresses emission characterization from the system point of view to optimally balance the emission source generation and utilization, the emission conscious plant-wide scheduling model is the major focus. Based on the collected data, a continuous time CUTP scheduling model will be developed. Then, to address the profit maximization and air emissions simultaneously, the emission characterization formulas should be coupled, which help generate emission conscious CUTP scheduling model, called ECUTP scheduling model. The model will be solved with OA based decomposition method (Duran and Grossmann, 1986) and comprehensive analysis based on the scheduling results will be conducted.

3. Emission characterized CUTP scheduling model

The whole scheduling model contains two main sections: (i) a front-end crude oil transfer model, which covers crude movement from vessels to portside storage tanks, inland charging tanks, and refinery CDUs; (ii) a refinery crude processing model, which addresses the detailed processing strategy based on the given input of crude blends. Since the integrated crude scheduling model is very complicated, including more than 200 equations, for conciseness, only emission characterization equations are elaborated in the following section. The detailed CUTP model is based on the previous paper (Xu et al. 2017a) with modest modifications.

A typical refinery process includes crude distillation, cracking, coking, reforming, hydrotreating, product blending, gas processing, and sulfur recovering facilities. To address air emission reductions in a refinery, spatial emission allocations in the plant need to be first identified. Then, emission species at each allocation and their emission quantities under process operating conditions have to be identified.

Emission from heating furnaces and processing units. Most refinery facilities employ furnaces for heating inflow streams. When fuel gas or heavy oil utilities are burned, a large quantity of air emissions will be generated, such as CO_2, VOCs, NO_X, and PM. Emissions may come from two types of processing facility:(i) The hydrogen plant generates high-purity hydrogen for hydrocracking and hydrotreating facilities. (ii) In the FCC unit, coke is generated and deposited on the catalyst surface. It has to be burnt immediately in a catalyst regenerator to recover catalyst functionality. This burning process will generate large amounts of waste gas containing CO_2, VOCs, NO_X, and PM.

Emission from flaring system and crudes storage and transport. Flaring system is designed to burn out waste gas under normal conditions, or large amounts of off-spec product stream for air pollution mitigation. In this study, we only consider normal operation conditions of refineries. We could expect the flaring emission amount is near zero based on optimized scheduling solutions. And the storage and transport of crudes produce emissions of VOCs and toxic air pollutants. Most VOC emissions from upstream processes may be from evaporation of storage tanks, hoses, and vehicles; equipment leakage; as well as wastewater treatment systems.

Table 1. Emission characterization under various sources

Emission Source	Main Emission Species	Emission Factor from Literature (lb/1,000 barrels)	Emission Factor (T/kT)
Dry gas	CO_2	---	2864.79
	VOCs	5.8	1.25269
	NOx	140	30.2372
	PM	3	0.64794
LPG	CO_2	---	3014.20
	VOCs	6.5	1.40387
	NOx	150	32.397
	PM	3.5	0.75593
Fuel oil	CO_2	---	3233.13
	VOCs	1.28	0.16763
	NOx	55	7.20265
	PM	7	0.91670
Burning coke	CO_2	---	3655.62
	VOCs	N/A	N/A
	NOx	71	9.50574
	PM	45	6.02476
Crude Oil	CO_2	N/A	N/A
	VOCs	12.12	0.042
	NOx	N/A	N/A
	PM	N/A	N/A

Overall, major emission sources under normal operation conditions include dry gas, LPG, fuel oil, burning coke and evaporation of crude oil. As reported, every emission source would have some emission factor, which can characterize its emission distribution with respect to per unit weight of the emission source (Kassinis, 1998).

Table 1 has summarized the emission characterizations under various emission sources. The unit of the original formulas is lb/1,000 barrels, which should be transferred to SI unit as T/kT. The transfer transformation is:

$$\phi_{SI} = \phi_{EN} \frac{0.45359 \times 10^{-3}}{1000 \times 0.15899 \times 10^{-6} \rho} \tag{1}$$

where ϕ_{SI} is the emission factor in SI unit as T/kT, ϕ_{EN} is the emission factor in lb/1,000 barrels. 0.45359×10^{-3} is the conversion factor from pound to ton. 0.15899 is the conversion factor from barrel to m^3. ρ is the density of emission source, and the unit of ρ is kg/m^3. The transferred emission factors are listed in the 4th column in Table 1.

It should be noted that the CO_2 emission is estimated through mass balance. Since CO_2 comes from burning of hydrocarbon (C_xH_y), the general reaction formula is:

$$C_xH_y + (x + \frac{y}{4})O_2 = xCO_2 + \frac{y}{2}H_2O \tag{2}$$

Meanwhile, a part of hydrocarbon will be transferred to VOCs or PM. The left amount is assumed to be totally transformed to CO_2. To calculate the CO_2 emission, the average

molecular weights of all emission sources, VOCs, and PM should be known. They can be estimated based on the plant data. Thus, the CO_2 emission factor can be obtained as:

$$\phi_{SI,\mathrm{CO}_2} = 1000\frac{44x}{12x+y} - \phi_{SI,VOCs}\frac{44x_{VOCs}}{12x_{VOCs}+y_{VOCs}} - \phi_{SI,PM}\frac{44x_{PM}}{12x_{PM}+y_{PM}} \tag{3}$$

where ϕ_{SI,CO_2} is the emission factor of CO_2 in SI unit as T/kT. $\phi_{SI,VOCs}$ and $\phi_{SI,PM}$ are emission factors of VOCs and PM in SI unit. The x and y values are based on the average molecular formula of C and H of different emission sources, VOCs and PM. In summary, the listed emission factors in Table 1 will be used to characterize various emissions in the ECUTP scheduling model. For every concerned emission species, m, the final emission amount is the summation of every stream containing m. It is formulated as:

$$FE^m = \sum_t \sum_{v\in V_{uo}} \sum_{s\in S_{v,uo}} \left(F_{v,s,uo,t}\ \phi^m_{v,s,uo}\right),\ \forall\, m \in ME \tag{4}$$

where FE^m is the amount of emission species m. $F_{v,s,uo,t}$ is the flowrate of emission source, and $\phi^m_{v,s,uo}$ is the emission factor.

The objective function of the ECUTP model is to maximize total process profit, which is defined in Eq.(5). It contains three main items. The first summation item represents the total refinery revenue, which includes plant product sale income and utility sale income. The second one represents the front-end crude operational cost. The third one is the total refinery cost, which includes feedstock costs, utility costs, and inventory costs.

$$\begin{aligned}\max Profit = &\sum_n \left(Sale\_PO_n + Sale\_UO_n\right) - \sum_n \left(Cost\_OP_n\right) \\ &- \sum_n \left(Cost\_FD_n + Cost\_UD_n + \sum_{u\in UINV} \Delta IV_{u,n}\right)\end{aligned} \tag{5}$$

4. Case study

4.1. Given data

The problem consists of three single-parcel vessels, one single docking berth, four storage tanks (ST), four charging tanks (CT), and a refinery plant starting from two CDUs. The refinery process includes crude distillation, cracking, coking, reforming, hydrotreating, gas processing, product blending, and component recovery. The specified scheduling time horizon is 15 days. For simplicity, the crude unloading and transferring data, crude property data, refinery product specifications, and economic data of the studied case are the same as those in the previous study (Xu et al. 2017a), and they have been provided by Tables 1-4 from that paper. Meanwhile, the detailed crude transfer network and schematic diagram of the studied refinery plant have also been provided by Figures 3 and 4 from that paper.

The whole scheduling problem is developed and solved in GAMS v24.8.3 on an Intel 3.6 GHz Windows PC with 12.0 GB memory. It is a mixed integer nonlinear programming (MINLP) problem. The nonlinearity mainly comes from crude oil and refinery product blending sections. We use DICOPT (CPLEX for the MIP subproblem and BARON for the NLP subproblem) as a solver to solve the MINLP problem.

4.2. Results of front-end crude scheduling

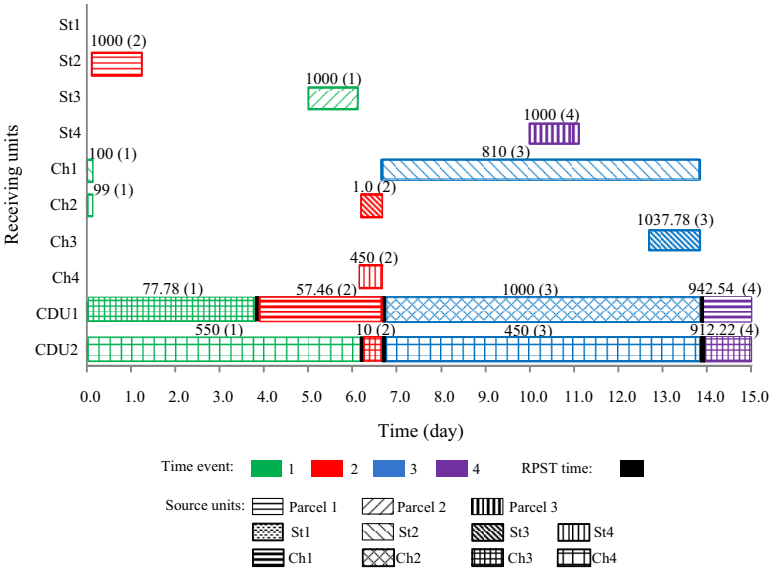


Figure 1. Front-end crude transfer scheduling results.

The optimal Gantt chart of front-end scheduling result for the ECUTP scheduling model is presented in Figure 1. As shown, the scheduling time is adopted as the horizontal axis, and crude "receiving units" are adopted as the vertical axis. The numbers close to the schedule bars represent the total volumes (Mbbl) transferred, while the number in parentheses denotes specific time events when particular operations occur. Here, various filling patterns indicate the source units of crude transfers. Overall, four time events are adopted for the ECUTP scheduling case. Small solid black schedule bars represent the RPST time (Xu et al. 2017b). It can be seen that two refinery CDUs have utilized all four time events, and they are continuously operated during the entire scheduling time horizon. RPST time is located where a crude changeover does occur.

4.3. Results of refinery processing

The refinery products include diesel, gasoline, petroleum coke, kerosene, utility fuel gas, liquefied petroleum gas, benzene, utility fuel oil, and naphtha etc. Diesel and gasoline are the main products of this refinery compared to others. Table 2 gives economic performance based on optimal scheduling results. Based on the proposed ECUTP methodology, the optimal solution gives the maximum total gross profit of \$104,844 K. In comparison, ratios of the gross profit, total net utility income, and crude purchasing cost over the total revenue are 33.16 %, 1.32 %, and 65.19 %, respectively. This matches the reality that the major expenditure of a refinery is for crude purchasing.

Table 2. Economic performance based on optimal scheduling results

Crude Oil Buy(k$)	206,094	Profit(k$)	104,844
Product Sell(k$)	316,152	Profit/Sell	33.16%
Operation Cost(k$)	5,214	Utility/Sell	1.32%
Front-end Operation Cost(k$)	1,059	Crude/Sell	65.19%

4.4. Emission results analysis

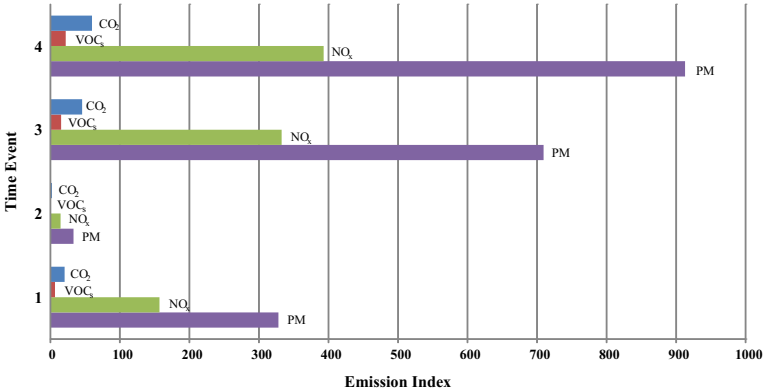


Figure 2. Refinery emission results

The refinery emission results are shown in Figure 2, where CO_2, VOCs, NO_X and PM emissions under current plant design and operations are shown, respectively. Note that the unit of CO_2 index is t while others are kg. It is understandable since the emission of CO_2 should be way more than other emissions.

5. Conclusions

In this paper, a systematic methodology has been developed for emission conscious crude unloading, transferring, and processing (ECUTP) system to achieve profit maximization. Compared to recent studies, the major contribution of this paper is that characterization and quantification of major air emissions from refineries are coupled with the model. The efficacy of the model has been demonstrated by a large-scale case study. As for our future work, we will propose a concept called profitable emission reduction (PER). Since the plant profit and emission can be quantitatively evaluated, alternative plant design and operations for PER opportunities might be raised.

Acknowledgements: This work was supported by the Center for Advances in Port Management, President Visionary Project, and Anita Riddle Faculty Fellowship from Lamar University.

References

Duran, M. A. and I. E. Grossmann (1986). "An outer-approximation algorithm for a class of mixed-integer nonlinear programs." Mathematical Programming 36(3): 307-339.

Kassinis, G. I. (1998). "A model for estimating pollution emissions for individual economic activities." Environmental Impact Assessment Review 18(5): 417-438.

Méndez, C. A., I. E. Grossmann, I. Harjunkoski and P. Kaboré (2006). "A simultaneous optimization approach for off-line blending and scheduling of oil-refinery operations." Computers & Chemical Engineering 30(4): 614-634.

Neiro, S. M. S. and J. M. Pinto (2004). "A general modeling framework for the operational planning of petroleum supply chains." Computers & Chemical Engineering 28(6): 871-896.

Xu, J., S. Zhang, J. Zhang, S. Wang and Q. Xu (2017). "Simultaneous scheduling of front-end crude transfer and refinery processing." Computers & Chemical Engineering 96: 212-236.

Xu, J., H. Qu, S. Wang and Q. Xu (2017). "A New Proactive Scheduling Methodology for Front-End Crude Oil and Refinery Operations under Uncertainty of Shipping Delay." Industrial & Engineering Chemistry Research 56(28): 8041-8053.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50199-3

Expanding RTN discrete-time scheduling formulations to preemptive tasks

Pedro M. Castro[a*], Iiro Harjunkoski[b], Ignacio E. Grossmann[c]

[a]*CMAFCIO, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal*
[b]*ABB AG, Corporate Research Center, Ladenburg, Germany*
[c]*Dep. Chemical Engineering, Carnegie Mellon University, Pittsburgh PA, USA*
pmcastro@fc.ul.pt

Abstract

This paper expands the Resource-Task Network (RTN) scheduling formulation to allow tasks to be interrupted when encountering a planned break period in production. The benefit from a more flexible mode of operation, is the improvement of overall equipment efficiency. This is illustrated by solving a benchmark problem from the literature. To address one of the major limitations of discrete-time approaches, we revisit solution strategies for the objective of makespan minimization, before proposing a method to reduce the number of iterations in the search for the optimal solution.

Keywords: Optimization; Mixed-integer linear programming; Algorithms.

1. Introduction

Enterprise-wide optimization (Grossmann, 2005) aims to simultaneously account for key performance indicators across multiple business units by looking into the integration of supply chain management, production control, planning and scheduling. To accomplish this goal, we need to be able to transfer data and information efficiently between different industrial production management systems. The ISA-95 standard can act as a data-exchange platform for production scheduling (Harjunkoski and Bauer, 2014) but it needs to be linked to a scheduling formulation that can cope with the wide variety of features that may be encountered at a process plant.

The most general scheduling formulations, based on unified frameworks for process representation (State-Task and Resource-Task Network), lack the possibility of processing tasks to be interrupted. Yet, preemption is often encountered in daily practice, with the ISA-95 input including planned break periods (e.g., due to preventive maintenance, weekends), number of stops allowed per task, minimum duration of each active partial task and possible penalties for interrupting a task.

Models that allow for preemption exist in the context of project scheduling. For instance, van Peteghem and Vanhoucke (2010) assume that an activity with discrete duration τ can be interrupted up to τ times, with the time(s) of interruption being determined by a genetic algorithm. In this paper, we consider a more constrained for of preemption, where tasks can be interrupted at given points in time, provided that they continue immediately after. We focus on the RTN discrete-time formulation (Pantelides, 1994), which is known to be very tight and better than continuous-time formulations at handling discrete events (Harjunkoski et al. 2014). Interestingly, the required changes occur at the level of the structural parameters and excess resource balance constraints, with the model variables

remaining the same. It is straightforward to extend a STN discrete-time formulation (Kondili et al., 1993; Lee and Maravelias, 2017) to handle preemption in a similar way.

2. Preemptive vs. non-preemptive scheduling

Non-preemptive scheduling is the standard mode of operation in PSE models, see Figure 1. Given a break period br occurring in time window $[b_{br}^{L}, b_{br}^{U}]$, task i can either be completely executed before the start of the break or start after the end of the break. In the context of a continuous-time model (Castro et al. 2014a), it can be formulated as an exclusive disjunction featuring starting variable Ts_i and duration parameter p_i. When using a discrete-time model, break periods help to reduce model size by restricting the domain of task extent 0-1 variables $N_{i,t}$. In the example in Figure 2, the break lasts for 5 time intervals, including slots $\{9, \ldots, 13\}$ (in red). Noting that durations p_i in minutes or hours are rounded up to multiples of δ (parameter specifying the length of every slot in the uniform grid), $\tau_i = \lceil p_i/\delta \rceil$, since the task lasts 5 slots (in grey), it can only start between slots 1 ($N_{i,1} = 1$) and 4 ($N_{i,4} = 1$), if the plan is to end before the break, or at slot 14, if started after the break, i.e. $T_i \in \{1, \ldots, 4, 14\}$.

The goal of preemptive scheduling is to reduce idle times by allowing part of the task to be executed before the break and part after the break, see middle of Figure 3. Since the location of break periods is known a priori, one can easily determine the duration of a task $\bar{\tau}_{i,t}$ as a function of its starting point t. In the alternatives illustrated in Figure 3, the duration of the task changes between 5 (non-preemption duration), 10 (one interruption) and 12 (two breaks). The domain of the task is thus wider when allowing for preemption.

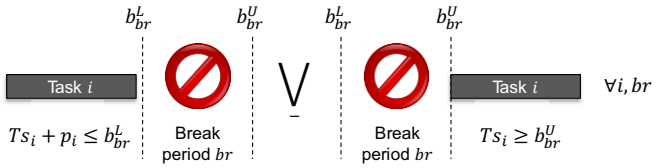


Figure 1. In non-preemptive scheduling, a task can either end-before or start-after a break.

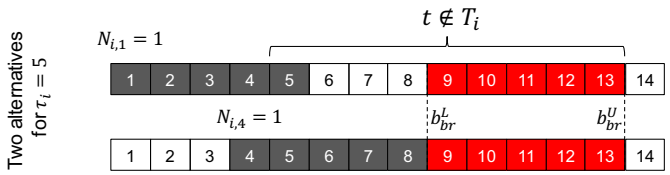


Figure 2. In non-preemptive scheduling, break periods allow reducing domain of processing tasks.

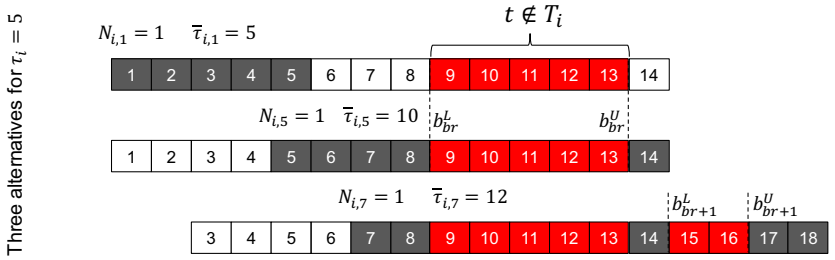


Figure 3. In preemptive scheduling, a task can be split over multiple periods.

3. New RTN discrete-time formulation

The remaining changes needed to extend the RTN discrete-time model to handle preemptive tasks occur at the level of the structural parameters and excess resource balances. Representing $\mu_{r,i,\theta}$ the amount of resource r consumed (-) or produced (+) by non-preemptive task i at a time θ relative to the start of the task, its preemptive counterpart $\bar{\mu}_{r,i,t,\theta}$ gains one index to identify the starting interval t. With the new structural parameters, resource consumption during task interruption can differ from normal execution. An example would be a residual energy consumption to compensate for heat losses during the break, which in our illustrative case study was set to 5%.

Considering that non-preemptive tasks are a special case, we simply need to make $\bar{\tau}_{i,t} = \tau_i \; \forall t \in T_i$, the excess resource balances of the new RTN formulation are given by Eq. (1). It features non-negative excess resource variables $R_{r,t}$ and parameters $\pi_{r,t}$, which account for the interaction with system boundaries (e.g. electricity purchase). Note that Eq. (1) is all that is needed to model preemption.

$$R_{r,t} = R_r^0|_{t=1} + R_{r,t-1} + \pi_{r,t} + \sum_{i'} \sum_{0 \le \theta \le \bar{\tau}_{i,t}: t-\theta \in T_i} \bar{\mu}_{r,i,t-\theta,\theta} N_{i,t-\theta} \; \forall r, t \tag{1}$$

4. Alternatives to minimizing makespan

Makespan minimization is perhaps the most difficult objective function in scheduling because it leads to poor linear programming (LP) relaxations of the mixed-integer linear programming (MILP) models. This is true for both discrete- and continuous-time formulations but it is more noticeable in the former for two reasons: (i) discrete-time formulations are very tight for cost-based objective functions (integrality gap very close to zero); (ii) problem sizes can be orders of magnitude larger, meaning more nodes to explore in the branch-and-bound tree, which is often translated into a poor computational performance. It is thus worth to revisit the alternatives for makespan minimization and evaluate their performance in the new preemptive tasks environment.

4.1. Most popular (option 1)

Makespan (MS) is defined as the maximum finishing time Tf_i over all tasks i, see Eq. (2). Assuming that all tasks are executed at most once during the time horizon, it can be expressed as a function of binary variables $N_{i,t}$ and parameters ft_t that define the position of slot t in minutes or hours. The standard way to minimize makespan is then to consider the linearization of the maximum function, Eq. (3). In cases where multiple instances of a task can be executed over the time horizon, one just needs to move index t from the domain of the summation in Eq. (3) to the inequality domain. Note that in a sequential facility, it suffices to include the subset of tasks linked to the last processing stage.

$$MS = \max_i Tf_i = \max_i (Ts_i + p_i) = \max_i (\textstyle\sum_t N_{i,t} \cdot ft_{t+\bar{\tau}_{i,t}}) \tag{2}$$

$$MS \ge \textstyle\sum_t N_{i,t} \cdot ft_{t+\bar{\tau}_{i,t}} \; \forall i \tag{3}$$

4.2. Defining a last task (option 2)

The second option was used by Castro et al. (2002) for an industrial problem with a superstructure with a task that was to be executed exactly once and was the last in the production sequence. In general, we can define an instantaneous last task i^* as one that consumes all given product demand. This is represented in Figure 4 for the case of a single batch per product (handled as unary resources). The makespan is then given by Eq. (4).

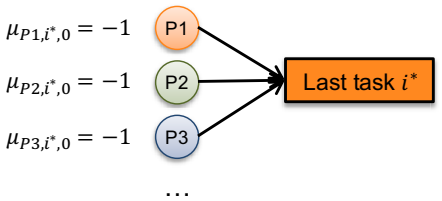


Figure 4. Structural parameters of the RTN associated to the definition of a last task.

$$MS = \sum_t N_{i^*,t} \cdot ft_t \quad (4)$$

4.3. Algorithmic search (option 3)

Maravelias and Grossmann (2003) proposed a very efficient iterative procedure for makespan minimization. Starting with a short time horizon H, insufficient to meet the demand, the problem is solved considering an objective function that is very sensitive to the schedule, e.g. total electricity cost. The solver will find very quickly that the problem is infeasible and so the next step is to keep adding one more time slot to the discrete-time grid per iteration until the problem becomes feasible. The number of slots then gives MS. The solution time per iteration will grow as we approach the optimal makespan (check chart in Castro et al., 2014b) and we want to start close enough. We thus propose to use the LP relaxation from option 2 (tighter than option 1), to set the initial number of slots. The drawback of option 3 is that the optimal solution is the first and lone feasible solution returned by the search procedure and is found only at the very end.

5. Computational results

We consider the demand side management problem of a steel plant with day-ahead, hourly electricity prices. The system and electricity price profile are taken from Castro et al. (2013). Note that to simplify the model, we allow idle times in the continuous caster units when handling heats belonging to the same group. The new RTN discrete-time scheduling formulation was implemented in GAMS 24.8.3 and solved with CPLEX 12.7 running in parallel deterministic mode using up to eight threads. The termination criteria were either a relative optimality tolerance of 10^{-6} or a maximum wall time limit of 7200 CPUs. The hardware consisted of a Windows 10, 64-bit desktop with an Intel i7-4790 (3.6 GHz) processor and 8 GB of RAM.

Table 1 presents the result for total cost minimization when considering breaks between [450, 510] and [930,990] (min). As expected, the wider domain of preemptive tasks, reflected in the larger number of discrete variables $N_{i,t}$, leads to a cost reduction. The difference to the non-preemptive case decreases from €16,253 (344,818-328.549) for time grids with δ=15 min slots, to a mere €147 for the more accurate δ=5 min. In the latter case, the optimal solution in Figure 5 features five tasks in preemptive mode. The computational time increases as δ decreases, as expected, but interestingly, only increases by a factor of two when considering preemptive tasks. Notice also that the formulation is very tight.

Table 2 and Table 3 present the results for makespan minimization. It can be seen that the standard way to minimize the makespan (option 1) is the worst since it can be as much as 5 times slower than the others and it returned a suboptimal solution of 1300 min for δ=5 min. The optimal solution of 1290 (Table 3) features 10 tasks being split over the two breaks to save 20 min of production time compared to the non-preemptive case

(makespan=1310 min, see Table 2). Option 2 was better for the non-preemptive mode, while option 3 prevailed otherwise. Overall, the impact of preemptive tasks in solution quality and computational time is similar for makespan and cost minimization.

Table 1. Computational statistics for total cost minimization

	Non-preemptive tasks				Preemptive tasks			
δ (min)	Discrete variables	RMIP (k€)	MIP (k€)	CPUs	Discrete variables	RMIP (k€)	MIP (k€)	CPUs
15	6012	344.817	344.818	4.96	6728	328.549	328.565	10.4
10	9042	314.691	314.844	22.1	10156	313.853	313.865	14.2
5	18072	310.367	310.436	3177	20400	310.271	310.289	6427

Table 2. Computational statistics for total makespan minimization (non-preemption)

		Option 1		Option 2		Option 3
δ (min)	Makespan (min)	RMIP (min)	CPUs	RMIP (min)	CPUs	CPUs
15	1440	872.15	27.4	937.81	**27.0**	65.0
10	1340	818.35	245	858.77	217	**151**
5	1310	808.11	2139	843.37	**666**	1026

Table 3. Computational statistics for total makespan minimization (preemption)

		Option 1		Option 2		Option 3
δ (min)	Makespan (min)	RMIP (min)	CPUs	RMIP (min)	CPUs	CPUs
15	1410	858.98	**86.5**	878.86	120	100
10	1340	816.16	1286	832.42	562	**260**
5	1290	789.44	7200	805.10	3535	**2951**

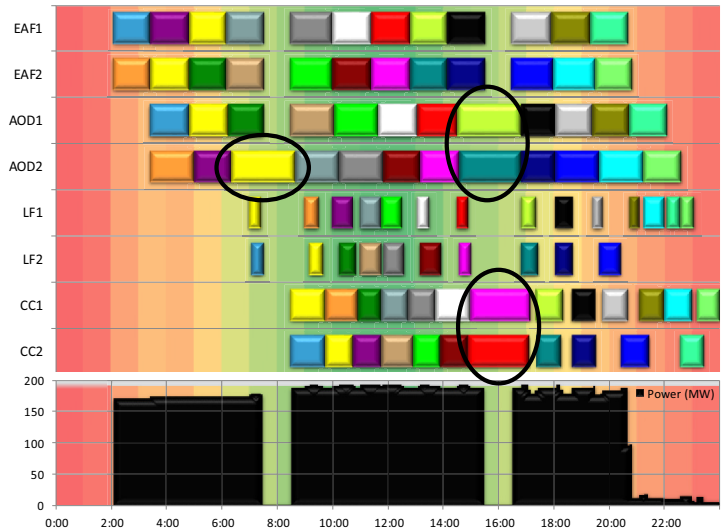


Figure 5. Optimal solution when minimizing electricity cost with δ=5 min time slots.

6. Conclusions

Whenever there is insufficient time to complete the execution of a processing task before a planned break, schedulers may be able to save time by splitting the execution of the task over the periods before and after the break. This preemptive mode of operation has mostly been neglected by scheduling formulations in process systems engineering. In this work, we have expanded a well-known and powerful discrete-time formulation to preemptive tasks. It involves acting at the level of structural parameters generation for the RTN, while keeping the same set of model variables and constraints. The benefits of considering preemption to improve solution quality where illustrated using an industrial case study from the literature, involving the demand side management of a steel plant. The results have also shown that while preemption makes the problem more difficult to solve, the computational time is of the same order of magnitude.

Acknowledgments

Financial support from FCT through projects IF/00781/2013 and UID/MAT/04561/2013.

References

P.M. Castro, I.E. Grossmann, P. Veldhuizen, D. Esplin, 2014a, Optimal Maintenance Scheduling of a Gas Engine Power Plant using Generalized Disjunctive Programming, AIChE J., 60, 2083.

P. Castro, H. Matos, A.P.F.D. Barbosa-Póvoa, 2002, Dynamic Modelling and Scheduling of an Industrial Batch System, Comp. Chem. Eng. 26, 671.

P.M. Castro, D. Rodrigues, H.A. Matos, 2014b, Cyclic Scheduling of Pulp Digesters with Integrated Heating Tasks, Ind. Eng. Chem. Res. 53, 17098.

P.M. Castro, L. Sun, I. Harjunkoski, 2013, Resource−Task Network Formulations for Industrial Demand Side Management of a Steel Plant. Ind. Eng. Chem. Res. 52, 13046.

I. Grossmann, 2005, Enterprise-wide Optimization: A New Frontier in Process Systems Engineering, AIChE J., 51, 1846-1857.

I. Harjunkoski, R. Bauer, 2014, Sharing Data for Production Scheduling using the ISA-95 Standard, Frontiers in Energy Research, doi: 10.3389/fenrg.2014.00044.

I. Harjunkoski, C. Maravelias, P. Bongers, P.M. Castro, S. Engell, I. Grossmann, J. Hooker, C. Méndez, G. Sand, J. Wassick, Scope for Industrial Applications of Production Scheduling Models and Solution Methods, Comp. Chem. Eng. 62, 161.

E. Kondili, C.C. Pantelides, R.W.H. Sargent, 1993, A General Algorithm for Short-term Scheduling of Batch Operations – I. MILP Formulation. Comp. Chem. Eng. 2, 211.

H. Lee, C.T. Maravelias, 2017, Discrete-time Mixed-integer Programming Models for Short-term Scheduling in Multipurpose Environments. Comp. Chem. Eng. 107, 171-183.

C.T. Maravelias, I.E. Grossmann, 2003, Minimization of the Makespan with a Discrete-Time State−Task Network Formulation. Ind. Eng. Chem. Res. 42, 6252.

C.C. Pantelides, 1994, Unified Frameworks for the Optimal Process Planning and Scheduling, In Proceedings of the Second Conferencce on Foundations on Computer Aided Operations, Cache Publications, New York, 253.

V.V. Peteghem, M. Vanhoucke, 2010, A Genetic Algorithm for the Preemptive and Non-Preemptive Multi-mode Resource-Constrained Project Scheduling Problem, European Journal of Operational Research, 201,409-418.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50200-7

Solving integrated ingot type selection and order batching problem in aluminum industry

Gongshu Wang, Shucheng Zhao, Qingxin Guo*

Institute of Industrial and Systems Engineering, Northeastern University, Shenyang, 110819, China
wanggongshu@ise.neu.edu.cn, *guoqingxin@ise.neu.edu.cn*

Abstract

This paper studies a planning problem encountered in the aluminum industry that is to select a pre-specified number of standard ingot types from the candidate pool to fill current orders and combine plates of different orders into batches, each corresponding to an ingot, so as to minimize the total weight of material waste. We formulate the problem as a novel integer programming model in which both *p*-median subproblem and cutting-stock subproblem are involved. The model is decomposed into a master problem and a set of subproblems using the Dantzig-Wolfe decomposition, and then column generation based branch-and-price solution approach is developed to obtain optimal solutions. Computational tests on a diverse set of random instances extended from the real production data show that our solution approach is much more effective than the direct MIP solver of CPLEX. In addition, our solution approach reduces the material waste by over 10.98 % as compared to the model in an existing article.

Keywords: aluminum industry, production planning, integer programming, column generation, branch-and-price

1. Introduction

Aluminum plates are widely accepted by the market due to their excellent properties. With the continuous innovation of manufacture technology and process, aluminum factories are capable of producing enormous variety of plates to meet the diversified and customized requirements. According to the past management mode, the aluminum factory first designs a dedicated ingot type for each order. Then, several rectangular ingots with same alloy and size are produced at the continuous caster for each order. The weight of an ingot is far greater than the weight of individual plate because the continuous caster prefers producing larger weight ingots to exploit economies of scale. Thus, each ingot will be cut into several pieces after casting. Each piece is rolled into a plate with required dimensions at the rolling mill. Based on the past management mode, the weight of the dedicated ingot for each order is an integer multiple of the weight of individual plate plus the necessary loss for cutting operation.

At the first glance, designing a dedicated ingot for each order can match customer requirements accurately and avoid material waste as well. However, with the increase of number of dedicated ingot types, it requires considerable investment and time to purchase and test moulds and related tools. Moreover, changing moulds at the continuous casting process incurs both a setup time resulting in productivity loss and a setup cost to repair the used tools. Thus, from the production control perspective, the aluminum factory prefers producing a fewer standard ingot types.

There is a serious conflict between customers' diversified requirements and the mass production model of the aluminum factory. This brings the new challenge to the production management. Faced with an increasingly competitive market environment, many Chinese aluminum factories have transited their production management mode from mass production to mass customization to enhance its core competence. Using a limited number of standard ingot types to meet a wide variety of orders is a key issue for the aluminum factory production planning under mass customization.

Selecting a given number of standard ingot types to produce in the following planning horizon is first decision in the production planning. With a given number of standard ingot types, several ingot types can be used to produce the finish plates for each order according to its specification on alloys and sizes. However, it is almost impossible to choose a suitable ingot type for each order such that the required plates for each order can be cut from ingots having same type without material waste. Thus, using a single ingot to jointly produce multiple plates from different orders with similar specifications on alloys and sizes is a more economic manner. How to combine plates of different orders into batches such that a batch of plates can be cut from a single ingot is the second decision in the production planning. It is obvious that decisions on selecting a set of standard ingot types and on packing the plates into ingots are actually intertwined, and thus it is beneficial to make these decisions jointly. The purpose of this paper is to find an efficient method to select standard ingot types and then generate production batches such that material utilization can be improved significantly.

There are only a few of studies available to solve the problem of selecting standard ingot types and combining orders into batches. Ventola (1991) discussed a set process rules for order batching involved in an aluminum rolling facility. Gopalan (1992) proposed the selection of standard pipes for the aluminum pipe production problem and classified it as facility location problem. Balakrishnan and Gopalakrishnan (1996) studied the problem of selecting standard ingot sizes to facilitate joint processing of finished sheets which is similar to our problem. However, they gave a set of strict rules for order batching which decreases the flexibility to combine different orders together. Similar studies have also been conducted in the steel industry. Tang *et al.* (2014) studied an integrated problem of charge batching and casting width selection arising in steelmaking and continuous casting production.

The rest of this paper is arranged as follows. Section 2 describes the problem studied and the mathematical model. The solution method based on column generation is introduced in section 3. In section 4, we present computational results and discuss the efficiency of proposed solution method. Finally, we draw conclusions in section 5.

2. Problem Description and Mathematic Formulation

The problem we study can be precisely described as follow. Let $N=\{1, 2, \ldots, n\}$ be a set of n orders, each order $i \in N$ contains d_i identical plates. The plate of order i is associated with several specific properties: alloy type a_i, weight q_i, width w_i and thickness t_i. Let $K=\{1, 2, \ldots, m\}$ be a set of m candidate ingot types, for each $k \in K$, let A_k, Q_k, W_k, and T_k respectively denote the alloy type, weight, width and thickness of ingot type k. We define N_k as a subset of orders that can be fulfilled by ingot type k. More precisely, for each order $i \in N_k$, the technological constraints including alloy, width and thickness compatibility should be satisfied, i.e. $a_i = A_k$, $w_i \leq W_k$ and $t_i \leq T_k$. Let l be the maximal number of standard ingot types that can be produced in the current planning horizon.

The problem is to select no more than l standard ingot types from the candidate pool K and then pack all ordered plates into ingots by using the selected ingot types, such that the waste material cut from the used ingots is minimized. To understand the decisions better, a general diagram representing the essential decisions is shown in Figure 1.

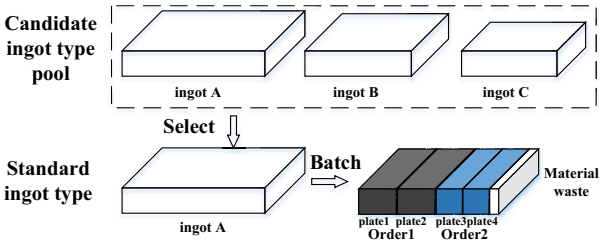


Figure 1. The integrated process of ingot type selection and order batching

For each ingot type k, the number of used ingots is unknown in advance. Let J_k be the set of ingots with type k to be used. By introducing three types of decision variables: x_{ijk} denotes the number of plate of order i packed in the j^{th} ingot of type k, y_{jk} denotes whether the j^{th} ingot of type k is used or not, z_k denotes whether type k is used or not. With the notations and variables described above, this problem is formulated as the following integer programming (IP).

$$\min \sum_{k\in K}\sum_{j\in J_k}(Q_k y_{jk} - \sum_{i\in N_k} q_i x_{ijk}) \tag{1}$$

$$s.t.\ \sum_{k\in K}\sum_{j\in J_k} x_{ijk} = d_i,\ \forall\ i\in N, \tag{2}$$

$$\sum_{i\in N_k} q_i x_{ijk} \le Q_k y_{jk},\ \forall\ k\in K, j\in J_k, \tag{3}$$

$$y_{jk} \le z_k,\ \forall\ k\in K, j\in J_k, \tag{4}$$

$$\sum_{k\in K} z_k \le l, \tag{5}$$

$$x_{ijk} \ge 0, \text{ and integer},\ \forall\ k\in K, j\in J_k, i\in N_k, \tag{6}$$

$$y_{jk},\ z_k \in \{0, 1\},\ \forall\ k\in K, j\in J_k. \tag{7}$$

The objective function (1) minimizes the material waste, *i.e.*, the total used ingots minus total customer demands. Constraints (2) imply that each order should be assigned with the demanded number of plates. Constraints (3) ensure that the total weight of plates cut from a single ingot should not exceed the weight of the ingot. Constraints (4) guarantee that a single ingot can be used if and only if the corresponding ingot type is selected. Constraint (5) limits the upper bound of selected ingot types. Constraints (6-7) define the value range of decision variables.

If the decision of selecting standard ingot types is given, the model is reduced to the cutting-stock problem. If a single ingot can only be used to fulfil a single plate, the problem is to select a given number of standard ingot types from the candidate pool to cover all orders, which becomes a p-median problem. Since both cutting-stock problem and p-median problem are NP-hard, so is our problem.

3. Solution Method

Since a direct commercial optimization solver such as CPLEX would fail to solve the IP model to optimality within limited CPU times, we decompose the IP model into a master problem and a set of subproblems using the Dantzig-Wolfe decomposition, and then column generation based branch-and-price algorithm is developed to obtain optimal solutions. In the literature, branch-and-price algorithm have been developed to

solve successfully practical instances of many batching problems arising in steel industry (Tang *et al.*, 2011, Tang *et al.*, 2014, Tang *et al.*, 2016).

For each ingot type k, we define a pattern as a combination of plates that can be cut from a single ingot of type of k. A pattern p for ingot type k is described by an integer array (a_{1kp}, a_{2kp}, . . . , a_{nkp}), where a_{ikp} gives the number of plates of order i that are contained in pattern p, and satisfies:

$$\sum_{i\in N_k} q_i\, a_{ikp} \le Q_k, \tag{8}$$

$$0 \le a_{ikp} \le d_i \text{ and integer}, \forall\ i\in N_k. \tag{9}$$

Let us introduce an integer variable λ_{kp} that gives the number of times pattern p for ingot type k is used. By using the Dantzig-Wolfe decomposition, the IP can be reformulated as the following master problem (MP).

$$\min \sum_{k\in K}\sum_{p\in P_k} c_{kp}\,\lambda_{kp} \tag{10}$$

$$s.t.\ \sum_{k\in K}\sum_{p\in P_k} a_{ikp}\,\lambda_{kp} = d_i,\ \forall\ i\in N, \tag{11}$$

$$\sum_{p\in P_k} \lambda_{kp} \le |J_k| z_k,\ \forall\ k\in K, \tag{12}$$

$$\sum_{k\in K} z_k \le l, \tag{13}$$

$$\lambda_{kp} \ge 0, \text{ and integer}, \forall\ k\in K, p\in P_k, \tag{14}$$

$$z_k \in \{0, 1\},\ \forall\ k\in K. \tag{15}$$

Note that $c_{kp} = Q_k - \sum_{i\in N_k} a_{ikp}\, q_i$ is the waste of material cut from pattern p for ingot type k. The objective function (10) is to minimize total material waste for all used patterns. Constraints (11) guarantee that the demand of each order is satisfied. Constraints (12) ensure that the patterns can be used if and only if the corresponding ingot type is select.

Since the MP usually contains an exponential number of patterns, it is impractical to solve MP directly. If the integrality requirement on λ-variables and z-variables is relaxed, we obtain the LP relaxation of the MP, denoted as (LMP). A lower bound for MP is generated by solving LMP using column generation. The column generation starts with a limited number of columns in LMP and generates necessary columns iteratively. In each iteration, we first solve a restricted master problem (which is a restricted version of LMP with the columns generated so far). Then we generate columns with negative reduced costs by solving a set of pricing problems. Let dual variable value be π_i for index i in (11), δ_k for index k in (12), the reduced cost of a column $p\in P_k$ for ingot k, denoted as rc_{kp}, is given by:

$$rc_{kp} = c_{kp} - \sum_{i\in N_k} \pi_i\, a_{ikp} - \delta_k = Q_k - \sum_{i\in N_k} (q_i + \pi_i) a_{ikp} - \delta_k \tag{16}$$

To generate a column with most negative reduced cost for each ingot type, we need to solve the following knapsack type subproblem (SP).

$$\min\ Q_k - \sum_{i\in N_k} (q_i + \pi_i) a_{ikp} - \delta_k, \text{ subject to constraints (8)-(9).} \tag{17}$$

For each ingot $k\in K$, if the objective value of SP is negative, this new generated column is added to the restricted master problem which is then updated. When no new columns with a negative reduced cost can be found in solving all pricing subproblems, the column generation procedure terminates and LMP is solved to optimality.

When the solution of LMP is fractional, we can embed the column generation lower bound into a branch and bound tree, thus obtaining a branch-and-price algorithm. We use a simple but efficient branching rule, that it, branching on a fractional λ_{kp} variable

and generating two children nodes by setting a lower bound $\lceil \bar{\lambda}_{kp} \rceil$ on λ_{kp} to right node and a upper bound $\lfloor \bar{\lambda}_{kp} \rfloor$ on λ_{kp} to left node, respectively.

4. Computational experiments

This section presents the result of the computational experiments designed for testing the effectiveness of developed model and algorithm. The column generation based branch-and-price algorithm was coded in C++ with all the LP problems solved by the LP solver of CPLEX.

We tested our model and algorithm on a set of 10 random instances extended from the real production data. Table 1 shows the computational results. The first few columns (first part) describe the problem structure, where n denotes the number of orders, m denotes the number of candidate ingot types, l denotes the number of standard ingot types to be selected, and $r=\sum_{k=1}^{m}|N_k|/(m\times n)$ represents the ratio of compatible order-ingot pairs to all order-ingot pairs. The second part of the table reports the computational results obtained by using the MIP solver of CPLEX to solve the existing model developed by Balakrishnan & Gopalakrishnan (1996). The third part of the table reports the computational results obtained by using the MIP solver of CPLEX to solve our IP model. The last part of the table reports the computational results obtained by branch-and-price algorithm.

To test the algorithm effectiveness, we compare solution of IP model solved by CPLEX with the solution obtained by branch-and-price algorithm. From the results presented in the third part and the last part of Table 1, we can observe that all the instances can be solved to optimality by branch-and-price algorithm within 2 minute, but CPLEX can't get optimal solutions for some instance within 1 hour. This indicates that the solution approach we developed is much more effective than the direct MIP solver of CPLEX. To test the model effectiveness, we compare the solution of the MP model solved by branch-and-price algorithm with the solution of B&G's model solved by CPLEX. By comparing the second part and the last part of Table 1, we can observe that both our MP model and B&G's model can be solved to optimality within acceptable computational time. However, the optimal solution of B&G's model is consistently larger than the optimal solution of our model. This is because the former deals with a more constrained version of the problem. In average, our model reduces the material waste by over 10.98 % as compared to B&G's model.

Table 1. Computational results for testing instances

Instances				B&G model		IP model				B&P Algorithm		
n	m	l	r	OPT	CPU (s)	LB_{best}	UB	Gap(%)	CPU (s)	LB_{root}	OPT	CPU/s
10	10	5	0.49	57,694	1.50	34,851	50,694	31%	3,600	49,194	50,694	1.48
20	10	5	0.40	125,207	0.92	99,924	146,207	32%	3,600	123,207	125,207	3.34
30	10	5	0.43	129,575	3.40	98,440	115,575	15%	3,600	111,908	114,575	12.85
40	10	5	0.45	35,969	4.89	10,014	37,969	74%	3,600	22,844	30,969	21.68
50	10	5	0.44	138,269	6.30	82,459	123,269	33%	3,600	109,102	117,269	22.48
60	10	5	0.44	133,011	11.38	103,482	129,011	20%	3,600	111,571	116,001	41.50
70	10	5	0.46	220,898	13.54	153,316	189,898	19%	3,600	166,862	175,898	50.76
80	10	5	0.45	251,331	16.02	206,025	305,331	33%	3,600	224,090	232,331	45.40
90	10	5	0.45	346,975	13.67	304,270	406,975	25%	3,600	319,030	323,975	74.54
100	10	5	0.46	260,265	9.66	206,061	479,265	57%	3,600	224,398	235,265	82.94

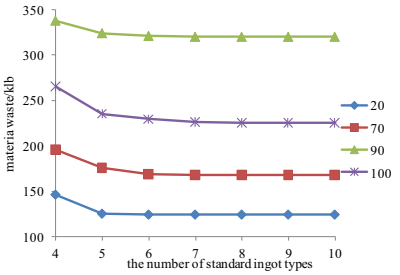


Fig. 2. The trendlines of the total waste material along with different l values

To analyze the impact of the number of selected standard ingot types (l) on the solution performance, we tested different values of $l \in \{4, 5, 6, \ldots, 10\}$ on 4 instances. The trendlines of the total waste material along with different l values are illustrated in Figure 2. As we can see, for all instances, the total waste material decreases with l increasing. However, there is a little reduction of the waste material after l reaches to 6 for all tested instances. The above sensitivity analysis can provide planners more support policy decisions about how to select the number of standard ingot types.

5. Conclusion

Selecting standard ingot types and batching plates of different orders together are very important planning decisions for daily operational management in the aluminum industry. This paper studies an integrated problem of ingot type selection and order batching. To address this problem, we formulated it as an integer programming model in which both p-median subproblem and cutting-stock subproblem are involved. To find solutions, a column generation based branch-and-price algorithm is developed. Computational results on a set of random instances demonstrate the effectiveness of developed model and algorithm.

Acknowledgements

This research is partly supported by National Key R&D Program of China (2017YFB0304100) and National Natural Science Foundation of China (71672032, 71602025).

References

A. Balakrishnan, S. Gopalakrishnan, 1996, Selecting ingot sizes for joint production of sheet products, INFORMS Journal on Computing, 8, 260–273.

R. Gopalan, 1992, Exploiting process flexibility in metal forming operations, Phd thesis, Massachusetts Institute of Technology, Cambrodge, CA.

L. Tang, G. Wang, J. Liu, J. Liu, 2011, A combination of Lagrangian relaxation and column generation for order batching in steelmaking and continuous-casting production, Naval Resaerch Logistics, 58, 370–388.

L. Tang, G. Wang, Z.-L. Chen, 2014, Integrated charge batching and casting width selection at Baosteel, Operations Research, 62, 772–787.

L. Tang, Y. Meng, Z.-L. Chen, J. Liu, 2016, Coil batching to improve productivity and energy utilization in steel production, Manufacturing & Service Operations Management, 18, 262–279.

D. P. Ventola, 1991, Order combination methodology for short-term lot planning at an aluminum rolling facility, Master degree thesis, Massachusetts Institute of Technology, Cambrodge, CA.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50201-9

Industrial Demand Side Management Formulation for Simultaneous Electricity Load Commitment and Future Load Prediction

Giancarlo Dalle Ave[a,b], Iiro Harjunkoski[a*], Sebastian Engell[b]

[a]*ABB Corporate Research Germany, Wallstadter Str. 59, 68526 Ladenburg, Germany*
[b]*Process Dynamics and Operations Group, Department of Biochemical and Chemical Engineering, Technische Universität Dortmind, EmilFigge-Str. 70, 44221 Dortmund, Germany*
Iiro.Harjunkoski@de.abb.com

Abstract

Large consumers' electricity bills depend on many factors including: different electricity purchasing contracts and markets, deviation penalties, and grid fees. Two key markets are the intraday and the day-ahead markets. On the day-ahead market, the consumer commits to a load of electricity that stacks on top of their longer-term contract commitments. The next day, the consumer must follow this demand profile to avoid paying deviation penalties. The demand curve can be modified on the existing day using the intraday market. A novel demand side response formulation is proposed that considers a two-day horizon for both the intraday and the day-ahead markets by a separate modeling approach. Results show that this formulation can effectively combine intraday and day-ahead market concerns and that the resulting demand profile from the two-day problem is more realistic than current models, which consider only a single-day problem.

1. Introduction

Demand Side Management (DSM) provides grid operators the opportunity to improve efficiency and stability of the power grid by flattening the electricity load curve and will play a crucial role in the improvement of grid efficiency and reliability for years to come. Simultaneously, DSM provides an opportunity for electricity consumers to lower their operating costs by responding to time-dependent electricity costs (Merkert, *et al.*, 2015).

The bill for an electricity consumer is dependent on many factors. One factor is the type of electricity purchase contract. These contracts include: long-term (up to a year) base load (BL) contracts, short-term (a few months) Time-of-Use (TOU) contract, and the day-ahead market in which consumers can purchase hourly electricity loads at hourly prices up to 12 hours in advance of the first delivery the next day. The total load from these contracts represents the hourly committed load curve for a consumer for the next day (24 hours). This load curve must then be followed to within a tolerance else so called deviation penalties must be paid. If a consumer experiences a major shift in electricity demand (for example, in the event of unit-breakdown) there exists an intraday market in which consumers can modify their committed loads to avoid paying deviation penalties. In addition, companies also need to pay various grid fees. One such grid free is a peak-pricing scheme in which a company must pay a fee based on their highest simultaneous electricity consumption over a given time-interval (Frontier Economics, 2016).

Due to the strong time-dependence of these electricity-related concerns, effective scheduling is essential in DSM, especially when a complex manufacturing process is involved. Several works have looked at DSM for scheduling, including Nolde & Morari, 2010 and Hadera, *et al.*, 2015 who investigate scheduling of steel manufacturing based on a multi-contract variable electricity pricing load commitment problem using continuous-time precedence-based scheduling models. Castro, *et al.,* 2013 also studied the load commitment problem for steel manufacturing but using a discrete-time Resource-Task Network (RTN) representation. The downside of only studying a 24-hour commitment problem is two-fold: the first is that this unrealistically makes the day-ahead and the intraday markets overlap by considering them to both be active at the same time, or by ignoring one of the markets. The second drawback is that the committed electricity load drops off towards the end of the schedule as the jobs are completed. This is unrealistic as production is performed on a continuous basis and is one of the factors hindering the application of DSM in industrial settings.

In this work, the current-day problem of load commitment (considering the intraday market) is combined with the problem of future load prediction (on the day-ahead market) in order to pay less deviation penalties and to achieve more realistic electricity-demand profiles that do not drop off towards the end of the scheduling horizon. In addition, this work considers penalties paid on the maximum electricity consumption. The problem is formulated as a discrete-time RTN (Pantelides, 1994) and is applied to the industrial problem of steel melt shop scheduling.

2. Problem Definition

This work looks at the problem of steel melt shop scheduling considering multiple electricity contracts and markets as well as peak-related grid-operator fees in order to minimize total cost. Stainless-steel production is an energy-intensive batch process with complicated processing constraints. The typical production process has four processing steps. The first involves melting the scrap in an Electric Arc Furnace (EAF) to form a so-called heat (a batch of liquid metal). This is a very energy intensive process and accounts for the majority of the electricity consumed. From there, the heat is transported to the next station where the carbon content of the steel is reduced in a process known as Argon Oxygen Decarburization (AOD). Next, the heat is transported to a Ladle Furnace (LF) to further adjust the chemistry and temperature of the heat. The final stage of the process is the Continuous-Casting (CC), where a group of several heats must be processed in an uninterrupted sequence according to a strict set of rules. The considered example here contains two parallel machines at each stage. A diagram of the process, including the RTN diagram of the process, can be viewed in Castro *et al.* 2013.

3. Model Formulation

A discrete-time model is chosen, as the location of every time point in the grid is known in advance allowing for straightforward modelling of intermediate events (such as changes in electricity pricing or power availability). Note that this formulation contains many different time intervals, but they are all matched to the underlying discrete grid.

An RTN model is applied, as it is general, yet simple, and has been shown to be both capable of handling problems of industrial relevance and expandable to account for DSM (Castro et al. 2013). The notation for the model can be viewed in Table 1. Note that all variables are non-negative continuous variables unless otherwise specified.

Table 1: Model notation

Index/Set	**Description**	**Set**	**Description**
$r \in R$	**Resources**	θ	**Relative time to start of task**
$i \in I$	**Tasks**	T_{hr}	**Set of hours in the horizon**
$h \in H$	**Set of steel heats**	T_{peak}	**Set of times over which the peak is calculated**
$g \in G$	**Set of steel groups**		
$k \in K$	**Set of processing stages**	T_{intra}	**Set of times over which electricity is sold**
$t \in T$	**Set of time points**		
$u \in U$	**Set of units**		
Parameter	Description	Parameter	Description
τ_i	Duration of task i in time slots	δ_T	Time grid discretization size for the RTN
$\mu_{r,i,\theta}$	Extent of discrete interaction of r with i at time point θ	$\delta_{T_{peak}}$, $\delta_{T_{intra}}$	Interval size for the peak and electricity selling times respectively
y_{hr}^{TOU}, y^{BL}	Respectively, amount of electricity load from a TOU or baseload contract	T^s	Time until the next trading on the intraday market
y_{hr}^{CL}	Total amount of current day committed electricity load	T^*	Hour marking the split between following a pre-committed electricity curve and predicting the future curve
$pw_{h,u}$	Power usage of h in u		
c_{pf}^{DA}	Penalty-free tolerance on load tracking		
Variable	Description	Variable	Description
N_{it}	Binary, execution of task i at time slot t	ξ	Maximum electricity peak
$R_{r,t}$	Amount of r available at t	ω_t^{DA}	Free var. penalty free zone from day-ahead contract
$\Pi_{r^{pw},t}$	Power consumption at t	$\sigma_{t_{intra}}^{DA}$	Free var. amount of day-ahead load sold or bought on the intraday market
y_{hr}^{PL}	Total amount of day-ahead predicted load		
Δ_t^{DA+}, Δ_t^{DA-}	Respectively positive and negative deviations from the day-ahead committed load	$\sigma_{t_{intra}}^{TOU}$, $\sigma_{t_{intra}}^{BL}$	Respectively amount of load sold from TOU or BL contract on the intraday market
Δ_t^{TOU-}, Δ_t^{BL-}	Respectively negative deviations from the TOU and BL contracts		

3.1. Model Constraints

Resource availability over the time grid is managed by the excess resource balance given in Eq. (1).

$$R_{r,t} = R_{r,t=1}^{0} + R_{r,t-1} + \Pi_{\mathrm{r^{pw}},t} + \sum_i \sum_{\theta=0}^{\tau_i} \mu_{r,i,\theta} N_{i,t-\theta} \quad \forall r, t \in T \tag{1}$$

Eq. (2) and (3) are used to ensure that all heats are processed exactly once at each stage, while Eq. (4) and Eq. (5) enforces transfer task tasks and times.

$$\sum_{u \in u_k} \sum_{i \in I_{h,u}} \sum_t N_{i,t} = 1 \quad \forall h, k = 1,2,3 \tag{2}$$

$$\sum_{u \in u_k} \sum_{i \in I_{g,u}} \sum_t N_{i,t} = 1 \quad \forall g, k = 4 \tag{3}$$

$$\sum_{u \in U_k} \sum_{u' \in U_{k+1}} \sum_{i \in hu,u'} \sum_t N_{i,t} = 1 \quad \forall h, k = 1,2,3 \tag{4}$$

$$\sum_{u' \in U_{k+1}} \sum_{r \in R_{h,u}^{IL}{}'} \sum_t R_{r,t} \leq \left\lceil \max_{u \in U} \frac{maxtr_{u,u'} - mintr_{u,u'}}{\delta_T} \right\rceil \quad \forall h, k = 1,2,3 \tag{5}$$

$$R_{r^{pw},t} = 0 \; \forall t \in T \qquad \forall t \in T \tag{6}$$

$$\Pi_{r^{pw},t} \leq \sum_h \sum_u pw_{h,u} \quad \forall t \in T \tag{7}$$

$$y_h^{PL} \leq \sum_h \sum_u pw_{h,u} \qquad \forall hr \in T_{hr} \tag{8}$$

$$\xi \leq \left\lceil \delta_T / \delta_{T_{peak}} \right\rceil \sum_h \sum_u pw_{h,u} \tag{9}$$

To prevent a buildup of electricity over time Eq. (6) is used. Eqs. (7), (8), and (9) set upper bounds on electricity consumption, predicted, and maximum electricity consumption. In order to calculate the deviations from the load that has been committed to within the current day, Eqs. (6) and (7) are used. Two equations are needed in order to account for the time before and after it is possible to trade on the intraday market.

$$\Pi_{r^{pw},t} = y_h^{CL} + \omega_t^{DA} - \Delta_t^{DA-} + \Delta_t^{DA+} - \Delta_t^{TOU} - \Delta_t^{BL} \;\; \forall\, t \in T_{hr}, T_{hr} < T^S \tag{10}$$

$$\Pi_{r^{pw},t} = y_h^{CL} + \omega_t^{DA} - \Delta_t^{DA-} + \Delta_t^{DA+} - \Delta_t^{TOU} - \Delta_t^{BL} + \sigma_{t_{intra}}^{DA} + \sigma_{t_{intra}}^{TOU} + \sigma_{t_{intra}}^{BL} \quad \forall\, t \in t_{intra}, t_{intra} \in T_{hr}, T^S \leq T_{hr} \leq T^* \tag{11}$$

The amount of load to commit on the day-ahead market is set by Eq. 12.

$$\Pi_{r^{pw},t} = y_h^{PL} + \omega_t^{DA} - \Delta_t^{DA-} + \Delta_t^{DA+} - \omega_t^{TOU} - \Delta_t^{TOU} - \omega_t^{BL} - \Delta_t^{BL} \quad \forall\, t \in T_{Hr}, \; T^* < T_{hr} \tag{12}$$

Since the penalty-free region is a percentage of the contracted load, the upper and lower bounds on the penalty free region for the day-ahead contract are enforced through Eqs. (9) and (10). Note that because the BL and TOU contracts represent loads that should always be met there is no penalty free region for these contracts.

$$-c_{pf}^{DA} * \left(y_h^{CL} - y_{hr}^{TOU} - y^{BL} - \sigma_{t_{intra}}^{DA}\right) \leq \omega_t^{DA} \leq c_{pf}^{DA} * \left(y_h^{CL} - y_{hr}^{TOU} - y^{BL} - \sigma_{t_{intra}}^{DA}\right) \forall\, t \in T_{hr} \;\; T_{hr} \leq T^* \tag{13}$$

$$-c_{pf}^{DA} * (y_h^{PL} - y_{hr}^{TOU} - y^{BL}) \leq \omega_t^{DA} \leq c_{pf}^{DA} * (y_h^{PL} - y_{hr}^{TOU} - y^{BL}) \forall t \in T_{hr}, T^* < T_{hr} \tag{14}$$

If the actual consumption strays outside of the penalty-free region defined above penalties must be paid. Eqs. (15) to (19) define the bounds on the magnitude of the deviations per contract. Note that because the contracts build on top of one another, deviations from 'lower' contracts are more costly than deviations from 'higher' contracts. Therefore, an extra variable is not needed to account for which contract is being violated, as the optimal solution will first penalize the higher contracts before proceeding to the lower ones.

$$\Delta_t^{DA-} \leq \left(1 - c_{pf}^{DA}\right)\left(y_h^{CL} - y_{hr}^{TOU} - y^{BL} - \sigma_{t_{intra}}^{DA}\right) \;\; \forall\, t \in T_{hr} \;\; T_{hr} \leq T^* \tag{15}$$

$$\Delta_t^{DA-} \leq \left(1 - c_{pf}^{DA}\right)\left(y_h^{PL} - y_{hr}^{TOU} - y^{BL} - \sigma_{t_{intra}}^{DA}\right) \;\; \forall\, t \in T_{hr}, T^* < T_{hr} \tag{16}$$

$$\Delta_t^{DA+} \leq \sum_h \sum_u pw_{h,u} \qquad \forall t \in T \tag{17}$$

$$\Delta_t^{TOU-} \leq y_{hr}^{TOU} - \sigma_{t_{intra}}^{TOU} \quad \forall t \in T \tag{18}$$

$$\Delta_t^{DA-} \leq y^{BL} - \sigma_{t_{intra}}^{BL} \qquad \forall t \in T \tag{19}$$

The last constraint given by Eq. (16) calculates the maximum peak power achieved.

$$\xi \geq \sum_{t \in t_{peak}} \Pi_{r,t} \quad \forall t_{peak} \in T_{peak} \tag{20}$$

The objective function of the problem is to minimize the total cost of production. This includes cost of the electricity from each of the contracts, the cost of deviations from each of the contracts, revenue or costs incurred by trading on the intraday market, and a flat fee paid on the maximum peak attained during production.

4. Results

Three case studies were solved using GAMS 24.7.4/CPLEX 12.6.3.0. They are defined as follows: Case 1 considers a two-day commitment to a predetermined load with no consideration of the intraday market or the maximum peak. Case 2 is similar to Case 1 but it includes the intraday market. The aforementioned case studies were run with a daily production of 12 and 17 heats as two separate sub problems (SP1/SP2), one for each of the two days considered. Lastly, Case 3 is the proposed novel approach that highlights the importance of considering a longer horizon comprising current day load commitment and day-ahead prediction. In this case, it is not possible to trade on the intraday market to highlight the fact that fewer penalties can be paid simply by shifting production between the two days and changing the day-ahead consumption. This case was run with a time horizon of 54 hours (two days + six hours) with the total two-day forecasted production, plus two additional heats to account for production on the third day. The additional heats are necessary to highlight the fact that production should not drop off towards the end of the day, resulting in an unrealistic load profile. To facilitate a fair comparison, the total-two day production (24 and 34 heats respectively) was constrained to be completed by the end of day two, with the additional two extra heats exiting the last stage on the final day. A summary of the results is presented in Table 2 and Figure 1.

Results indicate that the novel formulation is able to avoid deviation penalties simply by transferring production across the current/next day boundary to better track the current day committed load while modifying the day-ahead demand to avoid future penalties. In the 12-heat case, this was accomplished by pushing production forward to the current day in order to avoid paying negative deviation penalties. In the 17-heat case, it was more favorable to slow down production to avoid paying positive deviation penalties. That being said, Case 3 cannot avoid as many deviations as the case that has no limit on selling or buying on the intraday market, however the intraday market is often avoided by large consumers unless it is necessary. In addition, the new formulation is able to manage the largest peak within the flexibility of production; in Case 2-12H when peak management was not considered, both EAFs were operated at the same time, resulting in a peak nearly twice as high as when the maximum peak was penalized. In the 17-heat case, peak management was more difficult due to the higher production amount; however, a smaller peak was still achieved. Figure 1 indicates that the two-day formulation is also able to avoid a drop off in production towards the end of the day when compared to considering two one-day problems. This presents a more realistic production scenario as each stage of the plant is operated constantly, as opposed to operating the first stages only until the daily production has been met. The major drawback of the new approach is that it results in very large models and it has difficulty in converging to a provably optimal solution.

Table 2: Comparison of results for the different cases.

Case	Deviations Paid (€)	Max Peak (MW)	CPU Time (s) [SP1/SP2]	Rel. Gap (%) [SP1/SP2]
Case 1-12H	190,600	96	10,000/6,893	<0.1/0
Case 2-12H	120	177	3,505/10,000	0/0.3
Case 3-12H	25,100	94	10,000	8.6
Case 1-17H	305,400	185	265/10,000	0/0.1
Case 2-17H	7,217	192	10,000/10,000	2.6/1.3
Case 3-17H	16,500	177	10,000	19.3

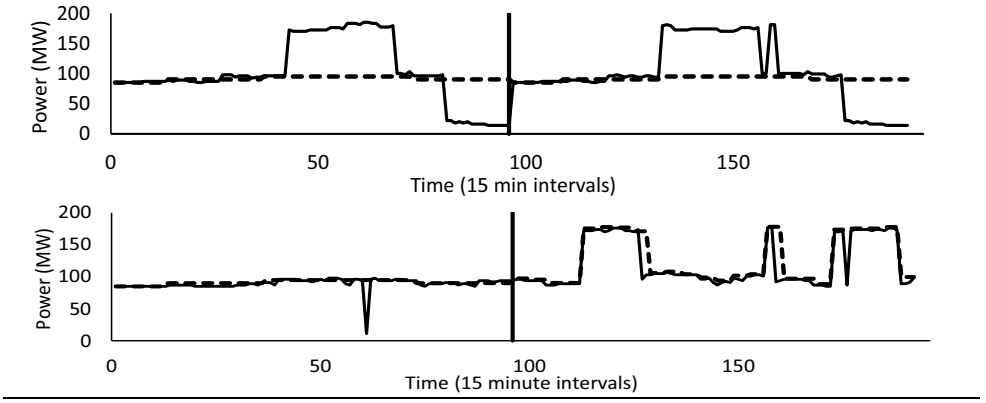


Figure 1: Comparison of committed load for Case1-17H (top) and Case 3-17H (bottom). Actual consumption is represented by the solid line and committed load by the dashed line.

5. Conclusions

A novel DSM model was presented in this work. The model differs from its predecessors, as it is able to effectively combine the problems of current day electricity load following considering the intraday market and day-ahead future load prediction considering the day-ahead market. Results show that this novel formulation is more realistic than its existing models as it is able to follow and predict more realistic load curves while minimizing maximum electricity peaks. The downside of this approach is that it results in very large, difficult to solve problems. Future work could investigate speed-up algorithms or decomposition approaches for such a formulation to ensure that good quality solutions can be found in industrially relevant time frames.

6. Acknowledgements

Financial support is gratefully acknowledged from the Marie Skłodowska Curie Horizon 2020 EID-ITN project “PROcess NeTwork Optimization (PRONTO)", Grant agreement No 675215.

References

P. M., Castro, L., Sun, & I. ,Harjunkoski (2013). Resource-Task Network Formulations for Industrial Demand Side Management of a Steel Plant. *Industrial and Engineering Chemistry Research*(52), 13046-16058.

Frontier Economics. (2016). *METIS Technical Note T4 Overview of European Electricity Markets.* Brussels: European Commission.

H., Hadera, I., Harjunkoski, G., Sand, I. E., Grossmann, & S. ,Engell (2015). Optimization of steel production scheduling with complex time-sensitive electricity cost. *Computers and Chemical Engineering, 76*, 117-136.

L., Merkert, I., Harjunkoski, A., Isaksson, S., Saynevirta, A., Saarela, & G. ,Sand (2015). Scheduling and energy - Industial challenges and opportunities. *Computers and Chemical Engineering*(72), 183-198.

K., Nolde, & M. , Morari (2010). Electrical load tracking scheduling of a steel plant. *Computers and Chemical Engineering, 34*, 1899-1903.

C. C. ,Pantelides (1994). Unified Frameworks for the Optimal Process Planning and Scheduling. *Proceedings of the Second Conference on Foundations of Computer Aided Operations.* Cache Publications: New York.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50202-0

A Mathematical Technique for Multi-period Planning of Unconventional Gas Field Development

Jui-Yuan Lee[a*], Raymond E. H. Ooi[b], Dominic C. Y. Foo[b], Raymond R. Tan[c]

[a]*Department of Chemical Engineering and Biotechnology, National Taipei University of Technology, 1, Sec. 3, Zhongxiao E. Rd., Taipei 10608, Taiwan, ROC*
[b]*Department of Chemical & Environmental Engineering/Centre of Excellence for Green Technologies, University of Nottingham Malaysua Campus, 43500 Semenyih, Selangor, Malaysia*
[c]*Chemical Engineering Department/Center for Engineering and Sustainable Development Research, De La Salle University, 2401 Taft Avenue, 0922 Manila, Philippines*
juiyuan@ntut.edu.tw

Abstract

This paper presents a multi-period mathematical programming model for the planning of unconventional gas field development, taking into account variations in the gas production rates and the gas demand. Based on a superstructure, the model consists of gas flowrate and impurity load balances and operational constraints. The proposed approach is demonstrated through a modified literature case study.

Keywords: gas sweetening, process integration, superstructure, optimisation.

1. Introduction

Human activities have caused serious environmental issues such as climate change and pollution. Despite increased use of renewable energy, fossil fuels will remain dominant although these conventional energy sources are a major contributor to greenhouse gas emissions. Among fossil fuels, natural gas (NG) is relatively clean with lower carbon dioxide (CO_2) intensity compared to coal and oil.

Development of unconventional gas fields has recently become common practice in the oil and gas industry, with the growing demand for NG as part of the global shift towards low-carbon energy sources. Unconventional gas fields produce low quality NG with high CO_2 content. Such sour gas needs removal of CO_2, or sweetening, in the offshore facilities before proceeding to the onshore gas processing plant (GPP), so as to meet the sales gas specifications for downstream applications (e.g. power generation). Removal of CO_2 offshore reduces the gas flowrates and hence the compressor duty, with reduced need for corrosion resistant pipes. The captured CO_2 may then be injected around the gas field for permanent geological storage or enhanced oil recovery (EOR). In addition, blending sour gas with sweet gas of lower CO_2 content to meet the specifications for gas processing is also a common practice in the industry. This strategy helps to reduce the dependence on gas sweetening and optimise the capacities of the units. However, sweet gas from conventional fields normally has limited availability; therefore, its utilisation also needs to be optimised.

Planning the development of gas fields thus involves determining: 1) which gas sweetening technique(s) should be used; 2) how much sour gas should be treated; and 3)

how much sweet gas should be used. Process integration techniques may be extended to address the gas field planning problem, which is analogous to the effluent treatment problem. Smith and co-workers (Wang and Smith, 1994; Kuo and Smith, 1997) developed graphical pinch analysis methods for the design of distributed effluent treatment systems, where impurity loads are to be removed from wastewater streams to comply with the environmental discharge limit. Other types of pinch-based techniques were also developed for effluent treatment networks (Ng et al., 2007; Bandyopadhyay, 2009). In these previous works, the main aim is to identify the minimum treatment flowrate, thus reducing the capital cost. Furthermore, the methods are designed to handle dilute waste streams with very low impurity content (typically at ppm levels), in which case flowrate losses due to impurity removal is negligible.

However, in the gas sweetening problem, sour NG streams can contain 30-50% of CO_2. This indicates that the sweetening process involves significant flowrate losses. Therefore, the earlier developed methods for waste treatment are not directly applicable before appropriate modifications. Foo et al. (2016) proposed pinch-based graphical and optimisation approaches to identify the minimum extent of CO_2 removal, based on the techniques for waste treatment systems (Ng et al., 2007) and carbon capture planning for the power generation sector (Ooi et al., 2013).

Apart from the aspects already addressed in the recent literature, variations in the gas production rates and the required flowrate for gas processing are also important operational issues to be taken into account. In this work, a multi-period mathematical model is developed for optimal planning of gas field development, considering the flowrate variability. Comprised of mass balance equations and operational constraints based on a superstructure, the time-dependent formulation can also be used to analyse the effect of development scheduling on the overall planning. A case study is solved to illustrate the proposed approach.

2. Problem statement

The problem addressed in this paper is formally stated as follows. Given:

- A set of high-CO_2 sources $i \in \mathbf{I}^{H}$, where CO_2 needs to be removed before the NG streams proceed to the downstream GPP, which tolerates a specified maximum CO_2 content. It is assumed that each high-CO_2 source produces gas at a constant CO_2 concentration but a varying flowrate based on the production forecast. A set of time periods $t \in \mathbf{T}$ are defined to capture such variability in gas flowrates.
- A set of gas sweetening technologies $k \in \mathbf{K}$ (e.g. membrane separation and amine absorption) of known performance for CO_2 removal (i.e. fixed product recovery factor and fixed CO_2 concentration of the purified stream).

Due to the necessary removal of CO_2 from sour NG streams to meet the requirement for gas processing, the total flowrate to the GPP will be lower than that originally from the high-CO_2 gas fields. Such flowrate losses will then be made up by introducing low-CO_2 sweet gas from an existing or new field. However, sweet gas has limited availability due to extensive exploitation and development of conventional gas fields over the years. The decision on the amount of CO_2 removal from sour gas is made by considering the cost of gas sweetening and/or the availability of sweet gas. The objective is to determine the minimum extent of CO_2 to be removed/captured to give the required NG quality, whilst minimising sweet gas supply.

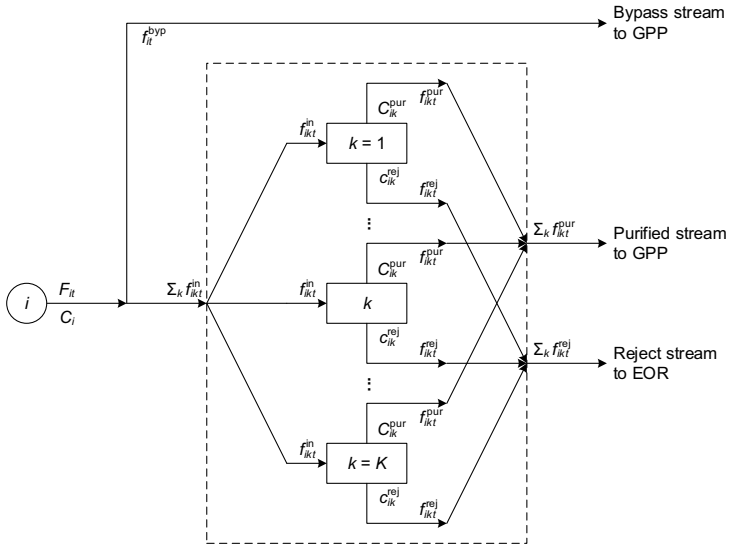


Figure 1. Schematic of Gas Sweetening

3. Model formulation

Figure 1 shows a schematic diagram of gas sweetening at a high-CO_2 source in a given time period. It can be seen that the sour gas from the source can be partially treated with available gas sweetening technologies. The bypass stream and the purified stream are then sent to the GPP, whilst the CO_2-rich reject stream may be used for EOR. Eq. (1) describes the flowrate balance for source *i* in time period *t*.

$$F_{it} = \sum_{k \in \mathbf{K}} f_{ikt}^{\text{in}} + f_{it}^{\text{byp}} \quad \forall i \in \mathbf{I}^{\text{H}}, t \in \mathbf{T} \tag{1}$$

where F_{it} is the flowrate of gas from source *i* in time period *t*, f_{ikt}^{in} is the inlet flowrate of gas sweetening unit *k* at source *i* in time period *t*, and f_{it}^{byp} is the bypass flowrate of gas from source *i* in time period *t*.

Eq. (2) states that only one technology can be selected for in-situ gas sweetening at each high-CO_2 source. Eq. (3) sets the upper limit on the inlet flowrate, whilst ensuring that no gas will be treated with unselected technology ($y_{ik} = 0$).

$$\sum_{k \in \mathbf{K}} y_{ik} \le 1 \quad \forall i \in \mathbf{I}^{\text{H}} \tag{2}$$

$$f_{ikt}^{\text{in}} \le F_{it} y_{ik} \quad \forall i \in \mathbf{I}^{\text{H}}, k \in \mathbf{K}, t \in \mathbf{T} \tag{3}$$

where y_{ik} is a binary variable indicating if technology *k* is selected for source *i*.

Eqs. (4) and (5) describe the flowrate and CO_2 load balances around gas sweetening unit *k* at source *i* in time period *t*, respectively.

$$f_{ikt}^{\text{in}} = f_{ikt}^{\text{pur}} + f_{ikt}^{\text{rej}} \quad \forall i \in \mathbf{I}^{\text{H}}, k \in \mathbf{K}, t \in \mathbf{T} \tag{4}$$

$$f_{ikt}^{\text{in}} C_i = f_{ikt}^{\text{pur}} C_{ik}^{\text{pur}} + f_{ikt}^{\text{rej}} c_{ik}^{\text{rej}} \quad \forall i \in \mathbf{I}^{\text{H}}, k \in \mathbf{K}, t \in \mathbf{T} \tag{5}$$

where f_{ikt}^{pur} and f_{ikt}^{rej} are the flowrates of the purified and reject streams, C_i is the CO_2 concentration of gas from source i, whilst C_{ik}^{pur} and c_{ik}^{rej} are the CO_2 concentrations of the purified and reject streams. The gas sweetening process is also characterised by the product recovery factor R_{ik} in Eq. (6).

$$f_{ikt}^{\text{pur}}\left(1-C_{ik}^{\text{pur}}\right)=R_{ik}f_{ikt}^{\text{in}}\left(1-C_i\right) \quad \forall i\in\mathbf{I}^{\text{H}},k\in\mathbf{K},t\in\mathbf{T} \tag{6}$$

Eqs. (7) and (8) impose the specification constraints on flowrate and CO_2 concentration for gas processing.

$$\sum_{i\in\mathbf{I}^{\text{H}}}\left(f_{it}^{\text{byp}}+\sum_{k\in\mathbf{K}}f_{ikt}^{\text{pur}}\right)+\sum_{i\in\mathbf{I}^{\text{L}}}f_{it}^{\text{SG}}\geq f_t^{\text{GP}} \quad \forall t\in\mathbf{T} \tag{7}$$

$$\sum_{i\in\mathbf{I}^{\text{H}}}\left(f_{it}^{\text{byp}}C_i+\sum_{k\in\mathbf{K}}f_{ikt}^{\text{pur}}C_{ik}^{\text{pur}}\right)+\sum_{i\in\mathbf{I}^{\text{L}}}f_{it}^{\text{SG}}C_i\leq f_t^{\text{GP}}C^{\text{GP}} \quad \forall t\in\mathbf{T} \tag{8}$$

where f_{it}^{SG} is the flowrate of sweet gas from low-CO_2 source $i\in\mathbf{I}^{\text{L}}$ in time period t, f_t^{GP} is the gas processing flowrate in time period t, and C^{GP} is the maximum CO_2 content for gas processing.

The objective may be to minimise the supply of sweet gas from low-CO_2 sources over the planning horizon, as given in Eq. (9), or alternatively to minimise the total operating cost associated with sweet gas supply and gas sweetening, as given in Eq. (10).

$$\min\phi_1=\sum_{i\in\mathbf{I}^{\text{L}}}\sum_{t\in\mathbf{T}}f_{it}^{\text{SG}}\Delta_t \tag{9}$$

$$\min\phi_2=\sum_{i\in\mathbf{I}^{\text{L}}}\sum_{t\in\mathbf{T}}UC_if_{it}^{\text{SG}}\Delta_t+\sum_{i\in\mathbf{I}^{\text{H}}}\sum_{k\in\mathbf{K}}\sum_{t\in\mathbf{T}}TC_{ik}f_{ikt}^{\text{in}}\Delta_t \tag{10}$$

where Δ_t is the length of time period t, UC_i is the unit cost of sweet gas, and TC_{ik} is the treatment cost for gas sweetening.

Due to the bilinear terms for CO_2 loads in Eq. (5), the overall model is a mixed integer nonlinear programme (MINLP), which is solved utilising global solver BARON in the GAMS environment on a Core i5-4340M, 2.90 GHz processor. Solutions for the case study presented in the next section were obtained with negligible processing time (< 1 CPU s), although global optimality may not always be guaranteed.

4. Case study

A modified literature example is solved to illustrate the proposed approach (Foo et al., 2016). Table 1 shows the data based on an actual commercial project in Southeast Asia. There are three high-CO_2 gas fields (A-C) to be developed. The gas production forecast over the 30-y planning horizon is shown in Figure 2(a). For onshore processing, the CO_2 content of the gas has to be reduced to 3% or below. A membrane separation unit with 95% product recovery and a fixed outlet CO_2 concentration of 2% may be used for in-situ CO_2 removal. Alternatively, an amine absorption unit with 90% product recovery and a 5% outlet CO_2 concentration can be used. Sweet gas is available from field D to compensate for the flowrate losses due to gas sweetening. It is assumed that the required total flowrate to the GPP is 1000 MMscfd in 1-10 y, 1200 MMscfd in 11-20 y, and 1500 MMscfd in 21-30 y to meet the growing energy demand. Two scenarios are analysed in this case study.

Table 1. Data for the Gas Sweetening Case Study

Gas field	Flowrate (MMscfd)	CO_2 content (%)
A	50-500	15
B	20-300	30
C	0-200	50
D (sweet gas)	To be determined	2

Scenario 1 aims to minimise the flowrate of sweet gas because of its limited availability. Solving the MINLP model (Eq. (9) subject to Eqs. (1)-(8)) gives the minimum sweet gas supply of 9111093 MMscf. The flowrates of the inlet streams for gas sweetening are shown in Figure 2(b), and the flowrates of the bypass streams in Figure 2(c). In addition, the sweet gas flowrate is shown in Figure 2(d). From these results, the design capacities of the gas sweetening units and the pipelines can be determined. For example, the CO_2 removal unit at field A requires a 423.1-MMscfd capacity, and its bypass line a 115.4-MMscfd capacity. Figure 3(a) shows the resulting gas sweetening system. It is worth noting that only membrane separation is selected for gas sweetening, because its higher product recovery helps to minimise the sweet gas flowrate.

In scenario 2, the objective is to minimise the operating cost of the system. The unit cost of sweet gas is taken to be 0.2 USD/MMscf, whilst the treatment costs for membrane and amine absorption units are taken as 0.5 and 0.3 USD/MMscf respectively (Foo et al., 2016). Solving the MINLP model (Eq. (10) subject to Eqs. (1)-(8)) gives the minimum operating cost of 3,878,226 USD. Figure 3(b) shows the corresponding gas sweetening system. Note that amine absorption is selected for fields B and C instead of membrane separation because it is less expensive, though this increases the sweet gas flowrate (due to the lower product recovery of amine absorption).

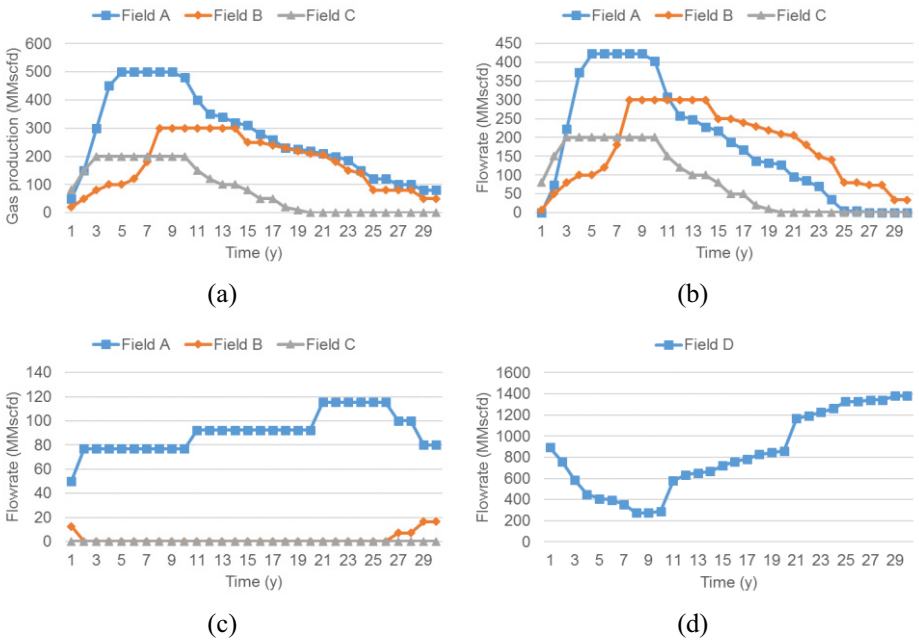


Figure 2. (a) Gas Production Rates; (b) Inlet Flowrates for Gas Sweetening; (c) Bypass Flowrates; (d) Sweet Gas Flowrates

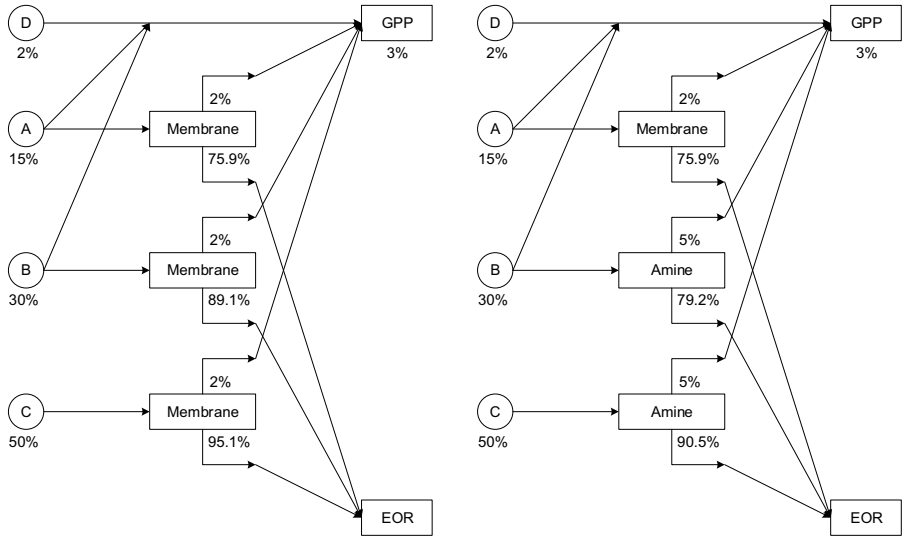


Figure 3. Gas Sweetening System for (a) Scenario 1 and (b) Scenario 2.

5. Concluding remarks

Based on a superstructure, a multi-period mathematical model for gas field development planning has been developed in this work. The formulation captures the variability in gas production and demand. A modified case study was solved to illustrate the proposed approach. Future work will consider capital investment for optimal cost trade-offs. The effects of development scheduling and parametric uncertainties on the planning may also be analysed.

References

S. Bandyopadhyay, 2009, Targeting Minimum Waste Treatment Flow Rate, Chem. Eng. J., 152, 2–3, 367-375.

D.C.Y. Foo, R.E.H. Ooi, R.R. Tan, J.-Y. Lee, 2016, Process Integration Approaches to Optimal Planning of Unconventional Gas Field Development, Chem. Eng. Sci., 150, 85-93.

W.-C.J. Kuo, R. Smith, 1997, Effluent Treatment System Design. Chem. Eng. Sci., 52, 23, 4273-4290.

R.E.H. Ooi, D.C.Y. Foo, R.R. Tan, D.K.S. Ng, R. Smith, 2013, Carbon Constrained Energy Planning (CCEP) for Sustainable Power Generation Sector with Automated Targeting Model, Ind. Eng. Chem. Res., 52, 29, 9889-9896.

D.K.S. Ng, D.C.Y. Foo, R.R. Tan, 2007, Targeting for Total Water Network. 2. Waste Treatment Targeting and Interactions with Water System Elements, Ind. Eng. Chem. Res., 46, 26, 9114-9125.

Y.-P. Wang, R. Smith, 1994, Design of Distributed Effluent Treatment Systems, Chem. Eng. Sci., 49, 18, 3127-3145.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50203-2

Multi-period operational optimization of natural gas treating, blending, compressing, long-distance transmission, and supply network

Wen W. Zhang, Bing J. Zhang, Chang He, Ming Pan, Qing L. Chen*

School of Chemical Engineering and Technology, Guangdong Engineering Center for Petrochemical Energy Conservation, Sun Yat-Sen University, No. 135, Xingang West Road, Guangzhou, 510275, China
chqlin@mail.sysu.edu.cn

Abstract

Natural gas has emerged as a promising energy source due to its high quality and massive reserves. Operational optimization of natural gas networks including gas treating, blending, compressing, long-distance transmission, supplying and power generation are quite challenging since it contains highly nonlinear processes, complex thermodynamics, and a set of partial differential equations (PDE). In this article, we first introduce a multi-period optimization framework for the large scale natural gas network. Then, we develop a nonconvex mixed-integer nonlinear programming (MINLP) model aiming for better economic performance for the network. Various levels of market price and demand for electricity are considered. The MINLP framework integrates the models of CO_2 removing processes, compressors, pipeline transmission and combined-cycle gas and steam turbines. The optimization framework and mathematical model are applied to a case study extracted from an industrial park at South China. Economy and energy performances are investigated and an optimum operating strategy is drawn out. The practicability of the optimization framework and mathematical model makes it more possible to apply to real world problems.

Keywords: Operational optimization, Natural gas, Network, Mathematical modelling.

1. Introduction

Natural gas is one of the cleanest energy sources on earth. Raw natural gas exploited from offshore reservoirs is pre-processed and dried at the wellheads before it enters onshore treating plants (Zhang *et al.*, 2017). The remaining CO_2 in dry natural gas (DNG) is further removed to a certain level at onshore treating plants due to downstream users' restrictions on fuel specifications. This brings attention to DNG network (DNGN), which is designed for a network of DNG treating, blending, compressing, long-distance transmission, supplying and power generating.

A DNGN receives DNG, which varies in pressure and chemical compositions from several terminals. The DNGN is employed to process and transmit DNG, and generate power to fulfil power consumers' requirements. It's very complicated to model a DNGN because it would involve bilinear, quadratic, polynomial, exponential, and other higher order terms. The multi-period operations make it even more challenging. Therefore, it is necessary to apply advanced techniques of multi-period MINLP models to the systems for economic performance and energy savings, and customer satisfaction improvements.

Complex chemical process systems including the optimization of natural gas industry has been studied. Most of these contributions in the natural gas industry focused on the natural gas production planning and control, gas pipeline network, and supply chains. Selot *et al.* (2008) developed a nonconvex MINLP short-term operational planning model of natural gas transporting integrated with processing to minimize the operational costs for the network. Cafaro & Grossmann (2014) developed a large-scale MINLP model for a shale gas supply chain network. Gao & You (2015) also studied operation and design problem of a shale gas supply chain. Recently, J. Li et al. (2016) develop a global optimization-based planning and modelling to predict overall product yields and properties in an integrated refinery-petrochemical complex.

To our best knowledge, very little work has been dedicated to the optimization planning of multi-period operational process and supply network for DNG. In this paper, the entire planning model is carried out as a large-scale nonconvex MINLP optimization problem. The global optimization approaches are applied to solve the model to global optimality.

2. Problem statement

A typical DNGN is illustrated in Figure 1. The DNGN receives DNG from several offshore terminals. DNG is processed, compressed and transmitted to power plants through pipelines to produce power which should meet the demand of downstream customers. It's not economical for power plants to work at full load persistently when demands fluctuate with seasons or other external reasons. We aim to optimize operation parameters for production and consumption to match in a period to avoid unnecessary energy consumption for the purpose of maximizing the total profit.

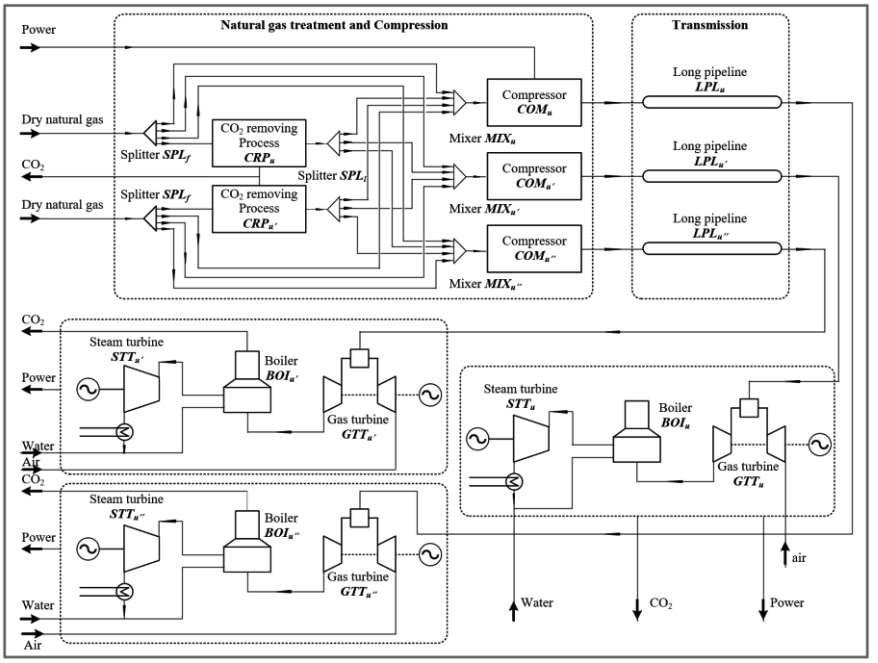


Figure 1. A schematic diagram of a typical DNGN

Methyl diethanolamine (MDEA) process was adopted for removing CO_2 in DNG. As being processed, DNG is then transmitted through a set of long pipelines to power plants in which gas turbines, boilers and steam turbines are installed. DNG can be stored on a short-term basis inside the pipelines due to its compressibility and the massive pipeline length. The challenges to model pipelines lie primarily in demand uncertainty and the dynamic behaviour of the long-distance pipelines.

3. Mathematic models

The whole DNGN is consisted of three parts: DNG treatment and compression, DNG transmission, and power generation. The mathematic models are presented as follows.

3.1. DNG treatment and compression

3.1.1. The model for CO_2 treating process

Absorption of CO_2 is achieved mainly by the use of amine in an absorber column. CO_2-rich amine solvents are regenerated in a stripper column and then pumped back to the absorber. a relationship between molar flow rate of solvent and CO_2 removal studied by *Mofarahi et al.* (2008) can be used to calculate the molar flow rate of solvent. Hence, the power consumption by pump can be calculated in Eq. (1), where h is a set of periods, u is a set of processes, Δp is pressure drop between absorbers and regenerators, $F_{h,u}^{sol}$ is mole flow rate of solvent, γ^{sol} is molar density of solvent, and η_u is pump efficiency. The sets in the domains are corresponded to the sets illustrated in Figure 1.

$$E_{h,u}^{pump} = \frac{\Delta p_u F_{h,u}^{sol}}{3600\gamma^{sol}\eta_u} \quad \forall u \in CRP \tag{1}$$

Eq. (2) expresses the mass balance for CH_4 before and after being treated by MDEA.

$$M_{h,u',u,c}F_{h,u} = M_{h,u,u',c}F_{h,u,u'} \quad \forall c = CH4, u \in CRP \tag{2}$$

where $M_{h,u,u',c}$ is the molar fraction of component c.

Low-pressure steam is used to meet the energy requirement of the reboilers in the regenerators. According to energy balance, Eq. (3) is presented to express that the heat provided to the reboilers $Q_{h,u}$ should equal the heat removed from the condensers, the reaction heat for CO_2 desorption, and the heat to increase the solvent temperature from the inlet to outlet of regenerator.

$$Q_{h,u} = \epsilon^{reg}\iota^{reg}F_{h,u}^{CO2} + \delta^{reg}F_{h,u}^{CO2} + \varepsilon^{sol}\Delta\tau^{reg}F_{h,u}^{sol} \quad \forall u \in CRP \tag{3}$$

where parameters ϵ, ι^{reg}, δ^{reg}, ε^{sol}, $\Delta\tau^{reg}$ are the specific latent heat of reflux stream in the regenerator, the reflux ratio of regenerator, the heat of desorption reaction, the specific heat capacity of solvent, and the temperature difference between solvent entering and exiting the regenerator, respectively.

3.1.2. The model for compressors

For isentropic compression, isentropic efficiency is defined to calculate the practical enthalpy change. Hence, the required power is computed using Eq. (4).

$$E_{h,u} = \frac{1}{\eta_{h,u}}\sum_{c\in C}\left(\varepsilon_c M_{h,u',u,c}F_{h,u}\right)\left[\left(R_{h,u}\right)^{\frac{\theta_u-1}{\theta_u}} - 1\right] \quad \forall u \in COM \tag{4}$$

where, variable $E_{h,u}$ is power consumed or generated, M is mole fraction of component c, isentropic exponent θ_u is the ratio of heat capacities of gas streams at constant pressure and temperature, ε_c is specific heat capacity, and c is a set of chemical components.

3.1.3. The model for mixers and splitters

Streams that comes from different processes are mixed together in the mixers. Eq. (5) expresses the pressure constraints in mixers. Binary variable $X_{h,u',u}$ is used to express whether a corresponding process is in operation. m is a large number.

$$P_{h,u} \le P_{h,u'} + m\,(1 - X_{h,u',u}) \quad \forall u \in MIX \tag{5}$$

Streams which go through splitters are divided into several streams with the same chemical components. Hence mass balance for the splitters and mass balance for individual chemical components can be easily modelled.

3.2. The model for transmission

A continuity equation across the pipelines is given in Eq. (6) under the assumption of ideal gas and isothermal conditions.

$$\frac{WA}{RT}\frac{\partial P}{\partial h} + \frac{\partial F_{h,u}}{\partial x} = 0 \quad \forall u \in LPL \tag{6}$$

where, A_u is cross-sectional area in pipeline u, $\frac{\partial F_{h,u}}{\partial x}$ is partial derivative of molar flow rate with respect to position, R is gas constant, T is temperature. The PDE above is derived from fundamentals of mass balance and is suitable for any given pipes. Baumrucker & Biegler (2010) proposed an approach that solves the problem using an integrated form of the continuity equation. Applying the approach to the continuity equation, the result takes the form in Eq. (7).

$$\frac{W_{h,u}A_uL_u}{RT_{ref}}\left(\bar{P}_{h+1,u} - \bar{P}_{h,u}\right) = \frac{1}{2}\left[\left(F^{in}_{h,u} - F^{out}_{h,u}\right) + \left(F^{in}_{h+1,u} - F^{out}_{h+1,u}\right)\right]\Delta h \quad \forall u \in LPL \tag{7}$$

where, variables $\bar{P}_{h,u}$, $W_{h,u,c}$ are average pressure, molecular weight of component c, parameters T_{ref} and L_u are reference temperature and length of pipeline u, respectively.

3.3. The model for power plant

A typical power plant consists of gas turbines, boilers and steam turbines. The model for power plant is represented as follows. Eq. (8) is the model of gas turbines (Smith, 2005), Eq. (9) is the model of boilers (Shang & Kokossis, 2004), and Eq. (10) is the model of steam turbines (Mavromatis & Kokossis, 1998). In Eq. (8), $J_{h,u}$ is work production, $\Delta H^{comb}_{h,u}$ is net heat of combustion, and $Q_{h,u,u''}$ is the heat of flue. In Eq. (9) variable $FB^{steam}_{h,u}$ is flow rate of steam produced by boilers, α_u and β_u are operational parameters for boilers, parameter $F^{steam,max}_{h,u}$ is the upper bound of boilers. $Q^{exh}_{h,u}$ is the heat of exhaust from boilers. In Eq. (10) $FT^{steam}_{h,u}$ is the molar flow rate of steam consumed by turbines, ϑ_u and σ_u are operational parameters for steam turbines, ϵ^{ise}_u is specific isentropic enthalpy change of streams in turbines.

$$J_{h,u} = F_{h,u',u}\Delta H^{comb}_{h,u} - Q_{h,u,u''} - J^{loss}_{h,u} \quad \forall u \in GTT \tag{8}$$

$$\alpha_u F^{steam,max}_{h,u} X_{h,u} + (1 + \beta_u) FB^{steam}_{h,u} = Q_{h,u,u''} - Q^{exh}_{h,u} \quad \forall u \in BOI \tag{9}$$

$$3600E_{h,u} = \frac{6}{5}\frac{1}{\vartheta_u}\left(\epsilon_u^{ise} - \frac{\sigma_u}{F_u^{max}}\right)\left(FT_{h,u}^{steam} - \frac{1}{6}F_u^{max}Y_{h,u}\right) \quad \forall u \in STT \tag{10}$$

3.4. Economic objective

The objective is to obtain maximum profit of the total DNGN, as is expressed in Eq. (11). The first item on the right side of Eq. (11) is product income, the second item is the costs of raw natural gas, the third item is the cost of electricity purchased for compressors, the forth item is steam consumed in CO_2 treating processes, ζ is the price per unit of power, DNG, or steam. π is other fixed costs.

$$OBJ = \sum_h \left[\sum_{u\in GTT\cup STT}(\zeta^s E_{h,u}) - \sum_s\left(\zeta^D F_{h,s}^{DNG}\right) - \sum_{u\in COM\cup CRP}(\zeta^p E_{h,u}) - \sum_{u\in CRP}\zeta^h Q_{h,u}\right] - \pi \tag{11}$$

4. Case study

A DNGN from South China is taken as an example and is investigated in this study. The data for the case is shown in Table1. 3 periods are considered and each period is 168h.

Table 1. Basic data for the case

Period	Maximum Flow (kmol/h)		Composition(CH_4)		Pressure(MPa)
	1#	2#	1#	2#	DNG
H1	3500	2000	0.95	0.90	2.48
H2	3500	2000	0.89	0.91	2.52
H3	3500	2000	0.87	0.90	2.50

The MINLP model includes a large number of bilinear items which makes it difficult to solve within limited computational time. For example, $(M_{h,u,u',c}F_{h,u,u'})$ in Eq. 2 is used to calculate the flowrate of component c from process *u* to *u'*. In order to reduce bilinear items, we initialize the variable $M_{h,u,u',c}$ at the outlet of mixers, and use $F_{h,u,u',c}$ to replace the item $(M_{h,u,u',c}F_{h,u,u'})$. As a result, the original MINLP model can be simplified and only includes the nonlinear equations for compressors. Hence, a strategic iteration approach is created to solve the problem. Through this method, a satisfying result within tolerance range is gained. The solution strategy is illustrated in Fig. 2. The simplified MINLP model is solved using global optimization solver BARON 16.3.4 in GAMS. The model involves 871 variables and 18 binary variables. The computational time is 2.33s.

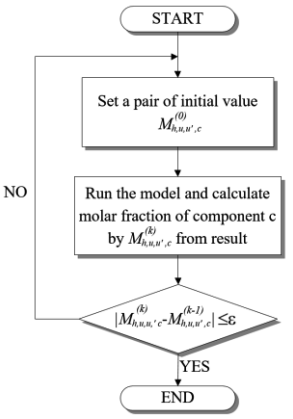


Figure 2 A solution strategy for the model

Table 2. Flowrate of DNG in sources

Period	H1	H2	H3
#1 (kmol/h)	2869	3400	2882
#2 (kmol/h)	2000	2000	2000

The result showed that the total profit reaches $ 5,419,052 within 3 periods. As is shown in Table 2, the DNGN received maximum flowrate from the DNG source 1, while varied the flowrate in different periods from source 2. Moreover, the flowrate in the processing section showed a different blending map and the network varied the power upload rate over elapsing time. The result indicates that the profitability can be improved by changing the operation parameters, the schedule of blending and the load of CO_2 removing process.

5. Conclusions

A large-scale nonconvex MINLP formulation is developed for a DNGN. The planning model integrates CO_2 removing processes, compressor models, transient pipeline transmission models and combined-cycle gas and steam turbine power plant models. Multiple DNG streams and different product specifications are taken into account in the MINLP formulation. A strategy for linearization is applied to the large-scale nonconvex MINLP optimization model and a reasonable result is gained.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (No. 21776323) and the project of Guangdong Provincial Natural Science Foundation of China (No. 2015A030313112).

References

B. T. Baumrucker, L. T. Biegler, 2010, MPEC strategies for cost optimization of pipeline operations, Computers and Chemical Engineering, 34(6), 900–913.

D. C. Cafaro, I. E. Grossmann, 2014, Strategic planning, design, and development of the shale gas supply chain network, AIChE Journal, 60(6), 2122–2142.

J. Gao, F. You, 2015, Shale Gas Supply Chain Design and Operations toward Better Economic and Life Cycle Environmental Performance: MINLP Model and Global Optimization Algorithm, ACS Sustainable Chemistry & Engineering, 3(7), 1282–1291.

X. Li, E. Armagan, A. Tomasgard, P. I. Barton, 2011, Stochastic pooling problem for natural gas production network design and operation under uncertainty, AIChE Journal, 57(8), 2120–2135.

S. P. Mavromatis, A. C. Kokossis, 1998, Hardware composites: A new conceptual tool for the analysis and optimisation of steam turbine networks in chemical process industries. Part I: Principles and construction procedure, Chemical Engineering Science, 53(7), 1405–1434.

M. Mofarahi, Y. Khojasteh, H. Khaledi, A. Farahnak, 2008, Design of CO2 absorption plant for recovery of CO2 from flue gases of gas turbine, Energy, 33(8), 1311–1319.

A. Selot, L. K. Kuok, M. Robinson, T. L. Mason, P. I. Barton, 2008, A short-term operational planning model for natural gas production systems, AIChE Journal, 54(2), 495–515.

Z. Shang, A. Kokossis, 2004, A transhipment model for the optimisation of steam levels of total site utility system for multiperiod operation, Computers and Chemical Engineering, 28(9), 1673–1688.

R. Smith, 2005, Chemical Process Design and Integration. John Wiley & Sons, Ltd.

B. J. Zhang, Q. L. Chen, J. Li, C. A. Floudas, 2017, Operational strategy and planning for raw natural gas refining complexes: Process modeling and global optimization, AIChE Journal, 63(2), 652–668.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50204-4

Product Decomposition in Supply Chain Planning

Braulio Brunaud[a], M. Paz Ochoa[a], Ignacio E. Grossmann[a*]

[a]*Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA*
**grossmann@cmu.edu*

Abstract

Lagrangean decomposition has been used to overcome the difficulties in optimizing large-scale supply chain planning models. Decomposing the problem by time periods has been established as a useful technique. In this paper a novel decomposition scheme by products is presented. The decomposition is based on a reformulation of knapsack constraints in the problem. The new approach also allows for simultaneous decomposition by products and time periods, enabling the generation of a large number of subproblems, with the potential benefit of using parallel computing. The case study shows that product decomposition shows a similar performance than temporal decomposition. Selecting different orders of products and aggregating the linking constraints can improve the efficiency of the algorithm.

Keywords: Supply Chain Planning, Lagrange Decomposition, Product Decomposition.

1. Introduction

Tactical planning of supply chains involves the decision of material flows and inventories throughout a network of manufacturing sites and warehouses to satisfy the demand of each customer. To ensure that best decisions are made for the mid-term, optimization models are employed. These models can be very large and hard to solve. To obtain the optimal solution in reasonable time, decomposition techniques such as Lagrangean decomposition (Guignard and Kim, 1987) can be used. In this type of decomposition, the model is broken into subproblems, and the constraints linking them are dualized, i.e. transferred to the objective function with a penalty term known as Lagrange multiplier. The objective is to find the multipliers that yield the tightest bound for the problem.

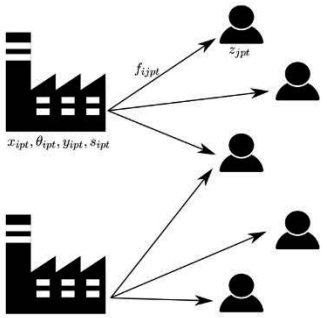
Jackson and Grossmann (2003) use temporal decomposition to solve a multi-site, multi-period planning problem. Terrazas-Moreno et al. (2011) compare spatial and temporal decomposition for the same kind of problem. They conclude that the temporal decomposition gives tighter bounds than spatial decomposition for some types of production planning problems. In temporal decomposition, the problem is decomposed by time periods, which are linked by an inventory balance constraint. There is a natural dynamic structure when decomposing by time periods. The solution of a subproblem at a given time period depends only on the inventory at the end of the previous period.

The horizon of a typical supply chain planning problem is between 6 months to 2 years, usually divided into monthly periods (Brunaud and Grossmann, 2017). At the same time the optimal decisions must be determined for a number of products that ranges typically from 50 to 1,000. However, products do not exhibit the same dynamic structure as time periods; it is not possible to say that one product comes before another. In this paper we show that such a dynamic structure can be exposed through reformulation allowing to decompose the problem by products and apply Lagrangean decomposition. Furthermore, we show that the problem can be decomposed simultaneously by products and time

periods. van Elzakker et al. (2014) propose a decomposition by SKU based on heuristics, different to the exact reformulation followed by Lagrangean decomposition from this paper. We evaluate each decomposition scheme, and explore the effects of subproblem aggregation, linking constraint aggregation, and order of products through a multi-period multi-product production planning example.

2. Problem Description

Given is a demand forecast for a set of products p ∈ P. It is required to determine the optimal production amounts, inventory levels, and shipments to markets j ∈ J at several multi-product facilities i ∈ I. The objective is to maximize the profit for a finite horizon divided in time periods t ∈ T of length L_t. For example, if a plan for 2 months is divided into weekly periods (t), each week has a length of 168 hours (L_t).Since the production rate is different for each product, the production amount is expressed in terms of the number of hours devoted to produce a product in a given period. Fig. 1 describes the problem and the main variables used.



Variables		Parameters	
x_{ipt}	production amount	β_p	price
θ_{ipt}	production time	γ_{ip}	production cost
s_{ipt}	stock	H_i	holding cost
y_{ipt}	production executed	σ_p	setup cost
f_{ijpt}	flow	TC_{ij}	transportation cost
z_{jpt}	Sales	SUT_{ip}	setup time
		R_{ip}	production rate
		D_{jpt}	demand

Figure 1. Problem diagram and nomenclature

The problem is formulated as an MILP from Eqs. (1)-(7), a standard supply chain planning problem (Brunaud and Grossmann, 2017).

$$\textit{Max} \quad \sum_i \sum_j \sum_t \beta_p z_{jpt} - \sum_i \sum_p \sum_t \left(\gamma_{ip} x_{ipt} + H_i s_{ipt} + \sigma_p y_{ipt}\right) - \sum_i \sum_j \sum_p \sum_t TC_{ij} f_{ijpt} \tag{1}$$

$$s.t. \quad s_{ipt} = s_{ipt-1} + x_{ipt} - \textstyle\sum_j f_{ijpt} \quad \forall ipt \tag{2}$$

$$\sum_p \theta_{ipt} + SUT_{ip} y_{ipt} \le L_t \quad \forall it \tag{3}$$

$$R_{ip}\theta_{ipt} = x_{ipt} \quad \forall ipt \tag{4}$$

$$\sum_i f_{ijpt} = z_{jpt} \quad \forall jpt \tag{5}$$

$$z_{jpt} \le D_{jpt} \quad \forall jpt \tag{6}$$

$$x_{ipt}, \theta_{ipt}, s_{ipt}, f_{ijpt}, z_{jpt} \ge 0, y_{ipt} \in \{0,1\} \tag{7}$$

3. Temporal Decomposition

In temporal decomposition, the goal is to divide the problem into individual time periods, which are linked by inventory variables. In Fig. 2 the inventory balance constraint is represented by the nodes, and the inventory continuity constraint ensuring that the final inventory at a given period is equal to the initial in the next period, is represented by the box. The problem can be decomposed dualizing the inventory continuity constraint.

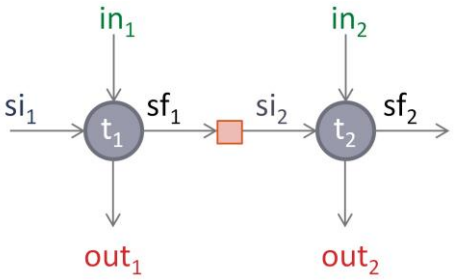


Figure 2. Representation of temporal linking constraints

Following the representation from Fig. 2 the previous model needs to be reformulated including variables to represent the initial (*si*) and final stock (*sf*). With these variables, Eq. (2) is replaced by Eqs. (8) and (9). Eq. (9) is then dualized to decompose the problem. The term $\lambda_{ipt}(sf_{ipt} - si_{ipt+1})$ is added to the objective function for each of the constraints represented by Eq. (9).

$$sf_{ipt} = si_{ipt} + x_{ipt} - \sum_j f_{ijpt} \quad \forall ipt \tag{8}$$

$$sf_{ipt} = si_{ipt+1} \quad \forall ipt \tag{9}$$

4. Product Decomposition

Unlike time periods, products do not have the same dynamic structure. They are usually sharing capacity in a knapsack constraint, whether is production, transportation, or inventory capacity. To reformulate the problem and obtain an equivalent dynamic structure, consider a planner that optimizes one product at a time. After the first flows and inventory for the first product have been optimized, the planner needs to know how much capacity is left for the next product. This process is shown in Fig. 3. We reformulate Eq. (3) according to the representation in Fig. 3 to obtain Eqs. (10) and (11), where ci_{ipt} represents the number of available hours for product p, and ci_{ipt} is the number of available hours after planning product p.

$$cf_{ipt} = ci_{ipt} - \theta_{ipt} - SUT_{ip}y_{ipt} \quad \forall ipt \tag{10}$$

$$cf_{ipt} = ci_{i(p+1)t} \quad \forall ipt \tag{11}$$

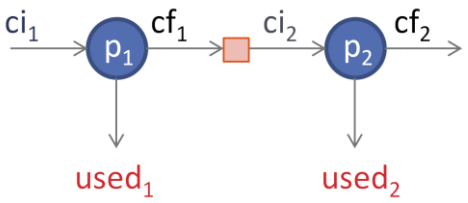


Figure 3. Product decomposition representation

The production capacity is a local variable for each product. The initial available hours for the first product is the entire length of the period L_t. Note that in order to write down Eq. (11) an arbitrary order is given to the products. When Eq. (11) is dualized, the problem is decomposed into subproblems for each product. The dualization adds the term $\mu_{ipt}\left(cf_{ipt} - ci_{i(p+1)t}\right)$ to the objective function for each constraint defined by Eq. (11).

Note that once the problem is decomposed, each subproblem is still a multi-period problem. Thus, it can also be decomposed by time periods following the reformulation from the previous section. This leads to a simultaneous decomposition by products and time periods where a subproblem can be as small as containing a single product and a single time period, opening the door to generating thousands of subproblems if desired.

5. Case Study

The computational performance of the proposed decomposition schemes is assessed through their application on the problem described in Section 2, motivated by an industrial application and described in Terrazas-Moreno, et al. (2011). The problem studied has 6 manufacturing sites and 10 markets. The number of products and time periods was set to 6 and 20, keeping the number of products equal to the number of time periods to ensure a fair comparison between product and temporal decomposition. Each time period represents 720 hours. The MILP formulations are modeled in JuMP and solved with Gurobi 7.5 in an Intel i7 machine with 16 Gb of RAM. Problem sizes are shown in Table 1.

Table 1. Case study problem sizes

Problem Size	Constraints	Variables	Binaries
6	2,232	4,032	216
20	24,800	44,800	2,400

5.1. Decomposition Schemes Comparison

Product decomposition (P), temporal decomposition (T), and simultaneous product and temporal decomposition (PT) were compared using two options to initialize the Lagrange multipliers: (1) zero ($\lambda = 0$), and (2) the multipliers obtained from solving the LP relaxation (λ = LP). The multipliers where updated using the subgradient method. The results for the wall clock time and optimality gap for 50 iterations are presented in Table 2.

Even though the performance of the product decomposition was superior in the examples presented, in other cases the temporal decomposition obtained a lower gap. It is not possible to conclude that one decomposition scheme is superior to the other. The simultaneous product and temporal decomposition does not result in an increased

performance for small problems, because with a larger number of subproblems there is a larger number of multipliers that need to optimized.

Table 2. Comparison of decomposition schemes for 50 iterations

	λ = 0			λ = LP		
	P	T	PT	P	T	PT
Size = 6						
Time (s)	65	144	4	288	12	17
Gap (%)	6.8	6.1	19	2.7	12	13
Size = 20						
Time (s)	106	905	18	132	131	152
Gap (%)	8.8	18	62	40	76	146

5.2. Aggregation of linking constraints

Since the difficulty of the Lagrangean decomposition algorithm lies in finding the optimal set of multipliers giving the tightest bound, the algorithm can benefit from reducing the number of multipliers. One way to accomplish this while keeping the size of the subproblems, is to aggregate the linking constraints before dualization. For example, Eq. (9) linking initial and final inventories can be replaced by the surrogate constraint from Eq. (12).

$$\sum_i sf_{ipt} = \sum_i si_{ipt+1} \quad \forall pt \tag{12}$$

The results of applying this aggregation for the three decomposition schemes is presented in Table 3. The column header indicates the summation indices of the aggregated constraint. For example, (ip) indicates sum Eq. (9) over sites and products.

Table 3. Effect of linking constraint aggregation

	P				T				PT	
	none	i	t	it	none	i	p	ip	none	ip/it
Size = 6										
Time (s)	288	6	2476	6	9	48	62	27	17	17
Gap (%)	2.8	0.0	1.28	0.0	12	0.6	0.3	2.4	13	9.2
Size = 20										
Time (s)	132	1,072	114	300	131	65	156	49	151	160
Gap (%)	40	20	24	41	76	124	42	2.8	146	57

The results show that the aggregation can help in improving the performance of the algorithm. The improvements are problem dependent and require experimentation.

5.3. Order of products

As mentioned before, in order to decompose the problem by products, an arbitrary order is given to the set. This effect was explored by solving the problem of size 6 for 1,000 random orderings of the products. The gap obtained after 10 iterations is presented in the histogram in Fig. 4. With most orderings, the gap obtained is less than 10%. However, there is a significant number of orderings that yield a large gap with this algorithm

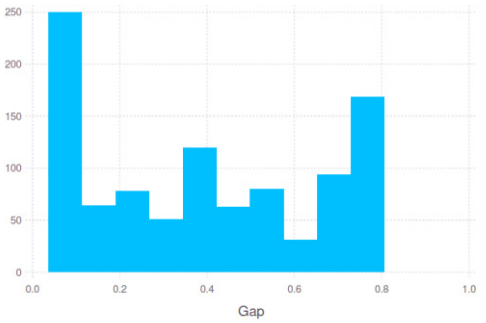


Figure 4: Effect of ordering of products

6. Conclusions

Product decomposition is a novel decomposition scheme that can potentially lead to improved results when applied to Lagrangean decomposition. Decomposing the problem by products is not consistently better or worse than decomposing by time periods. This is explained by the mathematical equivalency of both reformulations. This novel decomposition scheme is then presented as an alternative to modelers to experiment during the implementation of Lagrangean decomposition. Simultaneous product and temporal decomposition allows to generate a larger number of smaller subproblems that can be solved faster. The number of subproblems used is also a matter of experimentation in the implementation phase of the algorithm.

References

Brunaud, B., & Grossmann, I. E., 2017. Perspectives in multilevel decision-making in the process industry. Frontiers of Engineering Management, 4(3), 256-270.

Guignard, M., & Kim, S., 1987. Lagrangean decomposition: A model yielding stronger Lagrangean bounds. *Mathematical programming*, *39*(2), 215-228.

Van Elzakker, M. A. H., Zondervan, E., Raikar, N. B., Hoogland, H., & Grossmann, I. E., 2014. An SKU decomposition algorithm for the tactical planning in the FMCG industry. *Computers & Chemical Engineering*, *62*, 80-95.

Jackson, J. R., & Grossmann, I. E., 2003. Temporal decomposition scheme for nonlinear multisite production planning and distribution models. *Industrial & engineering chemistry research*, *42*(13), 3045-3055.

Terrazas-Moreno, S., Trotter, P. A., & Grossmann, I. E., 2011. Temporal and spatial Lagrangean decompositions in multi-site, multi-period production planning problems with sequence-dependent changeovers. *Computers & Chemical Engineering*, *35*(12), 2913-2928.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50205-6

Resilient scheduling under uncertain processing times: a hybrid CP/TOC approach

Franco M. Novara[a], Gabriela P. Henning[a,b,*]

[a]*FIQ, UNL, Santiago del Estero 2829, Santa Fe 3000, Argentina*
[b]*INTEC (UNL, CONICET), Güemes 3450, Santa Fe 3000, Argentina.*
ghenning@intec.unl.edu.ar

Abstract

Uncertainty and variability are inherent characteristics of industrial environments that affect production schedules and could turn them unfeasible or economically unattractive. Process-inherent uncertainty, one of the multiple stochasticity sources, influences task processing times, transforming them into one of the main uncertain parameters. This work proposes a scheduling approach to consider this situation in a proactive fashion at the decision stage, without resorting to the generation of scenarios. It relies on a Constraint Programming (CP) model that focusses on the Capacity Constrained Stage (CCS) to reduce its complexity and size. The proposal is tested with various instances of three case studies, showing that the approach is computationally efficient. The attained agendas can effectively absorb the processing times variabilities in an efficient way, exhibiting a resilient behaviour. To assess the schedules, they are compared with those reached by a deterministic CP formulation and with the agendas obtained by means of another proactive methodology proposed by the same authors.

Keywords: Scheduling, Uncertain processing times, Constraint programming, Theory of constraints, Multiproduct multistage batch plants.

1. Introduction

Industrial schedulers daily face uncertainty and unforeseen events that affect production agendas: incoming rush orders, changes in resource availability, raw materials late arrivals, stochasticity of processing times and other parameters, etc. Particularly, processing times variability is one of the most common issues that could lead to the generation of idle and wait times. While the first ones cause equipment under-utilization and reduce plant productivity, wait times can generate order delays and/or batch rejections due to material deterioration. Additionally, unexpected variations on processing times could lead to infeasible agendas and the need to frequently update them. Every time a schedule is revised, a phenomenon called shop-floor nervousness may occur and inefficiencies may be aggravated. To minimize this negative effect, a resilient agenda, able to absorb the influence of uncertain processing times without the need of introducing major changes, is pursued. To carry out this proactive scheduling, knowledge of uncertainty has to be considered at the decision stage.

Several academic approaches have been proposed to deal with uncertain processing times, such as fuzzy programming (Balasubramanian and Grossmann, 2003), genetic algorithms (Bonfill et al., 2008), two-stage stochastic programming (Bonfill et al., 2005), and constraint programming (Novara and Henning, 2017). Several techniques meant to cope with uncertainty (Li and Ierapetritou, 2008) rely on the generation of

scenarios, which increases the problem size, its computational expense, and prevents the solution of real-size problems. This contribution tackles uncertainties in an a priori manner, during the decision stage, by proposing a methodology able to generate resilient agendas that minimize the expected total tardiness. The short-term scheduling problem of multiproduct, multistage batch plants with uncertain operation times is addressed in a proactive fashion by means of a novel Constraint Programming (CP) model that makes use of Theory of Constraints (TOC) concepts. The model, which does not rely on the generation of scenarios or worst case formulations, takes advance of capacity limited resources; in particular, the capacity constrained stage (CCS), in order to minimize the complexity of the resulting CP model.

The remainder of this paper is organized as follows. The approach foundation is presented in the next section, including a description of how to capture uncertainties in the processing times. Then, the stochastic CP-TOC proposal is described, followed by a discussion of the results reached when applying the model to various instances of three case studies. Final remarks and future work are included in the last section.

2. Methodology

According to the TOC, every system has a capacity constrained resource (Spencer and Fox, 1995), whose operation subordinates that of the other resources. Tasks at this resource have to be carefully managed in order to prevent it becoming a bottleneck resource, with all its negative consequences. In this work it is assumed that multiproduct, multistage batch plants have a stage (CCS) that behaves as a capacity constrained resource. The proposed approach tries to prevent that delays due to processing time variabilities, taking place in units located upstream of the CCS, may introduce deferrals at this critical stage that can no longer be recovered.

Provided that processing times uncertainty directly affects the end time of each processing task, the proposed methodology associates a stochastic variable with the completion time of each batch. The value of this variable, which is named *end time subject to deviation* (*et-StD*$_b$), has a P probability of occurrence. For any given batch b, the accurate representation of *et-StD*$_b$ in a CP model requires a significant CPU effort during the solution stage. Nevertheless, such variable can be estimated in a practical way. The approximated variable that participates in the CP model is named *eetStD*$_b$, *estimated end time of batch b*. This variable is calculated as the expected end time of batch b plus n times the standard deviation of this estimated completion time, with n being the number of standard deviations associated with a P probability. The expected end time of batch b is computed by adding nominal processing times, that are assumed to have normal probability distributions, as well as idle times and changeover/setup times, which are supposed to have a deterministic behaviour. Therefore, the expected end time is also presumed to have a normal distribution. In order to estimate the *eetStD*$_b$ variable value with a reduced computational load, the CP-TOC model estimates the end time of batch b by taking into account the variance associated with the end time of batch b in a certain unit u belonging to the CCS, as well as the variance of the processing times of batch b in units located downstream the CCS.

3. New CP-TOC stochastic model

Uncertainties in the processing times are handled by means of two associated schedules that are simultaneously built. The first one, the production schedule, is created by just

using nominal processing times. It does not resort to buffer times or overestimated (conservative) processing times. Therefore, this agenda does not attempt to compensate possible delays that may result from the realizations of the uncertainties. Nevertheless, the assignment and sequencing decisions that are made try to better anticipate positive variabilities (greater than the nominal processing times) and their impact on total tardiness. The second schedule, which is not an operational agenda, is employed to capture the processing time variances of those tasks that are assigned to each equipment unit belonging to the CSS. When focusing on the CCS the structure of both schedules is similar; they have the same task-unit assignments and task sequencing. However, in the variance agenda, the length of any element takes into account the variance of the associated task ($task_{b,u}$) processing time, plus a fraction of the variance of the processing times of those activities required by batch *b* and located upstream of the CCS.

3.1. Model implementation

The CP model was developed using the OPL programming language, supported by the IBM ILOG CPLEX Optimization Studio environment (IBM ILOG, 2013).

Sets/Indexes. *B/b*: batches to be produced within the planning horizon. UC_u/-: units of stage s+1, which are unconnected to unit u, belonging to stage s. F_p/-: products that are forbidden as successors of product *p* when assigned to the same unit. *P/p*: products to be manufactured. *S/s*: processing stages. *U/u*: equipment units.

Parameters. $co_{u,p,p'}$: changeover time between products *p* and *p'* in unit *u*. dd_b: batch *b* due-date. *n*: number of standard deviations associated with probability *P*. $pt_{p,u}$: nominal processing time required by a batch of product *p* in unit *u*. $ptVar_{p,u}$: variance of $pt_{p,u}$.

Variables. $eetStD_b$: *et-*StD_b estimation (float). $nomEt_b$: nominal end time of batch *b* (integer). $stTask_{b,s}$: processing task of batch *b* at stage *s* (interval). $task_{b,u}$: processing task of batch *b* in unit *u* (interval). $unitBatchSeq_u$: sequence of the $task_{b,u}$ variables related to unit *u*, each one characterized by the *p* product concerning batch *b* (sequence). $var_{b,u}$: variance associated with the processing time of batch *b* in unit *u* and its preceding operations (interval). $varCCS_{b,u}$: variance of the processing time of batch *b* in unit *u* belonging to the CCS (interval). $varDS_b$: variance of the processing times of batch *b* in units located downstream of the CCS (integer). $varEnd_b$: variance associated with the end time of batch *b* in a given unit belonging to the CCS (integer). $varSeq_u$: sequence of the $var_{b,u}$ variables related to unit *u*, each one characterized by the *p* product concerning batch *b* (sequence). $varUS_b$: variance of the processing times of batch *b* in those units located upstream of the CCS (integer). *totalTardiness*: sum of the batches tardiness, each one computed as the non-negative difference between $eetStD_b$ and dd_b (float).

Constraints. Eq.(1) obliges each batch to be assigned to a single processing unit at each stage. Eq.(2) ensures the fulfillment of precedence relationships between consecutive operations of any batch *b*. Topology restrictions are captured by means of Eq.(3). Eq.(4) avoids overlapping the execution of tasks in any unit *u* and simultaneously inserts changeover times between successive tasks assigned to such unit. Eq.(5) avoids forbidden sequences with the aid of sequence variables and the *typeOfNext* construct (IBM ILOG, 2013).

$$alternative\left(stTask_{b,s}, all(u \in U_s)\ task_{b,u}\right), \qquad \forall\, s \in S, \forall\, b \in B \tag{1}$$

$$endAtStart\left(stTask_{b,s}, stTask_{b,s'=s+1}\right), \quad \forall\, b \in B, \forall\, s, s' \in S, s \neq Card(S) \tag{2}$$

$$minl\left(endOf\left(task_{b,\mathrm{u}}\right), endOf\left(task_{b,\mathrm{u}'}\right)\right) = 0, \forall\, b \in B, \forall\, \mathrm{u}' \in UC_u, \forall\, u, \mathrm{u}' \in U \quad (3)$$

$$noOverlap(unitBatchSeq_u, \mathrm{co}), \qquad \forall\, u \in U \quad (4)$$

$$typeOfNext\left(unitBatchSeq_u, task_{b,u}\right) \neq p', \forall (p, p')\ \in Fp, \forall\, u \in U, \forall\, b \in B_p \quad (5)$$

Eq.(6) and (7) express the calculation of the variables associated with the variance of the processing times of batch *b* in those units located upstream and downstream of the CCS, respectively, while Eq.(8) describes the variance related to the processing time of batch *b* in the assigned unit of the CCS. Eq.(9) and (10) capture the variance pertaining to the end time of batch *b,* when processed in unit *u* belonging to the CCS.

$$varUS_b = \sum_{\forall\, u \in U_s, \forall\, s \in \mathrm{S}, s < \mathrm{CCS}} presenceOf\left(task_{b,u}\right) \cdot ptVar_{b,u}, \forall\, b \in B \quad (6)$$

$$varDS_b = \sum_{\forall\, u \in U_s, \forall\, s \in \mathrm{S}, s > \mathrm{CCS}} presenceOf\left(task_{b,u}\right) \cdot ptVar_{b,u}, \forall\, b \in B \quad (7)$$

$$varCCS_{b,u} = presenceOf\left(task_{b,u}\right) \cdot ptVar_{b,u}, \quad \forall\, b \in B, \forall\, u \in U_{s=CCS} \quad (8)$$

$$sizeOf\left(var_{b,u}\right) = presenceOf\left(var_{b,u}\right) \cdot \left(maxl\left(0, varUS_b - startOf\left(var_{b,u}\right)\right)\right) + varCCS_{b,u}\ , \forall\, b \in B, \forall\, u \in U_{s=CCS} \quad (9)$$

$$varEnd_b = \max_{\forall\, u \in U_s, s \in \mathrm{S}, s = CCS}\left(endOf(var_{b,u})\right) \quad (10)$$

Eq.(11) - (13) relate the operational agenda to the variance schedule. Whenever the interval variable *task*$_{b,u}$, describing the processing task of batch *b* in unit *u*, participates in the solution, the corresponding interval variable capturing the variance of this task has to be incorporated too, as shown in Eq.(11). Eq.(12) prevents the overlapping of interval variables describing variances. Eq.(13) enforces those interval variables representing processing and variance tasks to follow the same sequence in each unit.

$$presenceOf\left(task_{b,\mathrm{u}}\right) = presenceOf\left(var_{b,u}\right), \ \forall\, b \in B, \forall\, u \in U_s, \mathrm{s} = CCS \quad (11)$$

$$noOverlap(varSeq_u), \qquad \forall\, u \in U_S, \mathrm{s} = CCS \quad (12)$$

$$typeOfNext\left(unitBatchSeq_u, task_{b,u}\right) = typeOfNext\left(varSeq_u, \mathrm{var}_{b,u}\right), \ \forall (p, p') \forall\, b \in B, \forall\, u \in U_s, \mathrm{s} = CCS \quad (13)$$

Eq.(14) captures the nominal end time expected for each batch *b*. Eq.(15) shows how the estimated end time of any batch *b* is calculated by combining deterministic and stochastic variables. Finally, Eq.(16) describes the expected total tardiness, which is the objective function to be minimized.

$$nomEt_b = \max_{\forall\, u \in U}\left(endOf(task_{b,u})\right), \qquad \forall\, b \in B \quad (14)$$

$$eetStD_b = nomEt_b + n \cdot \sqrt{(varEnd_b + varDS_b)}\ , \forall\, b \in B \quad (15)$$

$$totalTardiness = \sum_{\forall\, \mathrm{b} \in B} maxl(0, eetStD_b - dd_b) \quad (16)$$

4. Results

The CP-TOC model was validated by means of three case studies (C1-C3) available in the literature, which were slightly adjusted and are described in detail in Novara and Henning (2017). They correspond to the scheduling of multiproduct, multistage batch plants operating under a NIS-UW policy, and considering the expected total tardiness as the objective function. **C1:** It corresponds to a facility having 5 processing stages and 12 non-identical units. Sequence dependent changeovers and topology constraints are considered. **C2:** It is based on a multiproduct batch plant having 5-stages, 25 non-identical units and topological constraints. Product orders and processing units are characterized by release and ready times, respectively. In addition, some orders cannot be processed in certain units and there are forbidden processing sequences. **C3:** The facility has 5 stages with 20 dissimilar processing units. Fifty product orders are to be scheduled. Sequence dependent changeovers are considered.

The attained agendas have been contrasted by means of simulation to test their resilience to uncertainty (see Table 1). Stochastic processing times were described by means of various asymmetrical triangular distributions mimicking the behavior of real plants, with more delays than anticipations. Each distribution was obtained by adopting the deterministic processing time as the mode. The lower value was randomly created by subtracting to the mode the result of multiplying its value by an aleatory number that pertains to the [0, *inf*] interval. Likewise, the upper limit was obtained by adding to the mode the result of multiplying its value by a random number pertaining to [0, *sup*]. Different values to the *inf* and *sup* parameters have led to distinct problem instances.

Table 1. Comparison of deterministic and stochastic approaches by means of simulation. Expected total tardiness is reported

Case study	inf	sup	CP deterministic model (Novara and Henning, 2017)	Stochastic approaches	
				CP model (Novara and Henning, 2017)	CP-TOC approach
C1a	0.050	0.12	1.8	1.2	0.0
C1b	0.075	0.18	6.4	3.8	0.0
C1c	0.100	0.24	14.5	0.1	0.0
C1d	0.125	0.30	24.7	14.4	0.3
C1e	0.150	0.36	39.2	23.4	1.3
C1f	0.175	0.42	58.7	33.9	5.5
C1g	0.200	0.48	83.2	45.1	8.9
C1h	0.225	0.54	105.2	56.8	12.2
C1i	0.250	0.60	148.0	74.1	23.2
C1j	0.275	0.66	191.4	94.3	39.0
C2	0.200	0.50	739.1	693.2	0.0
C3	0.070	0.20	1307.4	112.8	223.2

As seen in Table 1, all the solutions obtained by means of the CP-TOC model exhibit a better behavior than the ones attained by the deterministic CP formulation. In addition, in 11 out of 12 examples the solutions reached by means of the current approach are of better quality (four of the resulting schedules are optimal) than the ones reported in the previous stochastic proposal (Novara and Henning, 2017) (see Fig. 1). The ten instances corresponding to Case 1 show that the new approach renders more resilient agendas in a wide range of processing time variabilities (see Fig. 2). The solution of Case 3 is the only one exhibiting worse quality. Nevertheless, it is a good solution that shows an improvement with respect the deterministic one, in a similar fashion to the all the other

examples. It is worth mentioning that the Case 3 solution obtained with the previous stochastic approach had an unusual high performance.

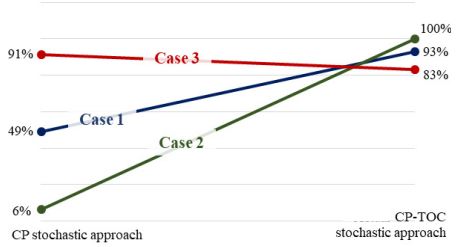


Fig. 1. Improvement percentage of the two stochastic approaches solutions compared with the deterministic one.

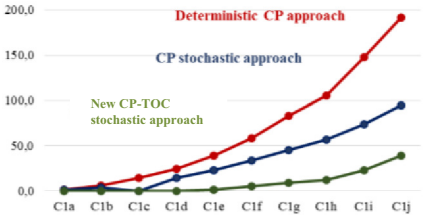


Fig. 2. Total tardiness for increasing processing times variability scenarios (10 instances of C1)

5. Conclusions and future work

This paper presents a novel CP-TOC stochastic scheduling methodology that addresses processing time uncertainty in an a priori fashion. When contrasted with another CP stochastic proposal (Novara and Henning, 2017), 11 out of 12 examples rendered better solutions (several of which are optimal) with less computational effort and without providing good quality initial solutions. It is worth mentioning that the lower computational expense is achieved due to a reduction in the model size. The CP-TOC model has $Card(B) \cdot (Card(U) - Card(U_{CSS}))$ less interval variables and $(Card(U) - Card(U_{CSS})$ less sequence variables than the CP stochastic model proposed by Novara and Henning (2017). In addition, the attained schedules exhibit more resiliency and are better prepared to soften uncertainty consequences. This conclusion was accomplished after simulating the agendas obtained with this proposal and the ones reached with a deterministic CP model and a stochastic one under various uncertainty conditions. The proposed CP-TOC model has the same inherent limitations that were discussed by Novara and Henning (2017). Future work will extend the approach to address other intermediate storage/inter-stage waiting policies.

References

J. Balasubramanian, and I. Grossmann, 2003, Scheduling optimization under uncertainty—an alternative approach, Comput. Chem. Eng., 27, 469–490.

A. Bonfill, A. Espuña, and L. Puigjaner, 2005, Addressing Robustness in Scheduling Batch Processes with Uncertain Operation Times, 44, 1524-1534.

A. Bonfill, A. Espuña, and L. Puigjaner, 2008, Proactive approach to address the uncertainty in short-term scheduling, Comput. Chem. Eng., 32, 1689–1706.

IBM ILOG, 2013. CPLEX - IBM ILOG - 12.5.1 http://www.ibm.com/support/knowledgecenter/SSSA5P_12.5.1/ilog.odms.studio.help/Optimization_Studio/topics/COS_home.html?lang=es.

Z. Li, and M.G. Ierapetritou, 2008, Robust optimization for process scheduling under uncertainty, Ind. Eng. Chem. Res., 47, 4148–4157.

F.M. Novara, and G.P. Henning, 2017, Scheduling of multiproduct multistage batch plants with uncertain processing times: an innovative constraint programming approach, Proceedings of Foundations of Computer Aided Process Operations/Chemical Process Control, Tucson, USA.

M.S. Spencer, and J.F. Fox, 1995, Optimum production technology (OPT) and the theory of constraints (TOC): analysis and genealogy, Int. J. Prod. Res., 33, 6, 1495–1504.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50206-8

Continuous-time and Precedence-based Modelling for Aluminium Electrolysis and Caster Scheduling Process

Qingxin Guo[a], Gongshu Wang[a,*], Xin Gao[b]

[a]*Institute of Industrial and Systems Engineering, Northeastern University, Shenyang, 110819, China*
[b]*College of Information Science and Engineering, Northeastern University, Shenyang, 110819, China*
wanggongshu@ise.neu.edu.cn

Abstract

Electrolysis is a high energy consumption production process in aluminium industry. This paper investigates the electrolysis cells scheduling problem considering the crucible and caster. A precedence-based model and a novel unit-specific event based continuous-time mixed integer programming model are proposed to describe the problem. The two models are solved by CPLEX. Computational experiments show that the proposed continuous-time formulation is efficient when comparing with a precedence-based model for medium and large size instances.

Keywords: aluminium electrolysis, scheduling, precedence-based, unit-specific event based, continuous-time.

1. Introduction

Aluminium is the most widely used nonferrous metal, the production process of aluminium includes bauxite mining, alumina production, electrolysis aluminium, castings, rollings, the production of consumer products and recycling. In the electrolysis aluminium production line, more than 99% pure molten aluminium is formed at the cathode deposited at the bottom of the electrolytic cell, and is tapped from the cell into a crucible by a vacuum siphoning system. The molten metal is transported and poured into a holding furnace and caster to be cast into ingots, extrusion or rolling ingots.

In a typical aluminium smelter, there may be hundreds of electrolysis cells. Each crucible taps the metal from up to three cells that contain different purity of the molten aluminium. The capacity of the holding furnace is more than three times the crucible, so there can be several different crucibles feeding one holding furnace. The molten aluminium which is mixed and stabilized synchronously in one holding furnace is called a charge. Therefore, both crucible and holding furnace are batch production mode. Therefore, in the aluminium electrolysis process, the production schedule is to form the batch tapping of the cells into crucibles and arrange batch feeding of the crucibles into holding furnaces, considering the constraints of the electrolysis and cast.

Ryan (1998) formulated the tapping of the cells problem as a set-partitioning model and solved it by the LP relaxation. Piehl (2000) used the revised simplex method and branch and bound with a constraint branching approach to solve the similar cell batching problem. Prasad et al. (2006) provided a MILP model for the scheduling of aluminium

casts of different alloys with respect to the actual number of batches to be processed in multistage. For the formulation of the scheduling problem, we refer to Floudas and Lin (2004) and Harjunkoski et al. (2014).

The remainder of this paper is organized as follows. In Section 2, a precedence-based model is presented for this problem. In Section 3, a novel unit-specific event based continuous-time mixed integer programming model are proposed to describe the problem. Section 4 reports the experimental results of the two models solved by CPLEX. Finally, Section 5 presents the conclusions.

2. Precedence-based model

To formulate the proposed problem, the following notations are given. The indices of electrolysis cells are i and j, m is index of crucible, r is index of the batches of a crucible, and k, h are indices of holding furnaces. N, C, Cl, and B are sets of electrolysis cells, crucibles, batches of crucible, and batches of holding furnace, respectively. $Cap1$ is the capacity of crucible, $Cap2$ is the capacity of holding furnace. p_{im} is the processing time of cell i on crucible m. t_{ik} is the processing time of cell i on furnace k. a_{ij} is the setup time of cells i and j in a furnace. And M is a large number.

For 0-1 decision variables, x_{imr} is whether cell i is assigned to batch r of crucible m, y_{ik} is whether cell i is assigned to furnace k, z_{kh} is whether k is followed by h directly. For continuous decision variables, S_{mr} is starting time of batch r of crucible m, C_{mr} is completion time of batch r of crucible m, R_k is release time of furnace k, Sl_k is starting time of furnace k, Cl_k is completion time of furnace k, pl_k is processing time of furnace k, d_{kh} is the setup time of furnaces k and h. Using the above symbols, the precedence-based model is constructed as follows:

Minimize $C_{\max}$

s.t.
$$\sum_{m\in C}\sum_{r\in Cl} x_{imr} = 1 \qquad \forall i \in N \tag{1}$$
$$\sum_{i\in N} x_{imr} = Cap1 \qquad \forall m \in C, r \in Cl \tag{2}$$
$$\sum_{k\in B} y_{ik} = 1 \qquad \forall i \in N \tag{3}$$
$$\sum_{i\in N} y_{ik} = Cap2 \qquad \forall k \in B \tag{4}$$
$$S_{m1} = 0 \qquad \forall m \in C \tag{5}$$
$$S_{mr} \geq C_{m,r-1} \qquad \forall m \in C, r \in Cl \tag{6}$$
$$C_{mr} = S_{mr} + \sum_{i\in N} x_{imr} \cdot p_{im} \qquad \forall m \in C, r \in Cl \tag{7}$$
$$R_k + (1 - y_{ik}) \cdot M \geq C_{mr} \cdot x_{imr} \qquad \forall i \in N, m \in C, r \in Cl, k \in B, h \in B \tag{8}$$
$$Sl_k \geq R_k \qquad \forall k \in B \tag{9}$$
$$p_k = \sum_{i\in N} t_{ik} y_{ik} \qquad \forall k \in B \tag{10}$$
$$Cl_k = Sl_k + p_k \quad \forall k \in B \tag{11}$$
$$z_{kk} = 0 \quad \forall k \in B \tag{12}$$
$$z_{kh} + z_{hk} = 1 \qquad \forall k \in B, h \in B, k < h \tag{13}$$
$$Sl_h + (1 - z_{kh}) \cdot M \geq Cl_k + d_{kh} \qquad \forall k \in B, h \in B, k \neq h \tag{14}$$
$$Sl_k + z_{kh} \cdot M \geq Cl_h + d_{hk} \qquad \forall k \in B, h \in B, k \neq h \tag{15}$$
$$d_{kh} \geq y_{ik} . y_{jh} . a_{ij} \qquad \forall i \in N, j \in N, k \in B, h \in B \tag{16}$$
$$C_{\max} \geq Cl_k \qquad \forall k \in B \tag{17}$$

$$y_{ik} \geq y_{jk} - M(2 - x_{imr} - x_{jmr}) \qquad \forall i \in N, m \in C, r \in Cl, k \in B, h \in B \tag{18}$$

$$y_{ik} \leq y_{jk} + M(2 - x_{imr} - x_{jmr}) \qquad \forall i \in N, m \in C, r \in Cl, k \in B, h \in B \tag{19}$$

$$x_{imr}, y_{ik}, z_{kh} \in \{0,1\} \quad i \in N, m \in C, r \in Cl, k \in B, h \in B \tag{20}$$

Constraints (1) ensure that each cell must be assigned to one batch of a crucible. Constraints (2) ensure that the total amount of aluminium assigned to every crucible is equal to the capacity *Cap*1. Constraints (3) ensure that each cell must be assigned to a furnace. Constraints (4) ensure that the total amount of aluminium assigned to every furnace is equal to the capacity *Cap*2. Constraints (5) ensure the starting time of the first batch in a crucible is zero. Constraints (6) ensure the starting time of a batch in a crucible is larger than or equal to the completion time of the previous batch in the same crucible. Constraints (7) represent the completion time of a batch in a crucible is equal to the starting time plus the processing time. Constraints (8) represent the release time of a batch in a furnace is larger than or equal to the completion time of each cell that assigned to this batch. Constraints (9) ensure the starting time of a batch in a furnace is larger than or equal to its release time. Constraints (10) represent the processing time of a batch in a furnace is equal to the processing time of every cell that assigned to it. Constraints (11) represent the completion time of a batch in a furnace is equal to the starting time plus the processing time. Constraints (12) and (13) represent the relationship between two directly adjacent batches in a furnace. Constraints (14) and (15) ensure the starting time of a batch in a furnace is larger than or equal to the completion time of the previous batch plus their setup time. Constraints (16) represent the setup time of two batches in a furnace is the largest setup time among the cells that assigned to them. Constraints (17) ensure the makespan is larger than or equal to each batch of the furnaces. Constraints (18) and (19) ensure two cells are assigned to the same batch of a furnace if they were assigned to the same batch of a crucible. Constraints (20) specify the integrity and continuous conditions on the decision variables.

3. Unit-specific event based model

The notations of unit-specific event based continuous time model are given as follows. *i* is the index of cells, *m* is the index of crucibles, *k* is the index of furnaces. *n* is the index of event point for crucibles, *n'* is the index of event point for furnaces. Correspondingly, *I* is the set of cells, *C* is the set of crucibles, *B* is the set of furnaces, *N* is the set of event point for crucibles, *N'* is the set of event point for furnaces.

For the parameter, $p(i,m)$ is the processing time when assigning *i* to *m*, $t(i,k)$ is the processing time when assigning *i* to *k*. $tTotal(k,n')$ is the processing time of a batch in a furnace, $a(i,i')$ is the setup time between two batches. *Cap*1 is the capacity of crucible, *Cap*2 is the capacity of holding furnace.

0-1 decision variables include: $w(i, m, n)$ is whether assign cell *i* to the event point *n* of crucible *m*. $w'(i, k, n')$ is whether assign cell *i* to the event point *n'* of furnace *k*. $v(m, n)$ is whether the event point *n* of crucible *m* is empty. $v'(k, n')$ is whether the event point *n'* of furnace *k* is empty. And the continuous variables include: $S(i, m, n)$ is the starting time of cell *i* in event point *n* of crucible *m*. $C(i, m, n)$ is the completion time of cell *i* in event point *n* of crucible *m*. $Sl(i, k, n')$ is the starting time of cell *i* in event point *n'* of furnace *k*. $Cl(i, k, n')$ is the completion time of cell *i* in event point *n'* of furnace *k*.

$d(n',n'+1)$ is the setup time between two batches of a furnace. Based on the above notation, the continuous time model is given as follows.

Minimize C_{max}

s.t.

$$\sum_{m\in C}\sum_{n\in N} w(i,m,n)=1, \qquad \forall i\in I \tag{21}$$

$$\sum_{i\in I} w(i,m,n)=Cap1, \qquad \forall m\in C, n\in N \tag{22}$$

$$v(m,n-1)\le v(m,n), \qquad \forall m,n>1 \tag{23}$$

$$C(i,m,n)=S(i,m,n)+\sum_{i\in I} w(i,m,n)\cdot p(i,m), \qquad \forall i\in I, m\in C, n\in N \tag{24}$$

$$S(i',m,n)\ge C(i,m,n-1)-M\cdot(2-w(i',m,n)-w(i,m,n-1)), \quad \forall i,i'\in I, i\ne i', m\in C, n\in N, n>1 \tag{25}$$

$$\sum_{i\in I} w(i,m,n)\le Cap1\cdot(1-v(m,n)), \qquad \forall m\in C, n\in N \tag{26}$$

$$\sum_{i\in I} w(i,m,n)+v(m,n)\ge 1, \qquad \forall m\in C, n\in N \tag{27}$$

$$S(i,m,n)=0, \qquad \forall i\in I, m\in C, n=1 \tag{28}$$

$$S(i,m,n)\ge S(i',m,n)-M(2-w(i,m,n)-w(i',m,n)), \quad \forall i,i'\in I, i\ne i', m\in C, n\in N \tag{29}$$

$$S(i,m,n)\le S(i',m,n)+M(2-w(i,m,n)-w(i',m,n)), \quad \forall i,i'\in I, i\ne i', m\in C, n\in N \tag{30}$$

$$C(i,m,n)\ge C(i',m,n)-M(2-w(i,m,n)-w(i',m,n)), \quad \forall i,i'\in I, i\ne i', m\in C, n\in N \tag{31}$$

$$C(i,m,n)\le C(i',m,n)+M(2-w(i,m,n)-w(i',m,n)), \quad \forall i,i'\in I, i\ne i', m\in C, n\in N \tag{32}$$

$$\sum_{k\in B}\sum_{n'\in N'} w'(i,k,n')=1, \qquad \forall i\in I \tag{33}$$

$$\sum_{i\in I} w'(i,k,n')\le Cap2\cdot(1-v'(k,n')), \qquad \forall k\in B, n'\in N' \tag{34}$$

$$\sum_{i\in I} w'(i,k,n')+v'(k,n')\ge 1, \qquad \forall k\in B, n'\in N' \tag{35}$$

$$v'(k,n'-1)\le v'(k,n'), \qquad \forall k,n'>1 \tag{36}$$

$$\sum_{i\in I} w'(i,k,n')=Cap2, \qquad \forall k\in B, n'\in N' \tag{37}$$

$$tTotal(\mathrm{k},n')=\sum_{i\in I}(w'(i,\mathrm{k},n')\cdot t(i,k)), \qquad \forall k,n' \tag{38}$$

$$Cl(i,k,n')\ge Sl(i,k,n')+tTotal(k,n'), \qquad \forall i,k,n' \tag{39}$$

$$Sl(i,k,n')\ge Sl(i',k,n')-M(2-w'(i,k,n')-w'(i',k,n')), \quad \forall i,i'\in I, i\ne i', k\in B, n'\in N' \tag{40}$$

$$Sl(i,k,n')\le Sl(i',k,n')+M(2-w'(i,k,n')-w'(i',k,n')), \quad \forall i,i'\in I, i\ne i', k\in B, n'\in N' \tag{41}$$

$$Cl(i,k,n')\ge Cl(i',k,n')-M(2-w'(i,k,n')-w'(i',k,n')), \quad \forall i,i'\in I, i\ne i', k\in B, n'\in N' \tag{42}$$

$$Cl(i,k,n')\le Cl(i',k,n')+M(2-w'(i,k,n')-w'(i',k,n')), \quad \forall i,i'\in I, i\ne i', k\in B, n'\in N' \tag{43}$$

$$Sl(i',k,n') \geq Cl(i,k,n'-1) + d(n',n'+1) - M \cdot (2 - w'(i',k,n') - w'(i,k,n'-1)), \quad \forall i,i' \in I, i \neq i', k \in B, n' \in N', n' > 1 \tag{44}$$

$$d(n',n'+1) \geq w'(i,k,n')w'(i',k,n'+1)a(i,i'), \quad \forall i,i' \in I, i \neq i', k \in B, n' \in N', n' > 1 \tag{45}$$

$$Sl(i,k,n') \geq C(i,m,n), \quad \forall i \in I, m \in C, k \in B, n \in N, n' \in N' \tag{46}$$

$$w'(i,k,n') \geq w'(i',k,n') - M(2 - w(i,m,n) - w(i',m,n)), \quad \forall i,i' \in I, i \neq i', k \in B, m \in C, n \in N, n' \in N' \tag{47}$$

$$C_{\max} \geq Cl(i,k,n'), \forall i \in I,, k \in B, n' \in N' \tag{48}$$

$$S(i,m,n), C(i,m,n), Sl(i,k,n'), Cl(i,k,n') \leq C_{\max}, \quad \forall i \in I, m \in C, n \in N, k \in B, n' \in N \tag{49}$$

Constraints (21)-(32) represent the assignment and scheduling of cells to crucibles. Constraints (33)-(45) represent the assignment and scheduling of cells to furnaces. Constraints (46) ensure the starting time of a batch in a furnace is larger than or equal to the completion time of a batch in a crucible. Constraints (47) ensure two cells are assigned to the same batch of a furnace if they were assigned to the same batch of a crucible. Constraints (48) and (49) represent the relationships of makespan and the starting/completion times.

4. Computational experiments

To evaluate the proposed models, randomly instances are generated and solved by CPLEX 12.5. The scale of problem mainly depends on the number of electrolysis cells, crucibles and furnaces. The process parameters are randomly generated according to the scope of the practical production data in an aluminium fabrication company of China. It is assumed that the number of the electrolysis cells in the crucible batch and the furnace batch is fixed. All computational experiments are performed on a PC with 64 bits of Intel Core i5 2.0GHz CPU. The maximum computation time of CPLEX is set to 5000 seconds. Table 1 gives 12 groups of instances. Table 2 are results that compare the computational time and the best value between precedence-based model and continuous time model.

Table 1 The scale of computation instances

Instances	Number of cells	Number of crucibles	*Cap*1	Number of batches of a crucible	Number of furnaces	*Cap*2
1	12	2	3	2	2	6
2	18	2	3	3	2	9
3	18	2	3	3	3	6
4	24	2	3	4	4	6
5	24	4	3	2	2	12
6	27	3	3	3	3	9
7	36	3	3	4	6	6
8	36	3	3	4	4	9
9	36	4	3	3	3	12
10	45	3	3	5	5	9
11	48	4	3	4	8	6
12	48	4	3	4	4	12

Based on the results shown in these tables, the following conclusions could be made. For small and medium-scale instances, the average computational time of continuous time model is smaller than the precedence-based model. However, precedence-based

model could obtain better LB for more instances and could obtain better solutions for some instances. When the size of the instances increases, for an instance of the same scale, the gap of continuous time model is smaller than precedence-based model, which indicates that continuous time model has a tighter lower bound than precedence-based model.

Table 2 The statistics of experimental result

Ins.	Precedence-based model				Continuous time model			
	opt	Time(s)	LB	Gap (%)	opt	Time(s)	LB	Gap (%)
1	98.6	20.75	97.7	0.93	97.3	0.31	97.2	0.09
2	186.5	24.91	182.7	2.01	185.7	0.57	185.4	0.15
3	157.3	206.06	150.8	4.12	158.4	10.56	158.3	0.04
4	110.6	657.37	106.9	2.78	109.9	57.36	109.7	0.21
4	257.9	1395.65	247.7	3.96	257.1	88.39	256.7	0.14
5	194.8	2458.87	170.2	12.6	195.4	239.98	194.7	0.35
6	453.9	3576.28	426.4	6.05	454.1	592.76	453.5	0.14
7	317.2	5000.00	301.4	4.98	318.5	1000.78	314.3	1.31
8	397.8	5000.00	371.2	6.68	395.4	500.81	392.8	0.65
9	234.3	5000.00	211.7	9.65	235.7	3765.59	233.4	0.98
10	309.6	5000.00	297.3	3.97	307.3	1075.65	307.0	0.09
11	672.0	5000.00	613.9	8.65	671.8	5000.00	666.5	0.79
12	568.3	5000.00	501.8	11.7	564.5	5000.00	555.8	1.54

5. Conclusions

This paper studies the batch scheduling problem of aluminium electrolysis and caster process. Precedence-based and unit-specific event based models are proposed to formulate the problem. Computational experiments are carried out using CPLEX solver. The experimental results show that both models can get the optimal solutions within a reasonable time for small and medium scale instances, and the proposed continuous time model could be used to obtain the optimal solutions for larger scale instances.

Acknowledgements

This research is supported by National Key R&D Program of China (2017YFB0304100) and National Natural Science Foundation of China (71602025, 71672032).

References

C. A. Floudas, X. Lin, 2004. Continuous-time versus discrete-time approaches for scheduling of chemical processes: A review, Computers and Chemical Engineering, 28: 2109-2129.

I. Harjunkoski, C. T. Maravelias, P. Bongers, P. M. Castro, S. Engell, I. E. Grossmann, J. Hooker, C. Méndez, G. Sand, and J. Wassick, 2014, Scope for industrial applications of production scheduling models and solution methods, Computers and Chemical Engineering, 62: 161-193.

T. S. Piehl, 2000, Cell batching optimisation for the New Zealand Aluminium Smelter, the 35rd Annual Conference of the Operational Research Society of New Zealand (ORSNZ).

P. Prasad, C. T. Maravelias, and J. Kelly, 2006, Optimization of aluminum smelter casthouse operations, Industrial & Engineering Chemistry Research, 45, 7603-7617.

D. M. Ryan, 1998, Optimized Cell Batching for New Zealand Aluminium Smelters Ltd, the 33rd Annual Conference of the Operational Research Society of New Zealand (ORSNZ).

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50207-X

Data-Driven Models and Algorithms for Demand Response Scheduling of Air Separation Units

Calvin Tsay[a], Michael Baldea[a*], Jun Shi[b], Ankur Kumar[b], Jesus Flores-Cerrillo[b]

[a]*McKetta Department of Chemical Engineering, The University of Texas at Austin, 200 E. Dean Keeton St., Austin, TX 78712, United States*
[b]*Smart Operations-Praxair Digital, Praxair, Inc., Tonawanda, NY 14150, United States*
mbaldea@che.utexas.edu

Abstract

Demand response (DR) scheduling of air separation units (ASUs) can be beneficial to both process and grid operations. It requires frequent changes in production schedules and the integration of dynamics and control in scheduling calculations. Model dimensionality and nonlinearity make this a challenging task, especially under practical execution-time constraints. This work focuses on DR operation of an industrial ASU system. Historical data are used to construct scheduling-relevant, Hammerstein-Wiener models of (MPC) closed-loop plant dynamics, which are then embedded in a DR scheduling calculation formulated as a nonlinear dynamic optimization problem. The approach ensures dynamic feasibility of the optimal schedule and exhibits excellent computational performance, demonstrated in an industrial case study.

Keywords: Integrated Scheduling and Control, Model Identification, Air Separation

1. Introduction

Air separation units (ASUs) provide important feedstocks to several major manufacturing industries. For example, argon is used in welding, oxygen is used in steel production, and nitrogen is used as an inert gas in food and metals processing. Cryogenic ASUs dominate bulk industrial gas production and are electricity intensive: large volumes of air are compressed to high pressures using compressors driven by electric motors. As a result of fast, significant electricity price fluctuations created by modern deregulation of electricity markets, process economics of power-intensive manufacturing processes can be improved by frequently adjusting process outputs in response to electricity price changes (Du et al., 2015). In the ASU case, this is achieved by ramping up production during low-price periods and storing excess products as cryogenic liquids, which can be vaporized to satisfy gas demand during high-price periods (Pattison and Baldea, 2014).

With utility prices and production targets changing at hourly or shorter intervals, ASU dynamics must be considered in scheduling calculations to ensure production rate target sequences are both dynamically feasible and economically optimal. Integrating knowledge about process dynamics in production scheduling calculations requires an approach that allows for fast solutions while considering a reasonably long time horizon (Baldea and Harjunkoski, 2014). Recent theoretical works have proposed using scale-bridging models (SBMs), or low-order representations of closed-loop process dynamics, to account for controlled process dynamics in scheduling calculations (Du et al., 2015; Touretzky et al., 2016). A novel framework was presented for selecting scheduling-relevant process variables and constructing SBMs from detailed, first-principles process

models, and a simulated ASU case study showed that the SBMs achieved a dynamically feasible optimal schedule with reasonable computational expense (Pattison et al., 2016).

In this work, we consider the DR operation of an existing industrial ASU. A framework is presented for creating SBMs using a historical industrial dataset to predict production rates, power consumption, and variables limiting the dynamic agility of the process. The controller setpoints and ambient temperature (a disturbance variable) act as SBM inputs. The framework accounts for the fact that the process operates under model predictive control (MPC). The SBMs can be easily updated if the controller or plant are modified. A test case considering DR operation over a four-day window shows that operating costs can be lowered compared with the constant-setpoint operation base case.

2. Process Description

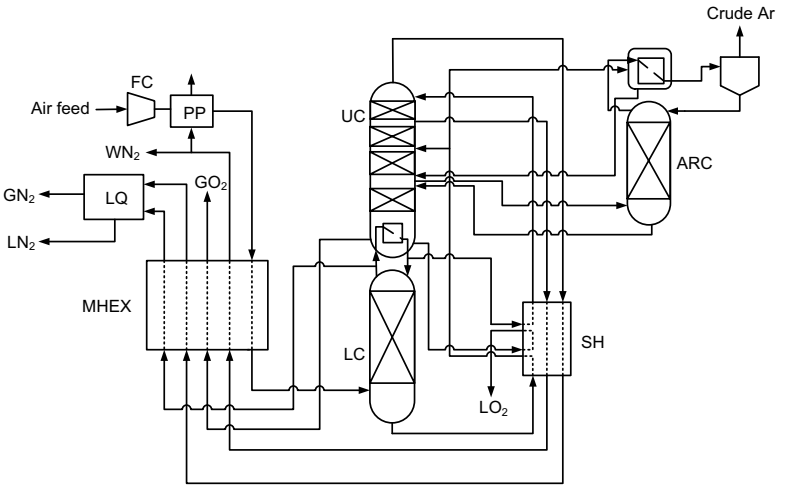


Figure 1: Industrial Air Separation Process Flowsheet.

The industrial cryogenic ASU process considered in this work is shown in Figure 1. Air is compressed in the feed compressor (FC) and is pre-processed to remove impurities (PP). The air stream is then cooled in the multistream heat exchanger (MHEX) by the cryogenic products and passed to the lower column (LC), where it is separated into a nearly pure nitrogen distillate (shelf gas) and an oxygen-rich bottoms stream. A portion of the shelf gas is used to boil the purified oxygen bottoms product of the upper column (UC) before being refluxed, and the rest of the shelf gas is passed through the MHEX and combined with the nitrogen product from the top of the UC. The bottoms stream of the LC passes through the nitrogen superheater (SH) and provides condensing utility to the argon column (ARC) before entering the UC.

The combined nitrogen stream is liquefied in the liquefier (LQ), resulting in both gaseous and liquid nitrogen products. A waste nitrogen stream is drawn at an intermediate stage of the UC to provide additional refrigeration in the MHEX and to purge pressure-swing adsorption beds in PP. Purified oxygen product is collected at the bottom of the UC: a portion of the gas vaporized by the shelf gas is passed to the MHEX, and the liquid portion is passed through the SH. Argon accumulates on an intermediate stage of the UC, where the ARC feed is drawn; argon is obtained as the ARC distillate.

3. Scheduling-Relevant Scale-Bridging Models

3.1. Model Identification Framework

Dynamic flowsheet optimization with detailed flowsheet models presents many mathematical challenges and can require extensive computational time and effort (Tsay et al., 2017). We therefore construct a set of low-order models that can accurately describe the closed-loop dynamics of scheduling-relevant variables and that can be used in scheduling calculations in lieu of the detailed dynamic model. Pattison et al. (2016) showed that the set of scheduling-relevant variables can be a small subset of the process variables, further reducing the scheduling problem size. In particular, only i) variables relevant to the scheduling objective/constraints and ii) variables near their bounds during historical static or transient operation should be modelled and included in scheduling optimization computations. The remaining variables are not considered relevant to optimal scheduling calculations as they do not hinder process agility.

The identification of such scale-bridging models (SBMs – Figure 2) can be carried out using historical operating data, system identification experiments, simulations using a detailed model, or a combination of these of sources. For the industrial ASU process in Figure 1, historical process data involving many production transitions are available and are used to construct an SBM for each output variable.

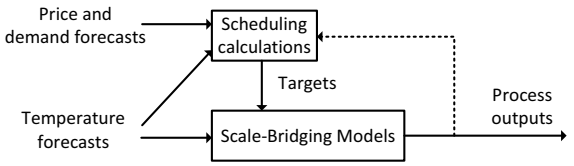


Figure 2: Scheduling Using Scale-Bridging Models (SBMs).

SBMs are identified in the Hammerstein-Wiener (HW) form (Pattison et al., 2016), with the linear dynamics represented as a state-space model:

$$h_i = H_i(u_i) \tag{1}$$

$$\dot{\vec{x}}_i = A\vec{x}_i + Bh_i$$

$$y_i = C\vec{x}_i$$

$$w = W(\sum y_i)$$

where H and W are nonlinear input and output transformations, A, B, and C are the matrices of the linear state-space system, u_i denotes the i^{th} model input (a setpoint of the MPC or a measured disturbance), and w is the system output.

3.2. Application to Industrial Air Separation Process

In deciding the input-output structure of the SBMs, the following factors were considered. On the input side, while the process is controlled by a model predictive controller (MPC) with six operator setpoints (targets), it was found that the production rates and scheduling-relevant state variables could be modelled accurately using only two operator setpoints, suggesting that the remaining four operator setpoints either have negligible effects on the modelled variables or are correlated with the two modelled setpoints. Additionally,

ambient temperature (a measured disturbance) has a large effect on process variables and was included as an input to the SBMs. Therefore each SBM has three inputs, [T SP1 SP2].

We focused on identifying models for operation during summer months, which tend to have larger fluctuations in electricity prices. 110 days of historical operating data, recorded at one minute intervals, were available for model identification after removing periods of start-up, shutdowns, and process or measurement faults. The first 88 days (80% of the data) were selected as training data, with the remaining 22 days used as test data. The nonlinear functions H and W in (1) are modelled as piecewise-linear functions, with the number of segments determined by minimizing the normalized Akaike information criterion (nAIC) for each SBM. The order of each linear state-space model is also determined using the nAIC, and all parameters are fitted using least-squares regression.

Although the industrial ASU produces five products, the liquid oxygen production rate was relatively constant throughout both the training and test datasets, indicating that a constant amount of liquid oxygen product is drawn from the UC (Figure 1) and only gas O2 production rate changes. Additionally, changing the operator setpoints has small effect on the argon production rate. Several variables are constrained in the MPC controller, but only two controlled process variables (CV1 and CV2) closely approach their bounds during operating point transitions, indicating they are the primary factors limiting the agility of the ASU. We therefore created SBMs for three production rates (gas O2, liquid N2, and gas N2), two state variables, and the total ASU power consumption. The SBM predictions for CV1 and CV2 for a week of test data are plotted in Figure 3, showing that the low NMSE values are largely due to high-frequency dynamics in the historical data. We note that all plotted variables have been scaled and filtered due to the confidential nature of industrial data. The number of piecewise-linear input segments, the order of the linear systems, and the number of piecewise-linear output segments are reported in Table 1 along with the normalized mean square error (NMSE) for each model.

Table 1: Identified Scale-Bridging Model Details

SBM output variable	Piecewise linear segments for inputs T, SP1, SP2	Linear system order for inputs T, SP1, SP2	Piecewise linear output segments	Training NMSE (%)	Test NMSE (%)
LN2 flow	[2,1,3]	[1,3,2]	5	91.56	69.15
GN2 flow	[4,3,3]	[3,2,1]	5	76.17	61.52
GO2 flow	[3,3,2]	[4,3,3]	3	66.39	65.55
CV1	[2,2,2]	[3,3,3]	4	67.53	61.86
CV2	[4,4,2]	[2,3,2]	4	42.22	42.65
Power	[4,3,3]	[3,3,1]	5	82.18	82.67

4. Optimal DR Operation with Time-Sensitive Electricity Pricing

The complete ASU scheduling problem using SBMs is formulated as a dynamic optimization problem. The objective is to minimize the total cost over a four-day horizon by manipulating the trajectories of SP1 and SP2, such that production constraints are met and scheduling-relevant CVs stay within their bounds. It is assumed that, i) the only production constraints are on total production (i.e. there is unlimited storage over four days) and ii) perfect forecasts of ambient temperature and electricity price are available for the four day period. While these assumptions may not hold true in practice, the approach provides valuable insights on the maximal benefits of DR scheduling of ASUs.

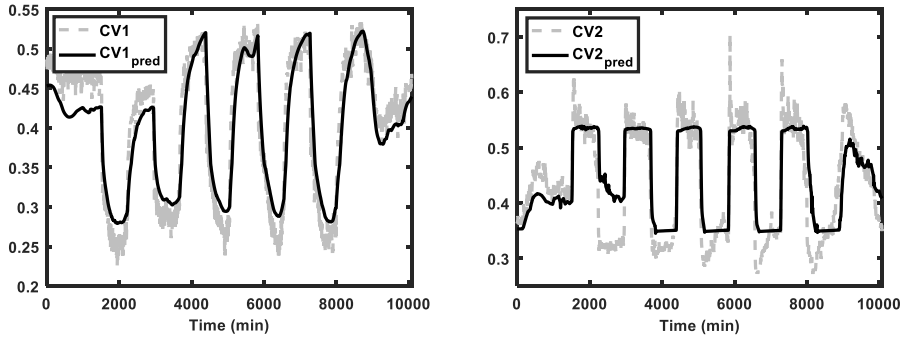


Figure 3: Controlled Variable SBM Predictions for One Week of Test Data.

A four-day period from the test data with constant operator setpoints is simulated with the SBMs as the base case, and the total amounts of each product produced are used as the lower bound for scheduling calculations. Historical ambient temperature profiles were obtained from the process historian, and historical electricity prices from the nearest location were used. The dynamic optimization is formulated with the SBM inputs (operator setpoints) parameterized as piecewise-constant functions over 2-hour intervals. The operator setpoints and production outputs are constrained to be within the range of operation found in the training dataset. The optimization problem was solved in two cases: Problem P1, where the CVs are constrained to be within their bounds as specified in the MPC settings. Here, the CVs did not reach their bounds in the optimal schedule. Because the SBMs for CVs cannot predict high-frequency dynamics, the optimization problem was then solved with the CVs constrained to be within their minimum and maximum values found in the base case simulation (Problem P2). The constraints for P2 are more conservative, and the solution revealed that the lower bound for CV1 limits the dynamic feasibility of the process, as shown in Figure 4.

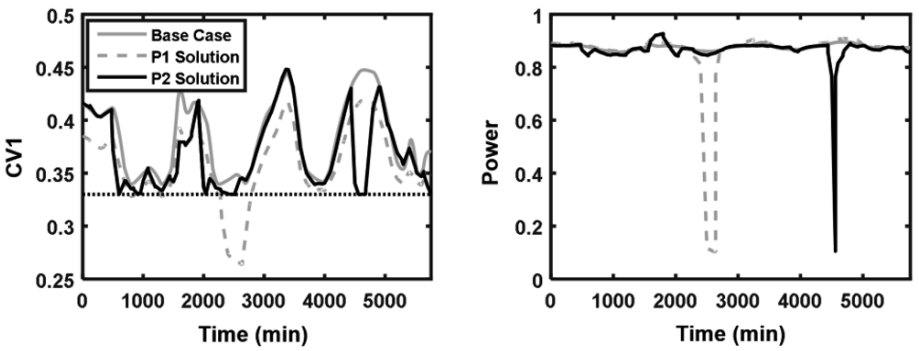


Figure 4: Selected Optimized Variable Profiles.

Both nonlinear optimization problems were solved to local optimality in gPROMS in less than one hour each, using reduced sequential quadratic programming (rSQP); it is expected that solution time can be further reduced using linearization and decomposition techniques (Kelley et al., 2018), noting that the employed objective function is linear. Tightening the bound for CV1 in P2 noticeably limits the dynamic agility of the process, and Figure 4 shows that the power consumption is decreased dramatically during peak electricity prices (near 2500 minutes) in the solution to P1, but not in that to P2.

Nevertheless, there are promising savings in total cost compared to the base case: a 16.0% reduction in operating cost (compared to the base case simulation) over the four-day time horizon for P1 and a 1.4% savings for P2. The significant savings found in the solution to P1 can be attributed to the ability to shut-down the plant during peak electricity price periods and overproduce during some periods with negative electricity price. Moreover, perfect forecasts for *both* price *and* ambient temperature (i.e., process disturbance) were included in the optimization problem. On the one hand, imperfect electricity price predictions and temperature forecasts, and storage constraints may reduce these values; on the other hand, DR benefits may be improved by designing the process for variable operation (Pattison and Baldea, 2014; Tsay et al., 2017). Note that, since all variables are scaled for confidentiality, a value of zero does not indicate zero power consumption, but rather a power consumption equal to the minimum value in the historical dataset.

5. Conclusions

In this work, we propose a framework for the identification of scheduling-relevant dynamic models for an ASU process based on industrial operating data. The models are in Hammerstein-Wiener form and implemented as scale-bridging models (SBMs) to integrate knowledge of process control and dynamics in scheduling calculations. We compute optimal schedules for the modelled industrial ASU process and find significant savings from DR operation, compared to the base case reflected in the historical dataset, where the operator setpoints remain relatively constant over long periods of time.

6. Acknowledgements

The authors gratefully acknowledge support from the National Science Foundation (NSF) through the CAREER Award 1454433 and Award CBET-1512379. The authors also gratefully acknowledge the support of Praxair, Inc.

References

M. Baldea and I. Harjunkoski. Integrated production scheduling and process control: A systematic review. *Comput. Chem. Eng*., 71:377-390, 2014.

J. Du, J. Park, I. Harjunkoski, and M. Baldea. A time scale-bridging approach for integrating production scheduling and process control. *Comput. Chem. Eng*., 79:56-69, 2015.

M.T. Kelley, R.C. Pattison, R. Baldick, and M. Baldea. An efficient MILP framework for integrating nonlinear process dynamics and control in optimal production scheduling calculations. *Comput. Chem. Eng.*, 110:35-52, 2018.

R.C. Pattison and M. Baldea. Optimal design of air separation plants with variable electricity pricing. *Comput.-Aided Chem. Eng.*, 34:393-398, 2014.

R.C. Pattison, C.R. Touretzky, T. Johansson, I. Harjunkoski, and M. Baldea. Optimal process operations in fast-changing electricity markets: framework for scheduling with low-order dynamic models and an air separation application. *Ing. Eng. Chem. Res.*, 55(16):4562-4584, 2016.

C.R. Touretzky, I. Harjunkoski, and M. Baldea. A framework for integrated scheduling and control using discrete-time dynamic process models. *Comput.-Aided Chem. Eng.*, 38:601-606, 2016.

C. Tsay, R.C. Pattison, and M. Baldea. A dynamic optimization approach to probabilistic process design under uncertainty. *Ind. Eng. Chem. Res.*, 56(30):8606-8621, 2017.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50208-1

Effective scheduling of complex process-shops using online parameter feedback in crude-oil refineries

Robert E. Franzoi,[a] Brenno C. Menezes,[a] Jeffrey D. Kelly,[b] Jorge W. Gut[a]

[a]*Department of Chemical Engineering, University of São Paulo, Av. Prof. Luciano Gualberto trav.3 n.380, São Paulo, SP 05508-010, Brazil*
[b]*Industrial Algorithms Ltd., 15 St. Andrew Road, Toronto, ON MIP-4C3, Canada*
robertfranzoijr@usp.br

Abstract

Integrated scheduling optimization comprising the battery limits of crude-oil refineries is a challenging problem to be solved as it includes decisions concerning the quantity and quality of the crude-oil feedstocks and final products (such as fuels and petrochemicals) as well as production, processing or make-side of the refinery flow network. So far, the literature on crude-oil scheduling optimization covered the problem from the crude-oil unloading and storage up to the distillation straight-run streams. To go further, this research extends the scope of the problem from the raw material deliveries up to product liftings through the refinery process-shop by using closed-loop, online and routine process feedback data from field and laboratory measurements for better process predictions, integrated within the scheduling cycle. For such engine, past routine operating data calibrates gains and biases as $y_{measured} = gain \times y_{model} + bias$, whereby the data updating in y considers both process yields and variables such as throughputs, flows, holdups and properties, whose effects propagate throughout the process network. Parameter feedback is applied, after data reconciliation computation, in a complete crude-oil refinery blend scheduling problem considering real tank topology, cascaded distillation towers, process-shops and blend-shops to effectively optimize the complex process system. The feedback strategy is solved within an iterative mixed-integer linear and nonlinear programming (MILP-NLP) decomposition by updating NLP results of process-shop's yields and properties, and recipes of blend-shops in the next MILP solution until convergence is achieved.
Keywords: Scheduling, parameter feedback, process-shops, blend-shops, crude-oil refinery.

1. Introduction

An effective optimization integrating the operations of scheduling in refining of crude-oils into fuels, lubes, asphalts and petrochemical feeds is a challenging problem to be solved. It involves decisions concerning logistics and storage of crude-oil feedstocks and final products as well as the processing and blending in the refinery network. Approaches to promote the integration of these multi-entity, multi-activity and multi-resource decision-making require efficient modeling and solving strategies that considers the difficulties posed by these logistics and quality aspects.

In terms of modeling, previous literature on crude-oil scheduling optimization covers the scheduling from the crude-oil unloading and storage up to the representation of the crude-oil atmospheric and vacuum distillation towers and their immediate straight-run streams (Kelly et al., 2017a). Predictions of product yields and properties from process-

shops, where the crude-oil blend *diet* feed is separated into straight-run distillation intermediate products for further downstream reactions, conversions, separations, treatments and blending throughout the remaining refinery unit-operations and tanks, remain a substantial limitation to better scheduling. Modeling such processes reliably and accurately is difficult due to the complexities and uncertainties in feed quality, operating conditions, process measurements, etc. and therefore parameter feedback is required.

In terms of solving, the search for optimized solutions in crude-oil scheduling may succeed with an MILP-NLP decomposition strategy instead of a full space MINLP approach, since in the latter a binary variable relaxation of the logistics problem is solved along with the non-convex blending and processing relationships in the NLP steps. Recently, Kelly et al. (2017a) solved a discrete-time benchmark for a 1-week time horizon with 2 h time-step involving the logistics details of ship arrivals, tanks and process units in an MILP as well as the quality calculations in an NLP for the blending of streams (e.g., feed *diet* of distillation towers) and for transformation of crude-oils into their main distillates. However, the benchmark excludes further physical and chemical transformations in other process units beyond the initial distillation towers. In this work a problem covering the whole scheduling process is modeled, from crude-oil unloading and storage up to fuel deliveries by including process- and blend-shops as well as the fine tune adjustments regarding on online measurement updates from the field.

With the evolution of solving capabilities determining crude-oil blend scheduling problems around 1 h as discrete time-step, there are reasons to evolve into a more complex scope and scale in the modeling frontier by including cascaded distillation towers, downstream process-shops and blend-shops (e.g., to specify products at their contracted deliveries). Furthermore, an online scheduling algorithm can better predict the model by including the valid plant state (measured) by using parameter feedback and operational updates from the field. According to Subramanian et al. (2012), disruptions or arrival of new information can make the incumbent schedule suboptimal or infeasible, motivating the need for online scheduling. Major considerations on strategies (Gupta and Maravelias, 2016; Gupta et al., 2016), design (Gupta and Maravelias, 2017a) and formulations (Gupta and Maravelias, 2017b) introduce the online scheduling modeling aspects, although other essential elements as re-scheduling activation and frequency as well as handling of process uncertainty are imperative.

For such online scheduling improvement, the use of parameter feedback is proposed considering the past routine operating data to calibrate *gains* and *biases* as: $y_{measured} = gain \times y_{model} + bias$. Data reconciliation and estimation steps are necessary for the next scheduling cycle. Two are the datasets for the parameter updating. First, by offsetting the lower, upper and target bounds of variables for output-uncertainty. Second, by creating new flow or quality variables for input-uncertainty overriding when these variables are propagating in the flowsheet.

2. Problem statement

The flowsheet in Figure 1 shows the operational scheduling optimization of a refinery with a real scale topology considering crude-oil supply, storage and feed tank operations, production in cascaded distillation towers, operations of process-shops and blend-shops of both crude-oils and products. The marine vessels or feedstock tanks

(CO1 to CO12) supply a crude-oil refinery with different quality of raw materials. The assignment of crude-oil feedstocks to storage tanks (S1 to S12) can be pre-defined by clustering similar quality raw materials (Kelly et al., 2017b). Storage tanks are connected to crude-oil blenders for preparation of distillation tower feeds or *diet* to be stocked temporarily in feed tanks (F1 to F4) before starting the crude-oil charging.

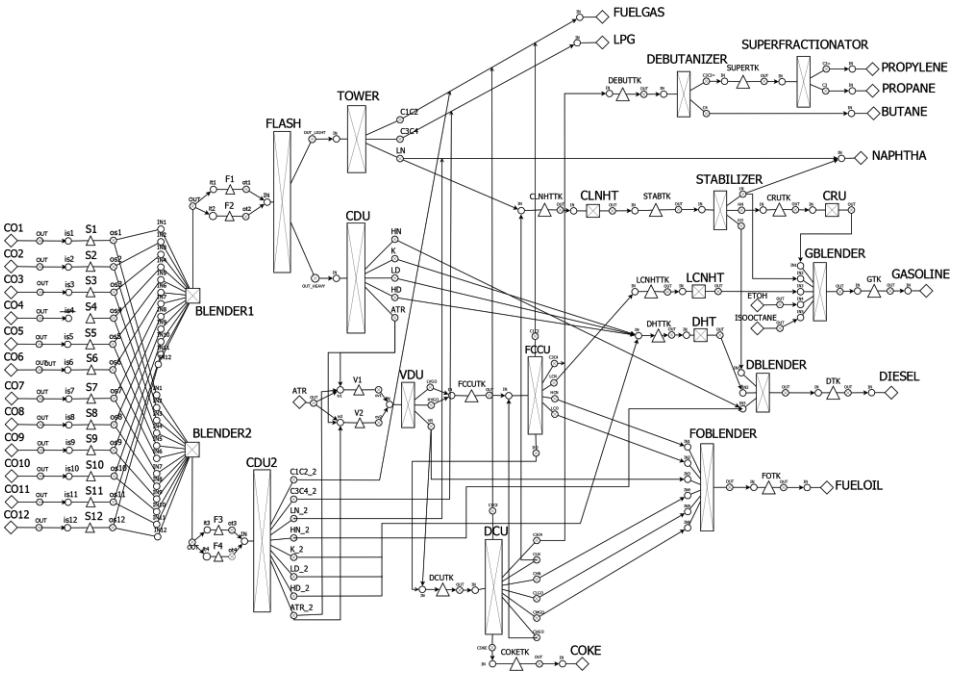


Figure 1. Crude-oil refining scheduling flowsheet.

The network in Figure 1 is constructed in the unit-operation-port-state superstructure (UOPSS) formulation (Kelly, 2005; Zyngier and Kelly, 2012). The UOPSS objects are detailed in the illustrative example in Figure 2 as: a) unit-operations m for sources and sinks (◇), tanks (△) and continuous-processes (⊠) and b) the connectivity involving arrows (→), inlet-ports i (○) and outlet-ports j (⊗). Unit-operations and arrows have binary and continuous variables (y and x, respectively) and the ports can hold the states as process yields or properties. The port-states j' and i'' represent upstream and downstream ports connected, respectively, to the in-port i and out-port j of a unit-operation m.

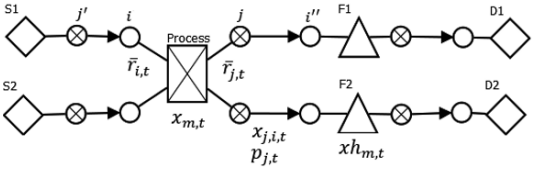


Figure 2. Illustrative example.

In the mixed-integer nonlinear (MINLP) problem (*P*), the objective function (1) maximizes the gross margin of fuels revenues by subtracting operational performance terms. Eq. (2) shows the types of performance constraints, considering deviations on

variables in consecutive time-periods (x_{t+1} and x_t) or around a target ($\bar{x}_t$). x_t^{LD} and x_t^{UD} are lower and upper deviation variables valid for any variable as flows of process-units $x_{m,t}$ and arrows $x_{j,i,t}$ as well as holdups $xh_{m,t}$ and properties $p_{j,t}$. The remaining logistics and quality calculations and the details of the MILP-NLP decomposition are presented by Kelly et al. (2017a). Unit-operations m for tanks, blenders, columns and final products belong, respectively, to the sets M_{TK}, M_{BL}, M_{CL} and M_{FP}. The deviation variables are represented in the set $DVAR$. For $x \in \mathbb{R}^+$, $p \in \mathbb{R}$ and $y = \{0,1\}$:

$$(P)\ \ Max\ Z = \sum_t \left(\sum_{m \in M_{FP}} price_{m,t}\, x_{m,t} - \sum_{x \in DVAR} weight (x_t^{LD} + x_t^{UD}) \right)\ \ s.t. \tag{1}$$

$$x_{t+1} - x_t + x_t^{LD} - x_t^{UD} = 0\ \ \forall\, t \quad \vee \quad x_t - \bar{x}_t + x_t^{LD} - x_t^{UD} = 0\ \ \forall\, t \tag{2}$$

$$\bar{x}_t^L\, y_t \le x_t \le \bar{x}_t^U\, y_t\ \ \forall\, t \tag{3}$$

$$\frac{1}{\bar{x}_{m,t}^U} \sum_{j'} x_{j',i,t} \le y_{m,t} \le \frac{1}{\bar{x}_{m,t}^L} \sum_{j'} x_{j',i,t} \ \wedge\ \frac{1}{\bar{r}_{i,t}^L} \sum_{j'} x_{j',i,t} \le x_{m,t} \le \frac{1}{\bar{r}_{i,t}^U} \sum_{j'} x_{j',i,t}\ \ \forall\, i,m,t \tag{4}$$

$$\frac{1}{\bar{x}_{m,t}^U} \sum_{i''} x_{j,i'',t} \le y_{m,t} \le \frac{1}{\bar{x}_{m,t}^L} \sum_{i''} x_{j,i'',t} \ \wedge\ \frac{1}{\bar{r}_{j,t}^L} \sum_{i''} x_{j,i'',t} \le x_{m,t} \le \frac{1}{\bar{r}_{j,t}^U} \sum_{i''} x_{j,i'',t}\ \ \forall\, j,m,t \tag{5}$$

$$xh_{m,t} = xh_{m,t-1} + \sum_{j'} x_{j',i,t} - \sum_{i''} x_{j,i'',t}\ \ \forall\, (i,m,j) \in M_{TK}, t \tag{6}$$

$$\sum_{j'} x_{j',i,t} = \sum_{i''} x_{j,i'',t}\ \ \forall\, (i,m,j) \in M_{CL} \wedge M_{BL}, t \tag{7}$$

$$\sum_{m \in U} y_{m,t} \le 1\ \ \forall\ t \tag{8}$$

$$H_{j,t}\, p_{j,t} = 0\ \ \ \forall\, j,t \tag{9}$$

The semi-continuous constraints to control the quantity-flows of the arrows $x_{j,i,t}$, the throughputs of the unit-operations $x_{m,t}$ (except tanks) and tank holdups or inventory levels $xh_{m,t}$ are bounded in Eq. (3). Eq. (4) imposes that the sum of the arrows arriving in the inlet-ports i (or uncontrolled mixers) of unit-operation m ($m \notin M_{TK}$) are bounded by their throughputs (flows) and connected inverse yields ($\bar{r}_{i,t}^L$ and $\bar{r}_{i,t}^U$), respectively, at the right and left side of Eq. (4). Similarly, in Eq. (5) the sum of the arrows leaving from the outlet-ports j (or uncontrolled splitters) of m ($m \notin M_{TK}$) must be between bounds of their throughputs and yields ($\bar{r}_{j,t}^L$ and $\bar{r}_{j,t}^U$). The unit-operations m ($m \notin M_{TK}$) can have more than one stream arriving in or leaving from their connected ports.

The quantity balance of inventory or holdup for unit-operations of tanks ($m \in M_{TK}$) in Eq. (6) considers the heels or initial inventories $xh_{m,t-1}$ and the material inlet and outlet of the tanks. Eq. (7) is a material balance in fractionation columns M_{CL} and blenders M_{BL} to assure that there is no accumulation of material in these types of units. In Eq. (8), for all physical units, at most one unit-operation m (as $y_{m,t}$ for procedures, modes or tasks) is permitted in U at a time t. Eq. (9) represents the NLP quality constraints to calculate the properties $p_{j,t}$ of the blends and certain process transformations as to convert crude-oil raw materials to the main distillates.

3. Online parameter feedback

The parameter updating of variables ($x_{m,t}$, $xh_{m,t}$, $x_{j,i,t}$, $p_{j,t}$) and hard bounds of yields ($\bar{r}^L_{i,t}$, $\bar{r}^U_{i,t}$, $\bar{r}^L_{j,t}$, $\bar{r}^U_{j,t}$) uses real, measured or valid values defined as $y_{measured} = gain \times y_{model} + bias$, considering *gain* and *bias* as the proportional and the linear deviations between the measured and model values. Two sets are defined for the data updating in the next scheduling cycle. For variables of unit-operation throughputs $x_{m,t}$, tank holdups $xh_{m,t}$ and parameters of inverse yields ($\bar{r}^L_{i,t}$, $\bar{r}^U_{i,t}$) and yields ($\bar{r}^L_{j,t}$ and $\bar{r}^U_{j,t}$), there is no need to create intermediate variables to update their ordinated values (or ordinates) in the model since the flowsheet structure is identical (same rows and columns in the modeling, although with different coefficients). For this type of updating, gains and biases are added directly in their soft bounds (target) and hard bounds of unit-operations variables ($x_{m,t}$ and $xh_{m,t}$) and their yields, procedure named as offsetting.

However, for flow of material and properties of unit-operation to unit-operation connections ($x_{j,i,t}$, $p_{j,t}$), there is a need to create intermediate variables to update their flowsheet propagation that overrides neighbor structures connected by arcs, paths or arrows (→). Therefore, the total number of variables increases by the additional flows and properties overriding the connected objects for the re-scheduling step with parameter feedback, representing new columns in the re-scheduled model. In such way, the effects of all measurement updates propagate throughout the network.

4. Examples

The examples run in an Intel Core i7 machine at 3.41 GHz (in 8 threads) with 64GB of RAM and uses the structural-based UOPSS framework found in the semantic-oriented modeling and solving platform IMPL (Industrial Modeling and Programming Language) from Industrial Algorithms Limited. By updating the overriding variables in the illustrative case (Figure 2), the number of continuous variables increases around 40%, hence for industrial cases this type of variable can be applied only in the initial time windows for less complex solving steps. For each scheduling cycle before the parameter feedback updating, a simultaneous data reconciliation and parameter estimation is performed considering the raw data collected in the past time window. The examples use the reconciled data of the measured variables as targets and the *gains* and *bias* are feedback in the next re-scheduling model to be executed continuously.

Figure 3 shows the measured and the optimized values with and without the parameter feedback (both using reconciled data as target) for the illustrative case in Figure 2. A time horizon of 8 hours considers fixed setups of unit-operations and arrow flows, resulting in an NLP problem. The plots for flow-in, flow-out, and holdup of the tank F2 consider 60 m^3/d and 90 m^3 as flow and holdup baseline, respectively.

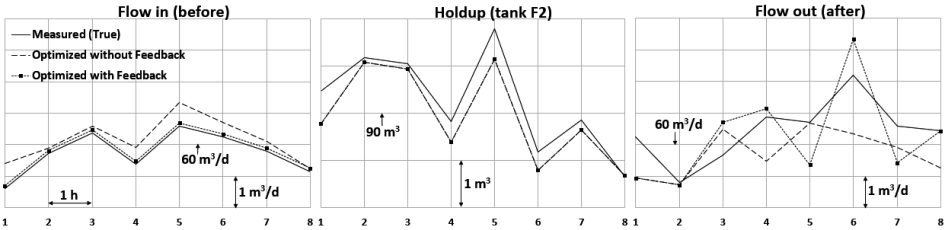


Figure 3. Measured and optimized solutions for tank F2 (see Figure 2).

For the optimized case with feedback (--■--), the flow after the tank F2 (on the right side) is the result of the effects of its flow-in (slightly) and holdup (significantly). The variations of tank flow-out delay one-time window to follow the new holdup value at each re-scheduling step and only the first time-window in the future of the optimized values are pointed in the Figure 3, therefore there are 8 cycles to schedule in this case.

In the industrial-sized example in Figure 1, the logistics MILP optimization for 5 days of time-horizon with 4 hours as time-step (30 time-periods) is solved in 86 seconds (with Gurobi 7.5.1) at 3.9% MILP gap. There are 12,568 constraints (2,545 equalities) for 5,777 continuous variables and 3,412 binary variables with 6,664 degrees-of-freedom. The quality NLP optimization is solved in 184 seconds (with IMPL's SLPQPE linked to Gurobi 7.5.1). There are 4,103 constraints (3,663 equalities) for 3,821 continuous variables with 278 degrees-of-freedom.

5. Conclusions

Efficient scheduling solutions in a near online fashion are becoming reality by the virtue of all advances in decision-making modeling, solving algorithms and computer-aided resources in terms of faster CPU clock speeds and memory. The major challenge now is to integrate proper or correct data (in timeliness and quality) to the decision automation core. An online scheduling engine with parameter updating might be useful to cope with uncertainties and to reduce the offsets or inaccuracies over the life-time of the problem as an effective way to close the gap among predictions and productions.

References

D. Gupta, C.T. Maravelias, 2016, On Deterministic Online Scheduling: Major Considerations, Paradoxes and Remedies. Computers & Chemical Engineering, v. 94, p. 312-330.

D. Gupta, C.T. Maravelias, J.M. Wassick, 2016, From Rescheduling to Online Scheduling. Chemical Engineering Research and Design, v. 116, p. 83-97.

D. Gupta, C.T. Maravelias, 2017a, On the Design of an Online Scheduling Algorithm. AICHE Meeting, Minneapolis, MN.

D. Gupta, C.T. Maravelias, 2017b, A General State-Space Formulation for Online Scheduling. Processes, v. 5, n. 4, p. 69.

J.D. Kelly, 2005. The Unit-Operation-Stock Superstructure (UOSS) and the Quantity-Logic-Quality Paradigm (QLQP) for Production Scheduling in The Process Industries. In Multidisciplinary International Scheduling Conference Proceedings: New York, United States, p. 327-333.

J.D. Kelly, B.C. Menezes, F. Engineer, I.E. Grossmann, 2017a, Crude-Oil Blend Scheduling Optimization of an Industrial-Sized Refinery: a Discrete-Time Benchmark. In Foundations of Computer Aided Process Operations, FOCAPO, Tucson, AR, United States, 10-13 January.

J.D. Kelly, B.C. Menezes, I.E. Grossmann, F. Engineer, 2017b. Feedstock storage assignment in process industry quality problems. In Foundations of Computer Aided Process Operations, FOCAPO, Tucson, AR, United States, 10-13 January.

J.D. Kelly, D. Zyngier, 2008, Continuously improve the performance of planning and scheduling models with parameter feedback. In Foundations of Computer Aided Process Operations, FOCAPO, Boston, MA, United States, 29-2 July.

K. Subramanian, C.T. Maravelias, J.B. Rawlings, 2012, A state-space model for chemical production scheduling. Computers & chemical engineering, v. 47, p. 97-110.

D. Zyngier, J.D. Kelly, (2012). UOPSS: A New Paradigm for Modeling Production Planning and Scheduling Systems. In European Symposium in Computer Aided Process Engineering: London, United Kingdom, p. 15146-56.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50209-3

Multistage Stochastic Models for Shale Gas Artificial Lift Infrastructure Planning

Zuo Zeng, Selen Cremaschi*

Department of Chemical Engineering, Auburn University, Auburn, AL 36849, USA
selen-cremaschi@auburn.edu

Abstract

Artificial lift methods (ALMs) are used to lift the accumulated fluids from wells and to help sustain their performances. This paper presents two multistage stochastic programming models (SMILP1 and SMILP2) to solve the artificial lift infrastructure planning problem by maximizing the expected net present value. The models incorporate endogenous uncertainty of production after a selected ALM is installed. The decisions are which ALMs to deploy, and their installation and removal times. Both models are applied to determine the artificial lift infrastructure plan (ALIP) for a hypothetical case study constructed based on Woodford Shale Horizontal Well for three different planning horizons. For the 16-month planning horizon case, the value of the stochastic solution is $163,831, which is a 5 % increase with respect to the deterministic solution. The solution times for SMILP2 are up to two orders of magnitude shorter when compared to those for SMILP1.

Keywords: Artificial lift methods, horizontal shale gas, stochastic programming, endogenous uncertainty

1. Introduction

Horizontal drilling has been around since 1950s. Recent advances in equipment and in hydraulic fracturing techniques allowed the production of natural gas from previously inaccessible sources such as shale formations (Robbins, 2013). This new and significant increase in shale gas production opened novel research areas for process systems engineering (PSE) community, and it, in turn, benefited from PSE contributions (e.g., Tavallali et al., 2014; Cafaro and Grossmann, 2016; Drouven and Grossmann, 2016; Gao and You, 2017). For example, Cafaro and Grossmann (2016) presents a continuous-time nonlinear programming model and a multi-period mixed-integer programming model to solve refracturing planning problem for horizontal shale gas wells. The models determine which wells to refracture, and the frequency of refracturing and its schedule.

Large amounts of liquids are injected into the shale formation during the fracturing process. The accumulation of liquids at the bottom of the well and gradual decline of well pressure, in general, halts the production prematurely. ALMs are used to lift the accumulated fluids from the well and to help sustain well performance. The ALIP is important for the long-term profitability of horizontal wells. A poor plan may reduce production and increase cost substantially. The ALIP depends on the production and reservoir conditions, the completion details, and field information, some of which change during production. The construction of ALIP is further complicated due to highly stochastic nature of well production: after an ALM is installed, the production rate is uncertain and depends on the chosen ALM. Hence, the production rate might violate the operating envelope of the installed ALM in the future.

This paper presents two multistage stochastic programming models (SMILP1 and SMILP2) to generate ALIPs for shale-gas producing horizontal wells. The solution determines: (a) which ALM(s) to install, and (b) their operating period. We incorporate the uncertain ALM-dependent production rates to the model, which yields multistage stochastic programs with Type II endogenous uncertainty (Apap and Grossmann, 2016).

2. Problem Statement

Nine candidate ALMs are considered: Plunger Lift, Foam Lift, Well Head Compression, Velocity Strings, Gas Lift, Electrical Submersible Pump (ESP), Sucker Rod Pump (SRP), Progressing Cavity Pump and Jet Pump. Givens are the characteristics of a candidate horizontal shale-gas well such as its geometry, operating conditions, and production history to date; and the economical parameters such as the prices of gas and natural gas liquids, and the costs of ALMs' equipment, installation, and operation. Each ALM has design and operational limitations obtained from Artificial Lift R&D Council guidelines, from previous expert experience (Wittfeld, 2005), and from typical attribute tables (e.g., Weatherford, 2013). The goal is to determine the optimal ALM(s) and their installation plan that yield the maximum expected net present value (ENPV) for the well.

2.1. Production decline curve of a horizontal well

The production rate is described using a hyperbolic curve (Fetkovich et al., 1996). The planning horizon is discretized into equal time periods $t = 1, 2, 3 \ldots T$, where period t starts at time t-1 and ends at time t. At first, the well is producing naturally, the solid curve in Figure 1. After liquid loading, the production becomes unstable, and is halted (at $t = p - 1$) unless an appropriate ALM is installed to stabilize it. Dashed line in Figure 1 represents a potential production rate after ALM i is installed at time p. We define the ratio of the production rate at time p (with ALM i installed) to the production rate at p-1 (the production right before ALM i installation) as the flowrate change ratio, Qrc_i, of ALM i.

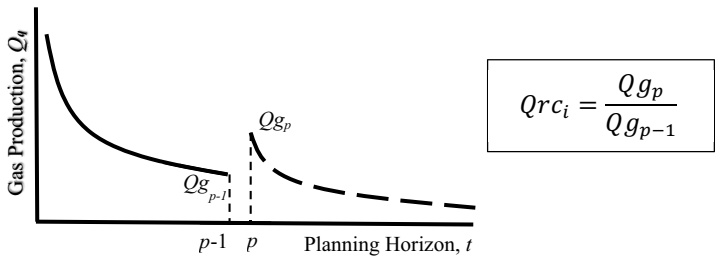


Figure 1. Production decline curve before and after ALM i is installed.

2.2. Flowrate uncertainty, scenario representation and future non-anticipativity

The flowrate change ratio, Qrc_i, is uncertain, and it is only revealed when ALM i is installed. Let ξ_i represent the random variable associated with the endogenous uncertain parameter Qrc_i. Assuming ξ_i has two outcomes, i.e., Ω_i = {High(H), Low(L)}, the scenario set, $s \in \boldsymbol{S}$, is generated as the Cartesian product of all possible outcome sets. With nine candidate ALMs, $|\boldsymbol{S}| = 2^9 = 512$.

Scenarios differ in the outcomes of Qrc_i. If ALM i is installed at period t, scenarios s and s' remain indistinguishable if they have the same outcome for Qrc_i. We define the subset $\boldsymbol{B}$ of $\boldsymbol{S} \times \boldsymbol{S}$ as scenarios s and s', which are still indistinguishable in the outcome of one

Qrc_i. The non-anticipativity constraints (NACs) should be enforced for scenario pairs $(s, s') \in \boldsymbol{B}$ that remain indistinguishable after ALM i is installed at period t.

3. Mathematical Programming Models

3.1. The first multistage stochastic mixed integer linear program (SMILP1)

The main decision variable is $w_{i,p,s}$, which is equal to one if ALM i is operational at the beginning of time period p under scenario s, and zero otherwise. We also introduce a binary variable $z_{i,t,s}$, which is equal to one if ALM i is uninstalled at period t in scenario s. Equations (1) and (3) state that each method can be installed and uninstalled at most once. Equations (2) and (4) limit the number of ALMs that can be installed and uninstalled at any given time period to a maximum of one. Equations (5) and (6) ensure that each ALM is only uninstalled once and the removal happens after its installation time. The last constraint, Eq. (7), prevents the overlap of two ALMs.

$$\sum_i w_{i,p,s} \leq 1 \qquad \forall p, s \tag{1}$$

$$\sum_p w_{i,p,s} \leq 1 \qquad \forall i, s \tag{2}$$

$$\sum_i z_{i,t,s} \leq 1 \qquad \forall t, s \tag{3}$$

$$\sum_t z_{i,t,s} \leq 1 \qquad \forall i, s \tag{4}$$

$$\sum_p w_{i,p,s} - \sum_t z_{i,t,s} = 0 \qquad \forall i, s \tag{5}$$

$$z_{i,t,s} w_{i,p,s} = 0 \quad \forall i, t, p, s, t \leq p \tag{6}$$

$$z_{i,t,s} y_{i,p,s} + z_{i,k,s} y_{j,l,s} \leq 1 \; \forall j \neq i, p+1 \leq l \leq t, t+1 \leq k \leq T, s \tag{7}$$

Once the well is unloaded, the production rates of gas ($Qg_{r,s}$), oil ($Qo_{r,s}$), and natural gas liquids ($Qn_{r,s}$) at period r in scenario s are defined by Eqns. (8) - (10), respectively. In Eqns. (8)-(10), D is the decline exponent and b is the nominal decline rate.

$$Qg_{r,s} = \sum_{i,t,p} z_{i,t,s} w_{i,p,s} \, Qg_{p-1,s} Qrc_{i,s} \big(1 + bD(t-r+1)\big)^{-\frac{1}{b}} \quad \forall r \geq 2, s \tag{8}$$

$$Qo_{r,s} = \sum_{i,t,p} z_{i,t,s} w_{i,p,s} \, Qo_{p-1,s} Qrc_{i,s} \big(1 + bD(t-r+1)\big)^{-\frac{1}{b}} \quad \forall r \geq 2, s \tag{9}$$

$$Qn_{r,s} = \sum_{i,t,p} z_{i,t,s} w_{i,p,s} \, Qn_{p-1,s} Qrc_{i,s} \big(1 + bD(t-r+1)\big)^{-\frac{1}{b}} \quad \forall r \geq 2, s \tag{10}$$

Design and operational envelopes of each ALM are incorporated as technical limitation constraints in the model (Eqns. (11) - (13)). Examples of these constraints can be found in Zeng and Cremaschi (2017). The most common constraints limit the ranges of liquid flow rates ($LFR_{r,s}$) that an ALM can operate within, and are given in Eqns. (11) and (12).

$$f_{i,p,r,s}\big(LFR_{r,s}, w_{i,p,s}\big) \leq 0 \qquad \forall i, p, r, s \tag{11}$$

$$f_{i,t,r,s}\big(LFR_{r,s}, z_{i,t,s}\big) \leq 0 \qquad \forall i, t, r, s \tag{12}$$

$$g_{i,p,s}\big(w_{i,p,s}\big) \leq 0 \qquad \forall i, p, s \tag{13}$$

The objective function maximizes the ENPV (Eq.(14)), which is calculated using minimum acceptable rate of return (*MARR*), gross income at period *r* in scenario *s* ($GI_{r,s}$), the taxable income at period *r* in scenario *s* ($TI_{r,s}$), the federal tax rate (*FT*), and the capital cost in scenario *s* (CC_s). The binary variable $x_{r,s}$ is equal to one if $TI_{r,s}$ is nonnegative.

$$ENPV = \sum_s P_s \left(\sum_r \left(GI_{r,s} - TI_{r,s} \cdot x_{r,s} \cdot FT \right) \left(\frac{1}{(1+MARR)^r} \right) - CC_s(1-FT) \right) \tag{14}$$

The first set of NACs (Eq. (15)) are for $p = 2$, the installation time of the first ALM. Our model assumes that the well is loaded at the first time period, and the first ALM is installed at the second time period. At this stage, all scenarios are indistinguishable because none of the ALMs are installed.

$$w_{i,2,s} = w_{i,2,1} \qquad \forall\, i, s \in \boldsymbol{S} \backslash \{1\} \tag{15}$$

The remaining NACs equate the decisions in scenarios *s* and *s′* for the later time periods, if $(s, s') \in \boldsymbol{B}$ are still indistinguishable after observing the outcome of only one ALM, denoted by $i^{s,s'}$. Because new ALMs cannot be installed before the existing ALM $i^{s,s'}$ is uninstalled, the remaining NACs are expressed in terms of variable $z_{i^{s,s'},t,s}$ as follows:

$$w_{j,l,s} - w_{j,l,s'} \leq 1 - z_{i^{s,s'},t,s} \qquad \forall l \geq t+1, (s,s') \in \boldsymbol{B} \tag{16}$$

$$w_{j,l,s} - w_{j,l,s'} \geq z_{i^{s,s'},t,s} - 1 \qquad \forall l \geq t+1, (s,s') \in \boldsymbol{B} \tag{17}$$

The model defined by Eqns. (1) – (17) yields a mixed-integer nonlinear program (MINLP). All nonlinear terms can be linearized using exact linearization (Oral and Kettani, 1990), and the detailed steps of this linearization can be found in Zeng and Cremaschi (2017). The resulting large-scale mixed integer linear program (MILP) is referred to as SMILP1.

3.2. The second multistage stochastic mixed integer linear program (SMILP2)

For this model, we introduce one more binary variable, $y_{i,t,p,s}$, which is equal to one if ALM *i* is installed on period *p* and uninstalled on period *t* in scenario *s*. The new binary variable satisfies the logical expression $y_{i,t,p,s} \Leftrightarrow w_{i,p,s} \wedge z_{i,t,s}$. Equations (18) - (20) translate this logical expression into constraints.

$$y_{i,t,p,s} \geq z_{i,t,s} + w_{i,p,s} - 1 \qquad \forall i, t, p, s \tag{18}$$

$$z_{i,t,s} \geq y_{i,t,p,s} \qquad \forall i, t, p, s \tag{19}$$

$$w_{i,p,s} \geq y_{i,t,p,s} \qquad \forall i, t, p, s \tag{20}$$

The planning constraints, Eqs. (21) – (25), replace Eqs. (1) - (7) of SMILP1.

$$\sum_{t,p} y_{i,t,p,s} \leq 1 \qquad \forall i, s \tag{21}$$

$$\sum_{i,t} y_{i,t,p,s} \leq 1 \qquad \forall p, s \tag{22}$$

$$\sum_{i,p} y_{i,t,p,s} \leq 1 \qquad \forall t, s \tag{23}$$

$$y_{i,t,p,s} = 0 \qquad \forall i, t, p, s \tag{24}$$

$$y_{i,t,p,s} + y_{j,k,l,s} \leq 1 \qquad \forall j \neq i, p+1 \leq l \leq t, t+1 \leq k \leq T \tag{25}$$

The multiplication of two binary variables ($z_{i,t,s} w_{i,p,s}$) in flowrate constraints (Eqns. (8) - (10)) are replaced by the new binary variable ($y_{i,t,p,s}$). Furthermore, instead of relating

$w_{i,p,s}$ and $z_{i,t,s}$ to the upper and lower bounds of liquid flow rate separately (Eqns. (11) and (12)), the constraint given in Eq. (26) is used to define these operational envelops.

$$f_{i,t,p,r}\left(LFR_{r,s}, y_{i,t,p,s}\right) \leq 0 \qquad \forall i,t,p,r,s \tag{26}$$

$$g_{i,t,p,s}\left(y_{i,t,p,s}\right) \leq 0 \qquad \forall i,t,p,s \tag{27}$$

The objective function and the NACs of SMILP1 are retained for this model, which, after linearization, is a large scale MILP, and is referred to as SMILP2.

4. Case Study

4.1. Field application: Woodford shale horizontal well

The well was hydraulically fractured in eight different stages. The production before the loading indicated a gas rate of 800 thousand standard cubic feet per day (Mscf/D), a liquid rate of 520 barrels per day (BPD) (of which 500 barrels were water and the remaining oil), a water cut of 96%, and a gas-to-liquid ratio of 1,950 standard cubic feet per stock tank barrel (scf/STB). The well is currently operated using gas lift with a well-site compressor (Valbuena, 2015). The case study assumes that the planning was carried out at the first loading prior to the installation of gas lift. For uncertain parameter Qrc_i, High outcome is assumed to be 20% above its nominal value, and Low outcome 20% below with equal probabilities. The planning horizon is 12 months, which is divided into 12 equal time periods. The mathematical programs are constructed in GAMS 24.6.1, and solved using CPLEX 12.6.3 on a Lenovo ThinkStation P900 with Intel 2.3 GHz CPU and 24 GB RAM. The solutions of SMILP1 and SMILP2 recommend installing SRP throughout the planning horizon for all scenarios. The ENPV is \$2,805,092, which is 8.8% higher than the ENPV if gas lift were used. For this problem, SMILP1 and SMILP2 are solved to 0.1% optimality gap in 7039 seconds and 27 seconds, respectively.

4.2. Comparison of the formulations for a hypothetical case study

We also constructed a hypothetical case study where the operating envelope of SRP is modified by setting the minimum allowable liquid flow rate to 300 BPD. The planning horizons are 12, 16 and 20 months. The maximum CPU time is limited to 20 hours. Table 1 summarizes the results.

Table 1. The ENPVs, optimality gaps, and solution times for SMILP1 and SMILP2

	SMILP1			SMILP2		
Months	ENPV (\$)	GAP (%)	Times (s)	ENPV (\$)	GAP (%)	Times (s)
12	2,783,894	0.08	4948	2,783,894	0.08	20
16	3,416,506	42.51	72000	3,416,507	0.1	1959
20	3,840,889	49.39	72000	3,916,996	11.89	72000

The SMILP1 cannot be solved to optimality for 16-month and 20-month planning horizons (Table 1), whereas SMILP2 cannot be solved to optimality only for 20-month planning horizon. We hypothesize that the additional binary variable $y_{i,t,p,s}$ helps CPLEX to branch better when solving SMILP2.

The solution of the deterministic model, using nominal values of Qrc_i, recommends installing and operating SRP throughout the planning horizon. If the deterministic solution were implemented, the ENPV would have been \$3,252,676. With the deterministic solution, under scenarios with Low outcome of $Qrc_{,SRP,}$ the production rate

violates the operational envelope of SRP in month 13, and the production is halted. The optimal plan obtained as the solution of SMILP2 installs SRP initially for all scenarios. At the 7th time period, the ESP is installed for scenarios with Low outcome of $Qrc_{,SRP}$ while SRP stays in operation for the rest of the scenarios. The difference between the ENPVs of the stochastic and deterministic solutions, i.e., the value of the stochastic solution, is $163,831, a 5% increase from the deterministic solution.

5. Conclusions and Future Directions

This paper contributes to the new and growing body of literature in PSE on shale-gas production and infrastructure planning from horizontal wells. We presented two stochastic models for generating ALIPs for shale gas producing horizontal wells. Both models incorporate the stochastic ALM-dependent well production, which is an endogenous uncertainty. The first model, SMILP1, has two decision variables, which track the installation and uninstallation times of each ALM separately. In the second model, SMILP2, a new binary variable that integrates both installation and uninstallation times of each ALM is introduced, and the production rate and planning constraints are expressed using it. The additional binary variable in SMILP2 reduced the solution times up to two orders of magnitude. We hypothesize that a similar reduction in solution times of traditional PSE planning problems may be expected with the addition of a similar binary variable, and our future work will focus on investigating this hypothesis.

References

R. M. Apap, I. E. Grossmann, 2016, Models and Computational Strategies for Multistage Stochastic Programming under Endogenous and Exogenous Uncertainties. Computers & Chemical Engineering, DIO: 10.1016/j.compchemeng.2016.11.011

D. C. Cafaro, M. G. Drouven, and I. E. Grossmann, 2016, "Optimization models for planning shale gas well refracture treatments." AIChE Journal 62.12: 4297-4307.

M. G. Drouven, and I. E. Grossmann, 2016, "Multi-period planning, design and strategic models for long-term, quality-sensitive shale gas development." AIChE Journal. 62, 2296−2323.

M. J. Fetkovich, E. J. Fetkovich, and M. D. Fetkovich, 1996, Useful Concepts for Decline Curve Forecasting, Reserve Estimation, and Analysis. Society of Petroleum Engineers 11(01): 13-22.

J. Gao, and F. You, 2017, "Design and optimization of shale gas energy systems: Overview, research challenges, and future directions." Computers & Chemical Engineering. DOI: 10.1016/j.compchemeng.2017.01.032.

M. Oral, O. Kettani, 1990, A linearization procedure for quadratic and cubic mixed-integer problems. Operations Research, Vol. 40, Supp. No. 1.

K. Robbins, 2013, Awakening the Slumbering Giant: How horizontal drilling technology brought the endangered species act to bear. Case West. Reserve Law Rev. 63 (4), 13–16.

M. Tavallali, Sadegh, and Iftekhar A. Karimi, 2014, "Perspectives on the Design and Planning of Oil Field Infrastructure." Computer Aided Chemical Engineering 34: 163-172.

J. D. Valbuena, 2015, Defining the artificial lift system selection guidelines in horizontal wells. The University of Tulsa.

Weatherford, 2013, Introduction to Artificial Lift Systems <http://www.alrdc.com/recommendations/Gas%20Well%20Deliquification/artificial%20lift%20selection/WFT%20Intro%20to%20Artificial%20Lift%20Systems.pdf> (accessed 04/14/2017)

C. Wittfeld, 2005, Foam Lifting Manual: Current Best Practices from the Land Asset, Nederlandse Aardolie Maatschappij B.V, Netherlands.

Z. Zeng, S.Cremaschi, 2017, Artificial lift infrastructure planning for shale gas producing horizontal wells. FOCAPO/CPC, Tuscan, AZ.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50210-X

Multi-stage integrated electricity procurement and production scheduling

Egidio Leo[a*] and Sebastian Engell[a]

[a]*Process Dynamics and Operations Group, Department of Biochemical and Chemical Engineering, Technische Universität Dortmund, EmilFigge-Str.70, 44221 Dortmund, Germany*
egidio.leo@tu-dortmund.de

Abstract

For the optimal operation of power-intensive plants, a challenge which is addressed in this work is to simultaneously determine the optimal production schedule and the optimal procurement of electric power. Electricity procurement decisions involve load commitment decisions, meaning that the consumers have to commit themselves to the amount of electricity they are going to purchase before the actual electricity demand is revealed. A major challenge here obviously is the uncertainty about the future demand. To cope with this uncertainty, a stochastic multi-stage mixed-integer linear programming model is proposed that considers uncertain product demands and equipment breakdowns to determine a solution to the integrated electricity procurement and production scheduling problem. The proposed model is applied to a continuous power intensive plant and the results show that there are significant improvements obtained from solving the stochastic model.

Keywords: Electricity procurement, Production scheduling, Stochastic programming, Demand Side Management, Load deviations.

1. Introduction

The increasing share of renewable energies, time-sensitive electricity prices and the flexibility of the production processes render industrial Demand-Side Response (iDSR) a promising strategy to improve profitability for the process industries (Merkert et al., 2014) and to simultaneously make production "greener". A challenge of iDSR as the integration of electricity procurement and production scheduling is to avoid the risk of having competing optimization solutions resulting in no improvements. This means the introduction of electricity purchasing decisions into the scheduling problem. For large consumers of electric power, the electricity purchase options include different types of power contracts:

- Long-term contract (base load) - constant price, constant amount of electricity delivered over time and a horizon of at least 1 year;
- Time-of-Use contract (TOU) - of the price is determined by a time-dependent component and an amount-dependent component; TOU contracts usually cover a period of time between 1 week and 3 months;
- Day-ahead market (spot market) - hourly-varying prices.

The day-head spot market is characterized by hourly varying electricity prices, according to the fluctuations of the availability of electric power. The day-ahead electricity price is assumed to be known 24 hours ahead. Apart from the volatility of the electricity price, a main feature of the day-ahead spot market is load commitment and

the related deviation penalties. For large consumers, the electricity suppliers impose a daily day-ahead hourly electricity commitment, and in case the actual consumption deviates from the pre-agreed values, financial penalties, which are often in the same range as the net electricity cost, are incurred. In (Hadera et al., 2015), the authors take into account these multiple electricity contracts and the load deviation problem to determine the optimal production schedule. However, the contract-related decisions (in particular the load commitments) are assumed as given and are not optimized. In this work, we address the challenge of determining simultaneously the optimal electricity purchase strategy and the production scheduling including the load commitments, i.e. the commitment of the consumers to the amount of electricity that they are going to use during a certain period of time. Depending on the time-scale, we can distinguish between the TOU commitment, which covers up to 3 months, and the daily day-ahead hourly commitment. Since decisions regarding the commitment to power contracts have to be made before the actual electricity demand is known for the time horizon of interest, it is crucial to account for uncertainties in the decision-making process. A recent work in this direction (Zhang et al., 2016) integrates production scheduling and TOU-contract decisions by applying two-stage stochastic programming (Birge and Louveaux, 2011) to model price and demand uncertainties. However, this work does not consider the day-ahead electricity commitment and the associated load deviation problem. To include the day-ahead commitment and the load deviation problem, we propose a three-stage stochastic programming approach. The first-stage variables represent the electricity commitment decisions related to the TOU power contracts. The second stage faces demand uncertainty and for each demand scenario the second-stage variables represent the day-ahead electricity commitment. The third stage takes into account plant capacity uncertainty and the third-stage variables for each scenario are the operating decisions of the plant (production levels, inventories etc.) and the deviations of the power demand from the day-ahead commitment. A similar approach has not been considered in this context before. In Section 2, the model of the uncertainty and the proposed three-stage MILP are presented. Section 3 applies the proposed framework to a power-intensive process and the results are evaluated in terms of the Value of the Stochastic Solution (VSS). The main outcomes are discussed in Section 4 before drawing a final conclusion in Section 5.

2. Problem statement

2.1. Uncertainty modeling strategy

To integrate electricity procurement and production scheduling, we propose a multi-stage stochastic programming approach (Birge and Louveaux, 2011). In stochastic programming, uncertainty is represented by discrete scenarios, and decisions are made at different stages which are defined by the realization of the uncertainty. In the proposed approach, we consider two sources of uncertainty: product demand uncertainty faced by the second stage and equipment breakdown faced by the third stage. For each source of uncertainty, we consider three different levels—low, medium, and high around the same expected value. The low, medium and high uncertainty levels represent a variation of 5, 10 and 20 percent of the product demand and a 10, 30 and 40 percent reduction of the maximum plant production capacity. The production capacity reduction is modeled by 8 breakdown scenarios: in scenario 1 no breakdown occurs, whereas the production capacity is reduced in periods 1-4 for scenario 2, in periods 5-8 for scenario 3, etc. The probabilities are 50% for scenario 1 and (50/7) % ≈ 7.14% for scenarios 2–8 since scenarios 2-8 are considered equiprobable.

2.2. Plant model

To demonstrate the main features of the proposed approach, we apply it to a continuous-production plant. The plant model is the same as in (Zhang et al., 2016). In the proposed mathematical formulation first-stage variables have no scenario subscript, second stage variables have the scenario subscript s and the third-stage variables have a double scenario subscript $s,\bar{s}$. The plant produces two products i (*P1* and *P2*) and it can operate in three different modes m: *off*, *startup*, and *on*. The possible mode transitions are *off* to *startup*, *startup* to *on*, and *on* to *off* and they can happen only after fixed periods of time that have been spent in the modes (*off*: 8 h, *startup*:2 h, *on*:6 h) (Eq.6). The transitions between the modes of operation and the relations with the active modes are modelled by Eq.(7). The binary variable (Eq.(4)) is 1 if mode m is selected in time period t of the horizon T. For each mode the operating conditions are expressed as a convex combination of the extreme points v_{mi} of the feasible region of operation (Eq.(2)). Eq.(1)-(2) define the hourly production levels $\overline{PD}_{mits\bar{s}}$ for each product i and operating mode m and the aggregated production $PD_{its\bar{s}}$.

$$PD_{its\bar{s}} = \sum_{m} \overline{PD}_{mits\bar{s}} \qquad \forall\, i,t \in T, s, \bar{s} \qquad (1)$$

$$\overline{PD}_{mits\bar{s}} = \sum_{m} \lambda_{mits\bar{s}} * v_{mi} \qquad \forall\, i,t \in T, s, \bar{s} \qquad (2)$$

$$\sum_{m} \lambda_{mits\bar{s}} = y_{mt} \qquad \forall\, i,t \in T, s, \bar{s} \qquad (3)$$

$$\sum_{m} y_{mt} = 1 \qquad \forall\, t \in T \qquad (4)$$

$$EU_{its\bar{s}} = \sum_{m} \left(\delta_m * y_{mt} + \sum_{i} \gamma_{mi} * \overline{PD}_{mits\bar{s}} \right) \qquad \forall\, i,t \in T, s, \bar{s} \qquad (5)$$

$$\sum_{k=1}^{\theta_{mm\prime}} z_{mm\prime t-k} \le y_{m\prime t} \qquad \forall\, (m,m') \in M, t \in T, s, \bar{s} \qquad (6)$$

$$\sum_{m\prime} z_{m\prime m,t-1} - \sum_{m'} z_{mm',t-1} = y_{mt} - y_{m,t-1} \qquad \forall\, m,t \in T \qquad (7)$$

Eq. (8) defines the inventory level $IV_{its\bar{s}}$ at time t as the sum of the inventory level at time period *t-1* and the production at time *t*, $PD_{its\bar{s}}$, minus the amount of product sold, $SL_{its\bar{s}}$, and the amount of products wasted, $PW_{its\bar{s}}$, at time period *t*. Note that all these quantities are defined for each scenario of the second and third stage. Eq.(9) sets upper and lower bounds of the inventory levels and Eq. (10) ensures that the demand of product i as the sum of the amount of product i sold, $SL_{its\bar{s}}$, and the amount of product i purchased from other sources, $PC_{its\bar{s}}$, is satisfied

$$IV_{its\bar{s}} = IV_{i,t-1,s\bar{s}} + PD_{its\bar{s}} - SL_{its\bar{s}} - PW_{its\bar{s}} \qquad \forall\, i,t \in T, s, \bar{s} \qquad (8)$$

$$IV_{it}^{min} \le IV_{its\bar{s}} \le IV_{it}^{max} \qquad \forall\, i,t \in T, s, \bar{s} \qquad (9)$$

$$SL_{its\bar{s}} + PC_{its\bar{s}} = D_{its} \qquad \forall\, i,t \in T, s, \bar{s} \qquad (10)$$

Eq. (11) - (14) provide the initial and final condition of the plant in terms of inventory levels and active operating modes.

$$IV_{i,0,s,\bar{s}} = IV_i^{initial} \qquad \forall\, i, s, \bar{s} \qquad (11)$$

$$y_{m,0} = y_m^{initial} \qquad \forall\, m \qquad (12)$$

$$IV_{i,t^{final},s,\bar{s}} \ge IV_i^{final} \qquad \forall\, i, s, \bar{s} \qquad (13)$$

$$z_{mm\prime t} = z_{mm\prime t}^{ini} \qquad \forall\, (m,m') \in M, s, \bar{s}, -\theta^{max} + 1 \le t \le -1 \qquad (14)$$

2.3. Power contract model

Eq.(15)-(21) define the amount of electric power $\widehat{EC}_c$ purchased from TOU contract c. B_c is the set of blocks of contract c, and $x_{cb}$is a Boolean variable that is true if block b of contract c is chosen. $\widehat{EC}_{c,b}^{max}$ is the amount of electricity that one has to purchase in block b before reaching the next block; β_{cb} is the quantity dependent price. Eq.(22) models the electricity commitments of the plant: EC_{ct} represents the amount of energy the consumer purchases from the TOU contract and ES_{ts} is the day-ahead electricity commitment. According to the timing of the decisions, TOU commitment decisions are first-stage variables and day-ahead commitment decisions are second-stage variables. Eq.(23) defines the over-consumptions ($\Delta e_{ts\bar{s}}^{+}$) and under-consumptions ($\Delta e_{ts\bar{s}}^{-}$) (third stage variables) to compute the penalty costs. The over- and under- consumption variables are the load deviations respect to the electricity commitments.

$$\widehat{EC}_c = \sum_t EC_{ct} \qquad \forall\, c \tag{15}$$

$$\widehat{EC}_c = \sum_b \widehat{EC}_{cb} \qquad \forall\, c \tag{16}$$

$$BC_c = \sum_b \beta_{cb} \widehat{EC}_{cb} \qquad \forall\, c \tag{17}$$

$$\widehat{EC}_{c,b-1}^{max} x_{cb} \le \widehat{EC}_{cb} \qquad \forall\, c, b \in B_c, b > 1 \tag{18}$$

$$\widehat{EC}_{cb} \le \widehat{EC}_{c,b}^{max} x_{cb} \qquad \forall\, c, b \in B_c \tag{19}$$

$$\sum_{cb} x_{cb} = 1 \qquad \forall\, c \tag{20}$$

$$EC_{ct} = \overline{EC}_{cp} \qquad \forall\, c, p \in P_c, t \in T_{cp} \tag{21}$$

$$Q_{ts} = EC_{ct} + ES_{ts} \qquad \forall\, t \in T, s \tag{22}$$

$$EU_{ts\bar{s}} - Q_{ts} = \Delta e_{ts\bar{s}}^{+} - \Delta e_{ts\bar{s}}^{-} \qquad \forall\, t \in T, s, \bar{s} \tag{23}$$

2.4. Objective function

The model minimizes the total expected operating costs, z, defined in Eq.(24) as the cost of purchasing energy from the TOU contract, the expected first-stage cost of purchasing electricity from the spot market and the expected second-stage cost of deviation penalties and purchasing of products on the market.

$$\begin{aligned} z = \sum_c BC_c + \sum_t \sum_c p_{ct}^{TOU} EC_{ct} \\ + \sum_s \varphi_s \sum_t \Bigg(p_t^{day-ahead} ES_{st} \\ + \sum_{\bar{s}} \varphi_{\bar{s}} \left(p_t^{+} \Delta e_{ts\bar{s}}^{+} + p_t^{-} \Delta e_{ts\bar{s}}^{-} + \sum_i p_i PC_{its\bar{s}} \right) \Bigg) \end{aligned} \tag{24}$$

where $p_{ct}^{TOU}, p_t^{day-ahead}, p_t^{+}, p_t^{-}, p_i$ represent the TOU electricity price, the day-ahead electricity price, the penalty cost for over-consumption and under-consumption and the product purchasing prices; φ_s , $\varphi_{s\bar{s}}$ denote the probability of scenario s and $\bar{s}$.

3. Results

To measure the improvement that can be achieved by solving the stochastic model instead of its deterministic counterpart, we compute the value of the stochastic solution

(VSS). The VSS (Eq. (25)) and the relative $\overline{VSS}$ (Eq. (26)) show the impact of the uncertainty on the first-stage variables:

$$VSS = z^*_{det} - z^*_{stoc} \quad (25) \qquad \overline{VSS} = \frac{z^*_{det} - z^*_{stoc}}{z^*_{det}} \quad (26)$$

Where z^*_{stoc} is the optimal solution of the stochastic problem and z^*_{det} is the optimal solution of the stochastic problem with first-stage variables fixed to the optimal deterministic solution (obtained with the expected value of the uncertain parameters). For the multi-stage formulation, the VSS_1 shows the cost reduction achievable solving the stochastic model instead of fixing the first-stage variables to the optimal deterministic solution. The VSS_2 shows the cost reduction achievable solving the stochastic model instead of fixing both the first- and the second- stage variables to the optimal deterministic solution. Table 1 shows that the VSS can be quite significant and that it grows with the uncertainty level. The VSS clearly shows the benefit of adopting the stochastic model instead of the deterministic one. Each model has up to 19,000 continuous variables, 212 binary variables, and 11,000 constraints. All models were solved to zero integrality gap in less than 20 s applying the solver CPLEX 12.6.3.0.

Table 1 Value of the stochastic solution for the proposed 3-stage stochastic model for different levels of uncertainty

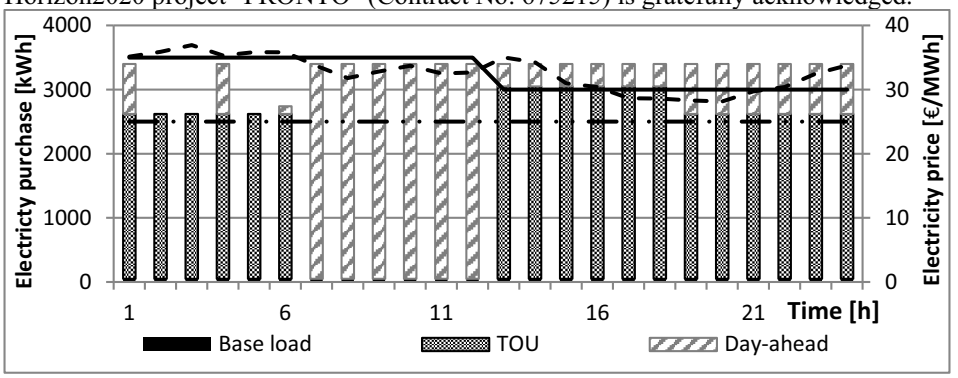
Demand Uncertainty level	**Breakdown uncertainty level**	$\boldsymbol{VSS_1}$	$\boldsymbol{\overline{VSS}_1}$	$\boldsymbol{VSS_2}$	$\boldsymbol{\overline{VSS}_2}$
Low	Low	105	4.3%	144	5.8%
Low	Medium	140.4	5.3%	168.5	6.3%
Low	High	130.3	4.7%	157.8	6.0%
Medium	Low	158.7	6.3%	229	8.9%
Medium	Medium	194.5	7.1%	242.4	8.8%
Medium	High	185.8	6.6%	230.1	8.0%
High	Low	222.8	8.4%	338.3	12.2%
High	Medium	249	8.7%	367.6	12.4%
High	High	245.2	8.3%	348.8	11.4%

4. Discussion

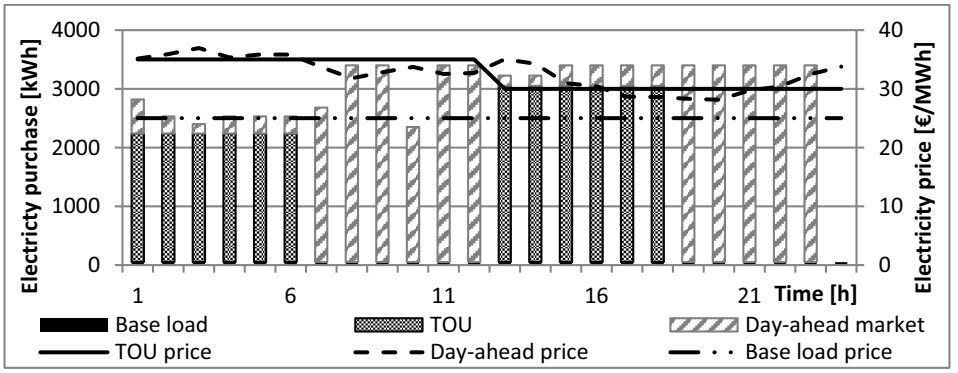
Fig.1 shows the electricity purchase profiles that were obtained for a medium uncertainty level when solving the proposed 3-stage stochastic model (Fig.1 a) and the 2-stage stochastic model (Fig.1 b) proposed in (Zhang et al., 2016), where the day-ahead electricity commitment and the related deviation problem are neglected. In the 2SSP model, the uncertainty source is the product demand, as in the second stage of the proposed 3SSP approach. Both approaches model the TOU contract decisions as first-stage variables. The base load profiles are supposed to be given and therefore not optimized since base load contract decisions have to be made before the time horizon of interest. The shown day-head electricity purchase profiles are the expected values computed over all scenarios. In the 3SSP solution, the plant operates longer in order to accommodate for the high-demand and breakdown scenario providing more flexibility for load shifting. Accounting for breakdown scenarios means purchasing more electricity (when convenient) in order to be able to compensate reductions of the production capacity. This renders the TOU contract (according to the electricity prices) more convenient than the day-ahead market and therefore the 3SSP solution purchase a higher amount of electricity from the TOU contract. It is remarkable how different the optimal electricity purchase strategy can be when considering the day-ahead commitment and the load deviation penalties.

5. Conclusion

This work addresses the integrated electricity procurement and production scheduling under uncertainty for continuous power-intensive processes. A three-stage stochastic programming model is proposed to model electric power purchase commitments accounting for uncertainties in product demands and equipment breakdowns. The application of the proposed approach instead of the deterministic counterpart shows that the total operating cost reduction for a power-intensive plant can be quite significant.
Acknowledgements- The financial support from the Marie Sklodowska-Curie Horizon2020 project "PRONTO" (Contract No: 675215) is gratefully acknowledged.



(a)



(b)

Figure 1 Electricity profiles obtained solving the 3-stage (a) and 2-stage (b) stochastic model

References

Q. Zhang, L. Jochen, L. Cremer, I.E. Grossmann, A. Sundaramoorthy and J.M. Pinto, 2016, , Risk-based integrated production scheduling and electricity procurement for continuous power-intensive processes, Comput. Chem. Eng., 86, 90-105.

H.Hadera, I.Harjunkoski, G.Sand, I.E.Grossmann and S.Engell, 2015, Optimization of steel production scheduling with complex time-sensitive electricity cost, Comput. Chem. Eng., 76.

JR. Birge, F. Louveaux.2011, Introduction to stochastic programming, Springer Science+Business

L.Merkert, I.Harjunkoski, A.Isaksson, S.Säynevirta, A. Saarela and G. Sand, 2014, Scheduling and energy – industrial challenges and opportunities, Comput. Chem. Eng., 72, 183-198.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50211-1

Refinery Operation Scheduling Considering Both Varying Feedstocks and Operating Conditions: An Industrial Data-based Modeling Method

Xiaoyong Gao[a, c*], Nan Ru[b], Yongheng Jiang[b*], Tao Chen[d], Dexian Huang[b*], Xin Zuo[c]

[a] *Institute for Ocean Engineering, China University of Petroleum, Beijing 102249, China*
[b] *Department of Automation and Tsinghua National Laboratory for Science and Information Technology, Tsinghua University, Beijing 100084, China*
[c] *Department of Automation, China University of Petroleum, Beijing 102249, China*
[d] *Department of Chemical and Process Engineering, University of Surrey, Guildford GU2 7XH, UK*
x.gao@cup.edu.cn, jiangyh@tsinghua.edu.cn, huangdx@tsinghua.edu.cn

Abstract

Due to the varying feedstocks and operating conditions in refinery routine operations, the current process models (i.e. outlets yield prediction and operating cost model) for scheduling mismatch much from the industrial reality. How to obtain the more accurate and feasible models to describe refining processes remains an open problem and is the key challenge for successful industrial application use. In this paper, an industrial data-based modelling method considering both varying feedstocks and operating conditions of process units is presented. In details, for a certain unit, feedstocks composition rates and outlets properties are introduced to indicate the varying feedstocks and operating conditions respectively. For catalytic reformer and diesel hydrotreater, outlets yield and operating cost models about feedstocks composition rates and outlets properties (here, specific properties variables are selected for each process) are determined by data-based regression. Furthermore, the resulted process models are embedded into the scheduling model. A computational case originated from a real refinery in China is solved by BARON in GAMS 24.0.2 using a CPU Intel Xeon E5-2609 v2 @2.5 GHz with RAM 32.0GB. Result shows that, (1) the operation schedule of units are well adjusted dynamically with the demand trend of different product oils, (2) the resulted schedules increase the outlet yield and reduce the operating cost of process units and thus, improve the economic profit of refinery.

Keywords: Refinery scheduling, Process model, Feedstock, Operating Condition

1. Introduction

Optimal scheduling will contribute to improving economic profit of refinery industry and addressing the growing challenges, such as increasing requirement for product quality, strict environmental regulations, frequent fluctuation of crude oils caused by tighter supply, changes in demand for product oils, and so on (Gao, et al, 2015).

A number of scheduling methods have been developed over the past few years. A novel refinery optimization modeling and decomposition strategy was proposed (Zhang and

Zhu, 2000; Pinto et al, 2000). Smania and Pinto (2003) pointed out that the operating conditions and inflow properties should be taken into consideration. However, how to model it remains an open problem. Continuous time formulation and decomposition method are proposed by many scholars. The yield model is usually described as constant regardless of operation and feed changes (Jia and Ierapetritou, 2004; Dogan and Grossmann, 2006; Wu and Ierapetritou, 2007). Shah and Ierapetritou (2011) early treated outlet yields of production units as optimized variable with predefined bounds. The predefined operation state with fixed feedstock and unit operation condition was used by Göthe-Lundgren (2002). However, this multi-fixed unit operation point method does not well cover the entire scheduling domain.

Although there are many valuable reports in refinery scheduling, very limited attention has been placed on the modeling of the yield and operating cost of refinery units (response variables) as function of decision variables. In the majority of existing studies, the yield and cost are fixed for each predefined operating mode, i.e. models resembling look-up tables (Göthe-Lundgren et al., 2002); this strategy does not well represent the real processes. However, if highly complex nonlinear models are adopted, such as neural networks or piecewise linear models (Gao et al., 2014a), the subsequent optimization becomes nonlinear and is hard to solve efficiently. More recently, Gao et al. (2014b) considered the impact of variations in crude oil on scheduling, and proposed a deep belief network based crude oil classification and multi-model method to describe the impact caused by crude oil changes. And also, Gao et al. (2015) proposed a piecewise linear modeling method to consider the operation condition changes using simulation data from Petro-SIM. How to model the impact caused by both feed and operating condition changes simultaneously is not considered yet. In this paper, we present a data-driven modeling method considering both varying feedstocks and operating conditions of process units. The operational data of process units are collected from a real-world refinery in China.

The rest of the paper is organized as follows. Section 2 provides the problem statement. The proposed process modeling method is presented in Section 3. After that, a case study is given in Section 4 to demonstrate the effectiveness of the proposed methodology. Finally, a brief conclusion is drawn in Section 5.

2. Problem statement

We consider a real-world refinery in China in this paper. For this refinery, the supply of crude oils keeps steady and the operation of the primary (including preliminary, atmospheric and vacuum distillation units) and secondary processing units (fluidized catalytic cracking (FCC), hydro-cracking, delayed coking, etc.) does not change. In this case, we take the downstream of the primary and secondary operations as the adjustable units including the hydro-upgrading processing units (HUPUs) and product oil blenders. The feedstocks and operation conditions of HUPUs observably influence the yield and quality of the outlet flows. Blenders making use of the outlets from HUPUs directly determine the amount of each product which impacts revenue (Singh et al., 2000). The flow diagram in Figure 1 illustrates these adjustable units.

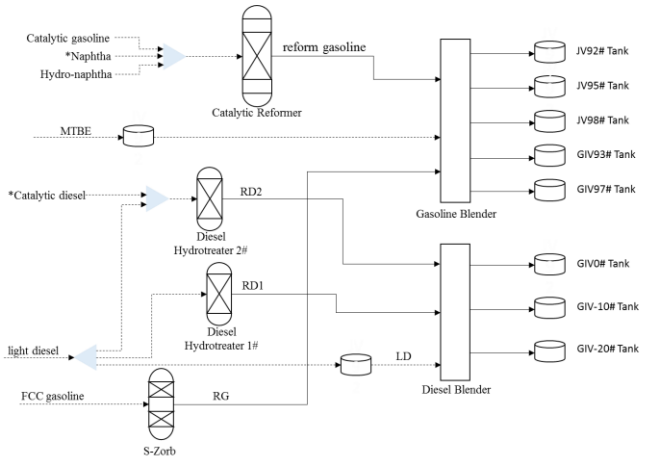


Figure 1. Flow chart of the refining units subject to scheduling

The HUPUs includes catalytic reformer, a FCC gasoline desulfurizer S-Zorb, a light diesel hydrotreater (Diesel Hydrotreater 1#) and a heavy diesel hydrotreater (Diesel Hydrotreater 2#) in this case. The dash lines represent material streams with fixed property from the upstream units. For catalytic reformer and diesel hydrotreater 2# having multiple inlets, key feed material (marked by * in Figure 1) is determined based on both mechanism and data analysis, which is then selected as the indicator to describe the feedstocks changes.

The objective is to maximize the profit subject to the process and operations constraints, and quality specifications. To formulate the optimal scheduling problem, models are needed to describe the yield and operating cost of each processing unit as a function of the feedstock decision variables and operating conditions decision variables. When HUPUs have more than one feed material, there is normally a key material which dominates the feedstock properties. Approximately, the percentages of the key material can be taken as an indicator of feedstock changes and thus introduced as the feedstock decision variables. In details, for catalytic reformer, the naphtha percentage is the feedstock decision variable, while for diesel hydrotreater 2#, the catalytic diesel percentage is treated as feedstock decision variable. Since the outlets of HUPUs will be blended as final products, the final product properties are key operation decision variables. Moreover, the outlet properties can also represent the operating conditions changes. In particular, for gasoline HUPUs (catalytic reformer and S-Zorb), the desulfurization amount (ΔSC) and delta research octane number (ΔRON) are the operation decision variables, whereas for diesel HUPUs (the two hydrotreaters), the desulfurization amount (ΔSC) and delta freezing point (ΔFP) are taken as operation decision variables. Besides these HUPU-specific decision variables, the inlet and/or outlet mass flows also need to be determined for each unit and blending/storage tank. For blenders, the outlet property is assumed as a linear function of those of the inlets. There are eight kinds of product oil, including five grades of gasoline and three grades of diesel. Dedicated tanks are assigned for each grade product.

3. Linear regression model

In this case, there are four HUPUs. For both outlet yield and operation cost model, in detail, decision variable refers to key feed material percentage $P_{u,n}$, desulfurization

amount $\Delta SC_{u,n}$, delta research octane number $\Delta RON_{u,n}$ for gasoline HUPUs, and delta freezing point $\Delta FP_{u,n}$ for diesel HUPUs, respectively. In order to obtain their process models, we have collected a whole year's operational data from real production processes of a refinery in China. For the sake of concision, we take outlet yield model of diesel hydrotreater 2# and operating cost model of catalytic reformer for examples to illustrate the effectiveness of the proposed data-based linear regression model.

3.1. Outlet yield model

For diesel hydrotreater 2# (HTU2 for short), the main product is refine diesel (i.e. RD2 in Figure 1) as one kind of diesel blending component oils. Here, the freezing point is firstly transformed into the freezing point factor in order to guarantee a linear relation for blending. Linear regression result for outlet yield model is shown in Eq. (1). The root mean square error (RMSE) and mean relative error (MRE) for this regression are 3.2905 and 4.505% respectively.

$$Yield_{HTU2,n} = 122.076 - 0.527 \cdot P_{HTU2,n} + 16.371 \cdot \Delta FP_{HTU2,n} + 0.0154 \cdot \Delta SC_{HTU2,n} \quad (1)$$

3.2. Operating cost model

For catalytic reformer (RF for short), the main product is reform gasoline, which is used to blend high quality gasoline as a gasoline component oil. Specifically, the pre-processing unit of RF desulfurize the feed material to 0.5ppm (sulfur content) to prevent the catalyst in reformer from poisoning. Linear regression model for the operating cost is written as Eq. (2). Similarly, the root mean square error (RMSE) and mean relative error (MRE) are 4.176 and 4.692% respectively.

$$OpCost_{RF,n} = -5.843 + 0.658 \cdot P_{RF,n} + 1.768 \cdot \Delta RON_{RF,n} + 0.048 \cdot \Delta SC_{RF,n} \quad (2)$$

4. Case study

Due to the frequent change of product oil demand in the investigated refinery, the scheduling problem is of short term with a horizon of 96 hours. The four orders are determined by monthly plan, as listed in Table 1.

This case is solved by BARON in GAMS 24.0.2 using a CPU Intel Xeon E5-2609 v2 @2.5 GHz with RAM 32.0GB. The obtained process model is embedded into the continuous time representation based scheduling model (Ierapetritou and Floudas, 1998a; Ierapetritou and Floudas, 1998b). The model involves 1473 continuous variables, 384 binary variables and 3328 constraints.

Table 1 Product Oil (due time & quantity)

Order	Due time/hours	Gasoline/ton					Diesel/ton		
		JV92	JV95	JV98	GIV93	GIV97	GIV 0	GIV-10	GIV-20
1	24	400	0	200	450	300	600	700	500
2	48	300	300	300	0	400	800	550	650
3	72	200	600	450	0	200	750	650	600
4	96	200	250	0	800	0	700	600	620

Clearly, the demand for premium and high grade gasoline (i.e. JV95#, JV98# and GIV 97#) firstly increases and then decreases. The schedule for the catalytic reformer is depicted in Figure 2 and 3. To satisfy the demand change of premium and high grade gasoline, the catalytic reformer first tries to increase the RON of refine gasoline (higher quality, higher operating cost and lower yield level), and then reduce the RON when demand for low-grade fuel is high. In addition, the desulfurization amount of S-Zorb also exhibits the same pattern, as shown in right part of Figure 2. Such fine tuning of the operation provides more flexibility, and thus opportunities to improve the profit while meeting the demand.

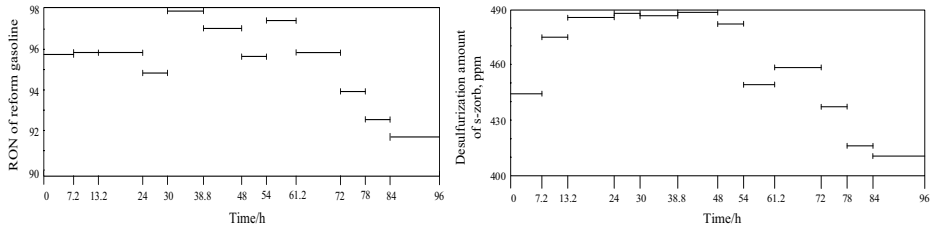


Figure 2. Left: refine gasoline RON of catalytic reformer; Right: desulfurization amount of S-Zorb

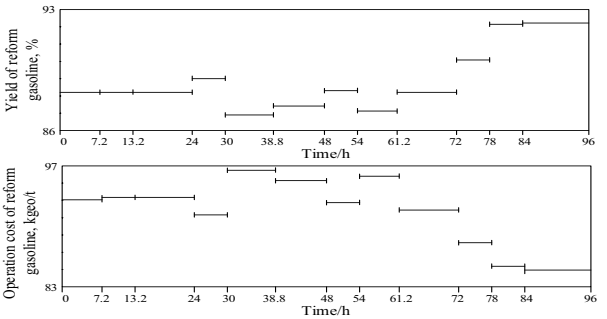


Figure 3. Yield and operation cost of Catalytic Reformer.

5. Conclusion

In this paper, both varying feedstocks and operating conditions are taken into consideration for refinery scheduling problem. To well model the impact caused by both varying feedstocks and operating conditions, the key feed percentage and outlet properties are introduced as independent variables. An industrial operational data based regression method is used and the model results prove its effectiveness. The resulted process model is then organized in the scheduling model and was further demonstrated through a scheduling case study, which was originated from a real refinery in China. The operation schedule of units are well adjusted dynamically with the demand trend of different product oils. This dynamic feature improves the operation ability to satisfy the quality specifications on edge, which is significant to increase the outlet yield and reduce the operation cost of process units and thus, improve the economic profit of refinery. Cleary, an updating strategy for the proposed linear regression model along the time is needed and it is further considered in our undergoing research.

References

Gao, Xiaoyong, et al. “Optimizing scheduling of refinery operations based on piecewise linear models.” Computers & Chemical Engineering 75 (2015): 105-119.

Zhang, N., and X. X. Zhu. “A novel modelling and decomposition strategy for overall refinery optimisation.” Computers & Chemical Engineering 24.2-7 (2000): 1543-1548.

Pinto, J. M., M. Joly, and L. F. L. Moro. “Planning and scheduling models for refinery operations.” Computers & Chemical Engineering 24.9 (2000): 2259-2276.

Smania, P., and J. M. Pinto. “Mixed integer nonlinear programming techniques for the short term scheduling of oil refineries.” Computer Aided Chemical Engineering 15 (2003): 1038-1043.

Jia, Zhenya, and Marianthi Ierapetritou. “Efficient short-term scheduling of refinery operations based on a continuous time formulation.” Computers & chemical engineering 28.6 (2004): 1001-1019.

Dogan, Muge Erdirik, and Ignacio E. Grossmann. “A decomposition method for the simultaneous planning and scheduling of single-stage continuous multiproduct plants.” Industrial & engineering chemistry research 45.1 (2006): 299-315.

Wu, Dan, and Marianthi Ierapetritou. “Hierarchical approach for production planning and scheduling under uncertainty.” Chemical Engineering and Processing: Process Intensification 46.11 (2007): 1129-1140.

Shah, Nikisha K., and Marianthi G. Ierapetritou. “Short‐term scheduling of a large‐scale oil‐refinery operations: Incorporating logistics details.” AIChE Journal 57.6 (2011): 1570-1584.

Göthe-Lundgren, Maud, Jan T. Lundgren, and Jan A. Persson. “An optimization model for refinery production scheduling.” International Journal of Production Economics 78.3 (2002): 255-270.

Gao, Xiaoyong, et al. “A novel high level canonical piecewise linear model based on the simplicial partition and its application.” ISA transactions 53.5 (2014a): 1420-1426.

Gao, Xiaoyong, et al. “Refinery scheduling with varying crude: A deep belief network classification and multimodel approach.” AIChE Journal 60.7 (2014b): 2525-2532.

Singh, A., et al. “Model-based real-time optimization of automotive gasoline blending operations.” Journal of process control 10.1 (2000): 43-58.

Shi, Lei, et al. “Refinery production scheduling involving operational transitions of mode switching under predictive control system.” Industrial & Engineering Chemistry Research 53.19 (2014): 8155-8170.

Ierapetritou, M. G., and C. A. Floudas. “Effective continuous-time formulation for short-term scheduling. 1. Multipurpose batch processes.” Industrial & engineering chemistry research 37.11 (1998a): 4341-4359.

Ierapetritou, M. G., and C. A. Floudas. “Effective continuous-time formulation for short-term scheduling. 2. Continuous and semicontinuous processes.” Industrial & engineering chemistry research 37.11 (1998b): 4360-4374.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50212-3

Multi-System Development Planning for Optimizing Shale Gas Production

Abigail Ondeck[a*], Markus Drouven[b], Nathan Blandino[b], Ignacio E. Grossmann[a]

[a]*Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA*
[b]*EQT Corporation, 625 Liberty Ave., Pittsburgh, PA 15222, USA*
aondeck @andrew.cmu.edu

Abstract

In this study, we present the simultaneous planning and scheduling of shale gas production using bi-criterion mixed-integer linear programming (MILP) to help guide the long-term corporate strategy for developing assets over multiple shale gas production systems. By analyzing the existing gas production and considering the economic costs of developing new prospective systems, the programming model will help determine the optimal asset development and gas production, subject to physical and financial constraints. We apply the planning and scheduling to a case study to quantify the value of utilizing multi-objective optimization for shale gas development and production.

Keywords: shale gas, bi-criterion optimization, planning, mixed integer programming.

1. Introduction

Shale gas production is expected to increase by over 50 % in the next 34 years, going from 14 trillion cubic feet (TCF) in 2016 to approximately 29 TCF in 2040 (EIA, 2016). This increase in production makes shale gas a major player in the natural gas field, with shale gas production fulfilling 69 % of the expected total dry gas natural demand (EIA, 2016). With this expected growth, all stages of the natural gas supply chain will need modification, either by adding new capacities or expanding existing capabilities (Drouven et al., 2016). To accommodate this expected growth, companies must make informed decisions of where, when, and how to develop their assets.

2. General problem statement

The problem can be stated as follows. Consider the representative core development area superstructure in Figure 1, which contains a collection of natural gas sales points, gathering and midstream pipelines, and producing and perspective pads and well groups. At each pad, all wells are bundled together into groups based on their physical characteristics and current production capabilities. All pads have been grouped together into systems, based on their physical proximity to a central midstream pipeline, and are connected, via gathering pipelines, to at least one midstream pipeline. If a well group is producing gas, then the associated pad is also producing. However, not all well groups at a pad must be developed to production. Utilizing this superstructure and knowledge specific to the core development area, the goal is to understand the long-term asset development needed to maximize asset value and production given available infrastructure and financial constraints.

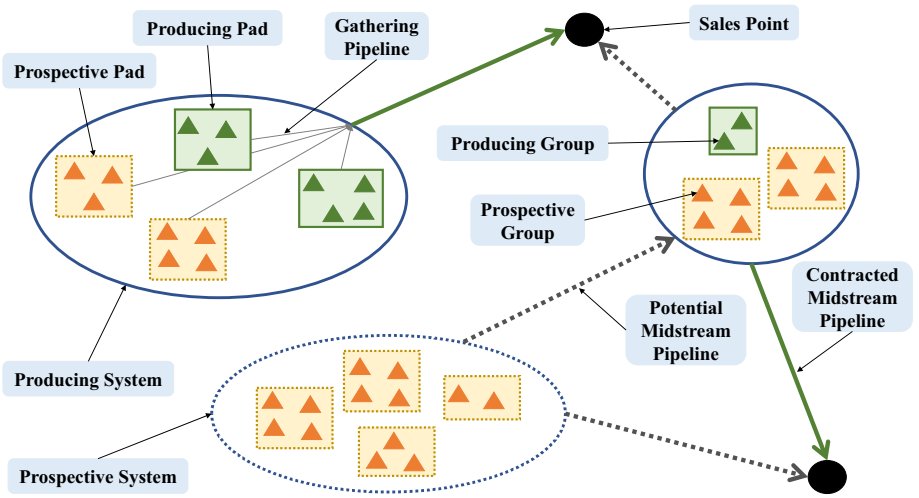


Figure 1. Representative illustration of the core development area superstructure.

From an executive level, the driving force behind maximizing long-term asset development may be divided, with shareholders expressing conflicting interests (Gao et al., 2017). While it is important to consider all conflicting interests when developing the assets, it can be impossible to maximize one interest without sacrificing the value of another. However, it is possible to identify an ideal compromise, where all interests sacrifice the least possible with respect to their maximum benefit.

The purpose of this work is to maximize the long-term asset value of a shale gas development area, considering two conflicting objectives that are to be maximized: the net present value (NPV) and the shale gas production volume. Intuitively, maximizing volume is thought to increase the NPV, but one goal of this research is to determine if this logic holds true for the shale gas sector. The NPV and production volume will both be determined via a long-term planning and scheduling, which identifies: a) when and which groups of wells should be developed, b) how much natural gas production should be curtailed over time, c) which midstream pipeline options should be selected, d) how to utilize the available midstream pipeline capacity, and e) when and how much gas should be delivered to each gas sales point. The planning and scheduling optimization, formulated as a bi-criterion mixed integer linear program (MILP), is solved using the ϵ-constrained method and finding the ideal compromise solution. The results will provide informed guidance for executive-level decisions in a sector where economic efficiency is imperative with the current low price of natural gas (Gao et al., 2017).

3. The bi-criterion development problem

The bi-criterion development problem consists of two parts: 1) the multi-system development planning and scheduling, and 2) the bi-criterion optimization. When combined, these parts will identify one solution where both interests will sacrifice the least that is possible with respect to their maximum benefit.

3.1. Multi-system development planning and scheduling problem

The first part of the multi-system development planning and scheduling is to transform the physical and operational aspects of the core development area superstructure into a discrete-time MILP formulation. The objective is two-fold: 1) maximize the NPV, and 2) maximize the production volume *(VOL)*:

$$\boldsymbol{NPV = Rev^{D} - \left(Cost^{DDev} + Cost^{DPipe} + Cost^{DGather}\right)} \tag{1}$$

$$\boldsymbol{VOL = \sum_{sp \in SP} \sum_{t \in T^{rev}} P_{t,sp}^{fin}} \tag{2}$$

where Rev^{D} is the discounted revenue from selling gas at the sales points, $Cost^{DDev}$ is the discounted cost of developing the well groups, $Cost^{DPipe}$ is the discounted cost of midstream pipeline contracts, $Cost^{DGather}$ is the discounted cost to collect and transport natural gas from the pads to the midstream pipelines, and $P_{t,sp}^{fin}$ is the volume of gas sold at sales point *sp* at time *t*. Due to the high cost of development and the long horizon needed to recover gas to cover these expenses, the development decisions (development and contracts) are only optimized during the first T^{dev} time steps, while the gas production is scheduled through the optimization horizon, T^{rev}, where $T^{rev} \gg T^{dev}$

The objectives are limited by both developmental and production-based constraints, equating to legal, resource, and physical limits. These constraints include:

- Permit dates: Well groups cannot be developed until all permits have been acquired
- Resource constraints: Limits on the amount of drilling allowed at the system and core development levels
- Pipeline contract terms: Midstream pipeline contracts can last over 10 years
- Capacity limits: Well group production (characterized by very high initial production rate, followed by a sharp decline within the first year (Drouven et al., 2016)) and pipeline capacity may not be exceeded
- Production repression: Well group production may be curtailed or stopped
- Gas balances: All gas produced from well groups must be delivered to a sales point

Both the objectives and constraints are driven by three sets of Boolean variables, $x_{t,j,k}^{con}$, $y_{t,g}^{start}$, and $z_{t,g}^{st}$, where the first set is *true* if a contract exists to allow transport of gas in a midstream pipeline from node *j* to node *k* at time *t*, the second set is *true* if well group *g* has been developed at time *t*, and the third set is *true* if well group *g*'s production is curtailed at time *t*, with the withheld gas saved in "underground storage".

By initially formulating the multi-system development planning and scheduling using General Disjunctive Programming (GDP), we can systematically derive mixed-integer inequalities for all disjunctions and logic propositions (Grossmann and Trespalacios, 2013) to obtain a discrete-time MILP formulation. This MILP formulation can then be used with the bi-criterion optimization to help understand and guide the long-term asset development needed to maximize asset value given available infrastructure and financial constraints.

3.2. Bi-criterion optimization

The second part of the multi-system development planning and scheduling is to identify the relationship between two conflicting interests, and determine one solution where

both interests will sacrifice the least that is possible with respect to their maximum benefit. Let f_1 represent the NPV and f_2 represent the total shale gas volume produced, where the objective is to maximize both f_1 and f_2. The multi-objective optimization, utilizing the ϵ-constrained method (Haimes et al., 1975) is formulated as follows:

$$\begin{aligned} \boldsymbol{max} \quad & \boldsymbol{f_1} \\ & \boldsymbol{f_2 \geq \epsilon} \end{aligned} \tag{3}$$

where ϵ is an adjustable parameter that lies in the interval $[f_2^L, f_2^U]$. By adjusting ϵ and maximizing formulation (3), a set of points is obtained, which defines a trade-off curve, demonstrating how improving one objective negatively impacts the other conflicting objective.

While the ϵ-constrained method gives a set of non-inferior solutions, an ideal compromise can be obtained by determining the closest non-inferior solution to the utopia point (i.e., the infeasible point where both the volume and NPV are at their maximum value) (Yu, 1973), (Grossmann et al., 1982). This is done by minimizing the distance between the utopia point and a non-inferior solution. To account for differences in magnitudes between f_1 and f_2, the objective functions are scaled based on their upper and lower limits. Selecting a particular norm of one, the following optimization is solved to obtain the ideal compromise:

$$\textbf{max} \quad \left[\frac{(\boldsymbol{f_1} - \boldsymbol{f_1^L})}{(\boldsymbol{f_1^U} - \boldsymbol{f_1^L})} + \frac{(\boldsymbol{f_2} - \boldsymbol{f_2^L})}{(\boldsymbol{f_2^U} - \boldsymbol{f_2^L})}\right] \tag{4}$$

where f_1^U is the upper bound obtained by maximizing f_1, and f_1^L is the lower bound obtained by selecting the minimal value of f_1 from the ϵ-constrained method and the trade-off curve.

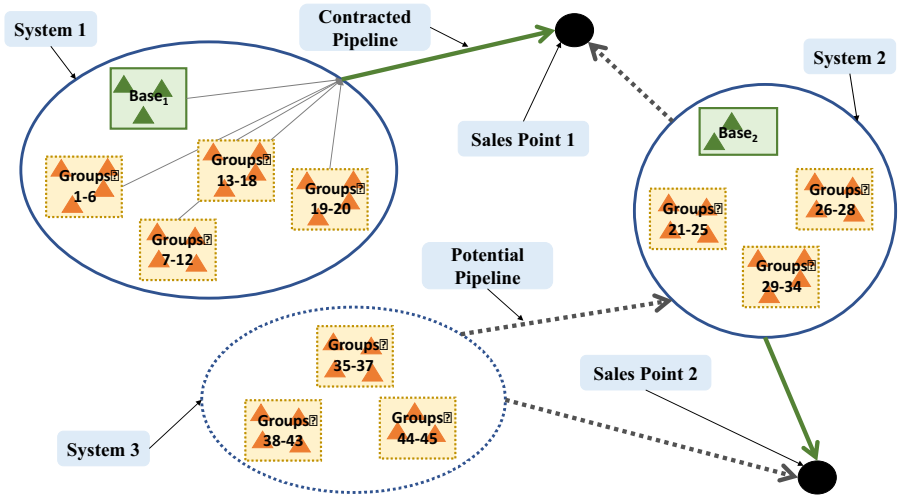


Figure 2. Prospective core development area used in the case study.

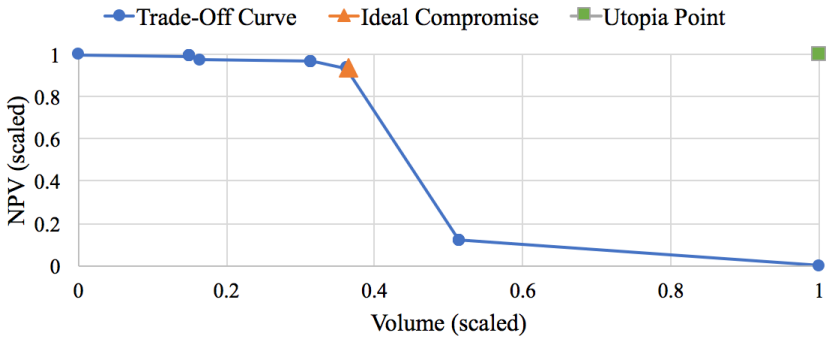


Figure 3. Scaled trade-off curve, created using the ϵ-constrained method.

4. Case study

In this case study, we apply the bi-criterion optimization and multi-system development planning and scheduling to an illustrative example (Figure 2) with a one-year planning horizon, a five-year revenue horizon, and a one month time step. The proposed multi-system development area includes two producing systems with potential for additional well group development (Systems 1 and 2), and a third undeveloped system (System 3). The two producing systems are connected to two sales points with different natural gas prices, but other midstream pipelines are available to be contracted. The prospective core development in Figure 2 was optimized with GAMS 24.8.5 using the commercial solver CPLEX 12.7.1.0 on a PC Intel Core i7, 3.60 GHz, 16 GB RAM, 64 bit, and Windows 10. The largest problem solved had just over 16,000 equations and 17,700 variables, of which 720 were binary, and took 10 seconds to solve.

The trade-off curve, obtained by using the ϵ-constrained method with forty volume-constraint parameters, is shown in Figure 3, where the volume production is scaled from

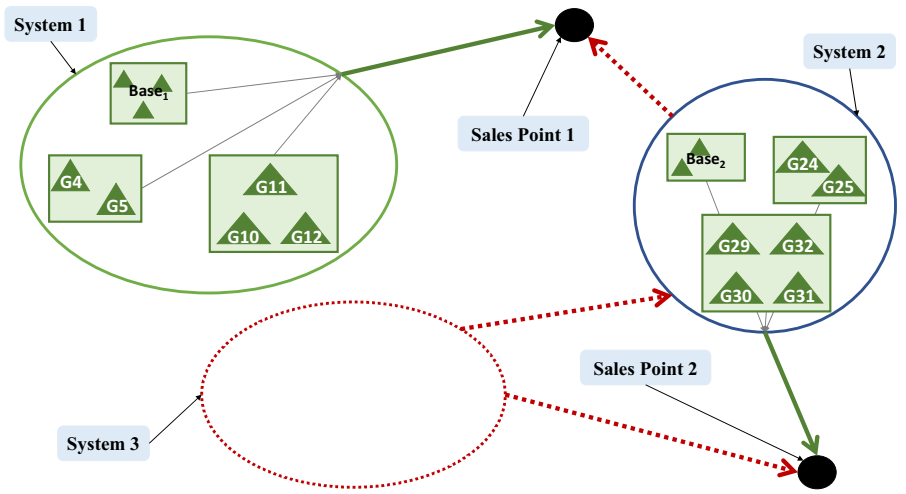


Figure 4. Optimal development of the core development area

267.2 BCF to 328.4 BCF, and the NPV is scaled from $ 25.9 million to $ 135.5 million. The trade-off curve indicates that maximizing volume production is not the best development policy. After producing 267.2 BCF, the NPV of the core development area will continue to decrease as additional well groups are developed and more natural gas is produced. A steep decline in the NPV is seen when the scaled volume production exceeds 0.37. Between 0.37 and 0.5 of the scaled volume production, natural gas production from Systems 1 and 2 is limited by pipeline capacity, so System 3 must be developed to produce additional natural gas. To develop any well group in System 3, a contract must be made to transport the gas from the system to Sales Point 2, introducing a very large development cost and a sharp decrease in NPV.

By scaling and optimizing both conflicting interests simultaneously, an ideal compromise solution was determined, where both the NPV and natural gas volume sacrifice the least, relative to the utopia point (Figure 3). The optimal advancement of the core development area in the ideal compromise is shown in Figure 4. No well groups were developed in System 3 due to the high cost to contract pipelines to deliver the gas to System 2 or Sales Point 2. However, the pipelines already under contract to deliver natural gas to Sales Points 1 and 2 have extra natural gas delivery capacity relative to the initial system base productions, allowing for multiple well groups in both systems to be developed to production.

5. Conclusions

In this paper, we optimized the design of a shale gas core development area taking into consideration developmental and production-based constraints and conflicting objectives to be maximized. Utilizing an illustration of a shale gas development area, we formulated a bi-criterion MILP planning and scheduling optimization and solved for the ideal development configuration using the ϵ-constrained method and compromise programming. The results demonstrated the impact of considering multiple driving forces when maximizing long-term asset development, and the value in using optimization to guide the long-term corporate strategy for developing multiple assets.

References

EIA, 2016. Most natural gas production growth is expected to come from shale gas and tight oil plays. https://www.eia.gov/todayinenergy/detail.php?id=26552.

M. Drouven, I.E. Grossmann, 2016. Multi-Period Planning, Design, and Strategic Models for Long-Term, Quality-Sensitive Shale Gas Development. AIChE Journal, 62 (7), 2296-2323.

J. Gao, F. You, 2017. Design and optimizatino of shale gas energy systems: Overview, research challenges, and future directions. Computers and Chemical Engineering, 106, 699-718.

I.E. Grossmann, R. Drabbant, R.K. Jain, 1982. Incorporating toxicology in the synthesis of industrial chemical complexes. Chemical Engineering Communications, 17, 151-170.

I.E. Grossmann, F. Trespalacios, 2013. Systematic Modeling of Discrete-Continuous Optimization Models through Generalized Disjunctive Programming. AIChE Journal, 59, 3276-3295.

Y.Y. Haimes, W.A. Hall, H.T. Freedman, 1975. Multiobjective Optimization in Water Resource Systems. Elsevier Scientific, Amsterdam.

P.L. Yu, 1973. A class of solutions for group decision problems. Management Science, 19 (8).

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50213-5

Expanding the Scope of Electric Power Infrastructure Planning

Cristiana L. Lara[a*], Ben Omell[b], David Miller[b], Ignacio E. Grossmann[a]

[a]*Department of Chemical Engineering, Carnegie Mellon University, 5000 Forbes Avuenue, Pittsburgh, PA 15213.*
[b]*National Energy Technology Laboratory, 626 Cochrans Mill Rd, Pittsburgh, PA 15236.*
grossmann@cmu.edu

Abstract

This paper addresses long-term generation expansion planning considering a high penetration of renewables and the possibility of investing in advanced fossil fuel energy systems and multiple energy storage technologies. We propose a deterministic multi-scale mixed-integer linear programming (MILP) formulation that includes annual investment decisions and hourly unit commitment. We adopt time sampling and clustering techniques to reduce the size of the model and improve its tractability. Additionally, we use the Nested Decomposition algorithm for mixed-integer multi-period problems proposed by Lara et al. (2017) to solve the problem more efficiently. We apply the proposed formulation to a case study in the ERCOT region, and show that for this case study it makes sense to invest in lithium-ion utility batteries and natural gas fuel cells.

Keywords: generation expansion planning, energy storage, advanced fossil fuel.

1. Introduction

Generation expansion planning models consist of finding the optimal investment strategy for building new generation capacity while meeting load demand and satisfying technical and economic requirements. Such models are used to support decision making in power systems, study the impact of new technology developments, identify resource cost trends, and evaluate the impact of policy shifts on the projected generation mix to meet future demand. The increasing penetration of renewable generation in the grid brought new challenges for the long-term planning models. Due to their intermittent and non-dispatchable nature, it becomes crucial to include operating details at the hourly level to ensure flexibility in the system. Therefore, recent publications such as Flores-Quiroz et al. (2016), Heuberger et al. (2017), and Lara et al. (2017) have focused on integrating unit commitment, ramping limits and operating reserves into long-term planning models.

A number of flexible, environmentally sustainable fossil energy systems are under development which can help maintain grid stability by providing baseload power as well as help manage fluctuations that occur with high penetrations of renewables. Such new systems are expected to have increasing flexibility and achieve lower costs over the next several years. In addition, energy storage can play a pivotal role in the future of renewable generation to smooth out the variability of wind and solar power output. Several stationary, large-scale energy storage technologies are under development and are projected (Schmidt et al., 2017) to significantly decrease in capital cost over the next few decades. Liu et al. (2017) considers a generic energy storage technology in their capacity expansion model, which was represented as a multistage stochastic linear programming model, and applied to a case-study in the ERCOT region.

In this paper, we extend the work by Lara et al. (2017) to address long-term planning of electric power infrastructure considering an increasing share of power generation from renewables and the possibility of having new, advanced fossil energy systems and/or electricity storage units incorporated to the grid. The modeling framework, which is based on mixed-integer linear programming (MILP), takes the viewpoint of a central planning entity whose goal is to optimize the generation expansion planning. This model is very comprehensive in terms of alternatives considered, identifying the source, generation technology, electricity storage technology, location, and capacity of future generation and storage units that can meet the projected electricity demand while taking into account detailed operational constraints (i.e., unit commitment), the variability and intermittency of renewable generation sources, and the power flow between regions. The major goal of this model is to help understanding the characteristics needed to develop new advanced energy generation and storage technologies that can be competitive in the anticipated future market considering all sources of competition.

In addition to the challenge of comprehensive grid modeling and solving the resulting MILP, another challenge lies in the multi-scale integration of detailed operating decisions at the hourly level with investment planning decisions over a few decades, significantly exacerbating the computational burden. Therefore, we adopt judicious approximations and aggregations, such as time sampling and generator clustering (Palmintier and Webster, 2014) to reduce the size of the model. In the algorithmic front, we use the Nested Decomposition algorithm for mixed-integer multi-period problems proposed by Lara et al. (2017) to solve the problem efficiently.

2. Problem Statement

For the proposed planning model, an area with a set of existing and potential generators is given. Regarding these generators, their source (nuclear, coal, natural gas, wind, and solar), and generation technology (e.g., steam, combustion and wind turbines, photo-voltaic, concentrated solar panels, solid oxide fuel cells) are known. Also known are: their nameplate capacity; expected lifetime; capital cost; fixed and variable operating cost; maximum yearly installation limit; start-up costs (fixed and variable); cost for extending their lifetimes; CO_2 emission factor; age in the beginning of the planning horizon; location (regions); and operating characteristics such as ramp rates, operating limits, maximum contribution to spinning and quick-start fraction for thermal generators, and capacity factor in an hourly basis for the renewable generators.

Also given is a set of potential storage units, with specified technology (e.g., lithium-ion, lead-acid, and flow batteries), capital cost, power rating, rated energy capacity, charge and discharge efficiency, and storage lifetime. Additionally, the projected load demand is given for each location on an hourly basis, as well as the distance between locations, transmission loss factor per mile, and transmission line capacity between locations.

The objective is to find the location, year, type and number of generators and storage units to install; when to retire the generators, and whether or not to extend their lifetime; the approximate power flow between locations; and the approximate operating schedule in order to meet the projected load demand while minimizing the overall operating, investment and environmental costs.

Modeling strategies are adopted to handle the problem size and its multi-scale nature. Regarding the time scale, we use *k*-means clustering approach to select representative days from historical data that best represent the load demand and renewable capacity

factor for each region (Lara et al, 2017). This reduced set of representative days with hourly resolution is multiplied by a weight to account for the entire year. For the spatial representation, the area considered is divided into regions that have similar climate and load demand profiles, and it is assumed that the potential locations are the midpoints in each region. Additionally, generators and storage units that have the same status (existing or potential) and the same technologies are grouped in clusters per region (Palmintier and Webster, 2014), such that the discrete decisions associated with generators and storage units are integer instead of binary variables, representing the number of units under a specific status in each cluster at each time period. To simplify even further the model, the transmission is determined by an energy balance between nodes, as in the "truck-route" model representation. This approximation ignores Kirchhoff's voltage law, but it is commonly used in long-term planning models that consider a large area.

3. MILP formulation and solution strategy

This section presents the proposed deterministic MILP formulation. For the sake of space, we only include in this paper the constraints that were modified or added to the original model by Lara et al. (2017). In addition to the set of regions $r \in R$, generator clusters $i \in I$, time periods (years) $t \in T$, representative days $d \in D$, and sub-periods of time (hours) $s \in S$, we also have a set of energy storage clusters $j \in J$.

The first modified constraint is the energy balance (1), which was Equation (1) in Lara et al. (2017). This constraint ensures that the sum of the instantaneous power $p_{i,r,t,d,s}$ by the generators in cluster *i*, regions *r*, plus the difference between the power flow from regions *r'* to region *r*, $p^{flow}_{r',r,t,d,s}$, and the power flow from region *r* to regions *r'*, $p^{flow}_{r,r',t,d,s}$, plus the power discharged from all the storage clusters *j* in region *r* $p^{discharge}_{j,r,t,d,s}$, equals the load demand $L_{r,t,d,s}$ at region *r*, plus the power being charged to the storage clusters *j* in region *r* $p^{charge}_{j,r,t,d,s}$, plus a slack for curtailment of renewable generation $cu_{r,t,d,s}$ at all times.

$$\sum_i p_{i,r,t,d,s} + \sum_{r' \neq r} \left(p^{\text{flow}}_{r',r,t,d,s} \cdot (1 - T^{\text{loss}}_{r,r'} \cdot D_{r,r'}) - p^{\text{flow}}_{r,r',t,d,s} \right) + \sum_j p^{\text{discharge}}_{j,r,t,d,s} = L_{r,t,d,s} + \sum_j p^{\text{charge}}_{j,r,t,d,s} + cu_{r,t,d,s} \quad \forall\, r,t,d,s \tag{1}$$

Constraints (2)-(24) from Lara et al. (2017) are also included in this formulation. The objective function (25) is modified to include a term for storage cost,

$$\min \quad \Phi = \sum_t \left(\Phi^{\text{opex}}_t + \Phi^{\text{capex}}_t + \Phi^{\text{PEN}}_t + \Phi^{\text{storage}}_t \right) \tag{25}$$

where the operating Φ^{opex}_t and investment cost Φ^{capex}_t of generators, and the penalties Φ^{PEN}_t are defined in equations (26)-(28) of Lara et al. (2017), respectively. The new term $\Phi^{storage}_t$ accounts for the costs associated with storage and is defined in (29).

$$\Phi^{\text{storage}}_t = If_t \cdot \sum_j \sum_r SIC_{j,t} \cdot Storage^{\max}_j \cdot nsb_{j,r,t} \qquad \forall\, t \tag{29}$$

Additionally, there are constraints related to the energy storage devices, which are assumed to be ideal and generic (Pozo et al., 2014). Constraints (30)-(31) compute the number of storage units that are ready to operate $nso_{j,r,t}$, taking into account the storage units already existing at the beginning of the planning horizon $Ns_{j,r}$, and the ones built

$nsb_{j,r,t}$ and retired $nsr_{j,r,t}$ at year *t*. Due to the flexibility in sizes for storage units, $nso_{j,r,t}$, $nsb_{j,r,t}$, and $nsr_{j,r,t}$ are relaxed to be continuous.

$$nso_{j,r,t} = Ns_{j,r} + nsb_{j,r,t} - nsr_{j,r,t} \qquad \forall\, j,r,t=1 \tag{30}$$

$$nso_{j,r,t} = nso_{j,r,t-1} + nsb_{j,r,t} - nsr_{j,r,t} \quad \forall\, j,r,t>1 \tag{31}$$

Constraint (32) enforces the retirement of storage units that have reached the end of their lifetime LT_j.

$$\sum_{t'' \le t - LT_j} nsb_{j,r,t''} = \sum_{t' \le t} nsr_{j,r,t'} \qquad \forall\, j,r,t \tag{32}$$

Constraints (33) and (34) establish that the power charge $p^{charge}_{j,r,t,d,s}$ and discharge $p^{discharge}_{j,r,t,d,s}$ of the storage units in cluster *j* $nso_{j,r,t}$ has to be within the operating limits $Charge^{min}_j$ and $Charge^{max}_j$, and $Discharge^{min}_j$ and $Discharge^{max}_j$, respectively.

$$Charge^{\min}_j \cdot nso_{j,r,t} \le p^{\text{charge}}_{j,r,t,d,s} \le Charge^{\max}_j \cdot nso_{j,r,t} \qquad \forall\, j,r,t,d,s \tag{33}$$

$$Discharge^{\min}_j \cdot nso_{j,r,t} \le p^{\text{discharge}}_{j,r,t,d,s} \le Discharge^{\max}_j \cdot nso_{j,r,t} \qquad \forall\, j,r,t,d,s \tag{34}$$

Constraint (35) specifies that the energy storage level for the storage units in cluster *j* $nso_{j,r,t}$ has to be within the storage capacity limits $Storage^{min}_j$ and $Storage^{max}_j$.

$$Storage^{\min}_j \cdot nso_{j,r,t} \le p^{\text{level}}_{j,r,t,d,s} \le Storage^{\max}_j \cdot nso_{j,r,t} \qquad \forall\, j,r,t,d,s \tag{35}$$

Constraints (36) and (37) show the power balance in the storage units. The state of charge $p^{level}_{j,r,t,d,s}$ at the end of hour *s* depends on the previous state of charge $p^{level}_{j,r,t,d,s-1}$, and the power charged $p^{charge}_{j,r,t,d,s}$ and discharged $p^{discharge}_{j,r,t,d,s}$ at hour *s*. The symbols η^c_j and η^d_j represent the charging and discharging efficiencies, respectively. For the first hour of the day, the previous state of charge (i.e., *s=0*) is the variable $p^{level,0}_{j,r,t,d}$.

$$p^{\text{level}}_{j,r,t,d,s} = p^{\text{level}}_{j,r,t,d,s-1} + \eta^{\text{c}}_j \cdot p^{\text{charge}}_{j,r,t,d,s} + \frac{1}{\eta^{\text{d}}_j} \cdot p^{\text{discharge}}_{j,r,t,d,s} \qquad \forall\, j,r,t,d,s>1 \tag{36}$$

$$p^{\text{level}}_{j,r,t,d,s} = p^{\text{level,0}}_{j,r,t,d} + \eta^{\text{c}}_j \cdot p^{\text{charge}}_{j,r,t,d,s} + \frac{1}{\eta^{\text{d}}_j} \cdot p^{\text{discharge}}_{j,r,t,d,s} \qquad \forall\, j,r,t,d,s=1 \tag{37}$$

Constraints (38) and (39) force the storage units to begin $p^{level,0}_{j,r,t,d}$ and end $p^{level}_{j,r,t,d,s=S}$ each day with 50% of their maximum storage $Storage^{max}_j$. This is a heuristic to attach carryover storage level form one representative day to the next (Liu et al., 2017).

$$p^{\text{level,0}}_{j,r,t,d} = 0.5 \cdot Storage^{\max}_j \cdot nso_{j,r,t} \qquad \forall\, j,r,t,d \tag{38}$$

$$p^{\text{level}}_{j,r,t,d,s} = 0.5 \cdot Storage^{\max}_j \cdot nso_{j,r,t} \qquad \forall\, j,r,t,d,s=S \tag{39}$$

The integrated planning and operations model for the generation expansion planning is then given by the multi-period MILP model defined by equations (1)-(39).

Even though this formulation incorporates the modelling strategies aforementioned, it can still be very computationally expensive to solve it depending on the size of the area, the length of the planning horizon, and the time resolution of the representative days. Therefore, we use the Nested Decomposition algorithm for deterministic multi-period problems proposed by Lara et al. (2017). This algorithm consists of decomposing the problem per time period (year) and solving Forward and Backward Passes iteratively. The

Forward Pass solves the problem for each time period sequentially, in a myopic fashion, and yields a feasible solution (upper bound). The Backward Pass solves a relaxed version of the subproblems from the last to the first time period, projecting the problem onto the subspace of the linking variables by adding cuts. These cuts provide approximations to predict the cost-to-go functions within the planning horizon, yielding a lower bound to the problem. The only difference between the algorithm framework presented by Lara et al. (2017) and the one used in this paper is that there are two additional state (i.e., linking) variables, $\boldsymbol{nso_{j,r,t}}$ and $\boldsymbol{nsb_{j,r,t}}$. For a detailed description of the algorithm, please refer to Lara et al. (2017).

4. Case study

We applied this formulation to a case study approximating the Texas Interconnection, managed by the Electric Reliability Council of Texas (ERCOT). Within the ERCOT covered area, we considered four geographical regions: Northeast, West, Coastal, and South, and also include a fifth region, Panhandle, which is outside the ERCOT limits but supplies electricity to ERCOT due to its renewable generation potential. Most of the data source information can be found in Lara et al. (2017), Section 6. In this case study we consider 3 types of utility batteries: lithium-ion, lead-acid and flow batteries, for which we use the capital cost forecast provided by Schmidt et al. (2017) and the technical information provided by Luo et al. (2015). We also consider two types of advanced fossil fuel energy systems: integrated gasification fuel cell (IGFC) (Iyengar et al. 2014) and natural gas fuel cell (NGFC) (Newby and Keairns, 2013). We assumed a carbon tax starting at \$10/tonne in year 5, and increasing linearly to \$100/tonne in year 14. The problem is solved for a 30-year planning horizon and 4 representative days per year. We implemented the monolithic formulation and the Nested Decomposition algorithm with Benders cut in Pyomo, and solved the LPs and MILPs using Gurobi version 7.0.1.

The full-space MILP model has 1,730,491 constraints, and 1,310,681 variables (810,181 continuous, and 500,500 integer variables). It takes 4.0 hours for Gurobi to solve the full-space MILP within 1% gap, while it takes 2.8 hours to solve the same problem within 1% gap using the Nested Decomposition algorithm with Benders cuts, which is a reduction of 30% in solution time.

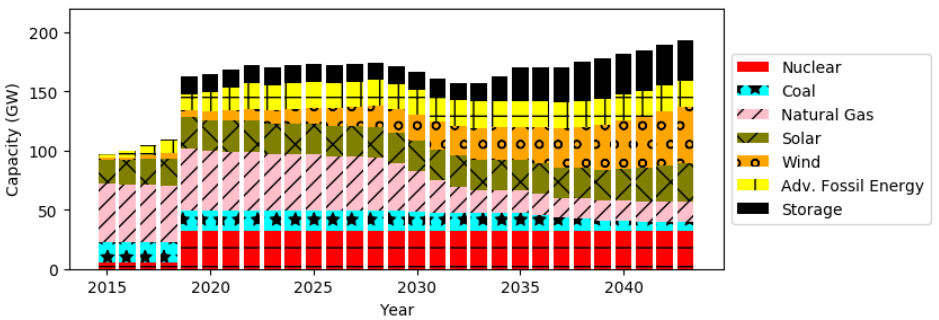


Figure 1: Generation and Storage Capacity: total ERCOT

The results displayed in Figure 1 show that most of the demand growth will be met by an increase in nuclear, solar photo-voltaic and wind capacity. It also shows that in the last year of the planning horizon there are 34.3 GW of lithium-ion batteries, and 22.0 GW of NGFC in the ERCOT region.

5. Conclusions

In this paper we propose an MILP model to solve power systems planning models considering increasing share of renewables and possibility of adding energy storage units. This deterministic formulation is applied to a case study in the ERCOT region. We show that by using the Nested Decomposition algorithm proposed by Lara et. al (2017) we are able to solve the problem more efficiently and predict which technologies show the highest likelihood of deployment and retirement under a variety of scenarios.

Acknowledgements

The authors gratefully acknowledge financial support from the Department of Energy as part of the IDAES Project.

References

A. Flores-Quiroz, R. Palma-Behnke, G. Zakeri, and R. Moreno, 2016, A column generation approach for solving generation expansion planning problems with high renewable penetration, Electric Power Systems Research, 136, 232-241.

C. F. Heuberger, E. S. Rubin, I. Staffell, N. Shah, N. M. Dowell, 2017, Power capacity expansion planning considering endogenous technology cost learning, Applied Energy, 204, 831-845.

A.K.S. Iyengar, R.A. Newby, D.L. Keairns, 2014, Techno-Economic Analysis of Integrated Gasification Fuel Cell Systems, DOE/NETL- 341/112613.

C. L. Lara, D. Mallapragada, D. Papageorgiou, A. Venkatesh, I. E. Grossmann, 2017, Electric Power Infrastructure Planning: Mixed-Integer Programming Model and Nested Decomposition Algorithm, Submitted for publication, url: http://www.optimization-online.org/DB_FILE/2017/08/6162.pdf

Y. Liu, R. Sioshansi and A. Conejo, 2017, Multistage Stochastic Investment Planning with Multiscale Representation of Uncertainties and Decisions, IEEE Transactions on Power Systems, 99, 1-1.

X. Luo, J. Wang, M. Dooner, J. Clarke, 2015, Overview of current development in electrical energy storage technologies and the application potential in power system operation, Applied Energy, 137, 511-536.

R. Newby, D. Keairns, 2013, Analysis of Natural Gas Fuel Cell Plant Configurations – Revision 1, DOE/NETL-2013/1593.

B. Palmintier, and M. Webster, 2014, Heterogeneous unit clustering for efficient operational flexibility modeling, 2014 IEEE PES General Meeting | Conference & Exposition, National Harbor, MD, 2014, 1-1.

D. Pozo, J. Contreras and E. Sauma, Unit commitment with ideal and generic energy storage units, 2016 IEEE Power and Energy Society General Meeting (PESGM), Boston, MA, 1-1.

O. Schmidt, A. Hawkes, A. Gambhir, I. Staffell, 2017, The future cost of electrical energy storage based on experience rates, Nature Energy, 2, 1-8.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50214-7

Optimal Front-end Crude Schedule for Refineries under Consideration of Inherent Upset Reduction

Honglin Qu[a], Jialin Xu[a], Sujing Wang[b], Qiang Xu[a*]

[a]*Dan F. Smith Department of Chemical Engineering, Lamar University, Beaumont, TX, 77710, USA*
[b]*Department of Computer Science, Lamar University, Beaumont, TX, 77710, USA*
Qiang.xu @lamar.edu

Abstract

This paper addresses the inherent upset (IU) issue, for instance, the flowrate fluctuation after switching charging tanks (CTs) for feeding crude distillation units (CDUs), during routine operations of crude supply process. A new MILP model with continuous time formulation is developed for the front-end crude scheduling (FECS) under consideration of inherent upset reduction (IUR). The optimal crude supply schedule can be determined via minimizing the overall operating cost and reducing IUs. Thus, both the process profitability and the operational stability of optimal schedules are guaranteed. In addition, the trans-mixing between adjacent crude slots inside LDPL is considered; and the inventory cost is precisely modeled compared with previous studies. All nonlinearities arising from the original mathematical problem are approached through piecewise linearization. The model efficacy and the IUR effect have been investigated through a refinery case study with a daily production of 160 kbbl.

Keywords: Front-end crude scheduling, Inherent upset reduction, Trans-mixing, MILP

1. Introduction

In recent decades, the front-end crude scheduling (FECS) has drawn considerable attentions from both academia and industry community. The entire scope of the FECS is concisely illustrated in Figure 1. In general, the FECS for inland refineries is to determine crude supply from vessel unloading through berths, temporarily storing in port-side storage tanks (STs), crude transferring via long-distance pipeline (LDPL), blending in refinery-side charging tanks (CTs) and continuously feeding to crude distillation units (CDUs) for processing. Significant profitability and operability benefits will be accomplished if the FECS could be optimally and timely performed. Therefore, lots of research efforts have been paid on the FECS (Li et al. 2012). It should be noted that inherent upsets (IUs) are inevitably generated during routine operations. For instance, the switch among CTs to CDUs may cause upsets in terms of flowrate and compositions. These IUs can affect local unit operations, also accumulate and propagate to downstream. In previous studies, flowrate IUs of feeding one CDU could be as high as 10 times. Conceivably, the implementation of such a schedule will cause huge losses and serious consequences, even shutdown to the entire refining plant. By now, most research works studying the FECS exclusively focused on the optimization of operating cost and gross margin (Xu et al. 2017a,b); few tackled IUs and their consequences. In this study, this tough concern has been taken into account along with other requirements, and simultaneously modeled and solved by the FECS. Thus, IUs can be reduced so as to help the smooth crude supply and stable CDU operations.

In this paper, a new MILP model with continuous-time formulation is developed for the FECS under consideration of inherent upset reduction (IUR). It is to explore the optimal crude supply schedule via minimizing the overall operating cost and reducing flowrate IUs of feeding CDUs. Thus, both the process profitability and the operational stability of the optimal front-end crude schedule can be guaranteed. In addition, the trans-mixing (TM) between adjacent crude slots inside LDPL is considered; and the inventory cost (IC) is also accurately modeled compared with previous studies. All nonlinearities arising from the original mathematical problem are approached through piecewise linearization. The model efficacy and the IUR effect have been investigated through a refinery case study with a daily production of 160 kbbl (1 kbbl=10^3 bbl).

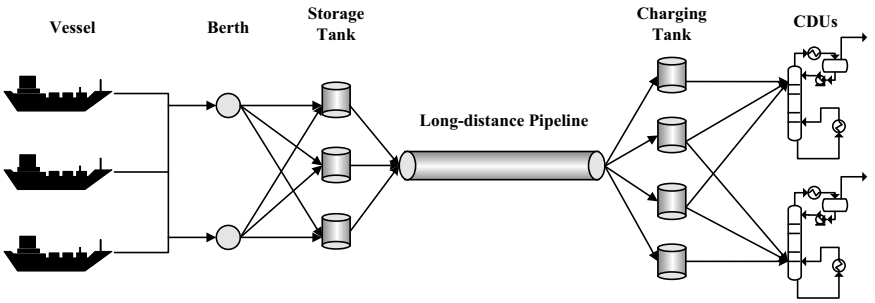


Figure 1. The entire scope of front-end crude scheduling for inland refineries.

2. Problem Statement

As shown in Figure 1, the studied problem includes three single-parcel vessels, two berths, three STs, one LDPL, four CTs and two CDUs. Arrows represent feasible crude transfer operations. Note that feeding operations of LDPL and CDUs are continuous during the entire time horizon. The objective is to determine the best crude supply schedule with minimal operating cost and reduced IUs. The assumption, given information, information to be determined and operating rules are summarized below.

Assumption: (1) no time cost for feed switching operation; (2) perfect blending within STs and CTs, and (3) fixed volume and ideal mixing for TM;

Given information: (1) capacity and initial inventory of vessels, STs, LDPL and CTs; (2) vessel arrival time; (3) allowable IU ranges for CDUs, (4) crude oil properties; (5) blending specs of CTs; (6) FECS time horizon; and (7) economic data;

Information to be determined: within the entire time horizon, (1) the detailed schedule of crude transfer from vessels, through berths, STs, LDPL and CTs, to CDUs; and (2) the inventory profile of the LDPL;

Operating rules: at one time event, (1) any unit can feed at most one downstream unit; (2) any unit except STs can be charged by at most one upstream unit; (3) STs can be charged by any available berths; (4) any STs and CTs cannot simultaneously receive and charge crudes; and (5) STs and CTs need BST right after receiving new crudes.

3. The FECS Model Extension

The FECS model is developed based on our previous work (Zhang and Xu, 2015). Note that the previous work mentioned below refers to this one. For the conciseness purpose,

only the newly developed constraints are provided in this section; other model information can be consulted from the previous work.

3.1. Reformulation of The Crude Transfer via LDPL

In this study, each feeding ST and charged CT is aligned with corresponding crude slots. Thus, unit set is enlarged to "*parcel*, *ST*, *L*, *CT*, *CDU*", where *L* denotes the set of slots and other elements represent the set of corresponding units. In this work, the crude transfer via LDPL is modeled as continuous within the entire time horizon. Hence, Eqs. (1) and (2) are employed.

$$\sum_{u\in ST}\sum_{l\in L^{new}}\sum_{n\in N, n\geq 1}(Te(u,l,n)-Ts(u,l,n))=H \tag{1}$$

$$\sum_{l\in L}\sum_{u\in CT}\sum_{n\in N, n\geq 1}(Te(l,u,n)-Ts(l,u,n))=H \tag{2}$$

where *Ts(u,u',n)* and *Te(u,u',n)* respectively represent the starting and ending time of crude transfer from unit *u* to unit *u'* at time event *n*; and L^{new} denotes new slots that are fed after the time horizon starts.

Due to the TM, the crude volume and its fraction of outlet slots are different with those when slots are injected. Based on the input volume *Vlc(l,c)* and volume fraction *flc(l,c)* of crude *c* in slot *l*, the discharged crude volume in slot *l*, *Vlcd(l,c)*, can be calculated by Eqs. (3) through (5), where *TV* refers to the TM volume of interchanged crudes at each interface. Note that the first and last slots have single interface to share with adjacent slots, while others have two. Thus, different formulations are resulted.

$$Vlcd(l,c)=Vlc(l,c)+TV\,(\,flc(l+1,c)-flc(l,c)\,),\quad \forall l\in L, l=1; c\in C \tag{3}$$

$$\begin{aligned}&Vlcd(l,c)=Vlc(l,c)+TV\,(\,flc(l+1,c)-2\,flc(l,c)+flc(l-1,c)\,),\\&\forall l\in L, 1<l<|L|; c\in C\end{aligned} \tag{4}$$

$$Vlcd(l,c)=Vlc(l,c)+TV\,(\,flc(l-1,c)-flc(l,c)\,),\quad \forall l\in L, l=|L|; c\in C \tag{5}$$

3.2. Inherent Upset Reduction for Feeding CDUs

To accomplish IUR for feeding CDUs, new constraints on the feeding rate are added based on the flexibility of the refinery operation processing capacity. Specifically, the IU is strictly constrained within 5% of the normal flowrate. This specification is enabled by Eqs. (6) and (7). Binary variable *X(u,u',n)* is used to represent the crude transfer operation from unit *u* to unit *u'* at time event *n*.

$$\begin{aligned}&V(u,u',n)\leq FRup\,(Te(u,u',n)-Ts(u,u',n))+M\,(1-X(u,u',n)),\\&\forall u\in CT; u'\in CDU; n\in N, n\geq 1\end{aligned} \tag{6}$$

$$\begin{aligned}&V(u,u',n)\geq FRlo\,(Te(u,u',n)-Ts(u,u',n))-M\,(1-X(u,u',n)),\\&\forall u\in CT; u'\in CDU; n\in N, n\geq 1\end{aligned} \tag{7}$$

where *V(u,u',n)* denotes the transferred crude volume; *FRup* and *FRlo* define the IU range of feeding CDUs; and *M* is for Big-M relaxation.

3.3. Reformulation of Inventory Cost (IC) in Objective Function

Unlike the approximation of IC in most studies, it is accurately modeled using the trapezoid area method following Eqs. (8) through (10). The IU is initially evaluated based on $Inv^0(u)\cdot H$; after the time horizon starts, for each input and output operation, extra cost will be "charged" and "credited" based on *Tin* and *Tot*, respectively. The accurate IU is determined by combing all these elements.

The reformulated IC expressions and crude blending equations contribute nonlinearities of this problem (Castro et al. 2014). In this study, these nonlinearities are approached following the effective piecewise linearization formulation in the previous work.

$$IC = \sum_{u\in ST\cup CT} C^{inv}(u)\left(Inv^0(u)\,H + \sum_{u'\in IU_u}\sum_{n\in N, n\geq 1} Tin(u',u,n) - \sum_{u''\in OU_u}\sum_{n\in N, n\geq 1} Tot(u,u'',n)\right) \tag{8}$$

$$Tin(u',u,n) = V(u',u,n)\,(H - 0.5\,(Te(u',u,n) + Ts(u',u,n))),\quad \forall u\in ST\bigcup CT; u'\in IU_u; n\in N, n\geq 1 \tag{9}$$

$$Tot(u,u'',n) = V(u,u'',n)\,(H - 0.5\,(Te(u,u'',n) + Ts(u,u'',n))),\quad \forall u\in ST\bigcup CT; u''\in OU_u; n\in N, n\geq 1 \tag{10}$$

where $C^{inv}(u)$ and $Inv^0(u)$ represent the IC coefficient and initial inventory of unit u; IU_u and OU_u denote the sets of possible input and output units of unit u.

Table 1. Problem data for charging tanks, flowrate range and economic cost

Charging tanks ID	Capacity (kbbl)	Charging Tanks: Initial Inventory		Spec on Key Component
		Amount (kbbl)	Composition	
CT1	[0, 1000]	300	100% C7	[0.025, 0.035]
CT2	[0, 1000]	500	100% C5	[0.045, 0.065]
CT3	[0, 1000]	300	100% C6	[0.06, 0.085]
CT4	[0, 1000]	500	100% C2	[0.065, 0.09]
Flowrate Range: (kbbl/day)				
Crude unloading	[0, 500]	Inter-tank transfer		[30, 35]
Feeding CDU	[78, 82]			
Economic Cost				
Demurrage ($/day)	30,000	ST inventory ($/day·kbbl)		40
Unloading ($/day)	30,000	CT inventory ($/day·kbbl)		80
Setup ($)	55,000			

4. Case Study

The developed FECS model is applied to a refinery with a daily production of 160 kbbl as shown in Figure 1. The normal daily flowrate of each CDU is 80 kbbl, thus the IU range is set as 78-82 kbbl per day. As comparison, rough flowrate bounds in the previous work are applied to the model without IUR. Parts of updated problem data are provided in Table 1. The time horizon in this case study is set as 12 days.

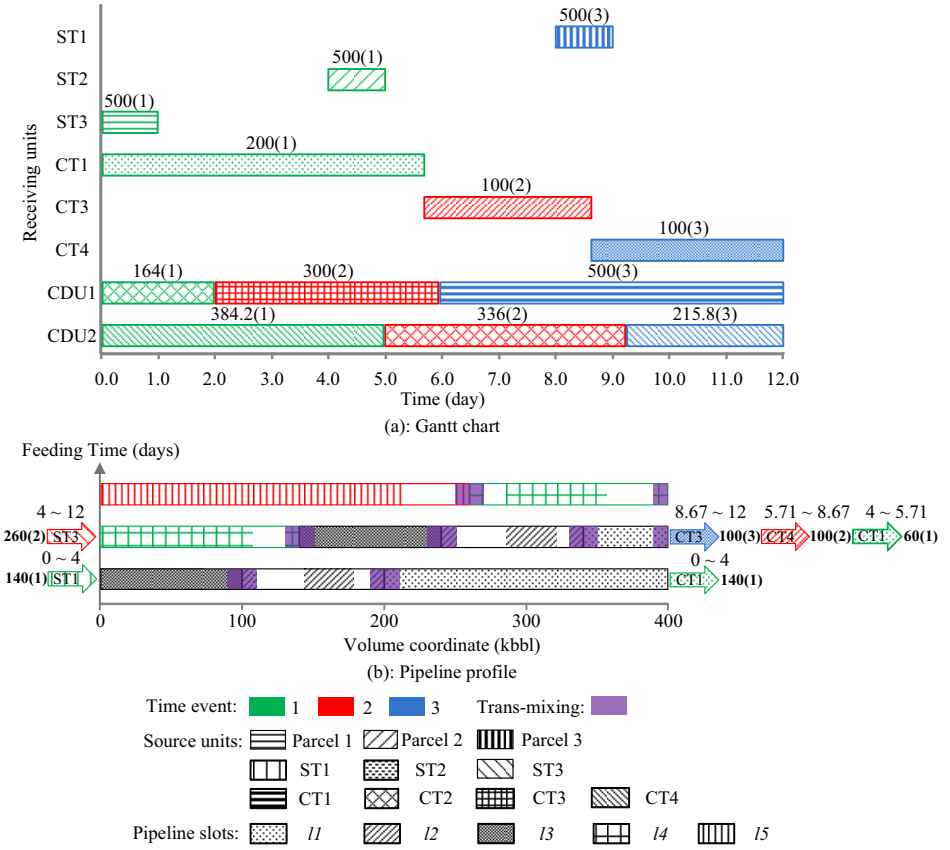


Figure 2. Gantt chart and pipeline profile for case study.

Through the developed model with IUR, the optimal crude supply schedule is successfully obtained. The optimal objective value is $2.08 million. The Gantt chart of optimal schedule and the pipeline profile are shown in Figure 2. With respect to receiving units, crude transfer operations are demonstrated in the Gantt chart. The timing can be read from x-axis, the volume is labelled nearby which is followed by the corresponding time event sequence, and the source is indicated using different filling patterns. As shown, all parcels are unloaded to STs as soon as vessels arrive at the port. The LDPL is continuously fed by STs, so that it keeps feeding CTs. The two CDUs are also continuously charged during the entire time horizon. Through the pipeline profile, information on charging of new slots, LDPL inventories and discharging of existing slots can be collected. TM between adjacent crude slots is marked at slot interface. ST1 feeds 140 kbbl of new crude into LDPL in 4 days, which forms slot *l4*. In the meantime, 140 kbbl of *l1* is charged into CT1. In the rest 8 days, ST3 feeds 260 kbbl of crude into LDPL, generating slot *l5*. From the LDPL outlet, the left 60 kbbl of *l1* is continuously charged into CT1 till 5.71 days; then *l2* is transferred to CT4 from 5.71 to 8.67 days. After that, *l3* is charged to CT3 for the rest 3.33 days. To demonstrate the effect of IUR, comparisons on feeding rate and IUs of CDUs from models with and without IUR are provided in Table 2. Through the model without IUR, all IUs are higher than 15% and it can be as high as -199%, which is for CDU2. Such huge IUs are not acceptable for any CDUs. On the other hand, through the model with IUR, IUs of both CDUs are strictly

kept within 5%. In fact, the highest IU is only 2.5%. From the operational stability point of view, the optimal schedule through the model with IUR is much better than that from the model without IUR.

Table 2. Comparison results of flowrate and IU

		Model without IUR				Model with IUR			
		Time events			Avg.*	Time events			Normal
		1	2	3		1	2	3	
CDU1	Flowrate	142.9	64.1	50	75.4	82	78.7	80.8	80
	IU (%)	89.5	-15	-34	--	2.5	-1.6	1	--
CDU2	Flowrate	247.5	50	--	82.9	78	78	78	80
	IU (%)	-199	-40	--	--	-2.5	-2.5	-2.5	--

The unit of flowrate is kbbl/day; the "avg." represents the average flowrate of feeding CDUs within the entire time horizon and it's calculated using the total crude volume divided by the time horizon.

5. Concluding Remarks

In this study, a new MILP model for optimal FECS is developed considering IUR. Through this model, the optimal crude supply schedule can be obtained via minimizing the overall operating cost and reducing IUs of feeding CDUs. Meanwhile, the TM inside LDPL is considered and the IC is also precisely modeled. The model efficacy has been demonstrated through a case study. IUs of feeding CDUs are strictly constrained within 5% of the normal flowrate, which is a significant improvement as compared with results through the model without IUR. Therefore, both the process profitability and the operational stability can be greatly enhanced through our development.

Acknowledgements

This work was supported in part by Graduate Student Scholarship and Anita Riddle Faculty Fellowship from Lamar University, United States.

References

J. Li, R. Misener, C.A. Floudas, 2012, Continuous-time modeling and global optimization approach for scheduling of crude oil operations, AIChE J., 58, 1, 205-226.

JL. Xu, SJ. Zhang, J. Zhang, SJ. Wang, Q. Xu, 2017a, Simultaneous scheduling of front-end crude transfer and refinery processing, Comp. & Chem. Eng., 96, 212-236

JL. Xu, HL. Qu, SJ. Wang, Q. Xu, 2017b, A New Proactive Scheduling Methodology for Front-End Crude Oil and Refinery Operations under Uncertainty of Shipping Delay, Ind. Eng. Chem. Res., 2017, 56, 28, 8041-8053.

P.M. Castro, I.E. Grossmann, 2014, Global Optimal Scheduling of Crude Oil Blending Operations with RTN Continuous-time and Multiparametric Disaggregation, Ind. Eng. Chem. Res., 53, 39, 15127-15145.

SJ. Zhang, Q. Xu, 2015, Reninery continuous-time crude scheduling with consideration of long-distance pipeline transportation, Comp. & Chem. Eng., 75, 74-94.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50215-9

A General Mixed-Integer Programming State-Space Model for Online Scheduling

Dhruv Gupta, Christos T. Maravelias*

University of Wiscosin-Madison, 1415 Engineering Drive, Madison, WI-53706 USA
christos.maravelias@wisc.edu

Abstract

We present a framework and a generalized state-space model formulation that allows modeling several features that are necessary when a scheduling model is used in an online setting. These features are: (1) task-delays and unit breakdowns (2) robust scheduling through the use of conservative yield estimates and processing times; (3) feedback on task-yield estimates before the task finishes; (4) task termination during its execution; and (5) unit capacity degradation and maintenance. Further, we propose a systematic scheme for updating the state of the process, as well as ability to modify states through disturbance parameters, based on feedback information.

Keywords: uncertainty, rescheduling, model predictive control, bio-manufacturing

1. Introduction

Scheduling is a key operations planning function in any enterprise. Thus, much work has been accomplished in building accurate models and obtaining high quality solutions in a reasonable time-frame. However, due to disruptions or arrival of new information, incumbent schedules can become suboptimal or even infeasible. Hence, the need for rescheduling (Li and Ierapetritou, 2008; Cui and Engell, 2010). Online scheduling is the generalization of rescheduling as an ongoing process (Gupta et al., 2016). It is now gaining increased traction in the industry. However, scheduling models as have been developed till now, require ad-hoc constraints which need to be modified on the fly, so as to represent any disturbances to the schedule. Thus, there is an important need to develop scheduling models with their application in an online setting in mind.

State-space based models, popular in process control, are a natural choice for online scheduling (Subramanian et al., 2012; Nie et al., 2014). There are six advantages of state-space models in scheduling over the conventional models: (1) the status of the plant can be interpreted from states at that moment in time; (2) it is easier to carry-over information from one iteration to next (online "ready"); (3) schedule disruptions can be conveniently modeled as linear additive disturbances; (4) an identical model (structure) is maintained in each online iteration, i.e., no ad-hoc constraints; (5) enables a natural way to integrate scheduling and control; (6) a tool for theoretical analysis to establish closed-loop properties.

2. Online scheduling framework

The standard discrete-time state-space models as popular in process control approaches have the following general form for the state-evolution and input-state constraints:

$$x(t+1) = Ax(t) + Bu(t) + B_d d(t) \quad (1)$$

$$E_x x(t) + E_u u(t) + E_d d(t) \leq 0 \quad (2)$$

where, A, B, and B_d are the state-space matrices, t is the index for time, and E_x, E_u, and E_d are the coefficient matrices of the states, inputs, and disturbances, respectively.

At any given time t, the scheduling decisions, or in other words the inputs, are whether to start a new task, the batch-size of this new task, and the amount of material to ship, if any. The state of the plant comprises of the amount of inventory and backlog of materials, the material in processing, and the resources engaged/available at that time. Typical disturbances are material demand, task delays, unit breakdowns, and material handling losses. Under nominal operation, there are no disturbances, and the inputs are all zero, maintaining the states at a steady value (Subramanian et al., 2012).

In online scheduling, a schedule for next H hours, is computed at every time-point t, but only the first scheduling decisions/inputs are implemented. The state of the plant at the next time-point results from a combination of the injected input and any disturbances. If we reset the time-grid in each online iteration, but label each iteration with index σ, we can still uniquely keep track of all scheduling inputs, states, and disturbances.

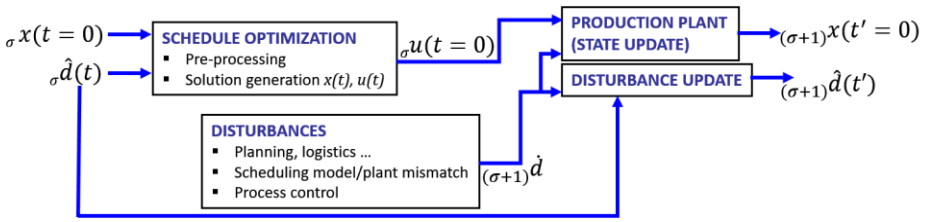


Figure 1: Online scheduling framework based on the state-space model.

In Fig. 1, we show one generic online (open-loop) iteration along with the plant state update step. To limit the scope of complexity in our work, we make a reasonable assumption that all scheduling-related plant states are observable and the scheduling disturbances are measurable. These scheduling disturbances originate due to changes in the planning and control level or due to scheduling model–plant mismatch. In iteration σ, a schedule $(x(t), u(t))$ is generated through the optimization model based on the initial state of the plant $_{\sigma}x(t=0)$ and the disturbances $_{\sigma}\hat{d}(t)$ as data. However, only the first input, $_{\sigma}u(t=0)$, is injected into the plant. Based on this input, and new materialized disturbances, $_{\sigma}\dot{d}$, states evolves to $_{\sigma+1}x(t'=0)$, where t' is the index for the new (reset) time-grid. In addition, as a separate computation, disturbances are updated to $_{\sigma+1}\hat{d}(t')$ based on previously known $_{\sigma}\hat{d}(t')$ and the newly observed disturbances $_{\sigma}\dot{d}$. With the new states and disturbances, the next online scheduling iteration is then carried out.

3. General state-space model

There are several novel features that are present in online scheduling, and not in offline scheduling. For example, terminating a task after it has started but before its completion. This is in particular common in bio-manufacturing. This decision to terminate a batch is typically taken when there is significant parallel growth of toxins along with the molecule of interest, rendering purification in downstream stages infeasible or

uneconomical (Martagan et al., 2016). Other common features are partial or complete yield-losses, and unit capacity degradation (e.g., due to gradual residue formation or sudden blockage) which requires maintenance. These features haven't been addressed in the literature. We model all the above features in our state-space model.

We use the State-Task Network (STN) framework to represent the different elements of the scheduling problem. It primarily comprises of tasks $i \in \boldsymbol{I}$, units $j \in \boldsymbol{J}$, and materials $k \in \boldsymbol{K}$. The processing time of task i, when executed on unit j, is denoted by τ_{ij}. Task i consumes/produces material k equivalent to $\rho_{ik}/\bar{\rho}_{ik}$ mass fraction of its batch-size (ρ_{ik}<0) for consumption and $\bar{\rho}_{ik} > 0$ for production). The subset of tasks that can be carried out on unit j are denoted by $\boldsymbol{I}_j$. An explanation of every variable and equation in the model is not possible within the limited space of this manuscript. Hence, here we present and briefly describe only selected components of the state-space model.

3.1. Disturbances

Single-period delays in task i running on unit j, which has progressed by n periods, are denoted by $\dot{Y}_{ij}^{n}$. Breakdown of unit j, while executing task i is denoted by $\dot{Z}_{ij}^{n}$. Yield-loss in a task is denoted by $\dot{\lambda}_{ij}^{n}$. Sudden partial loss in capacity of unit j to carry out task i is denoted by $\dot{\mu}_{ij}$. When delays span multiple periods, then disturbance parameters $\hat{Y}_{ijt}^{n}$ are needed that represent this extended delay information and, hence, help in correctly predicting when a task will end. Among others, these multi-period disturbance parameters require an update from one iteration to the next. For example, in iteration σ, for a ϕ period delay, ${}_{\sigma}\hat{Y}_{ijt}^{n} = 1$ for $t = \{0,1,2,\dots\phi - 2\}$. This delay decrements by one in the next iteration because the two iterations are spaced by 1-period. This combined with the time-grid shift, implies that now in iteration $\sigma + 1$, ${}_{\sigma+1}\hat{Y}_{ijt'}^{n} = 1$ for $t' = \{0,1,2,\dots\phi - 3\}$ and zero otherwise.

3.2. Inputs

The decision variables (inputs) include W_{ijt} which, when 1, denotes task i starts on unit j at time t; B_{ijt} the corresponding batch-size; T_{ij}^{n} which, when 1, implies that task i, on unit j, with progress index n, is being terminated; V_{kt}, the outgoing shipment of material k at time t. Certain inputs, such as, the task start binary variables W_{ijt}, have a lagged effect on future states. For example, these variables increase the inventory of material being produced after τ_{ij} periods, i.e., on task finish. This makes certain states $x(t)$ dependent on $u(t-\tau)$. This, however, does not conform to the standard state-space formulation (Eq. 1). Hence, these inputs are lifted forward and augmented to the future states, and the state-space is modified appropriately. For task start binary, new states $\bar{W}_{ijt}^{n}$ are defined which denote that task i, on unit j, has progressed by n periods at time t. When a task starts $\bar{W}_{ijt}^{n=0} = 1$, and when a task finishes $\bar{W}_{ijt}^{n=\tau_{ij}} = 1$. In the absence of any disturbances, this lifting is achieved through the following two constraints: $\bar{W}_{ijt}^{n=0} = W_{ijt}\ \forall\ i,j \in \boldsymbol{I}_j, t$ and $\bar{W}_{ij(t+1)}^{n} = \bar{W}_{ijt}^{n-1}\ \forall\ i,j \in \boldsymbol{I}_j, t, n \in \{1,2,\dots,\tau_{ij}\}$.

The lifting equations, for task-start binary, in the presence of disturbances are:

$$X_{ij(t+1)} = \hat{Y}^0_{ijt}\left(1 - T^0_{ij}\right) \forall j, i \in \boldsymbol{I}_j, t \tag{3}$$
$$\bar{W}^0_{ijt} = W_{ijt} + X_{ijt} \forall j, i \in \boldsymbol{I}_j, t \tag{4}$$
$$\bar{W}^n_{ij(t+1)} = \bar{W}^{n-1}_{ijt} - \hat{Y}^{n-1}_{ijt}\left(1 - T^{n-1}_{ij}\right) + \hat{Y}^n_{ijt}\left(1 - T^n_{ij}\right) \forall j, i \in \boldsymbol{I}_j, n \in \{1,2,\dots,\tau_{ij}\} \tag{5}$$

where, X_{ijt} is a dummy state defined to carry forward delays when task has zero progress, i.e., $n = 0$ (see Eq. 9).

3.3. States

The state of the plant is denoted by variables described as follows. $\bar{W}^n_{ijt} \in \{0,1\}$ and $\bar{B}^n_{ijt}$ are the lifted task-states which denote the progress status n and corresponding batch-size of the task. For example, if $\bar{W}^n_{ijt} = 1$, it implies that task i has progressed by n h on unit j at time t; its batch-size is $\bar{B}^n_{ijt}$. $\bar{L}^n_{ijt}$ is a lifted state that denotes the anticipated yield-loss as of time t in task i. S_{kt} and BO_{kt} represent the inventory and backlog level of material k in time-period $(t-1, t]$, while C_{ijt} is the capacity of the unit to perform task i on unit j at time t. With use, the capacity of the unit can degrade, but can then be restored through a maintenance task. ${}^{L}X_{ijt}$ denotes the yield-loss in delayed tasks with progress $n = 0$.

3.4. State update

From one iteration to the next, the plant states, such as, inventory and unit capacity are updated through Eqs. 6 and 7. Please note the use of $\dot{\mu}_{ij}$ disturbance parameter in Eq. 7. Further, the task-states, including yield-loss states, also require an update, to carry-over information about tasks already running. However, these update steps are softened, to allow task termination, and made part of the scheduling model (Eqs. 8-11).

$${}_{\sigma}S_{k(t=0)} = {}_{\sigma-1}S_{k(t'=1)} \forall k \tag{6}$$
$${}_{\sigma}C_{ij(t=0)} = {}_{\sigma-1}C_{ij(t'=1)} + \dot{\mu}_{ij} \forall j \in \boldsymbol{J}_{MT}, i \in \boldsymbol{I}_j \tag{7}$$
$${}_{\sigma}\bar{W}^n_{ij(t=0)} = ({}_{(\sigma-1)}\bar{W}^{n-1}_{ij(t'=0)} - \dot{Y}^{n-1}_{ij})(1 - T^{n-1}_{ij}) + \dot{Y}^n_{ij}(1 - T^n_{ij}) - \dot{Z}^{n-1}_{ij} \forall j, i \in \boldsymbol{I}_j, n \in \{1,2,\dots,\tau_{ij}\} \tag{8}$$
$${}_{\sigma}X_{ij(t=0)} = \dot{Y}^0_{ij}\left(1 - T^0_{ij}\right) \forall j, i \in \boldsymbol{I}_j \tag{9}$$
$${}_{\sigma}\bar{L}^n_{ij(t=0)} = ({}_{(\sigma-1)}\bar{L}^{n-1}_{ij(t'=0)} + \dot{\lambda}^{n-1}_{ij} - {}^{L}\dot{Y}^{n-1}_{ij})\left(1 - T^{n-1}_{ij}\right) + {}^{L}\dot{Y}^n_{ij}(1 - T^n_{ij}) - {}^{L}\dot{Z}^{n-1}_{ij} \forall j, i \in \boldsymbol{I}_j, n \in \{1,2,\dots,\tau_{ij}\} \tag{10}$$
$${}^{L}_{\sigma}X_{ij(t=0)} = {}^{L}\dot{Y}^0_{ij}\left(1 - T^0_{ij}\right) \forall j, i \in \boldsymbol{I}_j \tag{11}$$

3.5. State evolution

One of the state evolution equations, the inventory balance, is as follows:

$$S_{k(t+1)} = S_{kt} + \sum_j \sum_{i \in \boldsymbol{I}_j \cap \boldsymbol{I}^+_k} \theta^{\bar{\rho}}_{ikt} \left(\bar{B}^{\theta^{\tau}_{ijt}}_{ijt} - L^{\theta^{\tau}_{ijt}}_{ijt}\right) + \sum_j \sum_{i \in \boldsymbol{I}_j \cap \boldsymbol{I}^+_k} \rho_{ik} B_{ijt} - V_{kt} + \zeta_{kt} \; \forall j, i \in \boldsymbol{I}_j \tag{12}$$

where, $\theta^{\bar{\rho}}_{ikt} = \begin{cases} \bar{\rho}_{ik} \; \forall k, t = \{0\}; \\ \bar{\rho}^r_{ik} \forall k, t = \{1,2,\dots\} \end{cases}$ and $\theta^{\tau}_{ijt} = \begin{cases} \tau_{ij} \; \forall j, i \in \boldsymbol{I}_j, t = \{0\}; \\ \tau^r_{ij} \forall j, i \in \boldsymbol{I}_j, t = \{1,2,\dots\} \end{cases}$

Three important aspects in Eq. 12, are (1) $\bar{\rho}_{ik}^{r} < \bar{\rho}_{ik}$, hence, the use of $\theta_{ikt}^{\bar{\rho}}$, implies that batches bigger than the nominal size are started anticipating lower mass conversion than average, however, for tasks finishing at $t = 0$, the average mass conversion coefficient is used; (2) lifted yield-loss states ($\bar{L}_{ijt}^{\theta_{ijt}^{\tau}}$) are subtracted from $\bar{B}_{ijt}^{\theta_{ijt}^{\tau}}$ to reflect the actual yield; (3) finally, $\tau_{ij}^{r} > \tau_{ij}$, hence, tasks not finishing at $t = 0$ are anticipated to have a conservative processing time, but at $t = 0$ tasks finish at their average processing times plus any observed actual delays. Hence, we are able to carry out robust scheduling, but correctly account for feedback with regards to late/on-time task finish and actual yield-loss.

3.6. Computational remarks

Discrete time-grid models have been shown to scale well, especially, with the complexities of a larger system (Sundaramoorthy and Maravelias, 2011). When the model is solved, all the lifted variables and lifting equations are typically eliminated by preprocessing (Subramanian et al., 2012). Compared to traditional state-space model, for the generalizations, only a small number of variables are added to enable new features. For example, for maintenance, we only add one binary variable per unit per time. Through computational studies we see that the state-space model is comparable in performance to non state-space scheduling models (data not shown due to space limitation).

4. Case study

We present a simplified case study, from the field of bio-manufacturing, to illustrate the capabilities of our general state-space model, and present the results in brief due to space limitations. The medium size example we have chosen is typical of that found in a biomanufacturing setup and is of great interest to practitioners. For each open-loop iteration, we could solve our model to optimality in less than 0.1 seconds using CPLEX 12.6.1 (default options) via GAMS 24.4.3 on an Intel Xeon (E5520, 2.27GHz, 8 core) machine with 16 GB RAM and Linux CentOS 7 operating system. In the network in Fig. 2, material M3 is a desired pharmaceutical ingredient. As features, we allow for possible delays in the cell culture preparation (T1), and yield losses in the bio-reaction (T2). Thus, we carry out robust scheduling, using a conservative processing time (τ^{r}) for task T1 and a conservative mass-conversion coefficient ($\bar{\rho}^{r}$), against small yield losses, for task T2. Further, in the downstream purification stage, the chromatograph column (U3) loses capacity with usage. Executing a maintenance task (MT) restores capacity on the chromatographs. T4 is a dummy task to model storage of material M2 inside unit U2. The general state-space model we have developed enables us to address these new features in online scheduling (see Fig. 2).

5. Conclusions

We developed a general state-space model, particularly motivated by an online scheduling perspective, that allows modeling (1) task-delays and unit breakdowns (2) robust scheduling through the use of conservative yield estimates and processing times; (3) feedback on task-yield estimates before the task finishes; (4) task termination during its execution; and (5) unit capacity degradation and maintenance. Further, we propose a systematic scheme for updating the state of the process, as well as ability to modify

states through disturbance parameters, based on feedback information. These general features that we address are found in several industrial sectors, namely, pharmaceuticals, fine chemicals, pulp and paper, agriculture, steel production, oil and gas, food processing, bio-manufacturing, etc. Although, here we presented the model using STN based representation, these generalizations can also be adapted to RTN based representation.

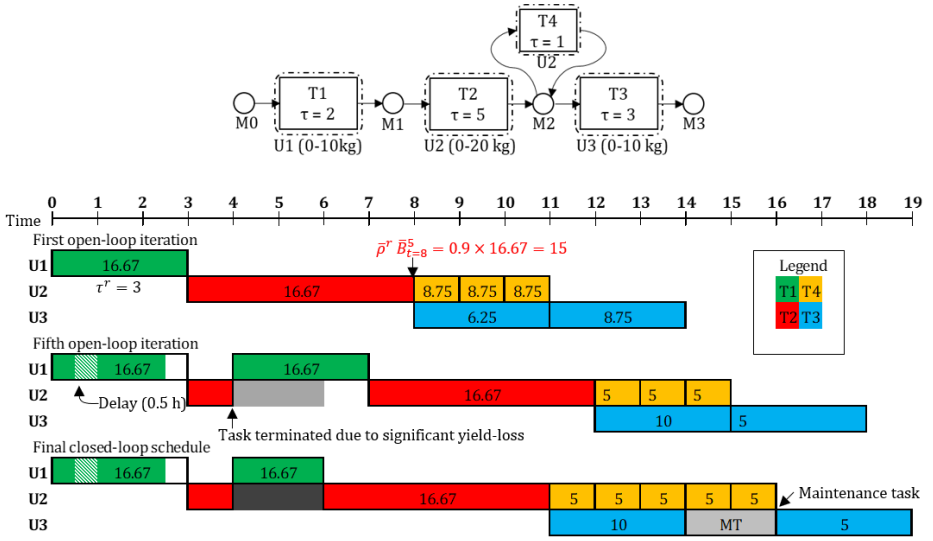


Figure 2: The network, the first and fifth open-loop iterations, and the final closed-loop schedule. The final closed-loop schedule is very different from the first open-loop schedule, due to observation of several disturbances and counter-actions. For example, in the fifth iteration, T2 is terminated due to observation of a significant yield loss.

References

Cui, J., Engell, S., 2010. Medium-term planning of a multiproduct batch plant under evolving multi-period multi-uncertainty by means of a moving horizon strategy. Computers and Chemical Engineering 34 (5), 598–619.

Gupta, D., Maravelias, C. T., Wassick, J. M., 2016. From rescheduling to online scheduling. Chemical Engineering Research and Design 116, 83–97.

Li, Z., Ierapetritou, M. G., 2008. Process scheduling under uncertainty: Review and challenges. Computers and Chemical Engineering 32 (4-5), 715–727.

Martagan, T., Krishnamurthy, A., Maravelias, C. T., 2016. Optimal condition-based harvesting policies for biomanufacturing operations with failure risks. IIE Transactions 48 (5), 440–461.

Nie, Y., Biegler, L. T., Wassick, J. M., Villa, C. M., 2014. Extended discrete-time resource task network formulation for the reactive scheduling of a mixed batch/continuous process. Industrial and Engineering Chemistry Research 53 (44), 17112–17123.

Subramanian, K., Maravelias, C. T., Rawlings, J. B., 2012. A state-space model for chemical production scheduling. Computers & Chemical Engineering 47, 97–110.

Sundaramoorthy, A., Maravelias, C. T., 2011. Computational Study of Network-Based Mixed-Integer Programming Approaches for Chemical Production Scheduling. Industrial and Engineering Chemistry Research 50, 5023–5040.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50216-0

A Novel Metaheuristic Framework for the Scheduling of Multipurpose Batch Plants

Matthew Woolway[a], Thokozani Majozi[b*]

[a]*School of Computer Science and Applied Mathematics, University of the Witwatersrand, Johannesburg, Private Bag 3, Wits 2050, South Africa*
[b]*School of Chemical and Metallurgical Engineering, University of the Witwatersrand, Johannesburg, Private Bag 3, Wits 2050, South Africa*
thokozani.majozi@wits.ac.za

Abstract

A genetic algorithm (GA) is proposed along with a novel general framework for the scheduling of a typical multipurpose/product batch plant. The majority of literature regarding these problems make use of mathematical programming methods. Modelling problems in this manner leads to numerous binary variables relating to material balance and sequence of batches along long-time horizons, thus resulting in large computational time. The proposed GA does not suffer the same scalability issues of mathematical programming approaches. The GA makes use of a coupled chromosome system with specific crossover and mutation functions utilised with the purpose of profit maximisation. Results show that optimal or close-to-optimal solutions can be achieved with a reduction of up to 98.53 % computational time in certain cases.

Keywords: Scheduling, Metaheuristics, Batch Plants

1. Introduction

Batch process research has seen significant increase in attention from both academic and industrial practitioners over the last two decades. This in part, is due to a modern demand for small volume, high value products of which batch processes are best suited. Batch plant flexibility provides means for the production of multiple products within the same facility. Unfortunately, this is coupled with an increase in complexity of facility scheduling. Much of the recent research has involved the development of optimisation models that provide schedules yielding improved objective values at reduced computational time. While metaheuristics have been well utilised in process engineering from many decades now, its application to multipurpose process scheduling is sparse. A comparison between metaheuristic techniques and MILP models in the scheduling of chemical batch plants can be found in Canton Padilla (2003). However, to date most approaches have generally been in the optimal design of batch plants or the scheduling of single-stage or multi-product batch plants. Work with genetic algorithms on the multipurpose batch plant scheduling with fixed processing times was utilised effectively to minimise makespan in He and Hui (2010), but the more complex variable processing time scenario was not investigated.

The focus of this work is to obtain close-to-optimal solutions at reduced computational times. This is done through a novel representation of integer coded tasks and real valued times generalizable to any particular batch plant configuration and optimizable with any suitable metaheuristic. A well-posed and accepted case study of a multipurpose batch

plant is considered. The literature example is introduced in Kondili et al. (1993) and contains batch mixing and splitting, and includes a recycling stream.

2. Problem description

Utilising a metaheuristic approach, specifically a genetic algorithm, two examples are considered: a motivating example and the aforementioned primary example.

Figure 1, shows the state-task-network (STN) for a motivating example involving a simple multiproduct batch plant (Ierapetritou and Floudas, 1998).

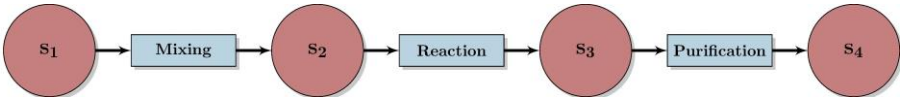


Figure 1: STN for Motivating Example

The corresponding data for Figure 1 is given in Table 1. This literature example has often been used for comparison when introducing new scheduling frameworks.

The primary example considered involves a relatively complex flowsheet introduced by Kondili et al (1993). The STN for the flowsheet is found in Figure 2. The problem data is given in Table 2.

In both examples, variable processing time is considered. Mathematical programming approaches have proven successful in managing constraints where the assignment of tasks to units is dependent on preserving correct material balances. The framework and GA introduced here satisfies all the recipe-dependent constraints in the accepted STN descriptions by Kondili et al. (1993) and Ierapetritou and Floudas (1998). In both cases, the objective is to maximise profit over a given fixed time horizon.

Table 1: Table showing capacity, suitability and mean processing time for the Motivating Example

Unit	**Capacity**	**Suitability**	**Mean Processing Time (τ)**
Unit 1	100	Mixing	4.5
Unit 2	75	Reaction	3.0
Unit 3	50	Purification	1.5
State	**Storage Capacity**	**Initial Amount**	**Price**
State 1	Unlimited	Unlimited	0
State 2	100	0	0
State 3	100	0	0
State 4	Unlimited	0	1

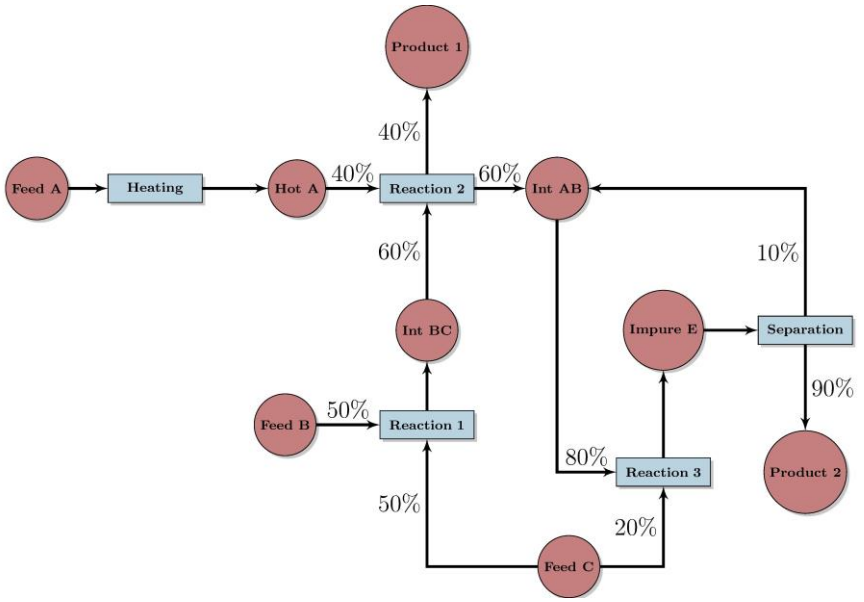


Figure 2: The STN of the Primary Example

Table 2: Table showing the capacity, suitability and mean processing time for the Primary Example

Unit	**Capacity**	**Suitability**	**Mean Processing Time (τ)**
Heater	100	Mixing	4.5
Reactor 1	50	Reaction 1, 2, 3	2.0, 2.0, 1.0
Reactor 2	80	Reaction 1, 2, 3	2.0, 2.0, 1.0
Still	200	Separation	1.0 Product 2, 2.0 Int AB
State	**Storage Capacity**	**Initial Amount**	**Price**
Feed A	Unlimited	Unlimited	0
Feed B	Unlimited	Unlimited	0
Feed C	Unlimited	Unlimited	0
Hot A	100	0	0
Int AB	200	0	0
Int BC	150	0	0
Impure E	200	0	0
Product 1	Unlimited	0	10
Product 2	Unlimited	0	10

3. Results

All results were obtained utilizing an Intel® Core™ i7-7700k @ 5.0 Ghz, 32 GB 3200 Mhz DDR4 RAM, running Windows 10 Professional x64 bit. For MILP models, GAMS 24.8.3 was used. The genetic algorithm was implemented in Matlab 2016b.

3.1. Motivating example

Seid and Majozi (2012a) showed that their model had the fastest solution time when compared to other MILP approaches in literature. Hence, the MILP comparisons to the implemented GA are drawn using this model. A GAMS cut-off time of 10,000 seconds was used. All GAMS models were solved to a 0 % relative gap unless otherwise specified. Time horizons ranging from 12 – 168 hours are considered since comparative results exist in Seid and Majozi (2012b). The CPU times listed are resultant from the averaging of 50 trials runtimes. The best obtained objective value from these trials is reported. A cut-off time of 10,000 seconds was imposed on the model by Seid and Majozi (2012b), with the model converging to a zero relative gap in all cases other than 168 hours, when the cut-off was reached. At this time GAMS reported a relative gap of 0.28 %. Considering the computational performance, it can be seen that the GA outperforms the model by Seid and Majozi (2012b) in all cases other than the short time horizons of 12 and 24 hours. A large increase in the time horizon, however, sees a massive increase in solution time for MILP formulation, while the GA does not exhibit this result. The scale of these growths in magnitude when considering solution time is illustrated in Figure 3. Importantly, the GA achieves this computational time reduction at minimal loss of objective value, with the worst performing case being 2.31 % less of the model by Seid and Majozi (2012b). The complete results are found in Table 3. The instance of the GA reporting an improved objective value is due to numerical instability and rounding between the two separate implementations.

Table 3: Results for Motivating Example

Model	Event Points	Time Horizon	Cpu Time (s)	Objective Value	Difference	Pop Size/Gen
S&M	5	12	0.240	71.4703	-	-
GA	5	12	1.035	71.4734	4.3×10^{-3} %	300/15
S&M	9	24	0.262	249.9390	-	-
GA	9	24	1.208	249.2559	-0.27 %	300/15
S&M	14	36	7.709	446.9474	-	-
GA	14	36	2.655	440.6822	-1.4 %	300/15
S&M	19	48	85.262	646.9474	-	-
GA	23	48	3.022	631.9465	-2.31 %	500/25
S&M	57	168	10,020.096*	2,645.4545	-	-
GA	60	168	62.883	2,607.2000	-1.44 %	1,000/100

*Cut-off time was set to 10,000 seconds. Relative gap at cut-off was 0.28 %
**S&M = Seid and Majozi (2012b)

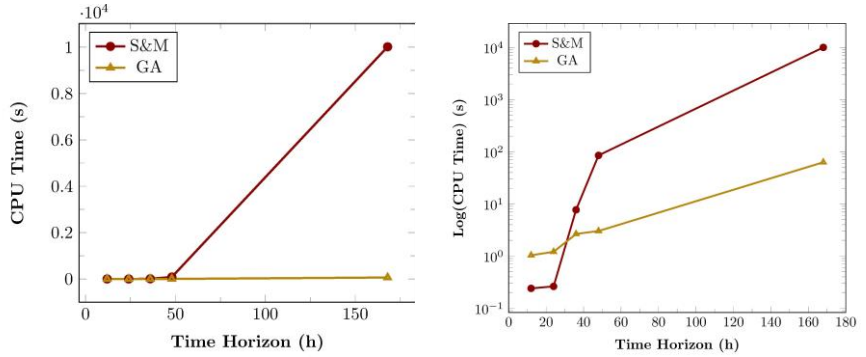


Figure 3: Computational time between Seid and Majozi (2012b) and GA for Motivating Example

3.2. Primary Example

The same time horizons from Seid and Majozi (2012a) were considered, with the additional time horizons of 18 and 20 hours respectively, to further explore longer time horizons. A GAMS cut-off time of 40,000 seconds was used, and models solved to a 0 % relative gap unless specified. As with the Motivating Example, the MILP model by Seid and Majozi (2012a) performed better on short term time horizons but scaled by orders of magnitude for longer time horizons, while the GA reports solutions within a single order of magnitude for all cases. Again, this speedup comes with an objective value reduction of no more than 2.39 %. The complete results are found in Table 4, while the computational scalability is reported in Figure 4. Similarly, the case where the GA outperforms the MILP implementation is due to the aforementioned numerical accuracy.

Table 4: Results for Primary Example

Model	Event Points	Time Horizon	Cpu Time (s)	Objective Value	Difference	Pop Size/Gen
S&M	5	8	0.265	1,489.8169	-	-
GA	5	8	3.949	1,498.1734	1.01 %	1,000/25
S&M	8	10	17.259	1,973.8273	-	-
GA	8	10	4.341	1,956.6194	-0.87 %	1,000/25
S&M	9	12	38.159	2,624.6639	-	-
GA	12	12	4.703	2,609.3645	-0.58 %	1,000/25
S&M	10	16	87.290	3,705.0358	-	-
GA	10	16	4.729	3,671.0000	-0.96 %	1,000/25
S&M	13	18	40,0001.265*	4,285.5545	-	-
GA	13	18	5.882	4,217.7454	-1.58 %	1,000/25
S&M	17	20	40,0003.301**	4,868.1123	-	-
GA	16	20	6.570	4,751.3228	-2.39 %	1,000/25

*Cut-off time was set to 40,000 seconds. Relative gap at cut-off was 0.88 %
** Cut-off time was set to 40,000 seconds. Relative gap at cut-off was 1.55 %
*** S&M = Seid and Majozi (2012a)

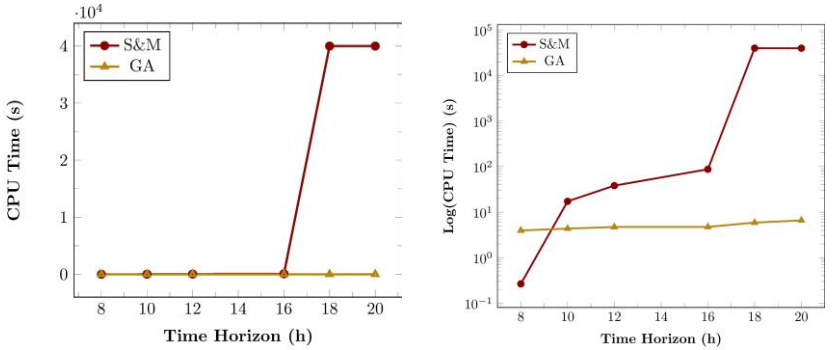


Figure 4: Computational time between Seid and Majozi (2012b) and GA for Primary Example

4. Conclusions

Here, a novel scheduling framework and corresponding metaheuristic technique for multipurpose batch plants is introduced. The technique satisfies all rules and constraints as that of MILP counterparts discussed in literature but does not suffer scaling enumeration from increased time horizons. The technique utilises a coupled chromosome genetic algorithm, where one chromosome is integer coded and the second chromosome is coded with real values. The introduced method does not suffer under increased granularity of the time horizons as seen by the results, especially on medium-to-long-time horizons. In almost all cases the GA outperformed the MILP implementation of Seid and Majozi (2012a) when considering computational time. In the case of long time horizons, solutions were achieved up to three orders of magnitude faster, yielding a 98.53 % reduction in solution time. Importantly, these computational speed-ups were attained at minimal loss of objective value, with the worst case returning an objective value of no less than 2.39 % of accepted literature.

References

J. Canton Padilla. “Integrated support system for planning and scheduling of batch chemical plants.” Universitat Politècnica de Catalunya, 2003.

Y. He, and C.W. Hui. "A binary coding genetic algorithm for multi-purpose process scheduling: A case study." Chemical engineering science 65.16 (2010): 4816-4828.

M.G. Ierapetritou, and C. A. Floudas. "Effective continuous-time formulation for short-term scheduling. 1. Multipurpose batch processes." Industrial & engineering chemistry research 37.11 (1998): 4341-4359.

E. Kondili , C. C. Pantelides, and R. W. H. Sargent. "A general algorithm for short-term scheduling of batch operations—I. MILP formulation." Computers & Chemical Engineering 17.2 (1993): 211-227.

R. Seid, and T. Majozi. "A robust mathematical formulation for multipurpose batch plants." Chemical engineering science 68.1 (2012a): 36-53.

R. Seid, and T. Majozi. "A novel technique for prediction of time points for scheduling of multipurpose batch plants." Chemical engineering science 68.1 (2012b): 54-71.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50217-2

A novel modelling approach to scheduling of multipurpose batch processes

Nikolaos Rakovitis, Jie Li*, Nan Zhang

School of Chemical Engineering and Analytical Science, University of Manchester, Manchester M13 9PL, UK
jie.li-2@manchester.ac.uk

Abstract

A great number of mathematical models have been developed for scheduling of multipurpose batch processes in the last three decades. None of them allow consumption and production tasks related to the same states to take place at the same event. It also seems that most unit-specific event-based models are task-specific instead. In this work, we introduce a new definition of recycling tasks and develop two novel mathematical models using unit-specific and task-specific event-based approaches. Both models allow consumption tasks to take place at the same event points with their related production tasks that are not recycling tasks. The computational results indicate that both models solve all examples to optimality using significantly less number of event points. Furthermore, it seems that the proposed unit-specific event-based model obtains optimal solutions in up to one magnitude less computational time compared to the proposed task-specific event-based model and the model of Shaik and Floudas (2009).

Keywords: Scheduling, Multipurpose, Batch process

1. Introduction

Scheduling of multipurpose batch processes has gained much attention during the last three decades (Harjunkoski et al., 2014). A great number of models including discrete-time (Lee and Maravelias, 2017) and continuous-time models have been developed (Li et al., 2010; Harjunkoski et al., 2014). The continuous-time models can be further divided into global event (Maravelias and Grossmann, 2003), unit-specific event-based (Shaik and Floudas, 2009), process slot (Sundaramoorthy and Karimi, 2005), unit-slot (Sursala et al., 2010) and sequence based (Mendez and Cerda, 2004) models. The advantages of unit-specific event-based models have been well established in the literature (Shaik and Floudas, 2009; Vooradi and Shaik, 2012). However, all models do not allow production and consumption tasks related to the same states to take place at the same event points. This may lead to an unnecessary increase of event points, limiting their capability to solve difficult and large-scale problems. In addition, it seems that many unit-specific event-based models are task-specific, which divides the scheduling horizon based on tasks instead of units.

In this work, we introduce a new definition of recycling tasks and develop two novel mathematical models using unit-specific and task-specific event-based approaches. Both models allow consumption tasks to take place at the same event points with their related production tasks (i.e., related to the same states) that are not recycling tasks. The

computational results demonstrate that the proposed models successfully solve all examples to optimality by using significantly less number of event points. It seems that the proposed unit-specific event-based model generates optimal solutions in up to one magnitude less computational time compared to the proposed task-specific event-based model and the model of Shaik and Floudas (2009).

2. Problem description

A multipurpose batch plant produces P (p = 1, 2, …, P) products in J (j = 1, 2, …, J) processing units with I (i = 1, 2, …, I) tasks. Each unit j is able to process one or more tasks. Tasks that can be processed in a unit j are included in the set $\mathbf{I}_j$. Such a facility stores S (s = 1, 2, …, S) states with different storage policies, which can be categorized in raw materials, intermediate states and products. Given the fixed ($\alpha_{i,j}$) and variable ($\beta_{i,j}$) processing time that a unit j requires to process a task i, unit capacities, the portion of state s produced/consumed from a task i ($\rho_{i,s}$) and product prices (P_s) as well as the scheduling horizon H, the scheduling problem is to determine optimal production schedule involving batch sizes, allocations, sequences and timings of processing units and inventory profiles.

3. Recycling tasks

To allow consumption and production tasks related to the same states to take place at the same event points, we introduce a new definition of recycling tasks, which is slightly different from that of Li and Floudas (2010). A task in a unit j is considered as a recycling task if it produces a state that can be consumed either by a task in its upstream units or by other tasks in the same unit (i.e., unit j). For instance, there are four tasks (i1-i4), two processing units (j1-j2) and three states (S1-S3) as shown in Figure 1. While tasks i1 and i3 can be processed in unit j1, tasks i2 and i4 can be processed in unit j2. Tasks i1 and i2 consume S1 and produce S2, whilst tasks i3 and i4 consume S2 and produce S1 and S3. Task i1 is considered as a recycling task because it produces S2 that can be used by task i3 in the same unit (i.e., j1). Similarly, tasks i2-i4 are also recycling tasks.

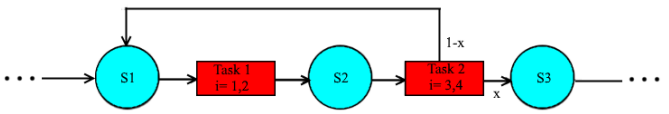


Figure 1. Illustration of recycling tasks. All tasks depicted are considered recycling tasks

In the Example 2 of Shaik and Floudas (2009), task i8 is also considered as a recycling task since it produces S5 that can be used by a task i6 in upstream units. All recycling tasks are included in a set $\mathbf{I}^{Re}$.

4. Mathematical formulations

4.1. Model 1

We develop our model 1 (denoted as **M1**) using unit-specific event-based modelling approach, where consumption tasks are allowed to take place at the same event points

with their related production tasks that are not recycling tasks. We define a new four index variable $y_{i,j,n,n'}$ to denote if a task i is processed in a unit j from event point n to n'.

4.1.1. Allocation constraints

At most one task can be processed in a unit at any time.

$$\sum_{i\in\mathbf{I}_j}\sum_{n-\Delta n\le n'\le n}\sum_{n\le n''\le n+\Delta n} y_{i,j,n',n''} \le 1 \qquad \forall j,\, n \tag{1}$$

4.1.2. Capacity constraints

We define $b_{i,j,n,n'}$ as the batch size of task i in unit j from event n to n'. Any batch size should not exceed the minimum [$B_{i,j}^{\min}$] and maximum [$B_{i,j}^{\min}$] capacities of a unit j.

$$B_{i,j}^{\min} y_{i,j,n,n'} \le b_{i,j,n,n'} \le B_{i,j}^{\max} y_{i,j,n,n'} \qquad \forall j,\, i\in\mathbf{I}_j,\, n \le n' \le n+\Delta n \tag{2}$$

4.1.3. Material balance constraints

The inventory of state s at n is equal to its inventory at $n-1$ plus batch sizes produced by a non-recycling task at n or a recycling task at $n-1$ minus batch sizes consumed at n.

$$ST_{s,n} = ST0_s + \sum_{i\in\mathbf{I}_S^P, i\notin\mathbf{I}^{\mathrm{Re}}} \rho_{i,s} \sum_j \sum_{n-\Delta n\le n'\le n} b_{i,j,n',n} + \sum_{i\in\mathbf{I}_S^C} \rho_{i,s} \sum_j \sum_{n\le n'\le n+\Delta n} b_{i,j,n,n'} \qquad \forall s,\, n=1 \tag{3}$$

$$ST_{s,n} = ST_{s,n-1} + \sum_{i\in\mathbf{I}_S^P, i\notin\mathbf{I}^{\mathrm{Re}}} \rho_{i,s} \sum_j \sum_{n-\Delta n\le n'\le n} b_{i,j,n',n} + \sum_{i\in\mathbf{I}_S^P, i\in\mathbf{I}^{\mathrm{Re}}} \rho_{i,s} \sum_j \sum_{n-1-\Delta n\le n'\le n-1} b_{i,j,n',n-1}$$
$$+ \sum_{i\in\mathbf{I}_S^C} \rho_{i,s} \sum_j \sum_{n\le n'\le n+\Delta n} b_{i,j,n,n'} \qquad \forall s,\, n>1 \tag{4}$$

4.1.4. Duration constraints

We define $Tsi_{i,j,n}$ and $Tfi_{i,j,n}$ to denote start and end times of task i in unit j at event point n. A task i in a unit j can finish after it is processed in this unit.

$$Tfi_{i,j,n'} \ge Tsi_{i,j,n} + \alpha_{i,j} \cdot y_{i,j,n,n'} + \beta_{i,j} \cdot b_{i,j,n,n'} \qquad \forall j,\, i\in\mathbf{I}_j,\, n \le n' \le n+\Delta n \tag{5}$$

Any task i processed in a unit j must start and finish within start ($Ts_{j,n}$) and end ($Tf_{j,n}$) times of this unit j.

$$Ts_{j,n} \le Tsi_{i,j,n} \le Tfi_{i,j,n} \le Tf_{j,n} \qquad \forall j,\, i\in\mathbf{I}_j,\, n \tag{6,7}$$

4.1.5. Sequencing constraints

A unit j can start at event point $n+1$ only after it finishes at time event n.

$$Ts_{j,n+1} \ge Tf_{j,n} \qquad \forall j,\, n<N \tag{8}$$

Since a consumption task is allowed to take place at the same event points with related production tasks that are not recycling tasks, the sequencing constraints for different tasks in different units are imposed for recycling and non-recycling tasks separately.

$$Tsi_{i,j,n} \ge Tfi_{i',j',n} \qquad \forall\, i\in\mathbf{I}_S^C,\, i'\in\mathbf{I}_S^P,\, i\in\mathbf{I}_j,\, i'\in\mathbf{I}_{j'},\, i'\notin\mathbf{I}^{\mathrm{Re}}, j\ne j',\, n \tag{9}$$

$$Tsi_{i,j,n+1} \geq Tfi_{i',j',n} \qquad \forall i \in \mathbf{I}_S^C,\ i' \in \mathbf{I}_S^P,\ i \in \mathbf{I}_j,\ i' \in \mathbf{I}_{j'},\ i' \in \mathbf{I}^{\mathrm{Re}},\ j \neq j',\ n < N \tag{10}$$

For finite or no intermediate storage policy, materials produced from a task should be immediately consumed.

$$Tsi_{i,j,n} \leq Tfi_{i',j',n} + M[1 - \sum_{n-\Delta n \leq n' \leq n} y_{i',j',n',n}]$$

$$\forall s \in S^{\mathrm{dis}},\ i \in \mathbf{I}_S^C,\ i' \in \mathbf{I}_S^P,\ i \in \mathbf{I}_j,\ i' \in \mathbf{I}_{j'},\ i' \notin \mathbf{I}^{\mathrm{Re}},\ j \neq j',\ n \tag{11}$$

$$Tsi_{i,j,n+1} \leq Tfi_{i',j',n} + M[1 - \sum_{n-\Delta n \leq n' \leq n} y_{i',j',n',n}]$$

$$\forall\ s \in S^{\mathrm{dis}},\ i \in \mathbf{I}_S^C,\ i' \in \mathbf{I}_S^P,\ i \in \mathbf{I}_j,\ i' \in \mathbf{I}_{j'},\ i' \in \mathbf{I}^{\mathrm{Re}},\ j \neq j',\ n \tag{12}$$

4.1.6. Objective function

The objective is to maximize productivity during a specified scheduling horizon.

$$z = \sum_{s \in S} p_s \sum_{i \in \mathbf{I}_S^P} \sum_{j \in J} \sum_{n \in N} \sum_{n \leq n' \leq n+\Delta n} \rho_{i,s} b_{i,j,n,n'} \tag{13}$$

This model is easy to extend for other objectives such as makespan (*MS*) minimisation.

4.2. Model 2

Our model 2 (denoted as **M2**) is developed based on task-specific event-based approach where consumption tasks are allowed to take place at the same event points with related production tasks that are not recycling tasks. Since this model is quite similar to that of Shaik and Floudas (2009), only different constraints are presented in the following.

4.2.1. Material balance constraints

Shaik and Floudas (2009) defined if a state *s* is produced at an event *n*, it should be consumed at *n*+1. In this model, a state *s* produced could be consumed at the same event point by a task that is not recycling task, which is similar to that in Model **M1**.

$$ST_{s,n} = ST0_s + \sum_{i \in \mathbf{I}_S^P, i \notin \mathbf{I}^{\mathrm{Re}}} \rho_{i,s} \sum_{n-\Delta n \leq n' \leq n} b_{i,n',n} + \sum_{i \in \mathbf{I}_S^C} \rho_{i,s} \sum_{n \leq n' \leq n+\Delta n} b_{i,n,n'} \qquad \forall s,\ n = 1 \tag{14}$$

$$ST_{s,n} = ST_{s,n-1} + \sum_{i \in \mathbf{I}_S^P, i \notin \mathbf{I}^{\mathrm{Re}}} \rho_{i,s} \sum_{n-\Delta n \leq n' \leq n} b_{i,n',n} + \sum_{i \in \mathbf{I}_S^P, i \in \mathbf{I}^{\mathrm{Re}}} \rho_{i,s} \sum_{n-1-\Delta n \leq n'' \leq n-1} b_{i,n',n-1} + \sum_{i \in \mathbf{I}_S^C} \rho_{i,s} \sum_{n \leq n' \leq n+\Delta n} b_{i,n,n'}$$

$$\forall s,\ n > 1 \tag{15}$$

4.2.2. Different tasks in different units

Similar to **M1**, the sequencing constraints for different tasks in different units are imposed for recycling and non-recycling tasks separately.

$$Ts_{i,n} \geq Tf_{i',n} - M[1 - \sum_{n-\Delta n \leq n' \leq n} w_{i',n',n}] \quad \forall i \in \mathbf{I}_S^C,\ i' \in \mathbf{I}_S^P,\ i \in \mathbf{I}_j,\ i' \in \mathbf{I}_{j'},\ i' \notin \mathbf{I}^{\mathrm{Re}},\ j \neq j' \tag{16}$$

$$Ts_{i,n+1} \geq Tf_{i',n} - M[1 - \sum_{n-\Delta n \leq n' \leq n} w_{i',n',n}] \quad \forall i \in \mathbf{I}_S^C,\ i' \in \mathbf{I}_S^P,\ i \in \mathbf{I}_j,\ i' \in \mathbf{I}_{j'},\ i' \in \mathbf{I}^{\mathrm{Re}},\ j \neq j',\ n < N \tag{17}$$

5. Computational results

We solve five examples to illustrate the capability of our proposed models. The first three examples are Examples 1-3 from Shaik and Floudas (2009). In Example 4, there are five states including one product and four processing units in series with each processing one task and producing one state. In Example 5, there are 27 states including three products and eight processing units in series with each processing 3 tasks and producing 3 states. All tasks in Examples 4-5 are non-recycling tasks.

Table 1: Comparative results for Examples 1-5

Example	Model	Event points	CPU time (s)	RMILP	MILP	Discrete Variables
1	S&F[1]	9	1.295	6601.65	5038.05	45
(H=16 h)	M1	7	1.310	6601.65	5038.05	35
	M2	7	1.544	6601.65	5038.05	35
2	S&F	8	5.974	4291.68	3738.38	64
(H=16 h)	M1	8	3.386	4291.68	3788.38	64
	M2	8	6.303	4291.68	3738.38	64
3	S&F	10	7.301	5225.86	4262.80	110
(H=16 h)	M1	10	24.27	5225.86	4262.80	110
	M2	10	6.989	5225.86	4262.80	110
4	S&F	10	24.18	6601.65	4305.46	60
H=16 h	M1	7	9.017	5404.31	4305.46	42
	M2	7	27.63	6601.65	4305.46	42
5	S&F	49	33.81	21000	20935.3	1176
H=144 h	M1	42	1.014	21000	20935.3	1008
	M2	42	43.54	21000	20935.3	1008

[1]Shaik and Floudas (2009) model. All examples are solved in a machine using GAMS 24.6.1. CPLEX 12 in an Intel® Core™ i5-2500 3.3 GHz and 8 GB RAM running Windows 7.

The comparative results are presented in Table 1. It seems that both proposed models successfully generate the same optimal solutions as those of Shaik and Floudas (2009) for all examples. The optimal schedule for Example 1 from the model **M1** is illustrated in Figure 2. It can observed that both production task i1 and consumption task i3 related to the same state take place at the same event N1 as depicted with red circles in Figure 2. Similarly, tasks i1-i2 and task i3 related to the same state take place at the same event N3 as depicted with blue circles. Both models require less number of event points than that of Shaik and Floudas (2009), leading to smaller model sizes. The model **M1** requires less computational time than other models. In some cases, it can generate the same optimal solution in one magnitude less computational time. Furthermore, it leads to much tighter MILP relaxation for Example 4. However, there is no decrease in the number of event points for Examples 2-3 because most of the tasks (6 out of 8 for example 2 and 9 out of 11 for example 3) are recycling tasks.

6. Conclusions and future work

In this work, we introduce a new definition of recycling tasks and develop two novel mathematical models using unit-specific and task-specific event-based modelling

approach for scheduling multipurpose batch processes. Both models allow consumption tasks to take place at the same event points with their related production tasks that are not recycling tasks. The computational results demonstrate that both models could significantly reduce the required number of event points. Furthermore, the proposed unit-specific event-based model generated the same optimal solution in up to one magnitude less computational time than the task-specific event-based model and the model of Shaik and Floudas (2009). Our future work is to propose a number of tightening constraints to further improve the efficiency of the proposed models.

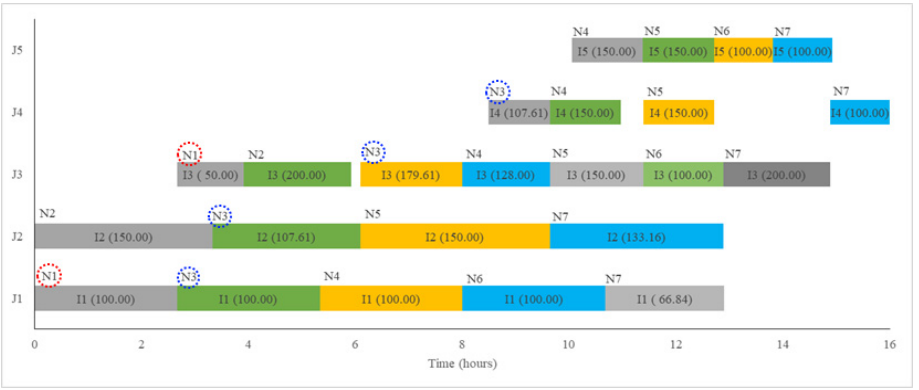


Figure 2. Optimal schedule for Example 1 using unit-specific model

References

I. Harjunkoski, C. Maravelias, P. Bongers, P. Castro, S. Engell, I. Grossmann, J. Hooker, C. Méndez, G. Sand, J. Wassick, 2014, Scope for industrial application of production scheduling models and solution methods, Computers and Chemical Engineering, 62, 161-193

H. Lee, C. Maravelias, 2017, Discrete-time mixed-integer programming models for short-term scheduling in multipurpose environments, Computers and Chemical Engineering, 107, 171-183.

J. Li, C. Floudas, 2010, Optimal event point determination for short-term scheduling of multipurpose batch plants via unit-specific event-based continuous-time approaches, Industrial & Engineering Chemistry Research, 49, 7446-7469

J. Li, N. Susarla, I. Karimi, M. Shaik, C. Floudas, 2010, An analysis of some unit-specific event-based models for the short-Term scheduling of noncontinuous processes, Industrial and Engineering Chemistry Research, 49, 633-647

C. Maravelias, I. Grossmann, 2003, New general continuous-time state-task network formulation for short-term scheduling of multipurpose batch plants, Industrial Engineering and Chemistry Research, 42, 3056-3074

M. Shaik, C. Floudas, 2009, Novel unified modeling approach for short-term scheduling, Industrial & Engineering Chemistry Research, 48, 2947-2964

A. Sundaramoorthy, I. Karimi, 2005, A simpler better slot-based continuous-time formulation for short-term scheduling in multipurpose batch plants, Chemical engineering science, 60, 2679-2702

N. Sursala, J. Li, I. Karimi, 2010, A novel approach to scheduling multipurpose batch plants using unit-slots, AlChe Journal, 56, 1859-1879

R. Vooradi, M. Shaik, 2012, Improved three-index unit-specific event-based model for short-term scheduling of batch plants, Computers and Chemical Engineering, 43, 148-172

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50218-4

Optimal multiperiod production planning in a sawmill

Nicolás Vanzetti[a], Diego Broz[b], Gabriela Corsano[a*], Jorge M. Montagna[a]

[a]*Instituto de Desarrollo y Diseño First affiliation, Avellaneda 3657, (S3002GJC) Santa Fe, Argentina.*
[b]*Facultad de Ciencias Forestales (FCF, UNaM), CONICET, Bertoni 124, (N3382GDD) Eldorado, Misiones, Argentina*
gcorsano@santafe-conicet.gov.ar

Abstract

Forest industry plays an important role in the national economic and social context, due to the region where it is located (northeastern of the country) and the production volume. The equipment used in the local facilities is not efficiently employed and, therefore, the process performance can be improved in order to reach higher productivity. Sawmill production planning involves determining which logs to process taking into account its diameter and length, with the aim of satisfying finished products demand. Logs are converted into lumber when they are cut according to a set of available cutting patterns (CP). One of the keys to improve sawmills efficiency is to develop a good production planning, and mathematical modeling is a good technique to attain it. Also, an efficient log procurement policy can be developed from the production planning, and therefore, raw material purchases and distribution can be anticipated. In this work, a mixed integer linear programming (MILP) model for the optimal planning in sawmills is proposed. A multiperiod formulation is presented, which involves a set of CPs for each type of log. As a result of this optimization, appropriate log procurement and distribution policy can be also defined which impact on the sawmill performance and profitability.

Keywords: production planning, multiperiod, sawmill, cutting pattern, mixed integer linear programming.

1. Introduction

The Argentinean forest industry is mainly located in the northeastern of the country (85% of the production, approximately 850,000 ha). More than one thousand forest factories (sawmills, plywood mills, pulp mills, and Medium Density Fibreboard factories) are concentrated in this region, where small and medium-sized enterprises represent 98% of these facilities (Broz et al., 2016). For these reasons, forest industry has a very important role in the economic and social development of this region. However, inefficient production and high transport cost have a negative impact on the competitiveness of this sector.

In the particular case of sawmills, efficient production can be achieved through optimal production planning considering raw materials (logs availability), final products (boards) and demands (customers), taking into account the logs diameter, length, grade of each log piece, characteristics of primary products, and industrial parameters, among others.

In a sawmill, logs are converted into boards when they are cut according to cutting patterns (CP). A CP is an arrangement of rectangles (thickness and width of the boards) within a circle. Planning operations at a sawmill is a complex task, since it involves several simultaneous decisions as raw material supply, cutting patterns (CP) allocation to different types of raw material, inventories management and demands satisfaction among others. Moreover, these decisions lead to various tradeoffs that complicate the problem resolution. Mathematical modeling appropriately allows addressing this problem. Also, many heuristics and simulation approaches have been used to face sawmills production planning. Maturana et al. (2010) present a mathematical model for determining the volume and type of logs to be processed according to six proposed CP. They consider the production over 6 weeks, with an ideal scenario of log supply which is then adjusted. Results are compared with a heuristic schedule used by a Chilean company. A similar approach is proposed by Alvarez and Vera (2014) but considering an annual planning period and solving the uncertainties through robust optimization, while in Zanjani et al. (2010) the planning horizon consists of 30 days, and the authors also propose robust optimization for solving the problem with random yields. In this last case, 5 CP are involved in the model. In Lobos and Vera (2016), a decomposition algorithm involving two levels is developed: a tactical planning horizon with monthly information and an operational level with detailed weekly production.

Although the production yield strongly depends on the CP selection and use, there are few practical tools that support the CPs generation and, to the best of our knowledge, there is no optimization model that involves all the possible CPs for the sawmill production planning. On the other hand, many final products can be obtained from different CPs, and therefore several tradeoffs can be assessed. Moreover, when planning is carried out, different decisions must be integrated, as raw material purchase, production, inventory, and demand satisfaction. Then, it is necessary to address this problem from a comprehensive point of view for attaining a good supply procurement policy along the time horizon, in order to reach the desired production for fulfilling the required demand.

In order to face this problem, a mixed integer linear programming (MILP) model for the optimal production planning in a sawmill is proposed in this work. The model embeds all the possible CP for different log diameters, given by an exhaustive generation algorithm. Also, in a multiperiod approach, all the decisions are simultaneously taken, and an appropriate log procurement policy, suitable production according to enterprise requirements, improved stock management, and accurate demand satisfaction are obtained in order to favour sawmill performance and profitability. Through an example, the approach capabilities are highlighted.

2. Problem description

In Fig. 1 the considered process is presented. The raw material (logs) arrives at the sawmill and is classified according to log length and diameter. Then, logs are transferred to the sawing sector where an initial cut is made through a primary saw in which a cant and two flitches are obtained. In the secondary saw, the cant is broken down into dimensional lumber pieces and flitches. Finally, flitches are processed in a re-sawing. This cutting process generates products of medium and large dimension. These products are boards with determined cross section area (characterized by width and thickness), and a length, given by the log length. The log cuts are made according to a selected CP.

Each CP is applied to a log with a determined useful diameter, and is valid for all the different log lengths with that diameter. The useful diameter corresponds to the diameter of the larger cylinder that can be obtained from the log.

At the beginning of the planning horizon, there is an initial logs stock at the sawmill, and daily, logs are bought from different suppliers, which have a limited capacity and different prices. Every day, a minimum production is required for some types of boards, while a final demand must be fulfilled at the end of the planning horizon. The objective is to determine the detailed weekly production planning, i.e. daily decisions about raw material purchases, how many logs of each type (diameter and length) have to be cut, the logs stock level, the finished product stock, demands fulfilment and the produced residues in order to minimize the sawing cost given by raw material, operation, and setups.

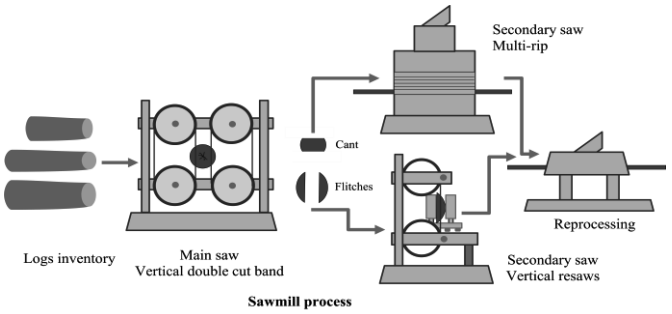


Figure 1. Stages considered in sawmill process

3. Proposed approach

The mathematical model involves, as input data, all the generated CP according to the available log diameters, demanded boards, and efficiency imposed by the enterprise. The objective of the CP generator is to systematically arrange rectangles (thickness and width of the boards) within circles. The insertion of the rectangles is done by simulating the cutting way of logs in sawmills. It divides the log into five sections: a central block, two equal lateral flitches, and upper and lower ones, equal to each other. From each of these sections, the different admissible boards are obtained. For each generated CP, data about the number of included boards, yield, processing time, and setup cost are provided to be used as input for the optimization model.

Following, the mathematical model for the multiperiod sawmill planning is briefly presented. It is assumed that the planning horizon is 5 days (Eq. (9)), but this can be easily generalized. For space reasons, some particular equations are omitted, but the detailed formulation is available for interested readers.

$$min\ Z = \sum_{d,l,a,t} Qb_{dlat}\, Cb_{dla} + \sum_{p,d,l,t} Qt_{pdlt}\, Cp_{pdl} + \sum_{p,t} Cx_p\, x_{pt} \tag{1}$$

$$\sum_{t} Qb_{dlat} \leq Dis_{dla} \qquad \forall d,l,a \tag{2}$$

$$\sum_{a} Qb_{dlat} + S_{dlt-1} = \sum_{p} Qt_{pdlt} + S_{dlt} \quad \forall d, l, t \tag{3}$$

$$\sum_{d,l} Qt_{pdlt} \leq M x_{pt} \quad \forall p, t \tag{4}$$

$$\sum_{\substack{p,d \\ p \in Rel(p,d)}} Qt_{pdlt}\ \rho_{pdi} = P_{ilt} \quad \forall i, l, t \tag{5}$$

$$\sum_{t \leq t'} P_{ilt} \geq \sum_{t \leq t'} Pmin_{ilt} \quad \forall i, l, t' \tag{6}$$

$$\sum_{t} P_{ilt} = VP_{il} + SF_{il} \quad \forall i, l \tag{7}$$

$$SF_{il} \leq Cap_{il} \quad \forall i, l \tag{8}$$

$$\sum_{p} TS_p\ x_{pt} + \sum_{p,d,l} Qt_{pdlt}\ Tp_{pdl} \leq Tmax_t \quad \forall t \tag{9}$$

The objective function considers the raw material cost, given by the amount of logs of diameter *d* and length *l* bought from different suppliers *a* at the day *t*, Qb_{dlat}, by its unit cost (Cb_{dla}), plus the production cost defined by the processed logs (Qt_{pdlt}) by the unit operation cost (Cp_{pdl}), plus the setup cost managed by the binary variable x_{pt} equal to one if the CP *p* is used at day *t*. Cx_p represents the cost for using *p* (setup cost). Eq. (2) limits the purchase of logs to each supplier, where Dis_{dla} represents the supplier capacity, while Eq. (3) defines the log stock management, where S_{dlt} represent the number of logs stored at the end of day *t*. Here, $S_{dl,0}$ are the available logs in the facility, prior to begin the production. Eq. (4) states that no logs of diameter *d* and length *l* are processed on *t* through CP *p*, if that pattern is not selected at day *t* (*M* is a big scalar). The number of boards with cross section *i* and length *l* produced in *t* (P_{ilt}) is defined in Eq. (5) through the parameter ρ_{pdi}, which represents the number of this type of boards generated by CP *p*. *Rel(p,d)* is the set that relates each generated CP *p* with the corresponding log diameter *d*. A minimum required production for boards is stated in Eq. (6), represented by the parameter $Pmin_{ilt}$. At the end of the planning horizon, the total stored boards of type (*i,l*) must fulfil the required demand VP_{il} and a final stock of these boards SF_{il} can be held (Eq. (7)), but with maximum capacity Cap_{il} (Eq. (8)). Finally, in Eq. (9), the operation time for each day *t* is restricted, where TS_p is the CP setup time, Tp_{pdl} the log processing time, and $Tmax_t$ the working day time.

4. Example

The proposed example considers logs classified into 4 diameters and 6 lengths. The logs may come from the sawmill stock or can be bought from 3 suppliers (A, B, and C) that have a limited number of logs. A total of 144 types of products, distributed among 24 cross sections and 6 lengths is considered.

For the CP generation, a yield greater than 80% is taken into account in order to make a better use of the wood. With this constraint and considering all the possible log diameters and demanded final boards, a total of 664 CPs are generated.

The total demand is equal to 32,590 boards, covering 71 different types of products (cross sections and lengths combinations). The first planning day, there are 290 logs of different diameters available at the sawmill. The adopted time horizon is equal to 5 working days with 8 hours each.

The example is implemented and solved in GAMS using CPLEX solver in an Intel(R) Core(TM) i7-3770, 3.40 GHz. The model consists of 86,764 equations, 163,407 continuous and 3,320 binary variables. The solution obtained in a limit of 300 seconds achieves an optimality gap of 1.07 %.

After solving the model, the weekly sawmill planning that fulfils the demand has a total cost equal to $ 895,464, which 58 % corresponds to the used raw material, 39 % represents the production cost and 3 % the setup cost.

Table 1. Planning results for each day

	Period				
	t1	**t2**	**t3**	**t4**	**t5**
Logs purchased [unit]	473	611	500	426	498
Used logs [unit]	484	637	505	426	576
Time used [h]	8	8	7.8	7.2	7.1
Produced boards [unit]	7223	7329	6867	6583	6421

Table 2. Processed logs according to CP and time period

Diameter	**CP**	**Period**				
		t1	**t2**	**t3**	**t4**	**t5**
d1	p3		539			508
d1	p7				104	
d2	p24	51				
d2	p30			247		
d3	p146	199				
d3	p147	176				
d3	p153				337	
d3	p161	24		212		
d3	p200		40			
d3	p311	9				
d3	p444					44
d3	p446			46		
d3	p505		18			
d4	p573		57			
d4	p1272				8	
d4	p1324				11	

In Table 1 general results for each planning day are shown. It can be noted that a total of 2,507 logs are purchased from the different suppliers, while 2,628 logs are used to meet the demand: 40 % from supplier A, 10% from supplier B, 45 % from supplier C, and 5 % from the initial stock. At the end of the planning horizon, 169 logs are stored.
From the production point of view and considering an 8-hour working day, 95 % of the available time is used, and a total of 34,422 final boards are produced. From this amount, at the end of the time period, 1,832 units are stored, which represents 5.6 % of the total demand. Taking into account that the model includes the raw material cost, the solution adjusts the production to reduce the stored units.
Table 2 shows the total logs that are cut in each period for each CP, where log diameter is included without mentioning the lengths due to space reasons. The CPs are identified with the name "*p*" plus a number, for example "*p3*" is the pattern 3 used for diameter d_1. Usually each CP is used with several lengths to reduce de setup time and cost. In this case, 16 CPs are used for producing the 71 different types (cross section area and length of the boards) of demanded products plus those that are generated when a CP is used and they are not demanded. These last boards are part of the final inventory (SF_{il}).
For space reasons, it is not possible to show all the results in detail, but the model determines, among other things, the number of logs, distinguished by diameter and length, that are used in each day and with each CP, the boards of each type and length that are produced and the inventory level at the end of each period of the time horizon.

5. Conclusion

In this work, a MILP model for the optimal multiperiod production planning in a sawmill was presented. Daily logs procurement, suppliers selection, quantity of processed logs, CPs assignment, log and boards inventories, and operation times are simultaneously obtained in order to satisfy the required demands in the proposed horizon time at minimum cost. The complete set of feasible CPs is generated through an exhaustive algorithm, which responds to the enterprise requirement about type of saw, available log diameters, log yields, and board dimensions. The large number of generated CPs complicates the model resolution due to the various tradeoffs among decisions variables. The same product can be obtained from different CPs and logs, and therefore inventory capacity, production times, raw material procurement, demand fulfilment, among others, are simultaneously assessed in order to obtain the best economical solution. Through the proposed example, these tradeoffs evaluations and the approach capabilities were highlighted.

References

P. P. Alvarez, J.R. Vera, 2014, Application of robust optimization to the sawmill planning problem, Annals of Operations Research, 219(1), 457-475.

A. Lobos, J.R. Vera, 2016, Intertemporal stochastic sawmill planning: Modeling and managerial insights, Computers & Industrial Engineering, 95(1), 53-63.

S. Maturana, E. Pizani, J. Vera, 2010, Scheduling production for a sawmill: A comparison of a mathematical model versus a heuristic, Computers & Industrial Engineering, 59(4), 667-674.

M.K. Zanjani, D. Ait-Kadi, M. Nourelfath, 2010, Robust production planning in a manufacturing environment with random yield: A case in sawmill production planning, European Journal of Operational Research, 201(3), 882-891.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50219-6

Optimal Short-Term Scheduling for Cascaded Hydroelectric Power Systems considering Variations in Electricity Prices

Pulkit Mathur[a], Christopher L. E. Swartz[a*], Danielle Zyngier[b], François Welt[b]

[a]*Department of Chemical Engineering, McMaster University, 1280 Main St. West, Hamilton L8S 4L8, ON, Canada*
[b]*Hatch, Renewable Power, 4342 Queen St., Suite 500, Niagara Falls L2E 7J7,ON, Canada*
swartzc@mcmaster.ca

Abstract

The efficient utilization of limited available resources plays a significant role in the operation of hydroelectric power systems. The purpose of scheduling in such systems is to determine the optimal power generation and generating unit commitment schedules so as to optimize an economic performance indicator subject to various system and external constraints. Fluctuations in electricity prices caused by the ever changing market conditions have a significant impact on the generation schedules for hydroelectric power systems. This may also give rise to a phenomenon of significant variation in optimal power generation and unit commitment schedules in response to slight fluctuations in pricing. These variations are undesirable since they warrant changes to the operational policy. This work focuses on the analysis of this "nervousness" phenomenon and the development of two strategies to mitigate its severity. This work involves the development of a mixed-integer nonlinear programming (MINLP) model for the short-term hydro scheduling (STHS) problem and its solution using a computationally efficient successive linear programming (SLP) technique. The effectiveness of the proposed methodology is demonstrated through a case study for daily generation scheduling in a cascaded hydroelectric power system.

Keywords: scheduling, hydroelectric, electricity price, "nervousness"

1. Introduction

Rapidly increasing global competitiveness and the highly dynamic nature of operational and marketplace conditions in the renewable energy sector motivate the need for development of efficient scheduling strategies for the optimal operation of hydroelectric power systems. Short-term hydro scheduling problems typically range from one day to a week and involve hourly or daily scheduling of generation on a system with the objective of optimizing a relevant economic performance indicator such as cost, revenue or generated energy for the given time horizon in view of all the various applicable constraints. These problems are usually composed of two main decision-making stages: unit commitment and economic load dispatch. The former is concerned with finding a subset of the set of available generating units which should be used to serve the forecasted energy demand at minimum operating cost. The latter is concerned with finding an optimal operational policy for a fixed number of units already connected to the system. The STHS problems are typically categorized mathematically as discrete,

nonlinear and non-convex (Wood and Wollenberg, 2012). The existence of hydraulically interconnected reservoirs, limited generating resources, variety of discrete and continuous constraints and decision-making over multiple stages, make these problems fairly large-scale and complex for cascaded hydroelectric power systems, motivating the need for development of effective modeling and solution approaches.

Mathematical programming approaches have been the centre of research in the area of STHS problems. These include methods ranging from linear programming (LP) to MINLP covering their entire spectrum. In order to improve computational efficiency, these approaches have often been applied within the frameworks of or in conjunction with dynamic programming, genetic algorithms, evolutionary computation, artificial intelligence methods, decomposition techniques and linear approximation based approaches such as SLP. A few nonlinear programming (NLP) and MINLP approaches for optimizing power generation efficiency considering hydroelectric power as a nonlinear function of water discharge and hydraulic head (Catalão et al., 2010a) subject to ramping constraints and start/stop of units (Catalão et al., 2010b) have been proposed in literature. SLP has been used to handle state dependencies and nonlinearities in STHS problems reported in literature (Belsnes et al., 2004; Fosso and Belsnes, 2004).

The highly dynamic electricity markets cause the prices to fluctuate frequently and deviate from their nominal profiles forecasted by power producers. These slight variations in pricing sometimes warrant drastic changes in their power generation and unit commitment schedules and they are undesirable from an operational standpoint. Therefore, effective strategies need to be developed to mitigate the negative impact of this "nervousness" phenomenon on system performance. This work focuses on formulating a comprehensive mathematical model for the STHS problem, solving it using a computationally efficient technique and developing two different strategies to effectively mitigate the ill effects of "nervousness". The remainder of the paper is organized as follows. Section 2 briefly describes the developed mathematical model. The employed SLP solution technique is summarized in Section 3. The methodology adopted for "nervousness" analysis and the strategies proposed for the mitigation of its undesirable effects are outlined in Section 4. The case study and its variations used for "nervousness" analysis are presented in Section 5. Finally, Section 6 provides concluding some remarks based on the results from the application of the proposed modeling and solution methods on the cases studied for the "nervousness" analysis.

2. Mathematical Formulation

The developed mathematical formulation takes the form of a non-convex MINLP model. The main assumptions made for its development include constant piecewise inflows to all the reservoirs during all time periods in the horizon, spill flows dependent on water levels in the reservoirs, multiple generating units associated with each reservoir in the system and no start-up/shut-down or operational costs for any generating resources. The objective function in Eq. (1) is the maximization of revenues from the sales of total generated powers from all generating units over the time horizon.

$$rev = \max\left[\sum_t \left\{ep_t . \left(\sum_k P_{k,t}\right)\right\}\right] \quad (1)$$

where k, t, ep_t & $P_{k,t}$ represent the reservoirs, time periods, electricity prices and total generated powers respectively. The major constraints imposed on the system include the multivariate cubic power generation function equation, hydraulic continuity equations,

multivariate nonlinear reservoir storage – water level relationship and spill flow equations. They are mathematically expressed using Eqs. (2), (3), (4) & (5) respectively.

$$p_{i,k,t} = \left(gus_{i,k,t}.A_{k,t}\right) + \left(B_{k,t}.q_{i,k,t}\right) + C_{k,t}.\left(q_{i,k,t}\right)^2 + D_{k,t}.\left(q_{i,k,t}\right)^3, \forall\, i \in I_k, k, t$$

$$\text{where } A_{k,t} = a_1 + a_2.h_{k,t} + a_3.\left(h_{k,t}\right)^2,\ B_{k,t} = b_1 + b_2.h_{k,t} + b_3.\left(h_{k,t}\right)^2,$$

$$C_{k,t} = c_1 + c_2.h_{k,t} + c_3.\left(h_{k,t}\right)^2, D_{k,t} = d_1 + d_2.h_{k,t} + d_3.\left(h_{k,t}\right)^2, \forall\, k, t \quad (2)$$

$$v_{k,t+1} = v_{k,t} - \left(Q_{k,t} + sp_{k,t}\right) + in_{k,t} + \left(Q_{k',t} + sp_{k',t}\right), \forall\, \{k\} \neq 1,\ \{k'\} = \{k\} - 1, t$$

$$v_{k,t+1} = v_{k,t} - \left(Q_{k,t} + sp_{k,t}\right) + in_{k,t}, \forall\, \{k\} = 1, t \quad (3)$$

$$v_{k,t} = C_0 + C_1.\left(l_{k,t} - C_2\right)^{C_3} + C_4.\left\{in_{k,t} + \left(Q_{k',t} + sp_{k',t}\right)\right\}^{C_5},$$

$$\forall\, \{k\} \neq 1,\ \{k'\} = \{k\} - 1,\ \{t\} \leq |T| + 1$$

$$v_{k,t} = C_0 + C_1.\left(l_{k,t} - C_2\right)^{C_3} + C_4.\left(in_{k,t}\right)^{C_5},$$

$$\forall\, \{k\} = 1,\ \{t\} \leq |T| + 1 \quad (4)$$

$$l_{k,min} \leq l_{k,t} \leq E_k^{Sill}.\left(1 - w_{k,t}\right) + l_{k,max}.w_{k,t}, l_{k,t} \geq E_k^{Sill}.w_{k,t},$$

$$0 \leq sp_{k,t} \leq w_{k,t}.sp_{k,max}, \forall\, k, t \quad (5)$$

where i, I_k, $|T|$, $p_{i,k,t}$, $q_{i,k,t}$, $gus_{i,k,t}$, $h_{k,t}$, $v_{k,t}$, $C_0 \ldots C_5$, $Q_{k,t}$, $sp_{k,t}$, $in_{k,t}$, $l_{k,t}$, $l_{k,min}$, $l_{k,max}$, E_k^{Sill} & $w_{k,t}$ are the generating units , generators downstream of reservoirs, set of time periods, powers generated from individual units, water discharges from each generator, generating unit status binaries, hydraulic heads, storage volumes, constant coefficients in storage – level relationship, total water discharges, spill flows, inflows, water levels and their minimum and maximum allowable values, sill elevations and spill flow controlling binaries respectively. The other constraints imposed on the system in the developed mathematical model are not presented here for the sake of brevity.

3. Solution Approach

The solution methodology adopted in this work to solve the developed non-convex MINLP formulation of the STHS problem is based on SLP. This technique was employed as it was found to perform with significantly greater computational efficiency compared to the commercial local (SBB & DICOPT) & global (ANTIGONE, BARON, COUENNE & SCIP) optimization solvers for MINLP problems, which failed to solve the considered STHS problem in reasonable time to within a prespecified optimality gap (1 %). The local and global MINLP solvers failed to solve the case problem within the stipulated time frames of 10^4 s and 1 day respectively whereas the SLP method solved it within 30 s to local optimality. SLP typically involves the iterative solution of a nonlinear problem using a first-order Taylor series based linear approximation around a chosen base point. Step size adjustment/trust-region bounding techniques (Palacios-Gomez et al., 1982), penalty function methods (Zhang et al., 1985) and their combinations (Zhang, 1989) have been used to guarantee feasibility and convergence.

The proposed SLP approach involves the linearization of the nonlinear power generation and reservoir storage – water elevation equations using first-order Taylor series approximations around chosen base points of hydraulic heads and discharge flows. The STHS problem under consideration takes the form of a non-convex MINLP. The novel iterative SLP algorithm employed here involves the solution of a sequence of mixed-integer linear programming (MILP) problems instead of the traditional LP subproblems. A step-size adjustment/trust-region bounding approach similar to the one in literature (Palacios-Gomez et al., 1982) is employed here in order to ensure the convergence of this algorithm in finite steps. Global optima cannot be guaranteed for

the non-convex MINLP problem under consideration here using this SLP technique. However, the convergence of this SLP algorithm to local optima is tested by fixing the binary decision variables in the MINLP using the solutions from the SLP approach, and comparing the solution from the SLP method to that obtained from a commercial local optimization solver for the resulting NLP problem. Local optima were confirmed using the above solution technique for all instances of the case study described in Section 5.

4. "Nervousness" Analysis & Mitigation

The "nervousness" phenomenon, characterized by drastic changes in the developed optimal schedules caused by even slight variations in electricity pricing, manifests itself in the form of significant differences in the power generation and generating unit commitment profiles in the context of the STHS problem. These variations are undesirable from an operational perspective. In order to investigate this phenomenon, a set of cases are developed by perturbing the hourly varying electricity prices in the nominal profile by random amounts within certain allowable ranges of perturbation.

Two different strategies, namely Penalty SLP (PSLP) and Start-Up Costs (STUC) are developed in order to curtail the undesirable effects of this "nervousness" phenomenon. The former involves the addition of a term to the objective function penalizing the sum of deviations in flows through individual generators from one time period to the next over all time periods in the horizon, except the last one. The latter involves the introduction of start-up costs in the objective function applicable only when a particular generator switches on from the switched off condition in any time period in the decision horizon. These "nervousness" mitigation strategies are mathematically expressed below.

$$rev = \max\left[\sum_t \{ep_t.\left(\sum_k P_{k,t}\right)\}\right] - \left[w.\left(\sum_{i,k,t<|T|} \left|q_{i,k,t+1} - q_{i,k,t}\right|\right)\right] \quad (6)$$

$$rev = \max\left[\sum_t \{ep_t.\left(\sum_k P_{k,t}\right)\}\right] - \left[\sum_{i\in I_k,t} \left(stu_{i,t}.stuc_i\right)\right], \text{ where}$$

$$stu_{i,t} = gus_{i,k,t},\ \forall\ k,\ i \in I_k,\ t = 1$$

$$stu_{i,t} \geq gus_{i,k,t} - gus_{i,k,t-1},\ \forall\ k,\ i \in I_k,\ t > 1 \quad (7)$$

where w, $stu_{i,t}$ & $stuc_i$ are the small penalty weight, binary variables signifying the changes in generating unit commitment and the constant start-up costs respectively.

5. Case Studies

The case study under consideration in this work involves the operation of a hydroelectric power system consisting of 3 cascaded reservoirs and the different numbers of multiple downstream generating units associated with them for a total decision horizon of 24 hrs. discretized into 1 hr. time periods. The inflows to the 3 reservoirs are assumed to be constant at 19012, 152 & 34 cfs respectively. The tail-water elevations are assumed to be constant at 801.23, 685.52 & 472.12 ft. respectively. The initial and final water levels upstream of the 3 reservoirs are fixed as the mid-points of their ranges at 1025, 801 & 684 ft. respectively. Sill elevations are assumed to be constant at 1025, 798 & 681 ft. for the 3 reservoirs respectively. The nominal electricity price profile with hourly varying prices is shown in Fig. 1 (EP_Case D). The cases developed for "nervousness" analysis are termed Cases D1, D2 & D3 for allowable random perturbations taken from uniform distributions within ±1, ±5 & ±10 %. The case with the nominal electricity price profile is termed Case D. Multiple scenarios are generated for each of these cases with perturbed profiles but results from only one representative scenario are illustrated here for the sake of brevity. The power generation

and generating unit commitment profiles for the first reservoir (designated as R1) and its first downstream generator (designated as G11) are compared for Cases D & D1 in the left-hand plots in Figs. 1 & 2. It is clear from these figures that the generation and commitment schedules for these cases differ from each other substantially for the small variation of pricing between them. The electricity price profiles in Fig. 1 overlap with no discernible differences due to the slight variation in pricing between Cases D & D1.

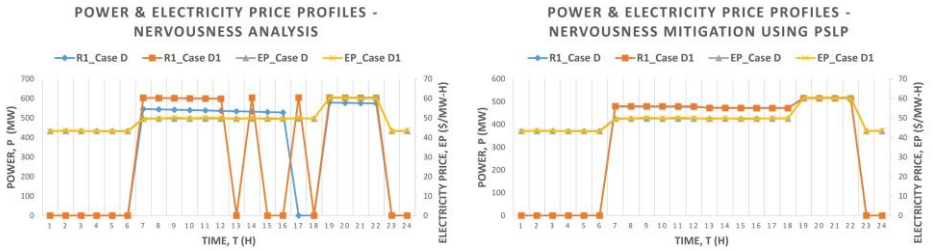


Figure 1. Comparison of power generation schedules between Cases D & D1 for R1 for "nervousness" analysis and mitigation using PSLP.

Table 1. Variation in economic objectives and generating unit commitment for "nervousness" analysis and mitigation strategies.

Case	"Nervousness" Analysis (NA)	PSLP	STUC
D (Nominal)	\$9.7585E5	\$9.7739E5	\$9.7740E5
	$\Delta_{NA} = 0$	$\Delta_{PSLP} = 0$	$\Delta_{STUC} = 0$
	$\Sigma_{NA} = 0$	$\Sigma_{PSLP} = 0$	$\Sigma_{STUC} = 0$
D1	\$9.5271E5	\$9.7615E5	\$9.7613E5
	$\Delta_{NA} = 2.37\ \%$	$\Delta_{PSLP} = 0.13\ \%$	$\Delta_{STUC} = 0.13\ \%$
	$\Sigma_{NA} = 29$	$\Sigma_{PSLP} = 0$	$\Sigma_{STUC} = 0$
D2	\$9.5322E5	\$9.6916E5	\$9.7045E5
	$\Delta_{NA} = 2.32\ \%$	$\Delta_{PSLP} = 0.84\ \%$	$\Delta_{STUC} = 0.71\ \%$
	$\Sigma_{NA} = 26$	$\Sigma_{PSLP} = 0$	$\Sigma_{STUC} = 0$
D3	\$9.6301E5	\$9.7679E5	\$9.7802E5
	$\Delta_{NA} = 1.32\ \%$	$\Delta_{PSLP} = 0.06\ \%$	$\Delta_{STUC} = 0.06\ \%$
	$\Sigma_{NA} = 39$	$\Sigma_{PSLP} = 0$	$\Sigma_{STUC} = 0$

The power generation and unit commitment schedules for the first reservoir (R1) and its first downstream generator (G11) are compared for Cases D & D1 in the right-hand plots in Figs. 1 & 2 upon the application of the PSLP approach for "nervousness" mitigation. Similar results are obtained from the application of STUC approach but are omitted for brevity. It is clear from these figures that this strategy results in smoother power generation profiles and eliminates the differences in the unit commitment schedules between them. A comparison of the economic objective values and two different performance metrics for all the cases studied for "nervousness" analysis and mitigation is given in Table 1. The first metric (Δ) is based on the differences in the objective values between the cases with perturbed and nominal electricity price profiles. The second metric (Σ) is based on the sums of differences in generator statuses for all generators over all time periods between the cases with perturbed and nominal price profiles. It is clear from Table 1 that both the PSLP & STUC approaches perform equally well in eliminating the variation in the generating unit commitment schedules. However, STUC performs slightly better compared to PSLP in terms of reducing the variation in the economic objective values, as evidenced from the results for Case D2 in Table 1. The PSLP method requires the specification of a penalty weighting parameter.

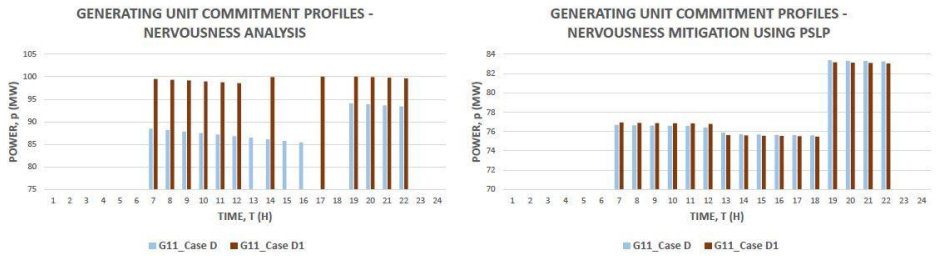


Figure 2. Comparison of generating unit commitment schedules between Cases D & D1 for R1 for "nervousness" analysis and mitigation using PSLP.

The optimal solution was found to be relatively insensitive to variations in the parameter value, provided that it remained within the same order of magnitude. The STUC approach, on the other hand, requires the specification of start-up costs for which reasonable estimates are typically available. All instances of the case study for both "nervousness" analysis and mitigation were solved using the commercial optimization solver CPLEX 12.6.3 within GAMS 24.7.4 on 3.60 GHz Intel Core i7 processor with 16 GB RAM running on Windows 10 operating system. The number of binaries, continuous variables & equations in the model for the "nervousness" analysis, PSLP and STUC case studies were 360, 2024 & 5405, 360, 2300 & 5957 and 648, 2312 & 5693 respectively. The computational times were lower than 30 s for all the cases.

6. Conclusions

The applicability and efficacy of a comprehensive MINLP formulation and computationally efficient iterative SLP solution technique for the STHS problem have been assessed in this work. The "nervousness" phenomenon has been investigated and two effective mitigation strategies (PSLP & STUC) have been applied to mitigate its severity for daily generation scheduling in a cascaded hydroelectric power system.

References

Belsnes, M. M., Roynstrand, J., Fosso, O., 2004. Handling state dependent nonlinear tunnel flows in short-term hydropower scheduling. In: Power System Technology, 2004. PowerCon 2004. 2004 International Conference on. Vol. 2. IEEE, pp. 1410–1415.

Catalão, J. P. d. S., Mariano, S., Mendes, V. M. F., Ferreira, L., 2010a. Nonlinear optimization method for short-term hydro scheduling considering head-dependency. International Transactions on Electrical Energy Systems 20 (2), 172–183.

Catalão, J. P. d. S., Pousinho, H. M. I., Mendes, V. M. F., 2010b. Scheduling of head-dependent cascaded reservoirs considering discharge ramping constraints and start/stop of units. International Journal of Electrical Power & Energy Systems 32 (8), 904–910.

Fosso, O. B., Belsnes, M. M., 2004. Short-term hydro scheduling in a liberalized power system. In: Power System Technology, 2004. PowerCon 2004. 2004 International Conference on. Vol. 2. IEEE, pp. 1321–1326.

Palacios-Gomez, F., Lasdon, L., Engquist, M., 1982. Nonlinear optimization by successive linear programming. Management Science 28 (10), 1106–1120.

Wood, A. J., Wollenberg, B. F., 2012. Power generation, operation, and control. John Wiley & Sons.

Zhang, J., 1989. Superlinear convergence of a trust region-type successive linear programming method. Journal of optimization theory and applications 61 (2), 295–310.

Zhang, J., Kim, N.-H., Lasdon, L., 1985. An improved successive linear programming algorithm. Management science 31 (10), 1312–1331.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50220-2

Reliability analysis associated with maintenance of online analyzers

Monique N. G. de Abreu[a]*, Karla Patricia S. O. R. Esquerre[a],
Ana Rosa C. de G. Massa[a], Robson W. S. Pessoa[a]

[a]*Universidade Federal da Bahia, Rua Prof. Aristídes Novis, 02, Salvador, 40210-630, Brazil*
mona.nunes@gmail.com

Abstract

This paper addresses the application of reliability analysis which has been widely accepted as an important tool in the strategy of maintenance management of systems. Reliability analysis of two online analyzers (A1 and A2) at a petrochemical company was carried out. The purpose of this work is to apply parametric and nonparametric methods to evaluate the reliability of two repairable systems and to estimate important parameters to maintenance management, such as mean time between failures (MTBF) and failure rate. The two sets of failure events underwent statistical tests in order to understand their behaviors and then their processes were modeled. It was observed that A1 presents a non-renewal process, better represented by the non-homogenous Poisson process (NHPP), while A2 is a renovation process consistent with the homogenous Poisson process (HPP). The study of A1 was complemented by the simulation the system's age at future times of consecutive repairs, and A2's study returned parameters such as MTBF and failure rate. It is expected that the results contribute to the adequacy of the planned maintenance, and that the presented tools continue to be applied, benefiting the organization maintenance management.

Keywords: reliability analysis, repairable systems, online analyzers, maintenance management.

1. Introduction

The maintenance management is important to ensure the constant improvement of equipment performance, ensuring quality operating within safe conditions for people and the environment (Jiang, 2015). According to Rausand and Hoyland (2004), with respect to a repairable system, the goal of maintenance is to restore it when necessary, but mainly to seek the improvement of the system reliability. Thus, the reliability analysis of a system provides information about its failure process, assisting in the comprehension of system operation and what to expect from its performance to better adjust the maintenance planning (Barringer, 2004). Birolini (2004) points out that reliability analysis is interrelated with cost and system effectiveness. It gives important indicators for the decision making in the maintenance management in order to attain the best possible usefulness of equipment and systems, as higher reliability in general leads to a lower operating cost.

2. A case study

The reliability analysis case study was carried out for two online analyzers (A1 and A2) which are gas chromatographs operating on a petrochemical company. Gas analyzers are of high importance for the quantitative analysis of contaminants (Thomas and Haider, 2013), in order to control and reduce the variability of the products, avoiding losses due to products out of specification.

The data mining was carried out through the company's shift book, by identifying all the failure events with respect to A1 and A2 between January 2012 and December 2016 (i.e. 1826 days). A failure event is given by any occurrence leading to the unavailability of the chromatographic analysis result.

Although A1 and A2 are essentially the same equipment operating in the same environment, they analyze distinct materials which incur different levels of aggression to the system, thus, the reliability analyses were performed separately.

3. Method

The reliability analysis procedure is structured in the following steps:

(1) Identification of all equipment failure events in the time of study

(2) Application of non-parametric graphic method for qualitative identification of possible trends in the system failure times

According to Tobias and Trindade (2012) a common data graph is the mean cumulative function (MCF), where the cumulative number of repairs is plotted against the system age t.

(3) Application of the parametric method (Poisson Process)

A number of observed systems ($m \geq 1$) is considered, with the i-th system being observed in the interval $(a_i, b_i]$, with n_i failures occurring at times T_{ij}, $j = 1, 2, \ldots, n_j$, $i=1, \ldots, m$. Defining $b_i = T_i$, if the case study is truncated at a specified time (t), n_i is the number of failures. $N_i(t)$ is the number of failures occurring in the i-th system in the interval (0, t], $t \leq b_i$ and $P[N_i(t)]$ it's the probability of occurrence (Jiang, 2015), then:

$$P[N_i(t) = \theta] = \frac{\left(\int_0^t \lambda_i(u)du\right)^{\theta} exp\left(\int_0^t \lambda_i(u)du\right)}{\theta!} \quad (1)$$

such that, $N_i(t) = \theta = 0,1,2, \ldots, n$. Thus, the average of $N_i(t)$ is given by Eq.(2):

$$\Lambda_i(t) = \int_0^t \lambda_i(u)du \quad (2)$$

where $\Lambda_i(t)$ is the accumulated intensity function until time t. Thus, $\lambda_i(t)$ is called intensity function of failures, which is required to specify the Poisson

process. A parametric form of $\lambda_i(t)$ widely used in the literature is called power law and is given by the following expression:

$$\lambda_i(t) = \frac{\beta_i}{\alpha_i}\left(\frac{t}{\alpha_i}\right)^{\beta_i - 1} \quad (3)$$

such that $t \geq 0$ and $\beta_i = 1$, where β_i is the reliability growth parameter and α_i is the scale parameter.

Thus, for a system to be characterized as a non-homogeneous Poisson process (NHPP), λ_i must obey the condition $\beta_i \neq 1$. Values $\beta_i > 1$ indicate that the i-th system is degrading, while values $\beta_i < 1$ indicate that the i-th system is improving (Birolini, 2004).

(4) Calculation of MTBF and failure rate (λ) when a process is characterized as a homogeneous Poisson process (HPP)

A HPP's failure rate corresponds to a constant intensity function ($\lambda(t) \equiv \lambda$) and is characterized as a renewal process (Lindqvist, 2007).

$$\text{MTBF} = \frac{T}{N} \quad (4)$$

$$\lambda = \frac{1}{MTBF} \quad (5)$$

with T as the final time of study and N the total number of failures at T.

(5) Simulation of future times of consecutive failures when a process is characterized as a non homogeneous Poisson process (NHPP)

A nonrenewal model, following a nonhomogeneous Poisson process, is a great challenge for maintenance planning, since costs and provisions for spare parts shall vary in time, due to nonstationary characteristics (Tobias and Trindade, 2012). It is expected a nonstable failure rate, so it not possible to obtain a single value for MTBF, as it would be possible to asses in a homogeneous Poisson process. In this way, it is intended to simulate the first N repair times for this NHPP.

An intensity function given by the Eq.(1).

$$\lambda(y) = \alpha\beta y^{\beta - 1} \quad (6)$$

It is possible to simulate as system ages in successive future repairs by Eq.(2, where $Y_1, Y_2, \ldots, Y_n$ are as system ages by generating random units $U_1, U_2, \ldots, U_n$, knowing that $Y_0 = 0$.

$$Y_i = Y_{i-1}^{\beta} - \frac{1}{\alpha}\ln(U_i)^{1/\beta} \quad (7)$$

4. Findings and discussion

4.1. Analyzer 1

Fig. 1 illustrates the mean cumulative function (MCF) for A1's failures with time. It allows us to see an indication of improvement on A1's system by the decreasing frequency of failures with time, observed by the smoothing of the curve's slope with time. From this perspective, it can be concluded that A1 has nonrenewal behavior. Table 1 leads to the same result of a nonrenewal process with respect to A1, since its failure time data failed the tests for randomness and stationarity, considering a significance level of 0.05, which led to the acceptance of the alternative hypothesis of a nonhomogeneous Poisson process.

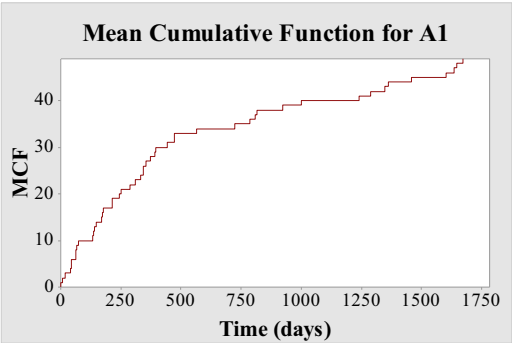


Fig. 1 – Mean Cumulative Function for A1

Table 1 - Randomness and stationarity tests (A1)

Test	H_0	T-test	p-value (α=0.05)	Process
MIL-Hdbk-189	HPP	180.30	$\ll$0.005	NHPP
La Place	HPP	-4.78	$\ll$0.005	NHPP
Anderson-Darling	HPP	14.25	$\ll$0.005	NHPP

Power relation model parameters were specified, by the performance of a parametric study on Minitab. Shape and scale parameters were obtained, by a power law process model: $\beta = 0.544$ and $\alpha = 1.299$. This result confirms the qualitative evaluation, that is, $0 < \beta < 1$ indicates that the system is improving with time. Then, using a unit uniform pseudorandom number simulator, available on Excel, it was possible to obtain the results shown on Table 2, for the first six repair times simulated.

Table 2 – Simulation of the first six repair times (A1)

A		B
1.29893		0.543532
N	U_i	t (days)
1	0.07385	3.69
2	0.46106	6.82
3	0.93614	8.77
4	0.02697	15.33
5	0.70363	18.61
6	0.09372	23.87

4.2. Analyzer 2

Fig. 2 illustrates the mean cumulative function (MCF) for A2's failures with time. Qualitatively, it is possible to say that there is no clear trending in the curve, but it is necessary to apply statistical tests to safely conclude if the variables are independent and identically distributed. Table 3 shows the results of the randomness and stationarity tests, leading to the conclusion that, at the significance level of 0.05, the null hypotheses of a homogenous Poisson process should not be rejected. Therefore, the analyzer 2 has a stable behavior and renewal process, better represented by the homogenous process of Poisson.

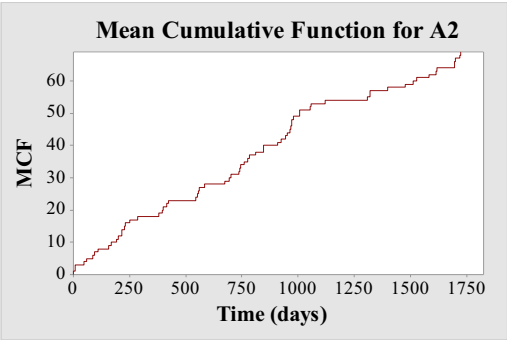


Fig. 2 – Mean Cumulative Function for A2

Table 3 - Randomness and stationarity tests (A2)

Test	Ho	Test statistic	p-value (α=0.05)	Result
MIL-Hdbk-189	HPP	165.15	0.052	HPP
La Place	HPP	-1.68	0.094	HPP
Anderson-Darling	HPP	2.05	0.086	HPP

A parametric analysis was performed on Minitab, applying the Poisson process model to calculate the MTBF and failure rate (see Table 4). Knowing these parameters is extremely important to optimize and improve maintenance operations, since it is an important indicator of the expected performance of the equipment.

Table 4 – Estimated parameters (A2)

Parameter	Estimate	Lower (95%CI)	Upper (95%CI)
MTBF (days)	25.29	19.94	32.08
λ (failures/day)	0.03953	0.03117	0.05014

Conclusions

Although the data was collected from two gas chromatographs with the same brand and model, they analyze distinct materials, which incur different factors of aggression to the equipment, requiring different maintenance planning in order to improve the performance of the analyzers. Thus two reliability analyses were carried out as a guide to maintenance planning and indeed reflected different failure process behaviors, A1 as a NHPP and A2 as a HPP. It is of upmost importance make these analyses a part of the planning management routine in order to try to control the deterioration process that leads to failures, as well as to identify future changes in the behavior patterns of the systems and uphold the consonance of the reliability analysis with maintenance management.

References

A. Birolini, 2004, Reliability Engineering: Theory and Practice, 4th ed., Scandicci, Springer.

B. Lindqvist, 2007, On the Statistical Modeling and Analysis of Repairable Systems. Statistical Science, v. 21, n. 4, p. 532–551.

M. Rausand; A. Hoyland, 2004, System Reliability Theory. 2th ed. Hoboken: Wiley-Interscience.

P. Barringer, 2004, Analyzer Reliability, In: 18th International Forum for Process Analytical Chemistry, Virginia, p. 2–9

P. Tobias; D. Trindade, 2012, Applied Reliability, 3th ed. Boca Raton, CRC.

R. Jiang, 2015, Introduction to Quality and Reliability Engineering, Berlin, Heidelberg: Springer Berlin Heidelberg.

S. Thomas; N. Haider, 2013, A Study on Basics of a Gas Analyzer, International Journal Of Advanced Research In Electrical, Electronics And Instrumentation Engineering, Chennai, v. 2, n. 12, p. 6016-6025.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50221-4

Precedence-based Lot-sizing and Scheduling Formulation with Mixing/Splitting and Shared Intermediate Vessels

Pablo A. Marchetti[a,b*], Jaime Cerdá[a]

[a]*INTEC, UNL-CONICET, Güemes 3450, (3000) Santa Fe, Argentina*
[b]*UTN Facultad Regional Santa Fe, Lavaisse 610, (3000) Santa Fe, Argentina*
pmarchet@intec.unl.edu.ar

Abstract

In this work, a new multistage batch plant lot-sizing and scheduling formulation considering mixing and splitting operations at shared intermediate storage vessels is presented. The model relies on a batch-oriented general precedence representation that handles not only sequencing decisions at processing units but also intermediate storage time and capacity constraints. The proposed representation requires a predefined set of proposed batches for each product and processing stage. From these sets, selected batches are determined by the model through binary decisions. Batch integrity is not preserved but, instead, batches can be mixed/split between consecutive stages. The model includes allocation, timing, and sequencing variables and constraints similar to previous general precedence formulations. New binary variables are only added to handle material storage and transfer decisions, where both direct transfer and transfer through multiple storage vessels is supported. Besides, the general precedence sequencing binary variables are reutilized to also handle sequencing constraints at intermediate tanks. Time and capacity constraints for intermediate vessels are defined using the sequencing and material transfer binary decisions. Small/medium size examples are presented to illustrate the application of the model.

Keywords: Batch scheduling, Lot-sizing, Intermediate storage, Mixing and splitting, MILP optimization.

1. Introduction

Scheduling plays a fundamental role in most manufacturing industries. Scheduling problems can be broadly classified in 1) sequential and 2) network processes, the latter more complex and with arbitrary topologies which allow mixing and splitting operations and even material recycle (Méndez et al., 2006). In sequential processes, with multiple consecutive stages, batches are usually preserved throughout the processing sequence. Thus, sequential methods decomposing the general problem into the lot-sizing (or batching) subproblem and the batch scheduling subproblem itself have been applied. However, when operating multistage facilities where equipment units differ in capacity or processing rate, it may be convenient to mix or split the batches to attain a better use of the resources. While monolithic methodologies (integrating batching and scheduling subproblems) have been applied to sequential processes using continuous times representations (Sundaramoorthy and Maravelias, 2008; Marchetti and Cerdá, 2012), they always assume the same batch size at every stage. In this work, the general scenario of allowing multiple batches of the same intermediate product to be

mixed at a given vessel and split among different batches to be processed in the next stage is considered. The appropriate representation of intermediate storage vessels, including storage capacity and time constraints, is critical to handle mixing and splitting operations. Shared vessels requiring sequence dependent cleaning times are considered. Additional problem characteristics include: (a) multiple orders per product with different due dates, (b) variable processing times, (c) waiting times at units, (d) direct transfer of material between consecutive stages, (e) plant topology constraints.

2. Problem Statement

In this work, the unified batching and scheduling problem for multistage facilities considering limited shared intermediate storage and mixing/splitting operations is tackled. Given are: (a) the processing stages $s \in S$ of the plant, equipped with parallel units $j \in J_s$ and storage vessels $k \in K_s$, (b) the capacity of equipment/tanks and topology limitations, (c) the required products $i \in I$, together with their due date $d \in D_i$ and demands r_{id}, (d) the set $J_{is} \subseteq J_s$ of units available for processing product $i \in I$, together with the minimum batch size, the processing time parameters, and the maximum waiting time ($wt_{ij}^{\max}$) of a batch of product i in unit j, (e) the set $K_{is} \subseteq K_s$ of tanks available to store product i between stages s and s+1, together with the minimum and maximum storage times ($rt_{ik}^{\min}$, $rt_{ik}^{\max}$), (f) the sequence-dependent changeover/cleaning times $\tau_{ii'j}/\sigma_{ii'k}$, and (g) the length of the time horizon H. The problem goal is to determine: (i) the number and size of batches to be processed, (ii) the unit and storage allocation decisions, (iii) the processing sequence at each unit and tank, and (iv) the starting and completion times of each processing task, such that production orders and operational/storage constraints are all satisfied at minimum total tardiness or makespan.

We assume that a batch can be transferred to one or more tanks, but just a single transfer operation for each batch-tank pair is allowed. Also, transfer times are negligible, and loading and unloading operations do not overlap. The storage time available when batches are mixed is limited to the lowest remaining storage time of the parts. Moreover, there is unlimited storage capacity at the last stage and enough capacity of additional resources at every stage.

3. Mathematical Model

The detailed lot-sizing and scheduling mixed-integer linear programming (MILP) model of Marchetti et al. (2012) is used as a basis for the proposed approach. Please refer to this work for details of the model and the precedence based representation used.

3.1. Proposed batches for each processing stage

A set of tentative batches B_{is} is proposed for each product i and processing stage s. The set size, $n_{is} = |B_{is}|$, is chosen large enough to guarantee the optimality of the best solution found. We first consider the last processing stage (s^{ℓ}), where the approach of Marchetti et al. (2012) is used. At this stage, a set of batches B_{id}^{ℓ} is proposed for each production order (i, d). Batches $b \in B_{id}^{\ell}$ can be used not only to fulfil order (i, d) but also, if they are large enough, to meet future ones. In contrast, a different set of batches is used for each intermediate stage $s \in S' = S - \{s^{\ell}\}$. In this case, the parameter n_{is} is defined by Eq. (1) based on the overall demand of product i, $R_i = \sum_{d \in D_i} r_{id}$.

$$n_{is} = \lceil R_i / bs_{is} \rceil \qquad \forall i \in I, s \in S' \tag{1}$$

3.2. Storage allocation

The binary variable Y_{bk}^{S} is introduced to denote the allocation of batch b to storage tank $k \in K_{is}$. Using this variable, Eq. (2) is required in the general case where each batch can be stored in multiple tanks. The binary variable SEL_b represents the selection of batch b.

$$Y_{bk}^{S} \le SEL_b \qquad \forall k \in K_{is}, b \in B_{is}, i \in I, s \in S' \tag{2}$$

3.3. Material transfer representation

Transfer operations are determined using binary variables $V_{bb'}$ and $W_{bb'k}$. These variables represent the direct transfer between units (when $V_{bb'} = 1$) and the transfer through storage vessel k (when $W_{bb'k} = 1$) of intermediate product from batch b to batch b'. Non-negative continuous variables $\beta_{bb'}$ and $\eta_{bb'k}$ account for the volumes being transferred in each case. Appropriate upper-bound constraints are also included to guarantee that $\beta_{bb'}$ and $\eta_{bb'k}$ are zero whenever $V_{bb'}$ and $W_{bb'k}$, respectively, are not selected. Provided that BS_b is the batch size, using these variables the required material balance equations are:

$$BS_b = \sum_{b' \in B_{i(s+1)}} \left(\beta_{bb'} + \sum_{k \in K_{is}} \eta_{bb'k} \right) \qquad \forall b \in B_{is}, i \in I, s \in S' \tag{3}$$

$$BS_{b'} = \sum_{b \in B_{is}} \left(\beta_{bb'} + \sum_{k \in K_{is}} \eta_{bb'k} \right) \qquad \forall b' \in B_{i(s+1)}, i \in I, s \in S' \tag{4}$$

Also, specific constraints are included to guarantee that $V_{bb'}$ and $W_{bb'k}$ are zero whenever batches b, b' are not selected. The same to ensure that $W_{bb'k}$ is zero when $Y_{bk}^{S} = 0$. Since either direct transfer or transfer through a storage tank can be selected, the following constraint is needed:

$$V_{bb'} + W_{bb'k} \le SEL_{b'} \qquad \forall b \in B_{is}, b' \in B_{i(s+1)}, k \in K_{is}, i \in I, s \in S' \tag{5}$$

3.4. Intermediate storage time constraints

Timing constraints for transfer operations are considered next. On the one hand, if some portion of batch b at stage s is directly transferred to batch b' at stage s+1 ($V_{bb'} = 1$), then b' should start anytime between CT_{bs} and $CT_{bs} + WT_{bs}$, where WT_{bs} is the associated waiting time at the unit.

$$\left.\begin{aligned} CT_{bs} - H\left(1 - V_{bb'}\right) &\le ST_{b'(s+1)} \\ ST_{b'(s+1)} &\le CT_{bs} + WT_{bs} + H\left(1 - V_{bb'}\right) \end{aligned}\right\} \quad \forall b \in B_{is}, b' \in B_{i(s+1)}, i \in I, s \in S' \tag{6}$$

On the other hand, in order to receive intermediate product from batch b stored at vessel k, each selected batch $b' \in B_{i(s+1)}$ should start anytime between $CT_{bs} + T_{bk}^{W}$ and $CT_{bs} + T_{bk}^{W} + T_{bk}^{S}$. The continuous variables T_{bk}^{W} and T_{bk}^{S} are the waiting time before and the storage time after, respectively, batch b is transferred to vessel k. A combined maximum storage time including both waiting and residence times is accounted for by Eq. (8). Besides, the condition $T_{bk}^{W} \le WT_{bs}$ is required to guarantee that WT_{bs} is large enough so that every transfer takes place.

$$\left.\begin{aligned} CT_{bs} + T_{bk}^{W} - H\left(1 - W_{bb'k}\right) &\le ST_{b'(s+1)} \\ ST_{b'(s+1)} &\le CT_{bs} + T_{bk}^{W} + T_{bk}^{S} + H\left(1 - W_{bb'k}\right) \end{aligned}\right\} \quad \forall b \in B_{is}, b' \in B_{i(s+1)}, k \in K_{is}, i \in I, s \in S' \tag{7}$$

$$T^W_{bk} + T^S_{bk} \le rt^{\max}_{is}\, Y^S_{bk} \qquad \forall b \in B_{is}, k \in K_{is}, i \in I, s \in S' \tag{8}$$

In order to calculate the storage time of mixed batches, a conservative approach is taken in this contribution. As stated in Section 2, if two or more batches are mixed at a given vessel, the maximum storage time of the mixture is the minimum time still available for any of the batches. This condition is expressed by ensuring that, for any pair of batches b and b' that are mixed in vessel k, the storage interval of both finishes at the same time (i.e., $CT_{bs} + T^W_{bk} + T^S_{bk} = CT_{b's} + T^W_{b'k} + T^S_{b'k}$). To detect if batches b and b' get mixed at tank k, it is enough to check if there is any batch b'' at stage s+1 to which both b and b' transfer intermediate product. To this end, big-M constraints activated by the condition $W_{bb''k} + W_{b'b''k} = 2$ are used, for all $b'' \in B_{i(s+1)}$. Additional constraints based on sequencing and transfer decisions are included to ensure that the condition represented by Eq. (7) does not hold when material transfer is not selected.

3.5. Sequencing constraints for storage vessels

Let the binary variable $Y^U_{b'k}$ represent that product from vessel k is transferred to batch b' at the next stage (when $Y^U_{b'k} = 1$). Besides, let the continuous variable $T^U_{b'k}$ be the residence time of the material in k before it is transferred to b'. Eq. (9) guarantees that the appropriate values for $Y^U_{b'k}$ and $T^U_{b'k}$ are chosen. When both b and b' are loaded from the same vessel k (i.e., $Y^U_{bk} = 1$ and $Y^U_{b'k} = 1$) the sequencing variable $X_{bb'(s+1)}$ determines the relative ordering of the storage tasks at stage s.

$$\left.\begin{aligned} W_{bb'k} &\le Y^U_{b'k} \\ ST_{b'(s+1)} - CT_{bs} - T^W_{bk} - H\left(1 - W_{bb'k}\right) &\le T^U_{b'k} \end{aligned}\right\} \quad \forall b \in B_{is}, b' \in B_{i(s+1)}, k \in K_{is}, i \in I, s \in S' \tag{9}$$

$$\left.\begin{aligned} ST_{b(s+1)} + \sigma_{ii'k} &\le ST_{b'(s+1)} - T^U_{b'k} + H\left(3 - X_{bb'(s+1)} - Y^U_{bk} - Y^U_{b'k}\right) \\ ST_{b'(s+1)} + \sigma_{i'ik} &\le ST_{b(s+1)} - T^U_{bk} + H\left(2 + X_{bb'(s+1)} - Y^U_{bk} - Y^U_{b'k}\right) \end{aligned}\right\} \quad \begin{array}{l} \forall b \in B_{i(s+1)}, b' \in B_{i'(s+1)}, \\ k \in K_{ii's}, (i,i') \in I, s \in S' : i < i' \end{array} \tag{10}$$

3.6. Storage capacity

In order to define the storage capacity constraints, we introduce the continuous variable $L_{bb'k}$ representing the volume of batch b that remains in tank k immediately after batch b' starts at stage s+1. Different constraints are used to set the value of $L_{bb'k}$. On the one hand, constraint (11) ensures that both $\eta_{bb'k}$ and $L_{bb'k}$ are driven to zero when $W_{bb'k} = 0$. On the other hand, if $W_{bb'k} = 1$, appropriate bounds are set for the storage level variables $L_{bb'k}$ using Eqs. (12). Because of Eqs. (12), the volume of batch b that remains at tank k will decrease while it is transferred to consecutive batches, according to the order determined by the sequencing decisions at stage s+1. Moreover, the batches of $B_{i(s+1)}$ are sequenced avoiding redundant solutions (using an approach similar to constraints 11 and 12 of Marchetti and Cerdá, 2009).

$$L_{bb'k} + \eta_{bb'k} \le q^{\max}_{ik}\, W_{bb'k} \qquad \forall b \in B_{is}, b' \in B_{i(s+1)}, k \in K_{is}, i \in I, s \in S' \tag{11}$$

$$\left.\begin{aligned} L_{bb''k} + \eta_{bb''k} &\le L_{bb'k} + q^{\max}_{k}\left(2 - W_{bb'k} - X_{b'b''(s+1)}\right) \\ L_{bb'k} + \eta_{bb'k} &\le L_{bb''k} + q^{\max}_{k}\left(1 - W_{bb''k} + X_{b'b''(s+1)}\right) \end{aligned}\right\} \tag{12}$$

$$\forall b \in B_{is}, (b', b'') \in B_{i(s+1)}, k \in K_{is}, i \in I, s \in S' : \left(b' < b''\right)$$

Finally, storage capacity constraints for stage s are verified at the starting times of batches $b \in B_{i(s+1)}$ (at the next stage). Since any batch $b' \in B_{is}$ can both transfer product to b through k ($\eta_{b'bk} > 0$), or remain at tank k after batch b starts at stage s+1 ($L_{b'bk} > 0$), then the total volume stored at k before time $ST_{b(s+1)}$ is given by the LHS of Eq. (13), which is bounded by the capacity of the tank, $q_k^{\max}$.

$$\sum_{b' \in B_{is}} \left(L_{b'bk} + \eta_{b'bk} \right) \le q_k^{\max} \qquad \forall b \in B_{i(s+1)}, k \in K_{is}, i \in I, s \in S' \tag{13}$$

The application of Eqs. (11)-(13) taking into account the bounds on variables $\eta_{b',b,k}$ and $L_{b',b,k}$ is depicted in Figure 1(a). In this schedule, three batches b_1-b_3 at stage s_1 are mixed at vessel k_1 and later split among batches b_4 to b_6 of stage s_2. A more detailed inventory balance for tank k_1 is shown in Figure 1(b). It is important to remark that the solution shown in Figs. 1(a-b) is one of many alternatives to distribute the mixed volume of batches b_1-b_3 among batches b_4-b_6.

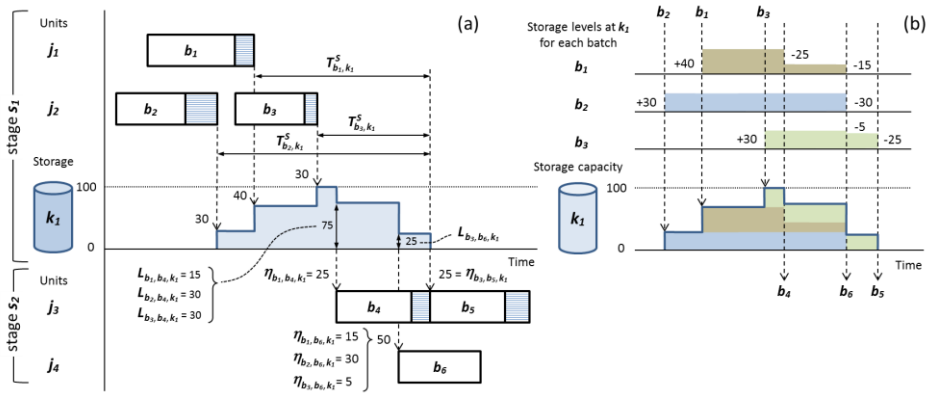


Figure 1: (a) Illustrative schedule considering multiple intermediate storage mixing and splitting operations; (b) Individual storage levels for each batch at k_1.

4. Case Studies

Both a small and a medium size example are presented to illustrate the application of the proposed model. A batch plant including two stages (S_1: mixing, S_2: reaction) with a total of 5 processing units and 2 intermediate vessels is considered. The problem goal is to minimize the overall tardiness. In example 1, two products (P_1 and P_2) are produced and a schedule for 3 orders is required. Different material handling alternatives for the intermediate products at stage S_1 are compared. Instance 1a assumes that batches are preserved between the consecutive stages and UIS is available. Instance 1b considers finite storage instead. In contrast, instance 1c includes both mixing and splitting and limited storage capacity (i.e., the proposed formulation). Finally, a medium size test case considering 3 products and 7 orders at different due dates (example 2) is solved with the proposed approach. Model size and computational results are reported in Table 1. The solution of instance 1c features the best overall tardiness of example 1. Since a more complex problem is addressed, 1c requires more computational effort than 1a and 1b. The medium size example 2 is considerably more complex than problem 1c. However, the required CPU time is still reasonable. The optimal solutions found for instance 1c and example 2 are shown in Figs. 2 and 3, respectively. Here, the relations between selected/proposed batches are 14/19 and 27/39, for each case.

Table 1. Test case statistics and computational results*

Example	Binary vars.	Cont. vars.	Constr.	Optimal solution (h)	CPU time (s)	Nodes	Iter./node
1a	136	116	568	3.42	0.66	343	4.99
1b	156	166	934	3.72	0.90	367	6.96
1c	300	497	3656	1.56	12.54	8277	28.68
2	903	1204	12206	0.70	280.58	35443	36.44

* Intel Core i7 3.2 GHz machine with GAMS/CPLEX 12.7.1.

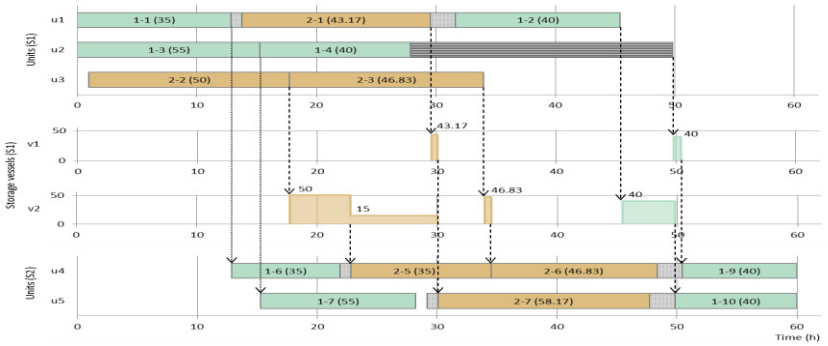


Figure 2. Small illustrative example optimal solution (Instance 1c).

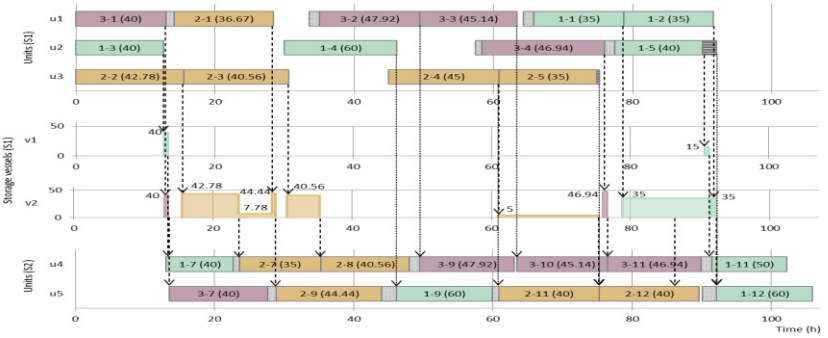


Figure 3. Optimal solution found for Example 2.

5. Conclusions

While the utilization of global-precedence sequencing decisions to set the relative ordering of processing tasks at equipment units is well known, this sequencing scheme has not been used at intermediate storage vessels to handle mixing and splitting operations before. Problems with increasing number of production orders, units or vessels are still difficult to solve. The proposed methodology bridges the gap to allow future applications of precedence-based sequencing schemes to general network-based multipurpose batch scheduling problems.

References

P.A. Marchetti, J. Cerdá, 2009, *Comput. Chem. Eng.*, 33(4), pp. 871-886.
P.A. Marchetti, J. Cerdá, 2012, *Ind. Eng. Chem. Res.*, 51, pp. 5762-5780.
C.A. Méndez et al., 2006, *Comput. Chem. Eng.*, 30, pp. 913-946.
A. Sundaramoorthy, C.T. Maravelias, 2008, *Ind. Eng. Chem. Res.*, 47(5), pp. 1546-1555.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50222-6

Strategic Planning of SAGD Reservoir Development under Exogenous and Endogenous Uncertainty

Farough Motamed Nasab, Hossein Shahandeh, Zukui Li*

Department of Chemical and Materials Engineering, University of Alberta, Edmonton, Alberta, T6G 1H9, Canada
zukui@ualberta.ca

Abstract

In this paper, a stochastic optimization framework for planning the development of SAGD drainage areas under uncertainty is presented. A multistage stochastic programming based optimization model is developed to arrange the multiperiod development plan of the drainage areas with consideration of reservoir production capacity and bitumen product price uncertainty. A linear decision rule based method is presented to handle the two different type of uncertainties. Results demonstrate that the method can effectively generate an economically optimal development plan that maximizes the profit under uncertainty.

Keywords: SAGD reservoir; strategic planning; endogenous and exogenous uncertainty; multistage stochastic programming; decision rule approximation

1. Introduction

Addressing uncertainty in reservoir development planning is critical since it significantly affects the economic performance of a project. Among the various sources of uncertainties, oil price and reservoir production capacity represent the major uncertain factors to be considered. In the past, stochastic programming has received lots of attention in the context of oil and gas investment and production planning under uncertainty. There are few number of publications addressing oil and gas planning problems including endogenous uncertainty. Jonsbraten (1998) discusses reservoir optimization and production uncertainty under reservoir size uncertainty and price uncertainty. Goel and Grossmann (2004) used a scenario-based and decomposing-based approximation approach for operational planning of gas field under uncertainty in gas reserves. Goel et al. (2006) discusses the problem of optimal investment and operational planning for development of gas fields under uncertainty in gas reserves and Lagrangean duality based branch and bound algorithm is used to solve the problem. Gupta and Grossmann (2014) proposed a new decomposition algorithm based on relaxation of non-anticipatively constraints that results in reduced computational expense while preserving the same optimality gap. All of the above mentioned methods required the discretization of uncertainty space which is computationally inefficient for large-scale problems. Recently Vayanos et al. (2011) proposed an algorithm for stochastic problems with endogenous uncertainty inspired by robust optimization methods that do not require discretization of uncertainty or relaxation of non-anticipativity constraints.

In this work, long term development planning of the oil sands steam assisted gravity drainage (SAGD) is addressed. The goal of this study is to maximize the net present value from the SAGD project during a long term time horizon while considering price and oil production uncertainties, capital restriction, processing capacity and steam generation limitation. We present a solution method that effectively incorporates the effect of price and oil production uncertainties. The problem is solved using linear decision rule technique.

2. Problem statement

In SAGD operations, usually several pairs of horizontal injector and producer well pairs are installed into a drainage area (DA) from a surface pad. Hot steam is injected into the upper injector well, as a result steam chamber grows and propagates into the reservoir which liquidizes the bitumen and subsequently bitumen is pumped out. The produced bitumen is processed in the central processing facility, where oil-water emulsion is separated and produced water is recycled and used for steam generation. In strategic planning of SAGD development, the development sequence and timing is optimized. In addition, the production rate plan is determined. To plan the development of the multiple drainage areas, we consider two important uncertainties: reservoir production capacity and oil price.

For reservoir production capacity uncertainty, it is assumed that for each DA, production capacity follows a uniform distribution around a nominal value. This reflects uncertainty in different reservoir geologic parameters such as porosity, permeability, saturation, etc. The uncertainty is modeled as $\tilde{q}_{i,k}^{maxoil} = q_{i,k}^{maxoil}(1+\zeta_i), \forall i \in \mathcal{I}, \forall k \in \mathcal{T}$, where $q_{i,k}^{maxoil}$ is the nominal value for oil production and $\tilde{q}_{i,k}^{maxoil}$ is the uncertain oil production capacity of i-th DA at year k. It is worth pointing out that the production capacity uncertainty is an endogenous uncertainty since it is revealed once the well is developed.

To model the oil price uncertainty, Auto Regressive Moving Average ARMA(p, q) is used: $OP_t = \{\phi_1 OP_{t-1} + \cdots + \phi_p OP_{t-p}\} + \{\epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}\}$, where $\phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q$ are model parameters that can be estimated from history data, ε_t represents uncorrelated white noise (with zero mean and constant variance). To predict the future oil price from time t+1 to t+T, we define $\epsilon = [\epsilon_{t+1}, \cdots, \epsilon_{t+T}]^T$. Then the future oil price can be compactly written as: $OP_{t+i} = A_{t+i}\epsilon + b_{t+i}, i \in \{1, \cdots, T\}$, where A_t is a vector, b_t is a scalar, both calculated using the model parameters. Oil price uncertainty is exogenous uncertainty since it is not depending on the decisions to be made.

For the stochastic planning problem, the following uncertainty set is constructed:

$$\Xi = \{\xi = [1, \zeta_1, \cdots, \zeta_N, \epsilon_1, \cdots, \epsilon_T]^T : \zeta_i \in [\underline{\zeta}_i, \overline{\zeta}_i], \forall i, \ |\epsilon_t| \le z_{1-\alpha}, \ \sum_t |\epsilon_t| \le \Gamma z_{1-\alpha}, A_t\epsilon + b_t \ge 10, \forall t\}$$

The constraints applied in the uncertainty set are explained as following: 1) For each drainage area i, $\zeta_i \in [\underline{\zeta}_i, \overline{\zeta}_i]$; 2) For oil price, the variation range of white noise is controlled by applying conditions on each ϵ_t such that $|\epsilon_t| \le z_{1-\alpha}$, with $[-z_{1-\alpha}, z_{1-\alpha}]$ being the $1-\alpha$ confidence level (CI) of ϵ_t, which follows zero mean normal distribution; 3) The total change of budget is controlled by: $\sum_t |\epsilon_t| \le \Gamma z_{1-\alpha}$, where Γ is

a scalar to control range of white noise change, t is time in project horizon; 4) Considering minimum oil price \$10/bbl: $OP_t(\xi) = A_t P\xi + b_t \geq 10, \forall t$. The above uncertainty set is a convex polyhedral, which can be compactly written as: $\Xi = \{\xi : W\xi \geq \nu\}$.

3. Multistage stochastic programming model

The proposed model aims to determine an optimal development sequence and production plan that maximizes the net present value (NPV) over a given time horizon.

$$\max \quad \sum_{t\in\mathcal{T}} \frac{1}{(1+DR)^t} \mathbb{E}\left[(1-RY)OP_t(\xi)\sum_{i\in\mathcal{I}} q_{i,t}^{oil}(\xi) - DA_{cost}\sum_{i\in\mathcal{I}} z_{i,t}(\xi) - FC\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}_i} z_{i,t-k+1}(\xi) q_{i,k}^{steam} N_w\right] \tag{1a}$$

$$s.t. \quad DA_{cost}\sum_{i\in\mathcal{I}} z_{i,t}(\xi) \leq CC_{DA} \qquad \forall t\in\mathcal{T}, \xi\in\Xi \tag{1b}$$

$$\sum_{t\in\mathcal{T}} z_{i,t}(\xi) \leq 1 \qquad \forall i\in\mathcal{I}, \xi\in\Xi \tag{1c}$$

$$q_{i,t}^{oil}(\xi) \leq \sum_{k\in\mathcal{K}_i} z_{i,t-k+1}(\xi)\, \tilde{q}_{i,k}^{maxoil}(\xi) N_w \qquad \forall i\in\mathcal{I}, t\in\mathcal{T}, \xi\in\Xi \tag{1d}$$

$$\sum_{i\in\mathcal{I}} q_{i,t}^{oil}(\xi) \leq Cap_{CPF} \qquad \forall i\in\mathcal{I}, \xi\in\Xi \tag{1e}$$

$$\sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}_i} z_{i,t-k+1}(\xi)\, q_{i,k}^{steam} N_w \leq Cap_{steam} \qquad \forall t\in\mathcal{T}, \xi\in\Xi \tag{1f}$$

$$q_{i,t}^{oil}(\xi) \geq 0 \qquad \forall i\in\mathcal{I}, t\in\mathcal{T}, \xi\in\Xi \tag{1g}$$

$$x_{1,t}(\xi) = 1 \qquad \forall t\in\mathcal{T} \tag{1h}$$

$$x_{i+1,t}(\xi) = \sum_{t'\leq t} z_{i,t'}(\xi) \qquad \forall i\in\mathcal{I}, t\in\mathcal{T}, \xi\in\Xi \tag{1i}$$

$$x_{l,t}(\xi) = 0 \qquad \forall l \geq N_{DA}+3, t < l-(1+N_{DA}), \xi\in\Xi \tag{1j}$$

$$x_{l,t}(\xi) = 1 \qquad \forall l \geq N_{DA}+2, t \geq l-(1+N_{DA}), \xi\in\Xi \tag{1k}$$

$$q_{i,t}^{oil}(\xi) = q_{i,t}^{oil}(x_{t-1}(\xi)\bullet\xi) \qquad \forall i\in\mathcal{I}, t\in\mathcal{T}, \xi\in\Xi \tag{1l}$$

$$z_{i,t}(\xi)\in\{0,1\}, x_{l,t}(\xi)\in\{0,1\} \qquad \forall i\in\mathcal{I}, t\in\mathcal{T}, \xi\in\Xi \tag{1m}$$

In the above model, index *i*, *t*, *k* refer to DA number, the year within the lifetime of the project, and the year within life time of a specific DA, respectively. Furthermore, $z_{i,t}$ is a binary variable indicating whether *i*-th drainage area is developed at year *t* or not, binary variable $x_{i,t}$ denotes the information revealing status, and $q_{i,t}^{oil}(\xi)$ represents the oil production plan. Objective function (1a) is the Net Present Value (NPV) calculated based on the total annual revenue, the capital cost on DA investment, and the operating cost from steam generation which requires fuel consumption. *DR* is the discount rate and *RY* is royalty parameter. Constraints (1b) enforce capital cost of the project in each year. Constraints (1c) enforce that each DA starts only once. Constraints (1d) enforce that the oil production rate is less than the upper bound computed based on the reservoir model parameters. Constraints (1e) enforces that the total oil production at each year must be less than the central processing facility capacity. Constraints (1f) compute the

total steam requirements and enforces it is less than the steam generation capacity. Constraints (1g) enforce nonnegativity on oil production rate. Non-anticipativity is modeled through equation (1h) to (1l). Constraint (1h) is used for the trivial element 1 in the uncertain vector ξ. The logic relation between information revealing status variable x and development decision z is modeled by constraint (1i). For price uncertainty, the information revealing status variable is only dependent on the time, this is modeled by constraint (1j) and (1k). Constraint (1l) represents non-anticipativity constraints for production decision variables (it is only decided based on information revealed based on Hadamard product. In this study, linear decision rule (LDR) based counterpart model derivation is applied to the constraints. For continuous variables, the following decision rule is applied: $q_{i,t}^{oil}(\xi) = Q_{i,t}\xi,\ \forall \xi \in \Xi, \forall i \in \mathcal{I}, t \in \mathcal{T}$. The LDR based counterpart of the stochastic model is summarized below:

$$\max \sum_{t\in\mathcal{T}} \frac{1-RY}{(1+DR)^t}\{tr(P^T A_t^T \sum_{i\in\mathcal{I}} Q_{i,t}\,\mathbb{E}(\xi\xi^T)) + b_t \sum_{i\in\mathcal{I}} Q_{i,t}\mathbb{E}(\xi)\}$$

$$- \sum_{t\in\mathcal{T}} \frac{DA_{cost}}{(1+DR)^t} \sum_i z_{i,t} - \sum_{t\in\mathcal{T}} \frac{N_w FC}{(1+DR)^t} \sum_i \sum_{k\in\mathcal{K}_i} z_{i,t-k+1} q_{i,k}^{steam} \quad (2a)$$

$$s.t.\ DA_{cost} \sum_{i\in\mathcal{I}} z_{i,t} \le CC_{DA} \qquad \forall t \in \mathcal{T} \quad (2b)$$

$$\sum_{t\in\mathcal{T}} z_{i,t} \le 1 \qquad \forall i \in \mathcal{I} \quad (2c)$$

$$-\nu\Lambda_{i,t} - \sum_{k\in\mathcal{K}_i} z_{i,t-k+1}\, q_{i,k}^{maxoil} N_w \le 0 \qquad \forall i \in \mathcal{I}, t \in \mathcal{T} \quad (2d)$$

$$-W\Lambda_{i,t} = Q_{i,t} - \sum_{k\in\mathcal{K}_i} z_{i,t-k+1}\, q_{i,k}^{maxoil} N_w e_i \quad \forall i \in \mathcal{I}, t \in \mathcal{T}$$

$$\Lambda_{i,t} \ge 0 \qquad \forall i \in \mathcal{I}, t \in \mathcal{T}$$

$$-\nu\lambda_t \le Cap_{CPF} \qquad \forall t \in \mathcal{T} \quad (2e)$$

$$-W\lambda_t = \sum_{i\in\mathcal{I}} Q_{i,t} \qquad \forall t \in \mathcal{T}$$

$$\lambda_t \ge 0 \qquad \forall t \in \mathcal{T}$$

$$\sum_{i\in\mathcal{I}} \sum_{k\in\mathcal{K}_i} z_{i,t-k+1}\, q_{i,k}^{steam} N_w \le Cap_{steam} \qquad \forall t \in \mathcal{T} \quad (2f)$$

$$\nu\mu_{i,t} \ge 0 \qquad \forall i \in \mathcal{I}, t \in \mathcal{T} \quad (2g)$$

$$W\mu_{i,t} = Q_{i,t} \qquad \forall i \in \mathcal{I}, t \in \mathcal{T}$$

$$\mu_{i,t} \ge 0 \qquad \forall i \in \mathcal{I}, t \in \mathcal{T}$$

$$x_{1,t} = 1 \qquad \forall t \in \mathcal{T} \quad (2h)$$

$$x_{i+1,t} = \sum_{t'\le t} z_{i,t'} \qquad \forall i \in \mathcal{I}, t \in \mathcal{T} \quad (2i)$$

$$x_{l,t} = 0 \qquad \forall l \ge N_{DA}+3, t < l-(1+N_{DA}) \quad (2j)$$

$$x_{l,t} = 1 \qquad \forall l \ge N_{DA}+2, t \ge l-(1+N_{DA}) \quad (2k)$$

$$|Q_{i,t,l}| \le M x_{l,t-1} \qquad \forall i \in \mathcal{I}, l \in \mathcal{L}, t \in \mathcal{T} \quad (2l)$$

$$z_{i,t} \in \{0,1\}, x_{l,t} \in \{0,1\} \qquad \forall i \in \mathcal{I}, l \in \mathcal{L}, t \in \mathcal{T} \quad (2m)$$

The above LDR based stochastic planning model are both mixed integer linear optimization (MILP) problem, which can be solved using stand MILP solver.

4. Case study

In this section, an example problem is solved to investigate the proposed planning method. There are 43 DAs to be planned and each DA includes 6 pairs of wells. The project horizon is considered to be 25 years. The life time of the DAs varies from 6 to 10 years. All the optimization problems are modeled and solved in GAMS using the CPLEX solver. Relative optimality gap of 0.1% was considered as the stopping criteria. In this case, central processing facility capacity = 5.25×10^6 m^3/year and steam capacity = 16×10^6 m^3/year. For stochastic planning model, in order to investigate the effect of price uncertainty on NPV, 15 experiments were conducted using different Γ and confidence interval (CI) values. Reservoir production capacity perturbation are fixed at ±10% of the nominal value. Figure 1 illustrates the stochastic planning solution with Γ=10 and CI=95%. The plan shows the development sequence, lifetime of each DA, the overall operating well number, the NPV and the total production rate from all the wells.

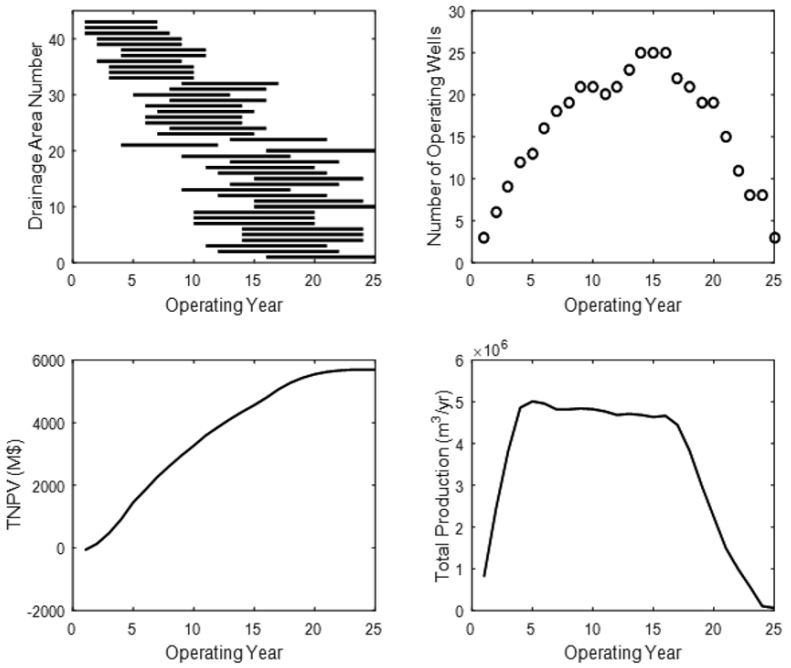


Figure 1. Stochastic planning solution

Table 1. Effect of price uncertainty set size on NPV

	CI=99%	CI=95%	CI=90%
Γ=25	6111.14	5692.72	5359.98
Γ=15	6096.67	5695.19	5365.10
Γ=10	6060.36	5546.57	5007.61
Γ=7	5475.82	4678.88	4216.29
Γ=5	4528.86	3897.52	3665.57

Table 1 illustrates the obtained results. For greater values of Γ and CI, the price uncertainty space is wider and since a minimum price constraint (bitumen price >10$/bbl) is applied in the uncertainty set, wider price uncertainty set results in greater bitumen price values and the NPV increases accordingly.

The performance of the stochastic planning solution was compared with the deterministic planning solution through sampling evaluation. Oil price uncertainty parameter confidence interval factor is fixed at 95% and Γ varies in the range of 5 to 25. For each Γ value, 1000 samples from the corresponding uncertainty set are generated and sampling method is applied on the results of stochastic and deterministic solutions. The advantage of stochastic programming is the feasibility of the solution. In stochastic solution, for all the different uncertainty sets (different Γ values), mean constraint violation is always zero which means stochastic optimization provides feasible solutions. While deterministic solution in general results in 16% of oil flow constraint violations among the 1000 samples. As a trade-off, the stochastic programming model is computationally more demanding, it takes 8 minutes to 4 hours for the various cases solved. While for the deterministic model, it is solved in 2 seconds.

5. Conclusions

In this work, multistage stochastic optimization model for SAGD reservoir development planning was developed. The objective of the models was to maximize the revenue obtained from the SAGD operation during the project horizon considering restrictions on capital investment, steam generation and oil processing capacities. The stochastic model was developed to incorporate oil price and oil production uncertainty. The principle advantage of the stochastic model is the feasibility of the solution subject to price and production uncertainty while the deterministic solution always resulted in infeasible plans violating realistic restrictions. Finally, since in this study a linear decision rule was employed in order to solve the nonlinear stochastic problem, the obtained solution was suboptimal. Future work includes improving the solution quality through better decision rule.

References

R. M. Z Butler, (1991). Thermal recovery of oil and bitumen.

V. Goel, I. E. Grossmann, (2004). A stochastic programming approach to planning of offshore gas field developments under uncertainty in reserves. Computers & Chemical Engineering, 28(8), 1409–1429.

V. Goel, I. E. Grossmann, A. S. El-Bakry, E. L. Mulkay, (2006). A novel branch and bound algorithm for optimal development of gas fields under uncertainty in reserves. Computers & Chemical Engineering, 30(6), 1076–1092.

V. Gupta, I. E. Grossmann, (2014). Multistage stochastic programming approach for offshore oilfield infrastructure planning under production sharing agreements and endogenous uncertainties. Journal of Petroleum Science and Engineering, 124, 180–197.

T. W. Jonsbr˚aten, (1998). Optimization models for petroleum field exploitation.

S. Rahim, Z. Li, (2015). Reservoir geological uncertainty reduction: an optimization-based method using multiple static measures. Mathematical Geosciences, 47(4), 373–396.

H. Shahandeh, S. Rahim, Z. Li, (2016). Strategic optimization of the oil sands development with sagd: Drainage area arrangement and development planning. Journal of Petroleum Science and Engineering, 137, 172–184.

P. Vayanos, D. Kuhn, B. Rustem, (2011). Decision rules for information discovery in multi-stage stochastic programming. In Decision and control and european control conference (cdc-ecc), 2011 50th ieee conference on (pp. 7368–7373).

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50223-8

A Closed-loop Dynamic Rescheduling Strategy Applied to Chemical Production Problems

Frederico S. de Paula[a], Sérgio M. S. Neiro[a], Valéria V. Murata[a], Luís C. Oliveira-Lopes[a*]

[a]*School of Chemical Engineering, Federal University of Uberlândia, Av. João Naves de Ávila, 2121, Santa Mônica, CEP 38400-902, Uberlândia, Minas Gerais.*
[*]*lcol@ufu.br*

Abstract

This work presents a rescheduling strategy based on a model predictive closed-loop algorithm with shrinking horizon based on uncertain models. This formulation can address most of the rescheduling needs of a chemical production operation. The strategy is based on a quantitative model to forecast the effects of process operation changes and disturbances in the production goal by using uncertain models. The studied problem is based on a model predictive control algorithm that uses a set of state-space models to describe all possible scenarios with uncertain properties. It also evaluates the properties and features of the closed-loop rescheduling algorithm based on closed-loop performance objectives. A comparison with standard scheduling strategies is performed showing the great potential of the closed-loop dynamic rescheduling algorithm.

Keywords: rescheduling, model predictive control, uncertain models.

1 Introduction

Scheduling has great importance due to the necessity of arranging, controlling and optimizing operations, machinery and human resources, product supply and material purchase for a chemical production. Furthermore, scheduling is responsible for ensuring that all plant limits, such as maximum volume of tanks and maximum feed flow rates are respected.

As scheduling problems are dynamic in nature, rescheduling is necessary when new information, such demands and process units availability change. A closed-loop approach is a great alternative to deal with the dynamic behavior, since scheduling becomes an iterative process. Subramanian et al. (2012) reported a new approach in which a general mixed-integer (MIP) scheduling model was represented in state-space form. Certain disturbances were also modeled in state-space form leading to rescheduling. Concepts of model predictive control (MPC) was used in a closed-loop approach in order to deal with disturbances over the scheduling horizon. The approach showed great potential, and it was capable of solving an addressed problem composed by a single reactor and two possible reactions.

Baldea and Harjunkoski (2014) brought forward a review about works addressing the integration of scheduling and control. Moreover, essential elements for the integration between control and scheduling were identified. Dias and Ierapetritou (2016) lead a study focused on analyzing uncertainties in scheduling and control problems.

The present work main goal is to introduce a model predictive closed-loop algorithm with shrinking horizon to deal with rescheduling The strategy uses uncertain models to predict and reject disturbances on the process and contemplate an interval for system preparation, for example with cleaning, which makes the use of receding horizon with shrinking not the option of choice. A comparison was carried out between standard and closed-loop approaches, indicating that the proposed technique is a promising rescheduling technique.

2 Problem Description

The scheduling model used herein is presented below:

- A set of tasks $i \in I$ that can be performed in a set $j \in J$ of units; $\bar{\tau}_i$ is the execution time of task I; B_{it} is the volume of each task batch i in time instant t. The subset I_j refers to tasks i that can be performed in unit j.
- A set of material $k \in K$; the subsets I_k^+ and I_k^- denote material k being produced/consumed by a task i; ρ_{ik} is the proportion of material k produced/consumed by task i.
- A set of time instants denoted by $t \in T$.

2.1 Scheduling model

A time-discrete model was considered, in which time horizon is divided in T periods with equal duration. It is important to highlight that in discrete models, task can only start or end at the beginning of each time instant t.

The objective function is represented by Eq(1), where γ_i is the production cost of task i, W_{ijt} is the binary variable responsible for indicating the start of task i in unit j at time instant t, ν_k is the inventory cost of material k, S_{kt} is the inventory level of material k at time period t, ϕ_k is the backlog cost and U_{kt} is the backlog of material k at time period t.. It is important to emphasize that backlog cost represents the most representative cost of the objective function, intending to avoid the backlog from occurring.

$$\min_{\Xi} \max_{\Pi \in \Lambda} Z = \sum_{i,t} \gamma_i \cdot W_{ijt} + \sum_{i,t} \nu_k \cdot S_{kt} + \sum_{i,t} \phi_k \cdot U_{kt} \quad i \in I, k \in K, t \in T \tag{1}$$

Where $\Xi \equiv \{W_{ijt}, S_{kt}, U_{kt}\}$ and $\{\Pi_I \dots \Pi_M\} \equiv \Pi \in \Lambda$ represents a set of scheduling models belonging to a compact set . The nature of the formulated problem is to perform a maximization over the set Λ and corresponds to choosing the model Π_i: $x(t+1) = A_i x(t) + B_i u(t) + B_{d_i} d(t); i = I, \dots, M$ for prediction that would lead to the largest or **worst-case** model among all plant models in Λ.

2.2 State-space model for scheduling problems

Scheduling model use binary variables to indicate the start of tasks, but during the operation of the unit j there is no variable responsible for indicating that a task is occurring in unit j. Therefore, it is necessary to create a state variable to designate that unit j is carrying a task i at time period t. The same idea can be used for variables B_{it}, U_{kt} and S_{kt}.

The state-space model can be written as Π_l: $x(t+1) = A_l x(t) + B_l u(t) + B_{d_l} d(t)$, in which $x(t) = [\bar{W}_{ijt}^n, S_{k,t}, U_{k,t}, \bar{B}_{it}^n]$ is the state vector, $u(t) = [W_{ijt}, B_{it}, V_{kt}]$ is the input

vector and $d(t) = [\xi_{kt}]$ and is represented by $[A_l\ B_l\ B_{d_l}] \equiv \Pi_l$, where A_l, B_l and B_{d_l} are the state-space matrices of Π_l and t is the time instant (number of time periods T). The state-space model derived from the scheduling model is used as the prediction model on the closed-loop scheduling strategy.

Model uncertainty is introduced as a finite set of models as $\Pi := \{\Pi_I \dots \Pi_M\} = \{[A_I\ B_I\ B_{d_I}] \dots [A_M\ B_M\ B_{d_M}]\}$. Task i yield ρ_{ik} varies in a certain viable range on the prediction models, and the worst case scenario will be chosen at each time instant in order to ensure that the problem is able to address the desired scheduling even in the worst case scenario for the prediction model. In addition, the actual yield of the plant is a random variable satisfying a normal distribution within the same range of the ρ_{ik} values chosen in the used uncertain model. Where Π is some prespecified set representing the multi-model paradigm.

A variety of disturbances can be modeled in the state-space model framework presented. Demand at each time instant t has been modeled as a disturbance, and it is part of the disturbance vector d, thus the nominal demand just changes over the time horizon.

2.2.1 Losses during reactors unloading

It is possible to foresee at the beginning of a task i in unit j the mass of the material k produced at the end of that task. However, unexpected events can happen during the unloading or transport of the product and some losses can be caused. Those losses can be modeled through the variable $\hat{\theta}^p_{ikt}$, which represent the mass product lost. This variable will be now part of the disturbance vector d. The inventory level can be calculated by Eq(3), with $t \in T, j \in J, k \in K, i \in I$

$$S_{k,t+1} = S_{k,t} + \sum_{i \epsilon I_k^+} \rho_{ik}(\bar{B}^{\tau_i}_{it+1} - \hat{\theta}^p_{ikt}) + \sum_{i \epsilon I_k^-} \rho_{ik} B_{it} + V_{kt} \tag{3}$$

3 Case study

The case study selected involves two tasks (reaction A and B) in a single reactor, and is a modification of the one proposed by Subramanian et al. (2012). Each task has a different duration, and only one task can be performed at each time instant in the reactor. Reaction A requires raw material (RM) and it produces A during three (3) period of time and has a conversion in the range of 90% to 100%. Reaction B requires raw material (RM), produces B in two (2) period of time and has a conversion in the range of 90% to 100%. Moreover, if there is a changeover between the tasks in reactor, it is necessary a setup time of two time periods for restarting the production. The initial inventory for RM, A and B are as follows 100, 15 and 20 units of volume, respectively. The time horizon is 48 hours divided equally in time periods of 1 hour, but two cycles with a duration of 24 hours each in a shrinking horizon was performed in order to schedule 48 hours of production. Shrinking horizon can be a good alternative for rescheduling, since it is easier to implement terminal constraints at the end of the shrinking horizon, such as a setup time before restarting the operation and a constraint to ensure that the equipment finishes their operation empty at the end of the shrinking horizon, as indicated in the case study with 2h for restarting the operation. Herein two different disturbances is addressed, first the change of the nominal demand and secondly events like losses during reactors unloading as follows:

- at t = 3, occurs a loss of 4 units of mass of product A;
- at t = 6, occurs a loss of 2 units of mass of product A;
- at t = 6 the nominal demand changes for both products (A and B), from 1.5 to 2 units of mass;
- at t = 15 occurs a loss of 3 units of mass of product B;
- at t = 30 occurs a change on the nominal demand for both products (A and B), from 2 to 3.5 units of mass remaining at that level until the end of the scheduling horizon

4 Results

All the simulations was performed on a i7/16GB computer. Furthermore, the software Python 3.4.0 integrated with GAMS 24.4.6 was used for all simulations. When not indicated otherwise the parameters used are as follows $\gamma = [2\ ,6]^T, \nu = [8,2\ ,10]^T,$ and $\phi = [100{,}100]^T$.

4.1 Case Study 1- Exact Model (no uncertainty)

In this case, the scheduling was performed considering a exact model (no model uncertainty). Yields of 100% for both reaction A and reaction B were used. Besides that, no plant-model mismatch was considered. In this case, it is possible to analyze the closed-loop disturbance rejection over the time horizon.

Fig(1) presents the Gantt Chart for classical scheduling and for rescheduling using the MPC closed-loop algorithm. When disturbances occur, rescheduling was necessary and the remaining previously calculated operation were not implemented. Instead, the remaining operations were replaced by the closed-loop MPC new scheduling. The inventory level for the problem without rescheduling and with rescheduling are presented in Fig(2) and Fig.(3), respectively.

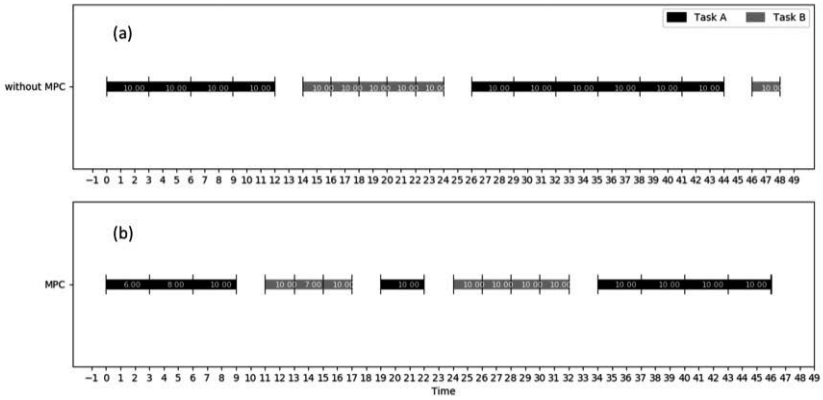


Figure 1- Case study 1 - Gantt chart: (a) Classical scheduling and (b) MPC rescheduling

The objective functions reached the values of \$178,898.00 and \$103,114.00, respectively. In this first case, at t = 6h the nominal demand changed, so the backlog started rising, since the scheduling did not take into account this change and did not plan shipments to meet the new demands.

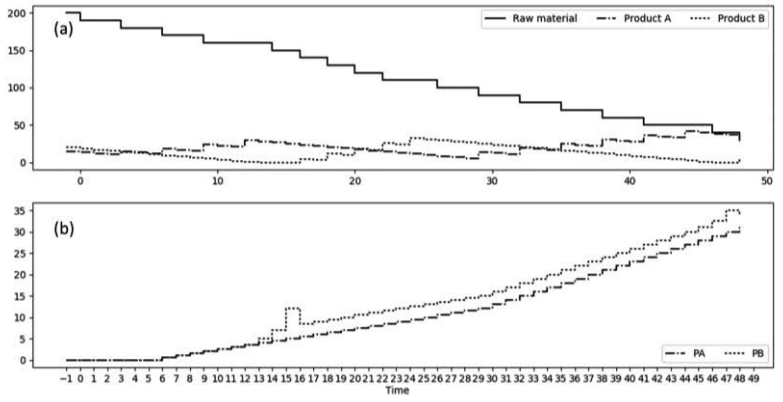


Figure 2 – Case study 1 without rescheduling: (a) Inventory level and (b) backlog

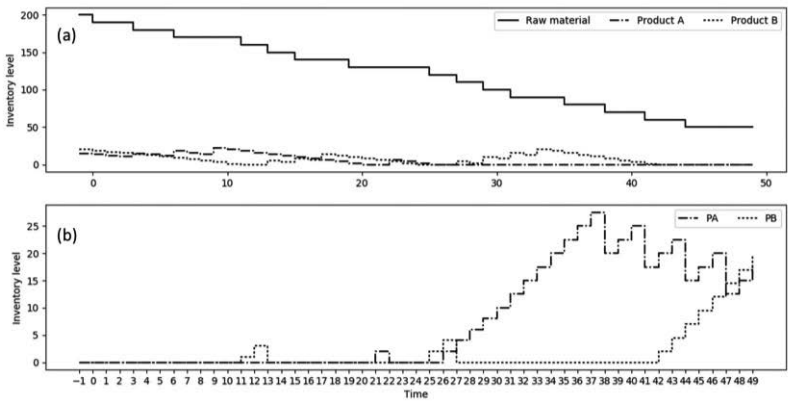


Figure 3 – Case of study 1 with rescheduling: (a) Inventory level and (b) backlog

The case with dynamic rescheduling performs new scheduling and shipments were changed so that the backlog occurred only when inventory was insufficient to meet nominal demand.

While backlog for the problem without rescheduling occurred at t = 6h, for the one using MPC rescheduling, it occurred at t = 9h. Moreover, reaction A was performed four times in a row because there were losses of product A over the time horizon. Even with losses, the backlog for product A occurred practically at the end of the time horizon ($t = 20$), showing how rescheduling can minimize the effect of disturbances.

4.2 Case Study 2 - Uncertain models

I this case study, a set of five models defined with reaction yields in the range of 0.9 to 1 were used at each iteration in the rescheduling multi-model problem. In addition, the actual yield of the plant was unknown and satisfied a normal distribution with value belonging to the range of specified yield for the plant. The set of yields used by the rescheduling in the case study was given by [0.9, 0.93, 0.95, 0.98, 1.0].

Fig(4) shows the inventory level for the problem with uncertainties. The objective function was $116,206.97. It is important to highlight that even when the decisions of

the worst case scenario was selected, the problem still presented satisfactory results. Moreover, in Fig(5) is pointed out that the solution is robust to uncertainties in the models, even when there exist plant-model mismatch, so it can be seen that the rescheduling based MPC algorithm is robust enough to reschedule the problem even under this range of uncertainties for this case study.

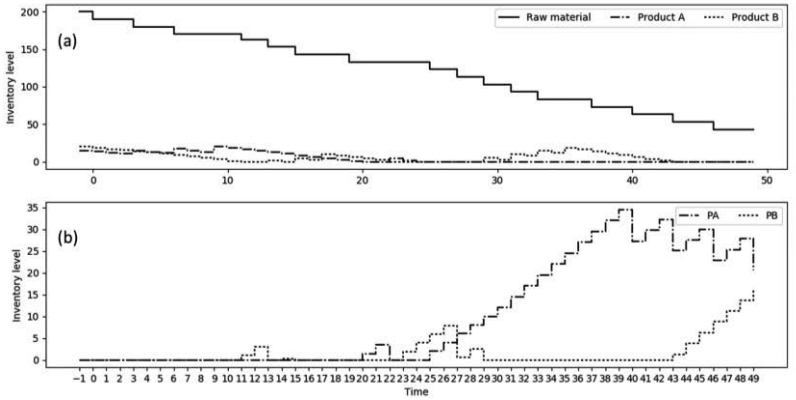


Figure 4 - Case study 2 - rescheduling with uncertainties: (a) inventory level and (b) backlog

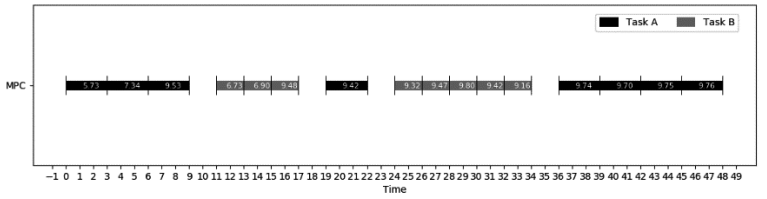


Figure 5 - Case study 2 - Gantt chart with rescheduling under uncertainties

5 Conclusions

A comparison of standard scheduling strategies and a closed-loop strategy for rescheduling showed that the later could reject disturbances and carry a lower backlog and objective function value. Shrinking horizon can be a good alternative for rescheduling allowing a realistic description of terminal constraints and objective function during setup times before restarting the process operation. The use of a set of uncertain models in a min-max framework applied to a plant with a normal distribution uncertainty on reaction conversion lead to satisfactory results. Finally, the closed-loop strategy showed great potential when applied to scheduling problems even when the used models are uncertain.

References

K. Subramanian; C. T. Maravelias; J. B. Rawlings; 2012, A State-space Model for Chemical Production scheduling, Computers and Chemical Engineering, 47, 97-110

M. Baldea; L. Harjunkoski; 2014, Integrated Production Scheduling and Process Control: A Systematic Review, Computers and Chemical Engineering, 71, 377-390

L. S. Dias; M. G. Ierapetritou; 2016, Integration of Scheduling and Control under Uncertainties: Review and Chalenges, Chemical Engineering Resarch and Design, 116, 98-113

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50224-X

Analysis of the Onset of Chaos for the Belousov-Zhabotinsky Reaction

Xi Wang[a], Chi Zhai[a], Zhijian Zhao[c], Ahmet Palazoglu[b], Wei Sun[a]*, Jiali Ai[a]

[a] *Beijing Key Lab of Membrane Science and Technology, College of Chemical Engineering, Beijing University of Chemical Technology, 100029 Beijing, China*
[b] *Department of Chemical Engineering, University of California, Davis, CA 95616, USA*
[c] *Research Center of the Ministry of Education for High Gravity Engineering and Technology, Beijing University of Chemical Technology, 100029 Beijing, China*
sunwei@mail.buct.edu.cn

Abstract

By utilizing Picard's iteration and Laplace-Borel transform, a preliminary analytical solution is obtained for the Arneodo model of the B-Z reaction. For purposes of comparison, the numerical simulation of this model is also conducted. Then, the occurrence of chaos in this model is recognized and characterized by the Lyapunov exponent method, correlation dimension method and power spectrum method, respectively.

Keywords: Chaos, Laplace-Borel transform, B-Z reaction.

1. Introduction

The Belousov-Zhabotinsky (B-Z) reaction is the first studied oscillating chemical reaction, which is a typical self-organized nonlinear system with a chemical wave spreading over time and across space (Belousov, 1959). The oscillation behaviour in B-Z reaction is quite similar to that in bio-systems, hence the study of the B-Z system could facilitate further investigations of general bio-systems. The self-organized oscillatory behaviour can be explained by the Hopf bifurcation theory, i.e., once a system evolves beyond a Hopf point, it oscillates into a limit cycle period, then, chaos may follow, which can be described by certain mathematical models for the B-Z reaction, such as the Arneodo model (Arneodo et al., 1998).

A chaos behaviour of a process can be identified by its sensitivity to initial conditions, topological mixing, dense periodic orbits (Boris et al., 2003), which could be characterized by the Lyapunov exponent method, correlation dimension method (Strogatz, 2000) and power spectrum method (William, 1992). This behaviour is usually observed through numerical simulations, which makes these characterization methods limited to simulation data, as the system is considered as a black box during the analysis.

Theoretically, chaos only emerges in nonlinear systems with no less than 3 dimensions (Pedro, 2006), or, infinite dimensional linear systems (Bonet et al., 2001). With the Laplace-Borel transform, a nonlinear system can be expanded into an infinite dimensional linear system (Batigun et al., 1997), which provides a possibility for the study of chaos analytically and a feasible way to understand the mechanism of the onset of chaos.

In this paper, the Arneodo model for the B-Z reaction system is solved by the Laplace-Borel transform, and then characterization of the onset of chaos is studied. For comparison purposes, numerical simulation results are also provided.

2. Chemical kinetics and Arneodo model for B-Z reaction

The B-Z reaction can be considered as an open system with a constant negative entropy flow, which is characterized by non-linearity and away from the equilibrium state. The reaction mechanism can be easily explained by the Oregonator model (Field et al., 1974),

$$\begin{aligned} &A+Y\rightarrow X+P && r_1=k_1AY\\ &X+Y\rightarrow 2P && r_2=k_2XY\\ &A+X\rightarrow 2X+2Z && r_3=k_3AX\\ &2X\rightarrow A+P && r_4=k_4X^2\\ &B+Z\rightarrow 0.5fX && r_5=k_5BZ \end{aligned} \tag{1}$$

where X is $HBrO_2$, Y is Br^-, Z is Ce(IV), A is BrO_3^-, B is organic, and P is HOBr. Assuming that reactants A and B are of high concentration and do not change significantly during several oscillations, the intermediate products X, Y and Z are generated and consumed alternately, causing Z to change back and forth between Ce^{4+} and Ce^{3+}. After adding the ferroin indicator to the system, the medium will repeatedly change color between blue and red. Figure 1 shows an open system, the overall reaction of the system is $A + B \rightarrow P$, the entropy change of the reactants and product exhibits a continuous negative entropy flow into the system, a part of which is used to maintain the chemical wave, i.e., a typical dissipative structure.

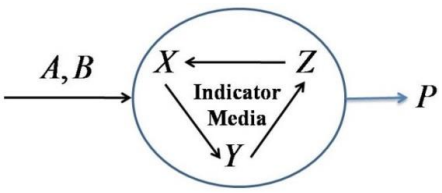


Fig. 1 The simplified structure of B-Z reaction.

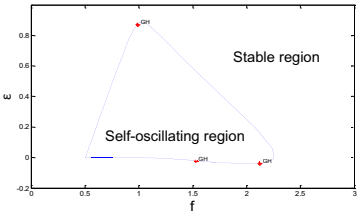


Fig. 2 The parameter space of Oregonator model. GH-generalized Hopf points.

From the dynamic system viewpoint (Fig. 1), the nonlinear term of the intermediate product leads to a supercritical Hopf bifurcation as shown in Eq. (1). According to the hypothesis, the dynamic model of the reaction intermediate of the dimensionless Oregonator model is displayed in Eq. (2) (Fig. 2), the blue line divides the parameter space of Eq. (2) into a self-oscillating region and a stable region,

$$\begin{aligned} &\varepsilon\frac{dx}{d\tau}=qy-xy+x(1-x)\\ &\delta\frac{dy}{d\tau}=-qy-xy+fz\\ &\frac{dz}{d\tau}=x-z \end{aligned} \tag{2}$$

where $x = X / X_0$, $y = Y / Y_0$, $z = Z / Z_0$, $\tau = t / T_0$, the steady-state value can be obtained by setting the right side of the system's dynamic equation to zero, thus X_0, Y_0, Z_0, $T_0 = F(k_1 \ldots k_5, A, B)$.

The B-Z reaction exhibits periodic oscillations, and sometimes chaos. The Oregonator model can only describe a limit cycle for this process, while more nonlinear behaviour is well illustrated by the Arneodo model. Thus the Arneodo model is studied here to identify the chaos in the B-Z reaction process,

$$\begin{aligned} &\partial x/\partial t = y \\ &\partial y/\partial t = z \\ &\partial z/\partial t = -\eta z - \upsilon y - \mu x + x^2 - 1.425y^2 + 0.2xz - 0.01x^2 z \end{aligned} \tag{3}$$

where $\mu = 1.38$, $\eta = 1.2$, $v = 1$.

2.1 Picard's iteration

The solution to ordinary differential equations can be expressed in terms of iterated integrals. Consider a differential equation given by

$$\frac{dX(t)}{dt} = f(X) \tag{4}$$

where *f(X)* is a continuous function. After expressing Eq. (3) in its integral form, successive approximations with the recurrence relations are given as,

$$X_k = X(0) + \int_0^t f(X_{k-1}, \tau)\, d\tau \tag{5}$$

If *f(X)* is a nonlinear function, the solution of this system can be expressed as

$$X(t) = X(0)\left[1 + \sum_{n=0}^{\infty} c(i_n \text{ K } i_0) \int_0^t d\xi_0 \text{ K } d\xi_n\right] \tag{6}$$

2.2 Integral algebra

It can be defined that

$$\int_0^\tau d\tau \Leftrightarrow x_0 \tag{7}$$

where x_0 is integral operator, and this operator satisfies the shuffle product operation which has two important properties. One is the distributive property described as,

$$\begin{aligned} &[F_1(x_0) + F_2(x_0)] \text{ C} [F_3(x_0) + F_4(x_0)] \\ &\quad = F_1(x_0) \text{ C } F_3(x_0) + F_1(x_0) \text{ C } F_4(x_0) + F_2(x_0) \text{ C } F_3(x_0) + F_2(x_0) \text{ C } F_4(x_0) \end{aligned} \tag{8}$$

The other is the multiplication of iterated integrals, which is expressed by the shuffle products as follows, where C is the shuffle product operation, on which the Laplace-Borel transform depends.

$$\begin{aligned} &(x_{\xi_j} \text{ L } x_{\xi_1}) \text{C} (x_{\xi_k} \text{ L } x_{\xi_1}) \\ &\quad = x_{\xi_j}[(x_{\xi_{j-1}} \text{ L } x_{\xi_1}) \text{C} (x_{\xi_k} \text{ L } x_{\xi_1})] + x_{\xi_k}[(x_{\xi_j} \text{ L } x_{\xi_1}) \text{C} (x_{\xi_{k-1}} \text{ L } x_{\xi_1})] \end{aligned} \tag{9}$$

2.3 Laplace-Borel transform

In Laplace-Borel transform, a domain change is employed to transform nonlinear differential equations to algebraic equations whose solution is represented by an infinite series. Consider a function *f(t)* that can be expanded in Taylor series,

$$f(t) = \sum_{n=0}^{\infty} a_n \frac{t^n}{n!} = \sum_{n=0}^{\infty} a_n \int_0^t d\xi_n \, \mathrm{K} \, d\xi_0 \tag{10}$$

Utilizing the relationship between iterated integrals and algebra, the Laplace-Borel transform is defined as

$$LB[f(t)] \equiv F(x_0) \equiv \sum_{n=0}^{\infty} a_n x_0^n \tag{11}$$

The explicit form of Laplace-Borel transform is derived as,

$$F(x_0) = \frac{1}{x_0} \int_{0^-}^{\infty} e^{-t/x_0} f(t) dt \tag{12}$$

The following is a relationship between Laplace transform and Laplace-Borel transform, which is the key to do inverse Laplace-Borel transform,

$$F(s) = x_0 F(x_0)_{x_0 = s^{-1}} \tag{13}$$

3. Results and discussion

In Eq. (3), variables y and z can be expressed by x, then the Laplace-Borel transform is:

$$x(t) \Leftrightarrow F(x_0) \tag{14}$$

Then an expanded solution is assumed in the form of

$$F(x_0) = F_0(x_0) + F_1(x_0) + F_2(x_0) + \mathrm{L} \tag{15}$$

where $F_i(x_0) = f(F_{i-1}(x_0), F_{i-2}(x_0), \mathrm{K}, F_0(x_0))$

By exploiting Eq. (13), performing an inverse Laplace transform, and setting *x(0)*, *x'(0)*, *x''(0)* to 0.1, the terms that define the time domain analytical solution can be obtained as,

$$\begin{aligned}
x_0(t) &= 0.07247e^{-1.269t} + (0.01376 + 0.09163i)e^{(0.03448-1.042i)t} \\
&\quad + (0.01376 - 0.09163i)e^{(0.03448+1.042i)t} \\
x_1(t) &= 1.524 \times 10^{-7} e^{-3.807t} + 0.0005225e^{-2.538t} + (7.722 \times 10^{-7} - 5.703 \times 10^{-10} i)e^{(-2.503-1.042i)t} \\
&\quad + (7.722 \times 10^{-7} + 5.703 \times 10^{-10} i)e^{(-2.503+1.042i)t} + 0.01331e^{-1.269t} \\
&\quad + (0.005697 + 0.007138i)e^{(-1.234-1.042i)t} + (0.005697 - 0.007138i)e^{(-1.234+1.042i)t} \\
&\quad + (0.00003872 - 4.118 \times 10^{-21} i)e^{-1.2t} - (2.919 \times 10^{-7} - 1.252 \times 10^{-7} i)e^{(-1.2-2.085i)t} \\
&\quad - (2.919 \times 10^{-7} + 1.252 \times 10^{-7} i)e^{(-1.2+2.085i)t} - (0.009917 - 0.005754i)e^{(0.03448-1.042i)t} \\
&\quad - (0.009917 + 0.005754i)e^{(0.03448+1.042i)t} + (0.001902 + 0.001588i)e^{(0.06896-2.085i)t} \\
&\quad - (0.009049 - 2.344 \times 10^{-19} i)e^{0.06896t} + (0.001902 - 0.001588i)e^{(0.06896+2.085i)t} \\
&\quad - (1.935 \times 10^{-7} - 2.178 \times 10^{-7} i)e^{(0.1034-3.127i)t} - (0.00009225 + 0.00004914i)e^{(0.1034-1.042i)t} \\
&\quad - (0.00009225 - 0.00004914i)e^{(0.1034+1.042i)t} - (1.935 \times 10^{-7} + 2.178 \times 10^{-7} i)e^{(0.1034+3.127i)t}
\end{aligned} \tag{16}$$

After evaluating the 2nd term in the frequency domain by 0th and 1st terms, it is found that the 2nd term cannot be computationally transformed back into time domain completely. The inverses in the time domain are omitted due to the page limitations. In time domain, the 0th term and most parts of the 1st term consist of conjugate pairs and convergent exponential terms, which exhibit an amplitude increasing oscillation over time, indicating that the oscillation of the state point starts with an amplitude increasing vibration, which can be regarded as the superposition of a series of periodic motions in 3D phase space

with gradually increasing linear velocities. Thus, the amplitude increasing oscillation can be assumed to be an instability process, which is accompanied by the onset of chaos. Fig. 3 exhibits that the 1st term contains a diverging exponential term, which leads to a symmetry breaking in the system, which is also an instability factor.

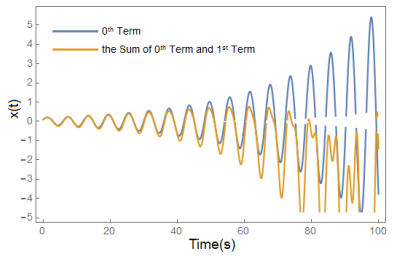


Fig. 3 0th term and sum of 0th and 1st terms.

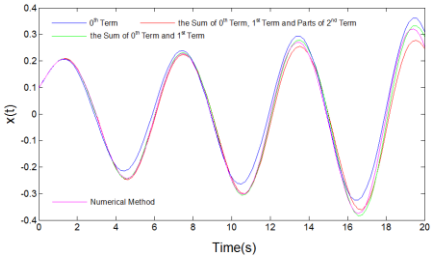


Fig. 4 Numerical and analytical methods.

It can be seen in Fig. 4 that the green line, representing the sum of 0th and the 1st terms, is closer to the purple line, representing the numerical method simulation, than the blue line, representing the 0th term. The red line represents the sum of 0th, 1st and part of 2nd term, so the green line still approximates the process better than the red line because of incomplete solution of the 2nd term in the time domain.

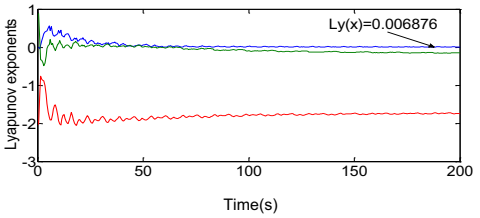


Fig. 5 Lyapunov exponents of Aronodo model.

Lyapunov exponents depict the degree that deterministic systems are affected by fluctuations in initial values. It can be seen from Fig. 5 that Lyapunov exponent of dimension x is positive, indicating that a slight fluctuation in initial value expands exponentially over time and the trajectory after a long time cannot be accurately predicted, which is a feature of chaos. To obtain the correlation dimension of this system, the phase space is first reconstructed based on the time sequence of a variable, which requires selecting the embedded dimension d and the delay time τ by seeking one point where curve is saturated in Fig. 6 and the other point at first valley of curve in Fig. 7, in this case, $d=6$ and $\tau=176$, then the slope of linear part of the double logarithmic curve of the correlation integral C and the radius parameter ε are calculated, whose fitting result in Fig. 8 shows that the system correlation dimension is 1.61, indicating that the system generates chaos for the dimension of this deterministic system being non-integer, which is also a feature of chaos.

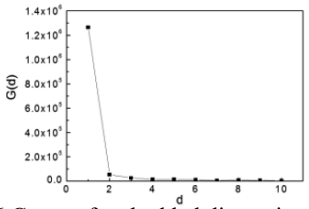


Fig. 6 Curve of embedded dimension.

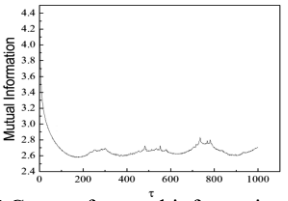


Fig. 7 Curve of mutual information.

In the power spectrum, the narrow peaks indicate the periodic behaviors, i.e., stable limit cycles (a kind of general attractor), and the broad peaks is the characteristic of the chaotic behaviors, i.e., or strange attractors. As shown in Fig. 9, there are five peaks in the power spectrum of the Arneodo model's attractor, which indicates that this attractor is not a stable limit cycle. The frequency interval of each of the two adjacent peaks is the same, which indicates that a regular pattern about frequency of sub-periodic behaviours exists. The first and second peaks are wider than others, which likely indicates a strange attractor, a feature of dissipative-system chaos.

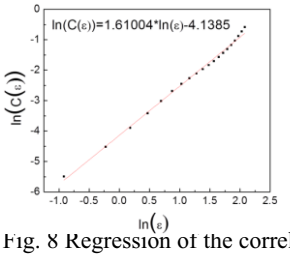


Fig. 8 Regression of the correlation dimension.

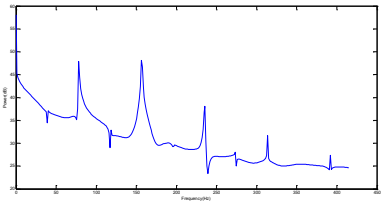


Fig. 9 The power spectrum of Arneodo model.

4. Conclusions

A preliminary analytical solution of Arneodo model of B-Z reaction was obtained for the Arneodo model of B-Z reaction by the Laplace-Borel transform. The numerical simulation for this model was also conducted. Then chaos occurring in this system was identified and characterized by Lyapunov exponent method, correlation dimension method and power spectrum method, respectively. We show that the nonlinear behaviour can be better studied and predicted if a complete analytical solution is available.

References

A. Batigun, K. Harris, A. Palazoglu, 1997, Studies on the analysis of nonlinear processes via functional expansions-I. Solution of nonlinear ODEs, Chemical Engineering Science, Vol. 52, No. 18, 3183-3195.

A. Espuña, M. Graells, L. Puigjaner, 2017, 27th European Symposium on Computer Aided Process Engineering.

A. Arneodo, P. Coullet, E. Spiegel, C. Tresser, 1985, Physica D, 14, 327.

B. Belousov, 1959, Периодически действующая реакция и ее механизм [Periodically acting reaction and its mechanism], Сборник рефератов по радиационной медицине, 147, 145.

H. Boris, A. Katok, 2003, A First Course in Dynamics: With a Panorama of Recent Developments, Cambridge University Press, ISBN 0-521-58750-6.

H. William, 1992, Power Spectrum Estimation Using the FFT, Cambridge University Press.

J. Bonet, F. Martínez-Giménez, A. Peris, 2001, A Banach space which admits no chaotic operator, Bulletin of the London Mathematical Society, 33, 2, 196.

R. Field, R. Noyes, 1974, Oscillations in chemical systems. IV. Limit cycle behavior in a model of a real chemical reaction, The Journal of Chemical Physics, 60, 5, 1877-1884.

R. Pedro, 2006, Underdetermination of Science: Part I, ISBN 1411693914.

S. Strogatz, 2000, Nonlinear Dynamics and Chaos, Westview Press, ISBN 0-7382-0453-6.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50225-1

Design concept for coal-based polygeneration processes of chemicals and power with the lowest energy consumption for CO_2 capture

Hong Huang, Siyu Yang *

School of Chemical Engineering, South China University of Technology, Guangzhou, 510640, PR China
cesyyang@scut.edu.cn

Abstract

Single coal based chemicals production processes emit large amount of CO_2 during conversion of the syngas to the high H_2/CO ratio feed gas for chemicals synthesis. However, the feed gas is with a low CO_2 molar fraction, leading to high energy cost for CO_2 capture. In this work, we try to reduce energy consumption for CO_2 capture by improving its molar fraction. A new methanol and power polygeneration process is designed and analysed based on process modeling and simulation. The hierarchical conceptual design methodology is introduced to design the polygeneration more reasonably. In this process, the shifted syngas exiting from the water gas shift unit first goes into the CO_2 capture unit to remove CO_2. Then, the purified H_2-rich syngas is mixed with the unshifted syngas and fed into the methanol synthesis to produce methanol. Then, unreacted syngas out from the methanol synthesis unit is moderately recycled to use, while the rest is used to generate power. Energy consumption for CO_2 capture of the polygeneration process is 0.7 GJ/t-CO_2, which is a 40.6% reduction compared to that of the single coal-to-methanol process and a 22.2% reduction to that of coal-to-hydrogen for power generation process. Energy saving ratio of the polygeneration process is 16.5%.

Keywords: polygeneration, process design, CO_2 capture, energy consumption

1. Introduction

Coal is the dominant resource for chemicals production or power generation in China. This is as known due to lack of natural gas and oil. This energy structure brings severe greenhouse gas (GHG) emissions since the production processes transfer carbon rich resource to hydrogen rich product or to power (Xiang et al., 2014; Man et al., 2014). Carbon capture and storage (CCS) technology is broadly accepted as a feasible approach to reduce CO_2 emission (Xiang et al., 2014). However, a CCS process adds large amount of energy consumption and increases operational cost much. In the CCS process, energy for CO_2 capture takes the largest proportion, accounting for 70-80% of the total.

Increasing CO_2 molar fraction can effectively reduce energy for CO_2 separation. Zhang et al. (Zhang et al., 2013) studied energy consumption for CO_2 capture using membrane separation process. Results showed that the energy for CO_2 capture was reduced from about 1.8 GJ/t to 0.1 GJ/t CO_2 as the molar fraction increases from 0.025 to 0.5.

Some research attempts to increase the CO_2 molar fraction in coal based chemical process, Li et al. (Li et al., 2013) proposed a coal based natural gas (SNG) and power polygeneration process. Coal based syngas is sent directly to the methanation reactor to produce SNG without H_2/CO ratio adjustment. The active gas is reacted and the molar fraction of CO_2 is increased after the reaction. The unreacted gas is separated and the corresponding energy for the CO_2 separation is reduced. This theoretical process seems reasonable to reduce the energy for CO_2 capture. However, the syngas with an insufficient H_2/CO ratio is not suitable for SNG synthesis, leading to catalyst deactivation and a low reaction rate. In this case, the loss in product yield outweights the energy saving from CO_2 capture.

In this paper, we introduce a hierarchical polygeneration process design concept for chemicals and power production based on the "onion" model proposed by Smith (Smith, 2005). It can pursue energy cascade utilization and resource efficient utilization. In this model, a chemical process is decomposed into four layers, starting from reaction, followed by separation process, heat exchanger network and utility system, as shown in Fig. 1.

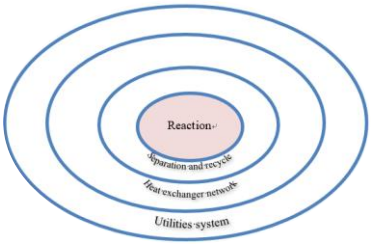


Figure 1. Onion diagram for process design.

For illustrative case study, a coal based methanol and power process is designed based on the concept, as shown in Fig. 2. Coal is converted into crude syngas in the gasifier. After heat recovery and scrubbring, the crude syngas is separated into two parts. One part is used for power generation and the other part is used for methanol synthesis. To maximize the CO_2 molar fraction, we place the CO_2 capture unit to the power generation side. In this branch, syngas is converted to hydrogen and the molar fraction is increased. The capture energy is lower than that in conventional chemical process. In the chemicals synthesis side, the H_2/CO ratio is increased by mixing with a part of H_2 from the power generation side. The key idea here is element /resource complementary. In addition, a part of unreacted syngas from the methanol synthesis reactor is recycled and the rest is used for power generation.

The key operating parameters and optimal process design are determined according to the simulation. Energy saving ratio of the polygeneration process is analysed to illustrate the advantage compared to single CTM and coal-to-hydrogen power generation processes.

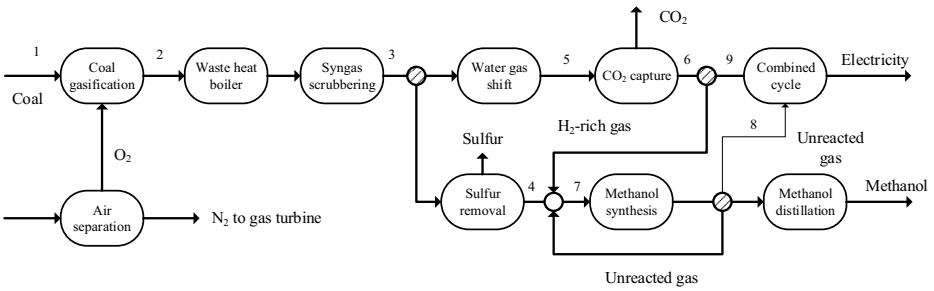


Figure 2. Schematic diagram of the methanol and power polygeneration process.

2. Process modeling and simulation

Different units are firstly modeled and simulated, including the ASU unit, the coal gasification unit, a WGS unit, a CO_2 capture unit, a methanol synthesis and distillation unit and a combined cycle power generation unit. In the following sections, we mainly describe the modeling and simulation of these units.

2.1. Coal gasification

For modeling the coal gasifier, coal is defined as unconventional solid using MIXCINC in Aspen Plus (He et al., 2013). RK-Soave is the property method. Coal gasification process is modeled as the combination of a pyrolysis, a gasification and a combustion stages. In the pyrolysis stage, coal is decomposed at 500 ℃. The pyrolysis reaction is modeled by the Rstoic block. In the gasification and the combustion stages，the main reactions include heterogeneous reactions. Due to the high temperature (>1000 ℃) of the gasification and the combustion stages, these reactions meet the equilibria and are modeled by RGibbs reactor according to Gibbs free energy minimization (Jarungthammachote et al., 2008).

2.2. Methanol synthesis

The crude syngas sent to the methanol synthesis and distillation firstly passes through a COS hydrolyzer where 100% of the COS is converted to H_2S via the hydrolysis reaction (Rhodes et al., 2000). After that, the syngas is cooled down to 40 ℃. H_2S is then absorbed by contacting with methanol solvent and subsequently released by heating the solvent. The CO-rich syngas ($H_2/CO \approx 0.7$) leaving the sulfur removal unit is mixed with the H_2-rich gas from CO_2 capture unit to adjust the H_2/CO molar ratio to 2.0-2.1. The mixed syngas is compressed to 8.0 MPa by a centrifugal compressor and then preheated to about 200 ℃ before going to the methanol synthesis reactor where an exothermic methanol synthesis reaction takes place. The methanol synthesis reaction converts CO and CO_2 into methanol over a $Cu/ZnO/Al_2O_3$ catalyst.

2.3. CO2 capture and power generation

In the power generation section, there are a WGS unit, a CO_2 capture unit and a combined cycle power generation unit. The syngas first goes into the WGS unit where CO and steam are converted to H_2 and CO_2. This unit has a two-stage configuration. The temperature of the crude syngas is increased to 250 ℃ before going into the first WGS reactor, according to the activation temperature of the CoMo catalyst (Carbo et al., 2009). In the reactor, axial temperature increases with the extension of the reaction.

Increase of temperature would limit the chemical equilibrium of the reaction and CO could not be further converted. The high temperature syngas leaving the first reactor is cooled down to 250 ℃ and goes into the second reactor for further conversion of CO.

The shifted syngas is sent to the CO_2 capture unit to remove acid gas such as H_2S and CO_2. The syngas from WGS unit is cooled down to 40 ℃ and fed into a Rectisol process. In this process, H_2S and CO_2 are sequentially absorbed in the H_2S absorber and CO_2 absorber by methanol solvent at a methanol/syngas molar ratio of about 0.85, and subsequently released by heating the solvent.

3. Energy consumption for CO_2 capture

Energy consumption for CO_2 separation and compression can be calculated by Eqs. (1) and (2) (Li et al., 2014).

$$E_{sep} = -RT_0 \frac{x_{CO_2} \cdot \ln x_{CO_2} + (1 - x_{CO_2}) \cdot \ln(1 - x_{CO_2})}{x_{CO_2} \eta_{cap}} F_{sep} \tag{1}$$

where R is the gas constant, 8.314 J/(mol K), T_0 is 298.15 K, x_{CO_2} is the CO_2 molar fraction in the gas before separation, η_{sep} is the separation efficiency and F_{sep} is the flow rate of the syngas.

$$E_{compr} = RT_0 \ln\left(\frac{P}{P_0}\right) \tag{2}$$

where P is the pressure after compression (MPa) and P_0 is the initial pressure (MPa) From the previous analysis, it is concluded that CO_2 molar fraction is a key factor affecting energy consumption. In this part, we calculate the energy consumption for CO_2 capture at different CO_2 molar fraction, as shown in Fig. 3. In the range of 0.3 to 0.46, energy consumption for CO_2 capture decreases from 1.2 GJ/t to 0.65 GJ/t. In this polygeneration process, CO_2 molar fraction is increased from 31% to 43.2%. Energy consumption for CO_2 capture stems from the compression refrigeration, and the steam consumption of the reboilers and CO_2 compression. Increasing CO_2 molar fraction brings reduction on the amount of methanol solvent. As a result, energy consumption for refrigeration and reboilers is reduced. On the other hand, the syngas flow rate in the polygeneration process is reduced compared to the single CTM process and coal-to-hydrogen for power generation process, and the amount of captured CO_2 is reduced. Thus, the energy consumption decreases to 0.7 GJ/t.

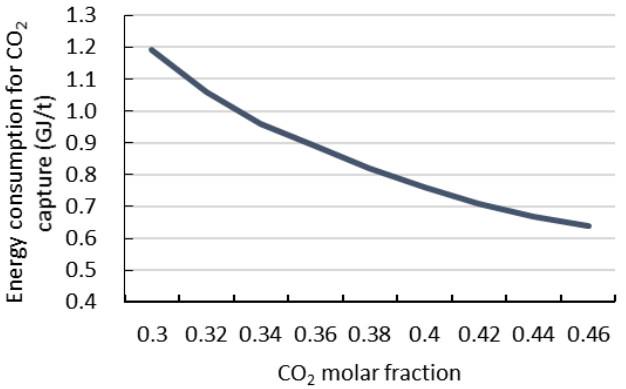


Figure 3. Energy consumption for CO_2 capture at different molar fractions.

4. Analysis of the polygeneration process

For analysing the polygeneration process, energy saving ratio ratio is the index for the analysis. This index is developed special for polygeneration process compared to pseudo single processes.

Energy saving ratio (ESR) is defined as the ratio of the differences between the single-product processes and polygeneation process to the single-product processes, as shown in Eq. (3) (He et al., 2012).

$$\mathrm{ESR} = \frac{E_{\mathrm{H,\,SG}} - E_{\mathrm{H,\,PG}}}{E_{\mathrm{H,\,SG}}} = \frac{\sum_{i=1}^{n}(E_{\mathrm{H,\,P}i} / \eta_i) - E_{\mathrm{H,\,PG}}}{\sum_{i=1}^{n}(E_{\mathrm{H,\,P}i} / \eta_i)} \tag{3}$$

where $E_{\mathrm{H,SG}}$ and $E_{\mathrm{H,PG}}$ represent the total fuel input based on higher heating value of the single-product processes and polygeneation process. $E_{\mathrm{H,Pi}}$ and η_i denote the higher heating value and energy efficiency.

The input-output material balance of the polygeneration process is listed in Table 1. Methanol output and power output are 75 t/h and 227 MW, consuming 200.7 t/h coal. Energy saving ratio is 16.5% according to Eq. (3). The power consumption of each unit is shown in Table 2. Compared to the CTM process, power is supplied by the combined cycle power generation unit in the polygeneration process instead of boiler, this contributes to higher energy efficiency. Due to less coal consumption, the production scale of the polygeneration process is reduced and therefore power consumption of each unit is reduced, especially the ASU. This leads to less energy consumption for methanol production. Moderate recycling of the unreacted syngas of methanol synthesis also contributes to the improvement of energy performance.

Table 1 Input-output of the polygeneration and the reference processes.

Item	CTME	CTM	IGCC
Input			
Coal (t/h)	200.7	147.1	93.4
Raw water (t/h)	570	375	309
Output			
methanol (t/h)	75	75	-
Power (MW)	227	-	227
Energy efficiency (%)	48.8	45	35.1
ESR (%)	16.5	-	-

Table 2 Energy consumption for the polygeneration process.

Unit	Electricity (MW)
Air separation	82.7
Gasification	4.3
Acid gas removal	25.1
CO_2 compression	8.3
methanol synthesis and distillation	17.5

5. Conclusions

In this work, we proposed a new methanol and power polygeneration process, aiming to reduce energy consumption for CO_2 capture. In this process, CO_2 capture is transferred from methanol production process to coal-to-hydrogen process to improve its molar fraction. Based on process modeling and simulation, energy consumption for CO_2 is calculated and energy performance is analysed. Main conclusions are as follows:

(1) The hierarchical conceptual design methodology is introduced to design the polygeneration process.
(2) Energy consumption for CO_2 capture is reduced from 1.18 GJ/t CO_2 in CTM process and 0.9 GJ/t in coal-to-hydrogen for power generation process to 0.7 GJ/t, showing advantage for CO_2 capture.
(3) The proposed polygeneration process presents 16.5% energy saving ratio.

Acknowledgments: The authors are grateful for the financial support from the China NSF project (21736004).

References

D. Xiang, Y. Qian, Y. Man, S. Yang, 2014, Techno-economic analysis of the coal-to-olefins process in comparison with the oil-to-olefins process, Applied Energy, 113, 639-47.
Y. Man, S. Yang, J. Zhang, Y. Qian, 2014, Conceptual design of coke-oven gas assisted coal to olefins process for high energy efficiency and low CO_2 emission, Applied Energy, 133, 197-205.
D. Xiang, S. Yang, X. Liu, Z. Mai, Y. Qian, 2014, Techno-economic performance of the coal-to-olefins process with CCS, Chemical Engineering Journal, 240, 45-54.
X. Zhang, X. He, T. Gundersen, 2013, Post-combustion Carbon Capture with a Gas Separation Membrane: Parametric Study, Capture Cost, and Exergy Analysis, Energy & Fuels, 27, 4137-49.
S. Li, H. Jin, L. Gao, 2013, Cogeneration of substitute natural gas and power from coal by moderate recycle of the chemical unconverted gas, Energy, 55, 658-67.
R. Smith, 2005, Chemical Process Design and Integration.
C. He, X. Feng, K.H. Chu, 2013, Process modeling and thermodynamic analysis of Lurgi fixed-bed coal gasifier in an SNG plant, Applied Energy, 111, 742-57.
S. Jarungthammachote, A. Dutta, 2008, Equilibrium modeling of gasification: Gibbs free energy minimization approach and its application to spouted bed and spout-fluid bed gasifiers, Energy Conversion and Management, 49, 1345-56.
C. Rhodes, S.A. Riddel, J. West, B.P. Williams, G.J. Hutchings, 2000, The low-temperature hydrolysis of carbonyl sulfide and carbon disulfide: a review, Catalysis Today. 59, 443-64.
M.C. Carbo, J. Boon, D. Jansen, H.A.J. van Dijk, J.W. Dijkstra, R.W. van den Brink, et al., 2009, Steam demand reduction of water–gas shift reaction in IGCC power plants with pre-combustion CO_2 capture, International Journal of Greenhouse Gas Control, 3, 712-9.
S. Li, H. Jin, L. Gao, X. Zhang, 2014, Exergy analysis and the energy saving mechanism for coal to synthetic/substitute natural gas and power cogeneration system without and with CO_2 capture, Applied Energy, 130, 552-61.
C. He, X. Feng, 2012, Evaluation indicators for energy-chemical systems with multi-feed and multi-product, Energy, 43, 344-54.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50226-3

Automatic Decomposition of Nonlinear Equation Systems for Improved Initialization and Solution of Chemical Engineering Process Models

Erik Esche[a*], Saskia Bublitz[a], Gregor Tolksdorf[a], Jens-Uwe Repke[a]

[a]*Process Dynamics and Operations Group, Technische Universität Berlin, Sekr. KWT 9, Str. des 17. Juni 135, D-10623 Berlin, Germany*
erik.esche@tu-berlin.de

Abstract

The initialization of process models for chemical engineering systems is often a challenge. While dedicated strategies exist for state of the art column models, novel models need to be initialized manually. In this contribution, we present a first application of these decomposition methods for supporting the initialization and solution of equilibrium- and non-equilibrium-type column models for distillation and compare them to well-known sequential solution methods for each. We show that these methods can supply viable solution procedures for nonlinear systems and are especially helpful in deriving new initialization techniques for novel process equipment.

Keywords: nonlinear system, initialization scheme, distillation, absorption.

1. Status Quo - Initialization and Solution of Column Models

State of the art process simulation tools feature dedicated initialization and solution schemes for distillation and absorption columns, and many other unit operations. Famous examples are Amundson and Pontinen's (1958) equation tearing method for equilibrium stage models, the sum-rates method tailored for absorption problems (Friday and Smith, 1964), etc. Of course, these methods have seen many evolutions over time and by now the inside-out methods (Boston and Sullivan, 1974) and their descendants are dominating the field of equilibrium-based column models.

1.1. Inside-Out Method for Column Models

The main advantage of the inside-out method lies in the introduction of an intermediate layer between thermodynamic calculations and the column model itself. This additional layer contains simplifying surrogate models for the phase equilibrium coefficients K and enthalpies h, which are successively retrained with complex thermodynamic models based on updated state variables from the column (J.D. Seader et al., 2011). The inside-out method requires the user to guess a temperature profile for the entire column and to estimate a vapor/gas flow for each stage. Based thereon the column model is split into an outer and an inner iteration loop. For the inner loop the stripping factors are taken as iteration variables ($S_{c,tr} = K_{c,tr} \cdot V_{tr} / L_{tr}$). In the outer loop calculations, the K and h correlations are then retrained and returned to the inner loop. Further details can be found in (R.A. Russel, 1983) and (J. Jelinek, 1988). The inside-out methods have proven so successful that they are nowadays implemented in most modern flowsheet simulators and applied for equilibrium and rate-based calculations of distillation, absorption, and extraction. Nevertheless, there are still many cases, in which solution schemes inside process simulators fail to reliably solve them. One example are complex

absorptions with electrolytes, which are difficult to initialize given the wide range of concentrations of all ionic species (Esche et al., 2014).

1.2. Initialization of Novel Models

For new unit operation models, tailored initialization and solution schemes usually become necessary and need to be implemented by the designer of each unit operation. Naturally, these steady-state models are systems of nonlinear equations and the main reason for not solving all equations simultaneously upfront lies in the ill-conditioning or in how bad a user-supplied initial guess is. In this contribution, we will investigate techniques from linear algebra to derive initialization techniques for nonlinear models. To benchmark these, their performance is compared against the inside-out method for the initialization and solution of column models.

2. Review of Methods in Linear Algebra

In linear algebra, several techniques are known, which by preconditioning, sorting of equations and variables, definition of loops, etc. improve the convergence of linear solution methods. Two famous examples are the Dulmage-Mendelsohn Decomposition (Dulmage and Mendelsohn, 1958) and the Bordered Block Transformation (Erisman et al., 1985).

2.1. Dulmage Mendelsohn / Block Diagonal Decomposition

The Dulmage Mendelsohn decomposition (DM) is a block diagonal decomposition, i.e. the nonzero elements of the Jacobian matrix are aligned along the main diagonal, so far that individually solvable blocks are created, which can be computed in sequence. This is achieved by permuting the original order of both rows and columns of the Jacobian. To this extent the DM decomposition can also be used to identify over- and underdetermined subsystems of a nonlinear equation systems. Further details hereon can be found in (Bublitz et al., 2017). The DM decomposition has two modes. In its “fine decomposition” the number of blocks along the main diagonal is maximized, leading to as small as possible subblocks. The block at the bottom right of the reorganized matrix needs to be evaluated first and then step by step each further block along the diagonal. The upper triangular section of the matrix contains additional occurrences of variables in equations, which are not sorted into any of these blocks. The respective variables are computed in prior blocks and have fixed values for their later occurrence. Several implementations of the DM decomposition are available. For this contribution, the implementation contained in (HSL, 2017) is used.

2.2. Bordered Block Transformation

Whilst the DM decomposition has the goal of a mostly empty lower triangular section of the Jacobian matrix, the bordered block transformation (BBTF) leads to a sparse upper triangular section. The system is split into a dense and a sparse part. The BBTF is a maximum matching algorithm, which primarily identifies variables, which appear in numerous equations. These variables appear as spikes above the main diagonal in the resorted Jacobian matrix. These spikes are subsequently moved to the right side of the Jacobian. Variables with few appearances on the other hand are sorted below the main diagonal. Similarly, equations with few variables are moved to the top left of the matrix. As a result, the BBTF lists variables with a great influence on the whole equation system, albeit from a linear perspective. The variables corresponding to the spikes are possible candidates for tearing the nonlinear system for an outer iterative loop. Therefore, initialization of these variables is essential and needs to be carefully done to

allow for successful solution of the whole nonlinear system. All other variables could theoretically be left uninitialized seeing that they are calculated in sequence by iterative solution of the resorted and teared system starting from the tearing variables. In this contribution, the BBTF is also taken from the Harwell Subroutine Library (HSL, 2017). As outlined above, the DM decomposition and the BBTF can be used to ease the computational effort for the solution of a nonlinear system, solely based on the structure of the Jacobian matrix. Given that the conditioning of the individual subsystems identified by DM decomposition or the BBTF-teared system with its corresponding subsystems cannot be worse than the conditioning of the entire nonlinear system, it can be expected that the solution strategies outlined above will show a better numeric performance than the equation-oriented solution of the original system. However, the question remains, whether these decompositions lead to better solution strategies compared to dedicated column algorithms, e.g. the inside-out method outlined above.

2.3. Implementation in MOSAICmodeling and Export to MATLAB

To support the application of both decompositions (DM & BBTF), they have been implemented in our department's equation-based modeling and code generation environment MOSAICmodeling (Merchan et al., 2016). Hence, initialization and solution sequences can be automatically generated for all equation systems implemented in MOSAICmodeling, which are tailored to the specific numerical needs of an individual system. To benchmark the solution strategies provided by DM decomposition and BBTF, all four different options have been implemented in MATLAB. As the baseline, the entire NLE model implemented in MOSAICmodeling is exported to MATLAB and solved with a standard Newton-Raphson algorithm. This solution path will subsequently be referred to as NR. The DM and BBTF options are similarly exported to MATLAB and the same Newton method is employed for the solution of all nonlinear subsystems. Finally, for the comparison against the state of the art column algorithms, an inside-out method is manually implemented in MATLAB, which also uses the same Newton method for all implicit, nonlinear solution steps on the inside. This will subsequently be referred to as IO. The resulting calculation sequences for DM, BBTF, and IO are sketched in Fig. 1.

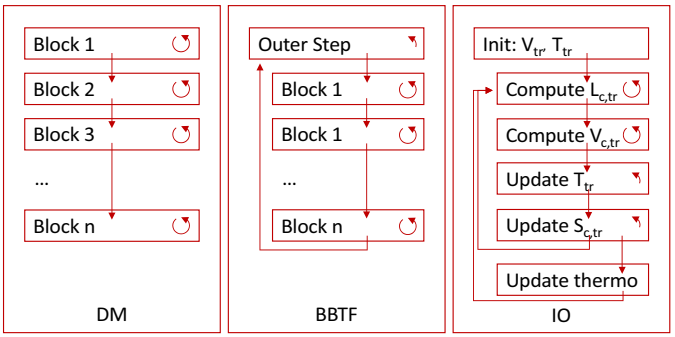


Figure 1: Graphical representation of calculation sequences based on DM and BBTF and for the implemented inside-out method. A full circular arrow implies a complete Newton-Raphson execution, a partial circular arrow depicts a single Newton step with dampening.

3. Case Studies

In this contribution two different applications of column models are investigated to benchmark the aforementioned implementations of DM and BBTF against an IO method and a basic NR algorithm.

3.1. Case Study 1 - Deisobutanizer

The first application is a distillation column for the separation of n- and i-butane. The column model consists of 20 trays using Raoult's law for the phase equilibrium and tray efficiencies based on DIPPR 102. Energy and molar component balances are formulated for each tray with enthalpies and densities based on further correlations from the DIPPR database (127 for enthalpy of vapor phase, 101 for the vapor pressure, 106 for the heat of evaporation, 105 for liquid phase densities). The model is further equipped with a total condenser and a common reboiler. The feed is positioned at stage 13 and consists of 40 % i-butane and 60 % n-butane. The pressure at the top of the column is set to 700 kPa and there is a constant pressure drop of 42 Pa per stage. Fig. 2 shows the incidence matrices of the distillation column model. As expected the DM decomposition (Fig. 2, left) shows a large block, which needs to be solved simultaneously. This block refers to the column itself and the interdependence of all trays given the countercurrent vapor and liquid flows. In any case, there is a sizeable reduction of the size of the largest sub-block compared to the overall system, by roughly 60 %. The BBTF (Fig. 2, right) generates a number of tear variables (spikes at the right above the main diagonal). Close inspection shows that these are temperatures on each tray and the vapor flows of i-butane leaving each tray. This is very similar to the guess values required by the IO methods, which demand temperature profiles and total vapor flow profiles entered by the user for the initialization of the IO procedure.

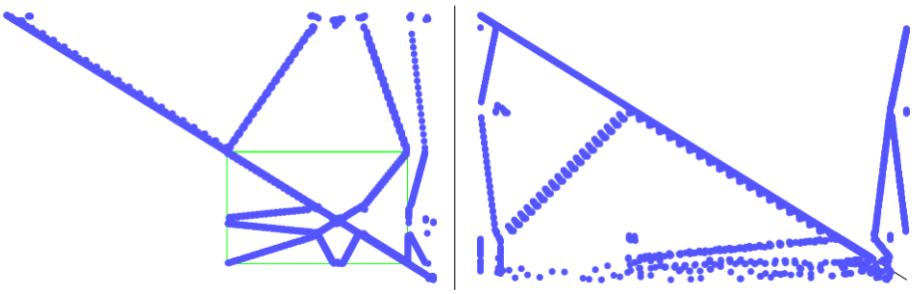


Figure 2: DM decomposition of the distillation column model (left) and BBTF (right).

For the subsequent case study the entire model is first of all manually initialized, i.e. all variables are given adequate guess values and then solved with the NR. The obtained solution is then used as a starting point for all further discussions. All state variables are then perturbed by ± 10, 20, and 30 % of their absolute values and the solution is then attempted with NR, DM, BBTF, and IO methods in turn. The results of this variation are summed up in Tab. 1. Apparently, the pure NR method has great issues with the variation of the initial points, which leads to large numbers of iterations and even failures to converge. The DM is coping better, despite the large block of equations it still has to solve simultaneously. Quite surprisingly the BBTF and the IO method show a highly similar behavior despite the fact that the IO performs a lot finer iterative computation compared to the BBTF with only two iterative layers. However, for both cases initializing temperature and vapor flows at reasonable values proved essential for

obtaining any solution at all, while the values of all other state variables can be left essentially uninitialized.

Table 1: Comparison of variation of initial point and its effect on the convergence of algorithmic solution methods of the deisobutanizer column in case study 1. For BBTF and IO the maximum number of iterations in inner (inn.) and outer (out.) iterations is given. For DM the maximum number of iterations in one subblock is given.

Variation	NR	DM	BBTF	IO
-30 %	failure	31 iterations	21 out. / 52 inn.	10 out. / 57 inn.
-20 %	failure	29 iterations	20 out. / 51 inn.	5 out. / 89 inn.
-10 %	4198 iterations	28 iterations	17 out. / 30 inn.	4 out. / 80 inn.
+10 %	4943 iterations	26 iterations	17 out. / 34 inn.	5 out. / 23 inn.
+20 %	failure	34 iterations	17 out. / 44 inn.	7 out. / 101 inn.
+30 %	failure	failure	19 out. / 68 inn.	7 out. / 75 inn.

3.2. Case Study 2 – Absorption Column

The second application is an absorption column for the removal of CO_2 from a flue gas stream using an aqueous solution of monoethanolamine (MEA) as scrubbing liquid. The column model has 35 theoretical equilibrium stages, which employ a tray efficiency correlation fitted to experimental data (Esche et al., 2014). Similarly, to the first example, molar component and energy balances are computed on each tray with correlations for enthalpies, heats of absorption, etc. The flue gas fed to the bottom of the column has a molar fraction of 18 % CO_2 and the liquid feed at the top a residual CO_2 load of 0.1 mol CO_2 per mol MEA. The column is operated at 2 MPa. The gas feed has a temperature of 20 and the liquid feed of 40 °C. Both models describe the same basic tray structure, the results regarding the decompositions are highly similar. For the individual absorption column, a large block will always remain based on DM. A slight change can be seen regarding the BBTF. Given that a ternary system is now in place, the tear variables change slightly: all stage temperatures and stage vapor flows for CO_2 an H_2O are recognized as important initials. Yet again, the model is initialized at a point close to the solution and then variations are carried out for the four different solution methods detailed above. The results are reported in Tab. 2.

Table 2: Comparison of variation of initial point and its effect on the convergence of algorithmic solution methods of the absorption column in case study 2 (same notation as in Tab. 1).

Variation	NR	DM	BBTF	IO
-30 %	107 iterations	21 iterations	15 out. / 60 inn.	3 out. / 77 inn.
-20 %	108 iterations	21 iterations	15 out. / 60 inn.	5 out. / 248 inn.
-10 %	86 iterations	20 iterations	14 out. / 60 inn.	5 out. / 82 inn.
+10 %	235 iterations	19 iterations	14 out. / 60 inn.	5 out. / 101 inn.
+20 %	727 iterations	21 iterations	15 out. / 60 inn.	5 out. / 52 inn.
+30 %	failure	529 iterations	15 out. / 60 inn.	4 out. / 56 inn.

The DM-based solution strategy outperforms the pure NR. Regarding the BBTF and the IO, variation of the initial temperatures for both by even 10% fails right away. Hence, the temperatures are initialized at reasonable values near the solution, while all other tear variables can be varied freely. For both latter techniques, the tear variables become absolutely essential for a successful solution. Of course, this is also a major advantage,

seeing that the set of variables to be initialized can be heavily reduced. It can be concluded that the maximum matching as implemented by the BBTF algorithm shows a reliable performance comparable to the IO method for complex column models. Seeing that the implementation of the solution strategy is quite straight forward based on the equation-oriented form of the nonlinear system, there is a clear advantage regarding novel, custom-made unit operations. The DM-based strategy appears robust and highly versatile for aiding in the solution, although more variables need to be initialized.

4. Conclusions and Outlook

The solution of nonlinear models remains an issue in chemical engineering. While dedicated solution strategies exist for common unit operations, these need to be designed for new models. Chemical engineering models can be reliably decomposed to improve their convergence compared to the monolithic solution with a single Newton-Raphson solver. In this contribution, we have presented an application of the DM decomposition and of the BBTF to find initialization strategies for custom unit operations. To evaluate their performance a comparison has been carried out against the inside-out method. The case studies of a deisobutanizer and of a CO_2 absorption column show the merit of the solution strategy in reducing the required effort on the initialization of all state variables. In future work, these decompositions are extended by new algorithms for maximum matching in nonlinear systems. In addition, the results of the decompositions will be made available inside MOSAICmodeling's user-defined language specification environment. Hence, the users will be able to design their own solution methods using the solution environment of their choice.

References

N.R. Amundson, A.J. Pontinen, 1958, Multicomponent Distillation Calculations on a Large Digital Computer, Ind. Eng. Chem., 50, 730-736

J.F. Boston, S.L. Sullivan, 1974, A new class of solution methods for multicomponent, multistage, separation process, Can. J. Chem. Engr., 52, 52-63

S. Bublitz, E. Esche, G. Tolksdorf, V. Mehrmann, J.-U. Repke, 2017, Analysis and Decomposition for Improved Convergence of Nonlinear Process Models in Chemical Engineering, Chem. Ing. Tech., 89, 11, 1503-1514

A.L. Dulmage, N.S. Mendelsohn, 1958, Coverings of bipartite graphs, Canad. J. Math., 10, 517-534

A.M. Erisman, R.G. Grimes, J.G. Lewis, W.G. Pool, 1985, A structurally stable modification of Hellerman-Rarick's P4 algorithm for reordering unsymmetric sparse matrices, SIAM J. Anal. 22, 369-385

E. Esche, D. Müller, R. Kraus, G. Wozny, 2014, Systematic approaches for model derivation for optimization purposes, Chem. Eng. Sci., 115, 214-224

J.R. Friday, B.d. Smith, 1964, An analysis of the equilibrium stage separations problem – formulation and convergence, AIChE J., 10, 698-707

HSL. A collection of Fortran codes for large scale scientific computation. http://www.hsl.rl.ac.uk, last accessed: 01.12.2017.

J. Jelinek, 1988, The calculation of multistge equilibrium separation problems with various specifications, Comp. Chem. Eng., 12, 2-3, 195-198

V.A. Merchan, E. Esche, S. Fillinger, G. Tolksdorf, G. Wozny, 2016, Computer-Aided Process and Plant Development. A Review of Common Software Tools and Methods and Comparison against an Integrated Collaborative Approach, Chem. Ing. Tech., 88, 50-69

R.A. Russel, 1983, A flexible and reliable method solves single-tower and crude-distillation column problems, Chem. Eng., 17, 52-59

J.D. Seader, E.J. Henley, D.K. Roper, 2011, Separation Process Principles, 3rd Edition, Wiley

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50227-5

Numerical simulations on direct contact condensation of saturated vapor to subcooled liquid spray

Hengtao Ding, Yiqing Luo*, Yangmin Pan, Xigang Yuan

School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, China
luoyq@tju.edu.cn

Abstract

In the present study, computational fluid dynamics (CFD) simulations were carried out to investigate the direct contact condensation heat transfer of saturated steam to subcooled liquid spray, where the subcooled liquid was sprayed into the saturated steam through a pressure-swirl nozzle. The characteristics of spray hydromechanics with and without heat transfer were compared and analysed, including spray cone angle, sheet breakup length and sheet thickness. It was found that heat transfer had a great impact on spray cone angle and sheet breakup length but little effect on sheet thickness. And heat transfer coefficient in the sheet region was much higher than that in the droplet region, which indicated that the sheet region played the dominant role in the heat transfer process.

Keywords: CFD, direct contact condensation, heat transfer, pressure-swirl nozzle.

1. Introduction

Direct contact condensation heat transfer (DCCHT) of saturated steam to subcooled liquid spray has been of great interest to researchers for its application possibilities in modern industries. Pressure-swirl nozzles are usually used to spray the subcooled liquid into the saturated steam. The liquid emerges from the pressure-swirl nozzle, then forms an annular sheet (sheet region) and finally breaks up into fine droplets (droplet region).

Experimental researches had been carried out in the past to investigate the heat transfer mechanism in the sheet region and droplet region. Weinberg (1952) studied heat transfer process in the sheet and droplet regions experimentally. It was reported that the ratio of temperature rise in the sheet region was 73~93 % of the entire heat transfer region. Takahashi et al. (2001) studied the condensation heat transfer of saturated steam to a hollow-cone spray experimentally. They found from their experiments that the heat transfer coefficient in the sheet region was significantly higher compared to that in the droplet region. It had been commonly accepted that dominant heat transfer took place in the sheet region. However, in the experimental study of Mayinger and Chavez (1992) for the direct contact condensation of refrigerant R113 spray with saturated vapour at higher pressure, it was found that temperature rise of the spray in the sheet region was not more than 50%.

Previous studies showed differences on the heat transfer coefficient of the spray. The differences can be explained by that each study had used different nozzles and operating conditions affecting the sheet and droplet behaviours substantially. Besides, the real-time temperature field measurement can’t be achieved and the flow field was affected by measuring equipment. Therefore, the mechanism of DCCHT was still limited due to the

poor repeatability and complex characteristics of the fluid flow. However, it is obvious that CFD numerical simulation can avoid these problems. Ibrahim (2006) performed simulations on the hollow-cone sprays produced by pressure-swirl nozzles. The comparison between the predicted and measured breakup length showed a good agreement. Mandal et al. (2008) studied the effect of nozzle geometry parameters on the external flow characteristics. The simulation results were proved to predict the isothermal two-phase flow through a pressure-swirl nozzle with reasonable accuracy.

During the last few decades, the internal and external flow characteristics of the pressure-swirl nozzle had been studied experimentally and numerically. Considerable effort had been made to study the initial sheet breakup process and spray evolution without heat transfer and phase change. However, there has been a lack in the CFD models to simulate the condensation phenomena in the spray process, which more often than not occurs between the gas-liquid countercurrent flow in industrial processes. In order to clarify the heat transfer mechanism and develop the heat transfer theory with phase change. It is significant to do research on DCCHT numerically.

The primary intention of this work is to develop a CFD based model that allows the predictions of heat and mass transfer as well as gas-liquid interface in the DCCHT. The heat and mass transfer due to phase change were modelled by adding the suitable source terms in the energy and continuity equations. The temperature contours and spray characteristics were predicted using the established CFD models. The results were compared to the experimental data reported by Takahashi et al. (2001). The heat transfer mechanism in the sheet and droplet regions was discussed and interpreted through CFD predictions and experimental results.

2. Computational domain

Figure 1 shows the geometry of the pressure-swirl nozzle studied in the present work.

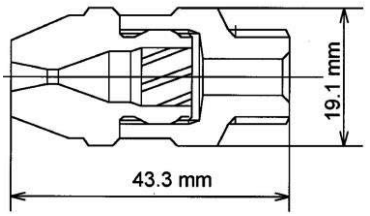
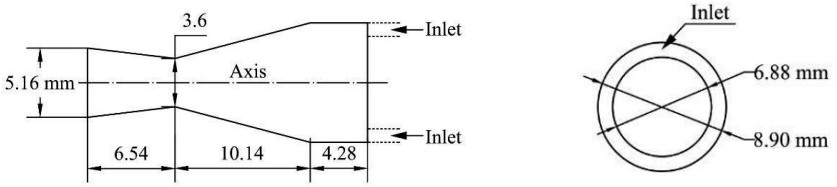


Figure 1: Hollow-cone spray nozzle

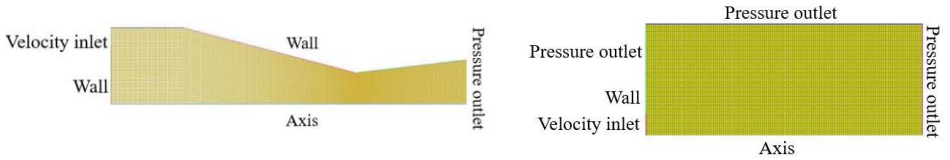
Because of the centrosymmetric structure of the pressure-swirl nozzle, the equivalent model was established as shown in Figure 2, which was used as the computational domain of the internal nozzle flow field. In addition, a cylindrical space with a height of 40mm and a radius of 16mm was determined to be the external flow field of the nozzle.



(a) Front view (b) Top view

Figure 2: Equivalent model of the nozzle

The software package ICEM 15.0 was used to generate rectangular mesh in the computational domain as shown in Figure 3. Based on the assumption that the internal and external flows of the nozzle were spatial symmetry, computational domains were set in 2-D axisymmetric models. In the velocity-inlet of the internal flow field, the volume fraction of the water was 1 and the velocity of the water stayed constant. The pressure was specified to atmospheric pressure and the temperature was 373.15 K in the pressure-outlet.



(a) Internal flow field (b) External flow field

Figure 3: Computational domain

3. CFD model

In the current study, the volume of fluid (VOF) model was used to model the multiphase flows in the DCCHT. The results of the momentum and energy equations were shared among phases. The tracking of the gas-liquid interface was accomplished by solving the volume fraction equation for the secondary phase (liquid phase):

$$\frac{\partial}{\partial t}\left(\alpha_l \rho_l\right)+\nabla\cdot\left(\alpha_l \rho_l \vec{v}_l\right)=\dot{m}_{vl}-\dot{m}_{lv} \tag{1}$$

Lee (1979) developed an evaporation-condensation model for the mass transfer between phases. Based on the assumption that the phase change occurs at a constant pressure and at a quasi thermo-equilibrium state, the mass transfer, as shown in the right side of Eq (1), can be described as follows:

Evaporation:

$$\dot{m}_{lv}=C_1\alpha_l\rho_l\frac{T_l-T_{sat}}{T_{sat}} \tag{2}$$

Condensation:

$$\dot{m}_{vl}=C_2\alpha_v\rho_v\frac{T_{sat}-T_v}{T_{sat}} \tag{3}$$

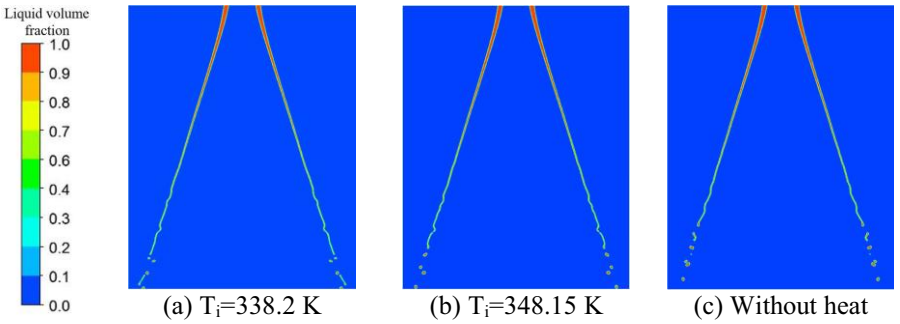
Where C is the so-called Lee model coefficient. Determination of the coefficient C requires the expressions for the interfacial temperature and the volumetric interfacial surface area of the liquid sheet. However, it is difficult to determine the volumetric interfacial surface area due to its dependence on the interfacial structure and phase change. The value of C must be fine tuned to numerically maintain the interface temperature within $T_{sat} \pm 1$ K. In the present study, C_1 was set to 0 and C_2 was set to 100. The source term in the energy equation can be obtained by multiplying the rate of mass transfer by the latent heat due to condensation.

The CFD analysis presented in this study was completed by ANSYS FLUENT version 15.0 using a pressure-based solver. In order to reach a low-resource-consumption

numerical calculation, the internal and external flow fields were calculated separately. The numerical results of the nozzle outlet boundary condition were used as the inlet boundary condition in the calculation of the external flow field.

4. Results and discussion

The external flow fields predicted by the CFD simulation with different water inlet temperatures T_i and without heat transfer are plotted in Figure. 4. It is measured that the spray cone angles are 34.6 °, 36.2 ° and 37.2 ° for water inlet temperature T_i of 338.2 K, 348.2 K and the case without heat transfer respectively. The corresponding measured spray cone angles from experiments (Takahashi et al., 2001) are 37.4 ° and 39.6 ° for water inlet temperature T_i of 338.2 K and 348.2 K. The relative errors between simulations and the experiments are 7.5 % and 8.6 %. It is found that spray cone angle increases with the increase of T_i.



(a) T_i=338.2 K (b) T_i=348.15 K (c) Without heat

Figure 4: Liquid volume fraction contour

The sheet breakup length is determined from the corresponding liquid volume fraction contour shown in Figure 4. The simulated sheet breakup length is 33.0 mm (without heat transfer), which is approximately the same as 33.2 mm predicted from the empirical correlation of Lee et al. (1984). In addition, sheet breakup length decreases from 35.4 mm to 34.3 mm when T_i increases from 338.2 K to 348.15 K.

As shown in Figure 5, the sheet thickness of T_i = 338.15 K is slightly higher than that without heat transfer. This is possibly because steam is condensed into liquid and then result in a small increase in the sheet thickness. However, both the two cases indicate that, as the sheet move downstream, it becomes thinner due to the steam shear exerted on the sheet.

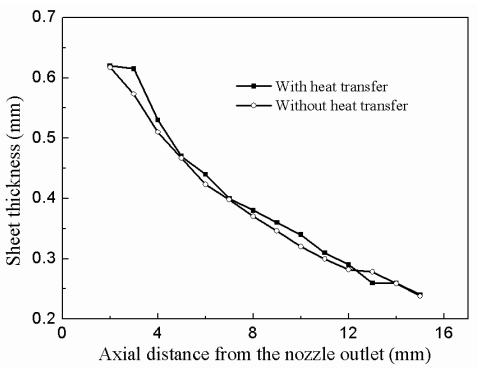


Figure 5: Profile of sheet thickness along the flow direction

Comparisons between the simulated and the measured radial temperature profiles from the thermocouples (Takahashi et al., 2001) have been performed in Figure 6. The mean deviation for temperatures at the center of the sheet is below 1.8 % with a maximum deviation of 2.8 % at x =20 mm. It can be found that with the increase in axial distance, the simulated and measured temperature profiles are broader. Overall, the temperature profiles obtained with the CFD model agree well with the experimental data. However, differences in the radial locations of the sheet can be found at the vicinity of the nozzle outlet. In the simulations, the resistances of the nozzle wall and steam on the liquid spray are not considered. As a result, the simulated radial locations of the sheet are larger than those in the experiments.

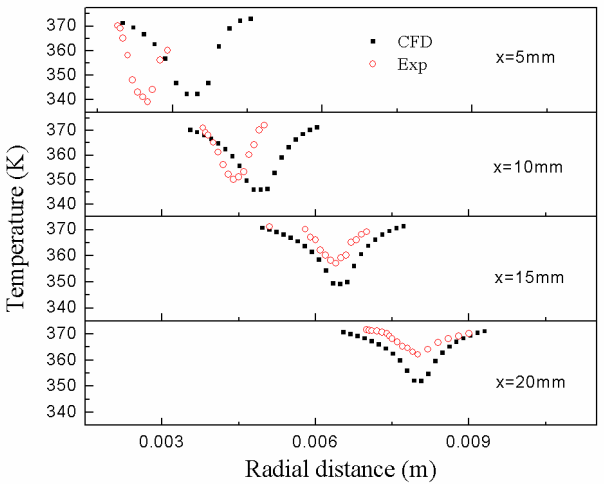


Figure 6: Comparison of the simulated temperature profiles with experimental data at various axial distance (x) from the nozzle outlet

The simulation results for the local heat transfer coefficient and the dimensionless temperature are shown in Figure 7. The local heat transfer coefficient decreases with the increase of the axial distance. It is obvious that the heat transfer coefficient in the sheet region is comparatively higher than that in the droplet region. The present data of sheet dimensionless temperature $(T-T_i)/(T_{sat}-T_i)$ where the sheet breakup is about 0.60, which means 60 % of heat transfer has been achieved in the sheet region. That is to say, dominant heat transfer takes place in the sheet region, even though the length of the sheet is small.

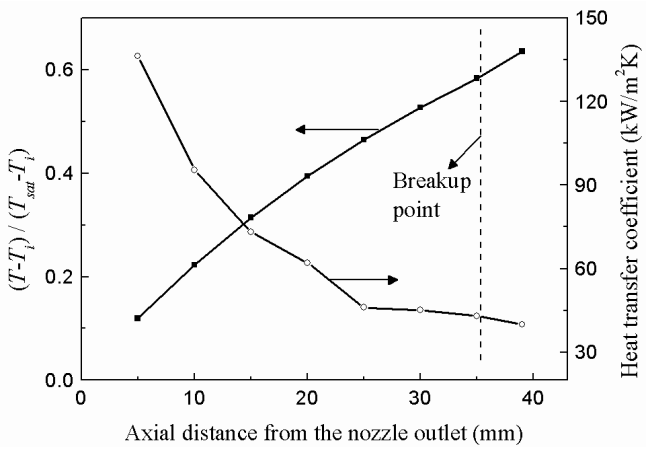


Figure 7: Variations of dimensionless temperature and heat transfer coefficient along the flow direction

5. Conclusion

A two-dimensional CFD model was developed to simulate the DCCHT. External flow characteristics of the pressure-swirl nozzle, as well as its performance parameters including spray cone angle, sheet breakup length, sheet thickness, temperature variations and heat transfer coefficient, were predicted. The computational model was verified by comparisons with experimental measurements and a good agreement was obtained. It was concluded that heat transfer in the sheet region was much more efficient than that in the droplet region. For the direct contact between the liquid spray and steam without heat transfer, results showed that sheet breakup length and sheet thickness were smaller while the spray cone angle was higher compared with DCCHT.

References

S. Weinberg, 1952, Heat transfer to low pressure sprays of water in a steam atmosphere, Proc Inst Mech Engrs, 06 (1B), 240-253.

M. Takahashi, A. K. Nayak, S. Kitagawa, et al., 2001, Heat transfer in direct contact condensation of steam to subcooled water spray, Journal of heat transfer, 123 (4), 703-710.

F. Mayinger, A. Chavez, 1992, Measurement of direct-contact condensation of pure saturated vapour on an injection spray by applying pulsed laser holography, International journal of heat and mass transfer, 35 (3), 691-702.

A. A. Ibrahim, 2006, Comprehensive study of internal flow field and linear and nonlinear instability of an annular liquid sheet emanating from an atomizer, Cincinnati: University of Cincinnati.

A. Mandal, M.A. Jog, J. Xue, et al., 2008, Flow of power-law fluids in simplex atomizers, International Journal of Heat and Fluid Flow, 29, 1494-1503.

W. H. Lee, 1979, A Pressure Iteration Scheme for Two-Phase Modeling, Technical Report LA-UR, 79-975.

S. Y. Lee, R. S. Tankin, 1984, Study of Liquid Spray (Water) in a Condensable Environment (air), International Journal of Heat and Mass Transfer, 27, 363-374.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50228-7

An Improved Recycle-loop Tearing Algorithm Based on Process Topology Information

Yi Ding[a,c], Weidong Zhang[a,b], Rafiqul Gani[d], Wei Sun[a,b*]

[a] *College of Chemical Engineering, Beijing University of Chemical Technology, Beijing, 100029, China*
[b] *Beijing Key Lab of Membrane Science technology, Beijing University of Chemical Technology, Beijing, 100029, China*
[c] *School of Chemistry and Materials Science, University of Science and Technology of China, Hefei, 230026, China*
[d] *PSE for SPEED, Skyttemosen 6, DK-3450 Allerod, Denmark*
**sunwei@mail.buct.edu.cn*

Abstract

Loop tearing is a conventional problem to solve in process simulation. Upadhye-Grens algorithm provided a solution to identify the family of all available tearing sets. However, the random starting point and unit-wise substitution may require more searching steps. In this work, information on both loop-stream and unit-stream are considered, and a logic matrix minus is defined and applied to replace the substitution based on individual unit in Upadhye-Grens algorithm. By utilizing the information on loop-stream in a loop matrix, the first non-redundant tearing set is easily identified. By defining a unit-stream matrix, the substitution in the improved algorithm is achieved by logic based matrix calculation, which could effectively speed up the identification of the family of all available tearing sets, as the substitution is simultaneously conducted at the process scale instead of simply unit-by-unit. In order to verify the proposed algorithm, an industrial process simulation is considered as an example. The searching process and result are compared with those obtained in Aspen plus.

Keywords: unit-stream matrix, logic minus for matrix, tearing algorithm

1. Introduction

Chemical process simulation technique is playing an important role in modern chemical process engineering design. Due to the inherent characteristics of chemical process, process model usually appears as a huge, sparse and nonlinear equation set, which is the key challenge for process simulation. Generally, all simulation problems can be converted into a series of sub-equation groups in certain order. It will get complicated when one or more recycle-loops (to be termed as "loop") exit in proposed process topology, where one or more starting points have to be identified first, called a tearing set. It has been proved that the convergence property depends on the selection of tearing sets (Gossner et al., 2003). Therefore, the efficiency of such a complex process simulation can be significantly influenced by the choice of the tearing sets.

Sargent and Westerberg (1964) proposed the computing time as a standard to evaluate the tear algorithm. An effective algorithm based on dynamic programming method was proposed by Upadhye and Grens (1972). In their later work (Upadhye and Grens, 1975), the non-redundant tearing set was defined to avoid repetitive tearing streams, and a

substitution rule was employed to identify all non-redundant sets. Subsequently, Westerberg (1979) proposed an algorithm to find an optimal tearing set based on the substitution rule. According to Yang and Ma (1988), the improved algorithm of Motard and Westerberg (1981) has been implemented in Aspen plus.

In general, the optimal tearing set should be at least a non-redundant tearing set and can be easily recognized if all the non-redundant tearing sets are identified, i.e. the non-redundant tearing family. Hence, after one of non-redundant tearing sets is found, it is easy to find all others, since all the non-redundant tearing sets are replaceable. Therefore, it is important to find the first non-redundant tearing set. However, for conventional Upadhye-Grens algorithm, it could be quite time consuming, when starting from a random valid tearing set, especially, when more units are included in one loop, where substitutions have to be conducted unit by unit, as the substitution rule is run in route, depending on the length of process system.

For a system with several loops, the information of loops and streams can be described by loop matrix, which consists of streams as columns and loops as rows. Topological information among streams are stored in a loop matrix and the redundant stream can then be easily identified via loop matrix. For the tearing of a loop, the concept of dependent stream was proposed by Lee (Lee et al., 1966), which is used to simplify calculations. However, in this approach, the logical relationship of unit-stream has not been considered, and the streams are simply removed.

In our work, a process topology information based algorithm is proposed. The tearing performance of the proposed algorithm is compared with that obtained from Aspen Plus for an industrial process.

2. Algorithm

2.1. Definition

A process system, as shown in Fig.1, is considered as a conceptual example, based on which definitions are given for the description of proposed algorithm.

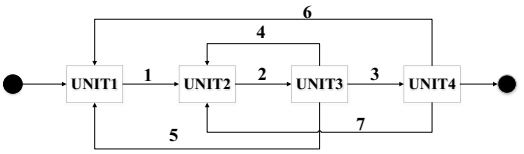


Figure 1 A simple process system

Loop matrix: It is a matrix with stream *j* as column and loop *i* as row, [a_{ij}]. If a loop *i* includes a stream *j*, the element in the loop matrix a_{ij}=1, otherwise a_{ij}=0. In the case of above, the loop matrix could be given as in Eqn.(1):

$$\begin{array}{c} \\ LOOP1 \\ LOOP2 \\ LOOP3 \\ LOOP4 \end{array} \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array} \tag{1}$$

Unit-stream matrix: It is defined as a matrix that shows streams information of output and input for every unit in a process, in which, column *j* is referred as stream and row *i*

corresponds to a loop. There should be n rows included in this incidence matrix for a process with n units [U_{ij}]. For a U_{ij}, the element for input stream is represented by 1, and that for output stream is represented by -1. In the case above, the unit-stream matrix could be given as in Eqn.(2):

$$\begin{array}{c} \\ UNIT1 \\ UNIT2 \\ UNIT3 \\ UNIT4 \end{array} \begin{array}{c} \begin{array}{ccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{bmatrix} -1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & -1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & -1 \end{bmatrix} \end{array} \tag{2}$$

Logical loop matrix minus: It is a logical operation, and can be implemented through the following three steps:

Step1. Give a valid stream-tearing set [X_k] for a process, in which the stream is torn if corresponding element is assigned to 1, otherwise to 0, i.e. it means that stream 2, 3, and 5 consist a valid tearing set if [X_k] is represented as [0 1 1 0 1 0 0].

Step2. Deduce every row of unit-stream matrix ([U_{ij}]) from the stream-tearing set [X_k] to obtain a new matrix with same size of [U_{ij}]. The element obtained in the new matrix [U_{ij}] can be -1, 0, 1, and 2. These elements have their specific physical meaning. If the element is -1, it means that the stream cannot be substituted via this unit, thus, the corresponding row shall be removed. If the element is 2, it means the stream is torn twice, the element shall be reset as 1. It can be seen that the element value is again either 0 or 1 in obtained matrix.

Step3. Put all obtained stream-tearing vectors together into matrix [Y_{ij}], keep all records of output [Y_{ij}].

The process of logical loop matrix minus is shown in Fig.2.

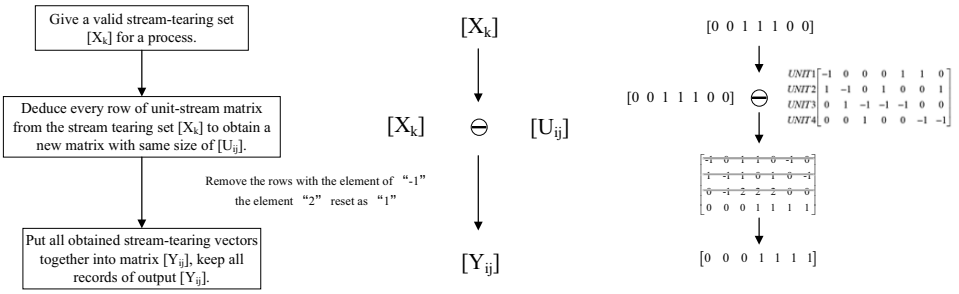


Figure 2 The schematic for logical loop matrix minus and an example

2.2. Method for identification of non-redundant stream tearing family

Given a process with a specified loop matrix and a unit-stream matrix, the non-redundant stream tearing family can be obtained through the following steps.

Step1. Randomly select a tearing stream in each row of loop matrix, to obtain a valid stream-tearing set, [X_k].

Step2. Check if there is a dependent stream among elements in the valid stream-tearing set, [X_k], obtained in Step 1, according to the loop-stream information described in loop

matrix, and set 0 to the element in $[X_k]$ corresponding to the dependent stream in the loop matrix, to obtain a new $[X_k]$.

Step3. Compare $[X_k]$ with each row in loop matrix $[a_{ij}]$, to see if there is a loop torn more than twice. If so, conduct a logical loop matrix minus between $[X_k]$ and unit-stream matrix, then perform Step 2 for each row; otherwise, go to Step 4, and in such case, the first non-redundant tearing stream set is obtained.

Step4. Conduct logical loop matrix minus, keep record of all rows in $[Y_{ij}]$ in the form of row vector.

Step5. Repeat Step 4 for each row vector in $[Y_{ij}]$ until the same non-redundant tearing stream set is obtained again. Collect all non-redundant tearing stream sets obtained from the row vector of $[Y_{ij}]$, i.e., the non-redundant tearing stream family. Stop searching if all row vectors in $[Y_{ij}]$ obtained in Step 4 have been considered. Put all non-redundant tearing stream sets together to form an overall non-redundant family for the process.

The corresponding block diagram is shown in Fig.3.

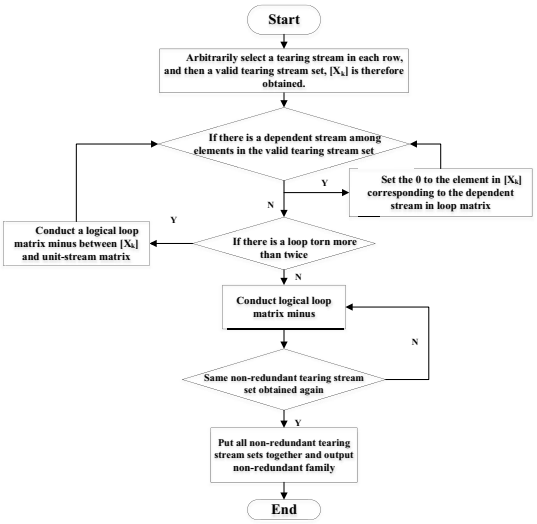


Figure 3 Algorithm block diagram

3. Application

3.1. A simple block diagram and discussion

The algorithm is applied to identify a tearing family for the process shown in Fig 1. The loop matrix and unit-stream matrix are given in section 2.1.

A valid tearing set can be obtained randomly for the system. For example, $[X_k]$ (3, 4, 5, 6) is a valid tearing set. It can be seen that stream 6 is a dependent stream relative to stream 3. Therefore, a new row $[X_k]$ (3, 4, 5) = [0 0 1 1 1 0 0] can be obtained.

In the Step 4, the stream tearing vector $[X_k]$ can be considered as non-redundant tearing set. Then, the process of logical loop matrix minus is calculated by proposed algorithm.

Finally, an overall non-redundant family can be obtained from the recorded row vectors. Non-redundant family: (3, 4, 5), (4, 5, 6, 7), (1, 4, 7) and (2).

For further discussion, an extreme starting point (1, 2, 3, 4, 5, 6, 7) is investigated. By the proposed algorithm, the non-redundant set (2) is obtained in one loop iteration. By contrast, at least three loop iterations are need to reach a non-redundant set by Upadhye-Grens algorithm for which the times of loop iterations are influenced by the choice of initial valid tearing sets. In the proposed algorithm, the introduction of stream-loop information allows the simultaneous consideration of both stream-unit and stream-loop, which makes the iteration is no longer completely conducted unit by unit along in the process. Consequently, the searching for the first non-redundant tearing set can be simplified significantly, and the initial value sensitivity can be reduced.

The algorithm is also applied on the well-known Cavett process (Cavett, 1963) to test the applicability of proposed algorithm. Details are not given because of the page limit. Compared to the conventional Upadhye-Grens algorithm, a non-redundant tearing set is obtained in less loop iterations, therefore, searching time is significantly reduced.

3.2. Application to an industrial process

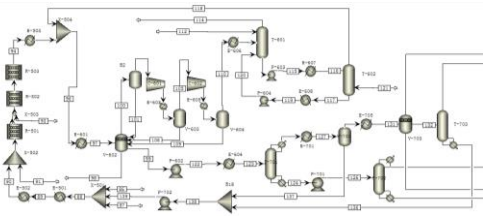


Figure 4 Flowsheet for an industrial process for the production of styrene through dehydrogenation of ethylbenzene dehydrogenation drawn in Aspen plus.

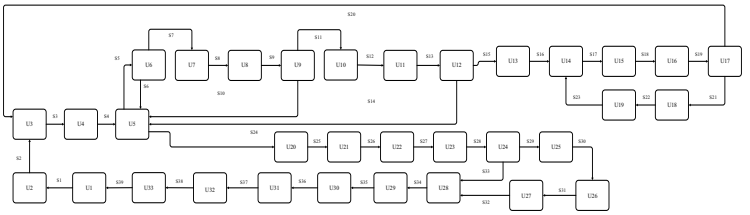


Figure 5 A simplified block diagram of the styrene production process

Figure 4 shows the flowsheet of an industrial chemical process where 250kt/y ethylbenzene is dehydrogenated to styrene. The proposed algorithm is tested on this process flowsheet. As shown in Figure 5, it is difficult to get the first non-redundant tearing set directly. The calculation results are obtained by proposed algorithm and compared with the result from Aspen plus.

A tearing set given by Aspen plus is (S5, S23, S34), which is similar to the first non-redundant tearing set (S5, S23, S27) we calculated by cutting all the streams in the process as the starting point. According to the process topology, they are equivalent in terms of the convergence in direct iteration in process simulation. Given the information available in the block diagram, it is hard to tell which one is better in simulation, therefore, another non-redundant tearing set (S5, S23, S39) is arbitrarily picked for comparison purpose. Process simulation is conducted with three tearing sets as the starting point respectively. The number of iterations needed to reach the convergence for each starting set is listed in Table 1. It can be seen that the least number of iterations is required when non-redundant

tearing set (S5, S23, S34) is adopted, which indicates that the one given by Aspen plus is not the best choice at least for the iteration algorithm employed by Aspen plus itself. With more stream information, the better or optimal tearing set can be easily identified based on the proposed algorithm.

Table. 1 No. of iterations required for three non-redundant tearing sets of Aspen plus

Non-redundant tearing set	No. of Iterations
S5, S23, S27	78
S5, S23, S34	112
S5, S23, S39	122

4. Conclusion

By introducing the topology information in a loop matrix and a unit-stream matrix, the non-redundant tearing sets can be identified more efficiently, compared to the Upadhye-Grens algorithm. Similar tearing sets are obtained with a default setting provided by Aspen plus, which shows the feasibility of proposed algorithm for the simulation of any chemical process. It has been noted that the optimal tearing set can be easily identified from the recognized non-redundant tearing sets by the proposed algorithm by providing more information on the units and streams of the flowsheet.

References

Gossner H, Esmark K, Stadler W., 2003, Simulation methods for ESD protection development[M]. Elsevier, Chapter 4 - Process simulation, 75-278.

Sargent R W H, Westerberg A W., 1964, Speed-up in chemical engineering design[J]. Trans. Inst. Chem. Eng, 42: 190-197.

Upadhye R S, Grens E A. , 1972, An efficient algorithm for optimum decomposition of recycle systems[J]. AIChE Journal, 18(3):533–539.

Upadhye R S, Grens E A. , 1975, Selection of Decompositions for Process Simulation[J]. AIChE Journal, 21:136-143.

Westerberg A W., 1979 Process flowsheeting[M]. Cambridge University Press, Chapter 7 – Flowsheeting by equation-solving methods based on tearing, 161-191.

Yihong Yang, Dexian Ma., 1988 An introduction to process systems engineering(Chinese)[M]. Hydrocarbon processing press：71-84.

Motard R L, Westerberg A W., 1979, Exclusive tear sets for flowsheets[J]. Aiche Journal, 27(5):725-732.

Lee W, Christensen J H, Rudd D F., 1966, Design variable selection to simplify process calculations[J]. AIChE Journal, 12(6):1104-1115.

R.H. Cavett, 1963, Application of numerical methods to the convergence of simulated processes involving recycle loops, Proc. Amer. Petroleum Institute, 43, 57–76.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50229-9

Comparative Study of Similarity Measures Used to Classify Residential Water Flow Pattern of Low-Income Households in Salvador - Brazil

Mariza Mello[a*], Karla Oliveira-Esquerre[a], Gabriella Botelho[b]

[a] *Universidade Federal da Bahia, Escola Politécnica, Rua Aristides Novis, nº 2, 6º Andar – Federação, EP-UFBA, Salvador, 40210-630, Brazil*
[b] *UFRB, Rua Rui Barbosa, 710 - Centro, Cruz das Almas-BA, 44380-000, Brazil*
mariza.smello@gmail.com

Abstract

Modelling residential water consumption is an important step towards knowledge about users' behaviour and later for the proposition of realistic goals that lead to the conscious consumption. The flows provided by smart meters facilitate the in-depth analysis of consumption and then it is possible to classify them into certain appliances using pattern recognition techniques. The research was conducted in low-income and "social interest" homes in Salvador, Brazil. The objective of this paper is to evaluate similarity measures using an adaptation of 1-nearest neighbour algorithm and to verify which stands out as a suitable option to classify disaggregated residential water flow. The measures chosen were dynamic time warping (DTW), longest common subsequence (LCSS), edit distance on real sequences (EDR) and edit distance with real penalty (ERP). It is also performed parameters optimization of each similarity measure to improve the classification accuracy.

Keywords: water flow, households, pattern recognition, similarity measures.

1. Introduction

Problems such as population growth, water pollution, urban disorderly development, intensive agricultural irrigation, climate change, degradation of spring areas and droughts contribute to disparities between the availability of quality water sources and consumption demand. In view of the above, a deeper understanding of residential water consumption is important tool for improvement of demand management strategies. Knowledge about how and when water is used can inspire viable solutions to decision makers facing water scarcity problems. In addition, the better management of water resources may indirectly influence the amount of energy needed for treatment of the sewage and reduction in the environmental impact caused by the discharge of the effluents in an inadequate manner.

The great availability of fresh water is a notable characteristic of Brazil, however, it is unevenly distributed (ANA, 2016). This study was carried out in the city of Salvador located in the northeast of Brazil, focusing on low-income households. In the socioeconomic aspect, they represent an expressive part of the population of Salvador and of the northeast of Brazil. The study of water consumption took place through the installation of smart meters and data loggers in some the residences.

The information coming from smart meters benefits the studies related to the characterization of indoor water consumption (NGUYEN et al., 2014; LIU et al., 2016). These are high-resolution electronic devices capable of measuring the consumption of water, energy or natural gas in short time intervals, jointly with equipment for almost instantaneous capture of emitted signal. The data collected provide detailed time series of consumption and it is possible to identify used appliances, through the application of a method for both consumption breakdown and pattern identification.

For low-income population, most of the appliances are taps (kitchen, bathrooms and outdoor area), showers and toilets. These devices depend exclusively on human handling and consequently it is associated to culture, water availability, plumbing and the level of awareness of proper water use. In addition, since the types of appliances is limited, they end up accumulating several functions, contributing to increase variation of duration and volume for the same type of equipment. Given the non-homogeneity of the data, it is necessary to develop a method that embraces this variability, without losing the quality of the classification. Among the techniques analysed, we propose a comparative study of similarity measures throughout an adaptation of 1-nearest neighbour (1NN) algorithm. The distances chosen were DTW, LCSS, ERP and EDR. According to Wang et al. (2013), the benefits in this type of approach using 1NN classifier are the ability of evaluate directly the efficacy of them since the performance of 1NN relies on the chosen distance and the simplicity of the algorithm that do not require parameter estimation.

The aim of this paper is to evaluate the selected similarity measures, using adapted 1NN algorithm, and to choose the one that presents the best potential to perform residential water flows classification. They will be tested in steps that includes use of prior restriction of equipment to be tested based on flow's volume and duration, parameter evaluation of each measure in order to find theirs optimal versions and establishment of a threshold to proper assign the water use into the chosen appliance.

2. Method

2.1. Similarity measures for time series

Berndt and Clifford introduced the DTW distance (KEOGH et al., 2005) in order to allow a time series to be "stretched" to provide a better combination with other time series. It is a widely employed and it makes use of distance L1 (also known as Manhattan) as the cost between elements. The EDR (CHEN et al., 2005), ERP (CHEN et al., 2004) and LCSS (VLACHOS et al., 2002) are adaptations for numerical series of the edit distance. It comes from string comparison field and computes the number of insertion, deletion and/or substitution operations that are required to make two series, of different possible lengths, identical to each other. The EDR looks for the minimum number of operations required to transform one series into another and removes noise effects by quantifying the distance between a pair of elements in two possible values: 0 and 1 (LCSS also performs the same quantization). The ERP uses a constant g to quantify the possible gaps in the time series and uses the distance L1 as the cost between the elements. The LCSS computes the distance between the series x and y, searching for the largest sequence in x that can also be found in y, that is, as the name itself says, seeks to find the longest common subsequence (not dealing with substitution operation). It has also a peculiarity, the biggest sequence does not depend on consecutive terms, they just need to occur in the same order (do not consider gaps). The

DTW and LCSS do not follow the triangular inequality and therefore are not metric distances.

The choice of the similarity measure demands that it can capture the dynamics between the time series and perform the classification in a coherent way. Wang et al (2013) presented an evaluation, among other techniques, of similarity measures (with those under evaluation here) by applying them in 38 real databases. One of the conclusions is that as the size of the database increases, the convergence between elastic measures and Euclidean's distance accuracies also increases. In addition, the LCSS, EDR and ERP presented accuracies very close to the DTW and they presented the best results in comparison with the elastic measures addressed.

2.2. Parameter optimizations

The best measure evaluation also depends on the choice of the best parameters through: limitation of the deformation window, minimum distance between two elements and gaps penalty (WANG et al., 2013, KURBALIJA et al., 2014). The limitation of the deformation window (local and/or global constraints) prevent misalignment, slightly accelerates the calculation and can be applied via methods suggested by Sakoe-Chiba, Itakura, among others. The minimum distance between two elements can be varied through the parameter ε for LCSS and EDR. For the ERP, the parameter g has the role to penalize the gaps and is adjustable however, the zero value is appropriate because it is a reference with the horizontal axis of the Cartesian coordinate system (CHEN et al. 2004).

3. Model construction

3.1. Collected data

The database used came from two projects CT-HIDRO (Sectorial Fund for Water Resources) and PROSAB (Basic Sanitation Research Program). The first had the period monitored from December 2010 to January 2012, and analyses intra-households water consumption of two residences in level of end-uses by flowmeter record and disaggregation of consumption using Trace Wizard® software. An important information is that one of the families move out from the house and another one started to reside, towards to end of the project. The new family agreed to continue the research and since they presented another behaviour of water consumption, three residences were considered instead of two. The second project occurred between December 2012 and April 2014, was monitored seven residences with the same methodology to obtain data. The flowmeter resolution was of 0.1L/pulse measured in 10s intervals. During recorded data period occurs a week period to investigate water consumption patterns, in first project these identifications were made with the residents help (by notes with time of consumption and appliance utilized) and in second project was utilized simplified flowmeters attached on each appliance of household. Trace Wizard® software processed these data, splitting simultaneous uses and classifying single uses. Thus, for the evaluation of this method, previously classified consumption data were used.

3.2. Initial tests

Initially, it was studied the application of prior restrictions to limit the number of equipment to be tested. In the model unrestricted (MUR), each unknown time series to be classified was compared with all references of each equipment. With the restriction (MR), duration and volume of time series were used to define which appliances should

be selected for the classification step. The volume and duration criteria were chosen because it is expected these information to be representative of each appliance and therefore to lead to the correct match. If at least one of these criteria fit, the equipment was chosen. Hence, this form of choice is very flexible, since the focus is the reduction of computational time. In general, the usages of taps in bathrooms and kitchen have very similar behaviours and they are therefore easily confused. Thus, they were simplified as taps.

After the initial tests, the most promising models went through parameter optimization step. The parameter ε (LCSS and EDR) suffered a variation with range from 0ml/s increasing in 10ml/s until 30ml/s. The parameter g (EDR) did not suffer modification. The local constraint used was established for Sakoe-Chiba and the chosen were trivial case of no constraint (NC), symmetric with slope constraint condition P=0 (SP0) and asymmetric with P=0 (ASP0). The global restrictions can be applied to all measures evaluated here. However, the size of the window is a function of the length of the time series and since there is a wide variety of lengths, the values must be large to avoid errors. Knowing that there is no limit to the size of the time series, the use of global restrictions becomes unfeasible. At the end of this phase, the final result will be calculated by choosing the most promising model and measures from the initial tests.

4. Preliminary results

4.1. Initial tests

The overall results are in figure 1. The columns indicate the accuracy for the tie results and in this case, it is the maximum value that can be reached, requiring one or more classifications steps to select a single class. The area graph show the accuracy for water flows classified to a single appliance. The line indicates the percentage of the tie results per model per distance. The first point of the analysis is the percentage of tie results and it is required that the value is near zero thus not requiring new classifications steps. Second, the accuracy of the single results is also an important factor and the results show that MR has present less tie results, but MUR is better in accuracy. The next analysis was performed by evaluating the classification in equipment level.

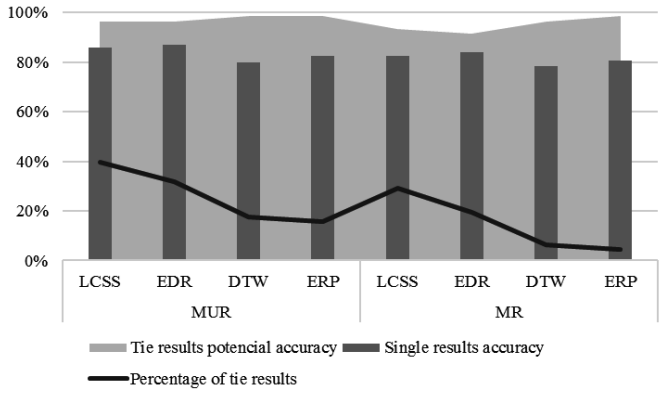


Figure 1 - Overall results of the suggested models.

The results (figure 2) relate the models with the similarity measures and the accuracy by equipment, disregarding the tie results. For both models, the EDR and ERP present in

general the best accuracies. The EDR and ERP have similar form to calculate the distances and because of this, the results are similar; the best outcomes happen mostly without restriction. The LCSS, in searching for the longest common subsequence without taking into account the gaps, has greater possibility of confusing appliances since there are large variability of user's behaviour. For the DTW, the restriction application seems to lead to the correct answer but the classification percentage is lower than the ERP and EDR.

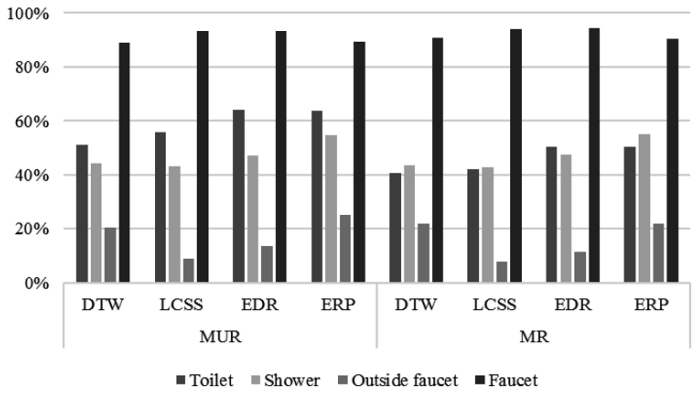


Figure 2 - Accuracy in equipment level.

According to the results, the contribution of the constraints is not clearly positive. Due to this problem, an evaluation of the pre-selection efficiency was carried out and the results indicate that 65% of the outside taps were pre-selected correctly; for the others, the correct pre-selection was higher than 74%. These values limit the accuracy and since the imposition of restriction criteria has the main function of optimizing the computational time, the next step was to perform parameter optimization using MUR.

4.2. Evaluation of measures parameters

The evaluation of the parameters measures variation occurs in the same way that initial tests. For the LCSS, as the threshold ε increased, an increasing of the percentage of tie results was also observed, with variations between 40% and 60%; in addition, there was a slight increasing in the accuracy of the single results. The EDR, which went through the same variation as the LCSS, initially had 31% of tie results and rose to 50%. For both, these variations barely affect the accuracy. With the DTW, the percentage of tie results was very low for SP0 and NC, about 17%, and their accuracy is inferior to the others. Based on all prior information, the final step was test data using EDR with ε = 10ml/s and ERP, in a MUR model.

The classification accuracy was also attributed to the diversity of flow references. Consistently, the faucet is the equipment that holds the highest percentage of flow references, followed respectively by the toilet, shower and outside faucet. This difference in quantities is associated with the frequency of equipment usage, suggesting that for those with low frequency it is necessary more time for data collection. Thus, it is suggested an increase in the flow references. However, this also causes an increasing in computational time.

5. Conclusions

The initial evaluation of the chosen similarity measures combined with adapted 1NN showed that there is an interesting potential of good results. The unrestricted model proved to be the ideal. The accuracy of each similarity measure pointed to interesting features of the modelled data. The LCSS demands homogeneity inside groups, and since the database has a heterogeneity characteristic, its tie results were high. The other measures presented acceptable performance, especially ERP and EDR. The variation of ε showed that, for LCSS and EDR, as it increased, higher was the percentage of tie results. As for the DTW, the best outcome was when the deformation window had no restriction; however, its results were less accurate. The next steps are testing with the major part of the database, determination of thresholds and an attempt to reclassify correctly the misclassified and tie results.

Acknowledgments

We would like to thank TECLIM Network for the data and Adelmo Aguiar for data organization.

References

A. Liu, D. Giurco & P. Mukheibir, 2016, Urban water conservation through customised water and end-use information, Journal of Cleaner Production, 112, 3164-3175.

E. Keogh & C. A. Ratanamahatana, 2005, Exact indexing of dynamic time warping, Knowledge and information systems, 7, 3, 358-386.

Informe 2016, 23. Available in: http://www3.snirh.gov.br/portal/snirh/centrais-de-conteudos/conjuntura-dos-recursos-hidricos/informe-conjuntura-2016.pdf .

K. A. Nguyen, A. R. Stewart & H. Zhang, 2014, An autonomous and intelligent expert system for residential water end-use classification, Expert Systems with Applications, 41, 2, 342-356.

L. Chen & R. Ng, 2004, On the marriage of lp-norms and edit distance, In Proceedings of the Thir tieth international conference on Very large data bases, 30, 792-803.

L. Chen, M. T. Özsu & V. Oria, 2005, Robust and fast similarity search for moving object trajectories, In Proceedings of the 2005 ACM SIGMOD international conference on Management of data, 491-502.

M. Vlachos, G. Kollios & D. Gunopulos, 2002, Discovering similar multidimensional trajectories, In Data Engineering Proceedings. 18th International Conference, 673-684.

V. Kurbalija, M. Radovanović, Z. Geler & M. Ivanović, 2014, The influence of global constraints on similarity measures for time-series databases, Knowledge-Based Systems, 56, 49-67.

X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann & E. Keogh, 2013, Experimental comparison of representation methods and distance measures for time series data, Data Mining and Knowledge Discovery, 1-35.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50230-5

Process design and simulation of a novel wet Claus desulfurization technology

Yang Yang, Xiao Feng and Qiao Zhang*

School of Chemical Engineering and Technology, Xi'an Jiaotong University, Xi'an, 710049, China
.zhang1986@stu.xjtu.edu.cn

Abstract

As the utilization of fossil fuels such as coal, oil and the waste discharged by chemical production process, the environment has been seriously polluted. A large amount of SO_2 emitted from flue gas in thermal power plants is an important cause of environmental pollution. In order to reduce such emission, this paper presents a new process design of liquid-phase Claus recovery process design based on traditional Claus technology. The process is composed of four parts: H_2S absorption and desorption, SO_2 absorption, Claus reaction and separation and recycling. Sodium citrate is employed to absorb sulfur dioxide in the flue gas and MDEA to absorb hydrogen sulfide in shale gas, hydrogen sulfide and sulfur dioxide reacts in the sodium citrate solution to form sulfur. Afterwards, sulfur separation is followed to achieve high efficient sulfur recovery. Meanwhile, absorbents of MDEA and sodium citrate are regenerated to reduce the cost. The simulation of the process is carried out by ASPEN PLUS 10.2 and the results show that hydrogen sulfide and sulfur dioxide can obtain high removal rate. After desulfurization, the sulfur content of purified gas is much lower than that of environmental tolerance level. Such sulfur recovery process is not only theoretical improvement but also useful to future application.

Keywords: Sulfur recovery; Shale gas; Flue gas; Liquid Claus reaction

1. Introduction

The exploitation, processing and utilization of fossil fuels have led to serious environmental pollution, which seriously threatens daily life and sustainable development of human society. Among traditional fossil fuels, because of the complex composition and the high ash, sulfur and nitrogen content, coal will cause dust, sulfur oxides and nitrogen oxides emissions in thermal power plants. Haze, acid rain caused by coal utilization has become urgent environmental issue. Generally, desulfurization processes are classified to wet and dry methodologies. The effect of toluene and carbon dioxide in the sulfur recovery process was investigated through experimental and simulation by Chardonneaua et al. (2015) and they revealed the importance of temperature and its excursion for enhanced sulfur recovery in a Claus process. A modified Claus process, enhanced by the use of High Temperature Air Combustion (HiTAC) technology in the Claus furnace was introduced by Mohamed and Ashwani (2008). Dowling et al. (1990) explored kinetics of the reaction between hydrogen and Sulfur under high-temperature Claus furnace conditions. S.Ibrahim et al. (2013) presented experimental investigations on the effect of different amounts of toluene addition to H_2S combustion in oxygen under Claus condition (pH = 3). In order to

increase the overall sulfur recovery as well as to decrease the costs, El-Bishtawi and Haimour (2004) developed a new modification on conventional Claus process and simulated and studied for various N_2/O_2 ratios with and without using double combustion technique. Chen et al.(2007) reviewed present status of Claus process based sulfur recovery units in domestic refineries. The super Claus sulfur recovery process and characteristics were introduced, and the parameters of device and process gas were calculated and analysed through simulation by Zhang et al.(2016). Cao et al. (2014) simulated the sulfur recovery project of Two oil refineries in China, and the results were in accordance with that of SULSIM software by and large, and the calculation error was within the allowable range. Chen (2003) reviewed the desulfurization and decarbonization by oxyamine process and the results indicates that the wet process has better application prospects. Li et al. (2006) compared several desulfurization process. Li (2008) used the traditional Claus process for desulfurization. Ghahraloud (2017) modeled and optimized the Claus process in thermal and catalytic section. These studies above show that the applications of Claus process in desulfurization are really common. However, traditional Claus process still have some disadvantages. i.e. The reaction temperature is 240 ℃ and the conversion rate is about 65 % in Cao et al. (2014) and the reaction temperature is 1061 ℃ in Zhang et al. (2016). Those process requires higher reaction temperatures and thus they are not conducive to the exothermic reaction. Moreover, since the reaction takes place in gas phase, the reaction rate and the conversion rate are limited. In order to deal with these problems in Claus process, this paper presents a new process design of liquid phase Claus reaction based on traditional Claus sulfur recovery technology. Compared with the traditional Claus process, this reaction is a gas-liquid phase reaction, which can fix gaseous pollutants in the liquid phase, easy to handle, mild reaction conditions, and can simultaneously treat two kinds of exhaust gases and has a high desulfurization rate.

2. Process design

According to existing Claus sulfur removal technology and absorption and desorption recycle, this paper puts forward a novel process design for recovery of sulfur from H_2S in shale gas and SO_2 in flue gas through wet Claus reaction. The novel process is mainly composed of four parts: H_2S absorption and desorption, absorbent recycling, SO_2 absorption and Claus reaction and separation. The flow diagram of the process is shown in Figure 1. This process contains a reactor(R_1), a centrifuge(CF_1), a pump(P_1), a compressor (C_1), 2 mixers (M_1 and M_2), 2 flashes (F_1 and F_2), 3 absorption columns (T_1,$T_2$and T_3) and 4 heat exchangers (H_1,H_2,H_3 and H_4). The model of the reactor we use is RCSTR and the model of the columns T_1-T_3 is RadFrac. Since the reaction system is an electrolyte-containing system, The property method we choose is ELECNRTL. In H_2S absorption and desorption part, shale gas is pressurized by C_1 and heated up by H_1 to arrive at required temperature and pressure of absorption column T_1. Meanwhile, makeup and recycled MDEA is mixed and pumped to T_1 for H_2S absorption. After H_2S absorption, the desorbed methane is sent out as natural gas and the H_2S concentrated MDEA absorption liquid from the bottom of T_1 is sent to desorption column T_2 for H_2S purification and MDEA regeneration. As for absorbent recycling part, the regenerated MDEA is recycled back to pump P_1 and H_2S is sent to flash 1 for further separation of entrained MDEA and H_2O. Afterwards, the separated MDEA is sent to T_2 for reuse and H_2S is purified and sent to reactor for sulfur generation.

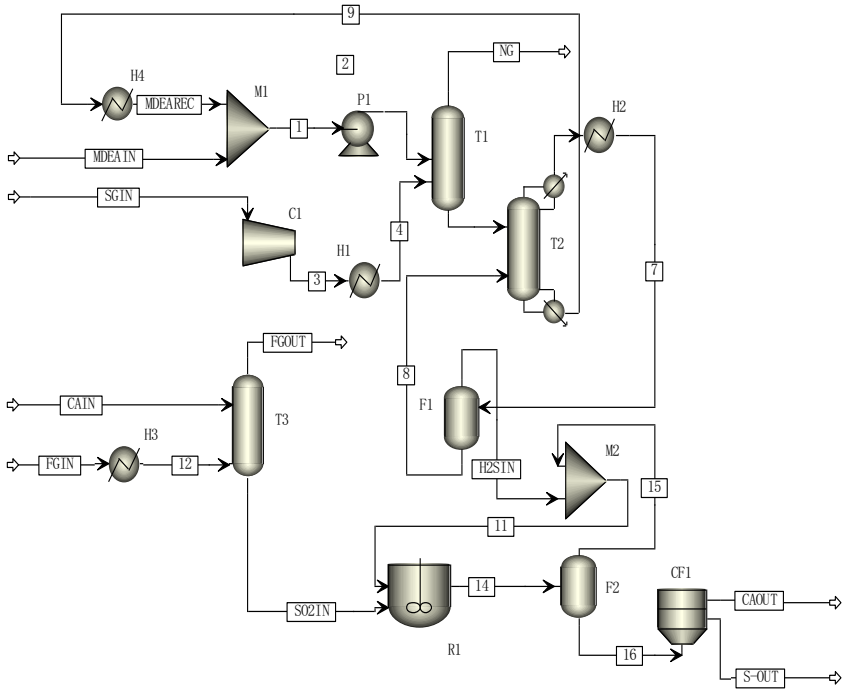


Figure 1 Novel wet Claus sulfur recovery process flowsheet

In SO_2 absorption section, flue gas and sodium citrate are sent to absorption column T_3 for SO_2 separation. After absorption, the residual gas is directly discharged to atmosphere and SO_2 concentrated sodium citrate solution is sent to R_1 for Claus reaction. In Claus reaction and separation section, H_2S and SO_2 react in CSTR reactor R_1 in the existence of acidic sodium citrate solution (pH=4.5) to generate sulfur content. The reactants and products, as well as sodium citrate solution are separated in flash F_1 and centrifuge CF_1 to reuse unreacted H_2S and collect the sulfur product.

3. Simulation and case study results

The shale gas and flue gas data is from practical petroleum enterprise and given by Table 1, where the flow rate and composition of main components are listed. The simulation in this article is carried out by ASPEN PLUS. All computational studies were performed on a HP 288 Pro G2MT desktop with Intel(R) Core(TM) i7,6700 CPU @3.40GHz and 8 GB RAM, using windows 7, 64 bit operating system.

Table 1 Shale gas and flue gas feed composition

	FGIN	SGIN	FGIN	SGIN
	Mole Flow (kmol/h)		Mole Frac	
H_2S	0	6.2	0	0.012
CH_4	0	493.8	0	0.988
N_2	1197	0	0.798	0
O_2	300	0	0.2	0
SO_2	3	0	0.002	0

The composition of shale gas is simplified as 1.24 % (mol) H_2S and 98.76 % (mol) CH_4 and the flue gas is 79.8 % (mol) N_2, 20 % (mol) O_2 and 0.2 % (mol) SO_2. This data comes from the actual data of a shale gas field in China. The 28% (wt) MDEA solution is used as the absorbent to absorb H_2S and the 0.5 mol/L sodium citrate is used as the absorbent to absorb SO_2. The reaction equation (Zhang et al., 2003) is given by Eq. (1).

$$H^+ + 2H_2S + HSO_3^- = 3S + 3H_2O \tag{1}$$

The reaction rate and kinetic equations(Zhang et al., 2003) are given by Eqs. (2) and (3):

$$r = kC_{H_2S}C_{HSO_3^-}^{0.5} \tag{2}$$

$$k = 1.531\times10^4 \exp(-\frac{1.778\times10^4}{RT}) \tag{3}$$

Where temperature and pH are key parameters to both conversion performance. Low temperature and proper acidity are in favor of the reaction. Besides, since the reaction is gas-liquid phase and the mass transfer is controlled by liquid film, high pressure is also an important driving force.

3.1. H_2S absorption and desorption

The mole flow rate of the shale gas is 500 kmol/h with 1.2% H_2S. The temperature and pressure of the absorption process are 25 ℃and 15 atm. Table 2 is the result of the H_2S absorption and desorption. The stream *SGIN* is the feed of the shale gas, the stream *SGOUT* is the gas phase after absorption and the H_2S feed is the H_2S after absorption and desorption. As shown in this table, the absorption rate of H_2S is about 100 %, and the desorption rate is about 99.97 % because MDEA is alkaline and thus has high efficient chemical absorption effect to H_2S. the absorption rate of H_2S is more than 98% at 0.004 MPa (Lin, 2013). The higher absorption pressure, 15 atm, which is more conducive to the absorption of H_2S, makes H_2S absorption rate can reach nearly 100 %. Absorbent MDEA can be recycled after desorption, and the recovery can be more than 98 %. Such high recovery rate of absorbent indicates the process can be economical.

Table 2 Shale gas data and stream compositions after H_2S separation

Streams	SGIN	SGOUT	H_2S feed	SGIN	SGOUT
		Mole Flow (kmol/h)			
H_2O	0	1.129	0.5	0	0.002
H_2S	6.2	0	6.198	0.012	0
CH_4	493.8	493.723	0.077	0.988	0.998

3.2. SO_2 absorption.

Table 3 is the result of the SO_2 absorption. Sodium citrate ($C_6H_5O_7Na_3$) is the absorbent. The stream *FGIN* is the feed of the flue gas and its flow rate is 1500 kmol/h with 0.2 % SO_2. The stream *FGOUT* is the gas phase after absorption and the SO_2 feed is the liquid phase after absorption. From the composition of SO_2 feed in sodium citrate solution, the absorption of SO_2 is about 100 % because of similar reason for H_2S absorption and such absorbent has been industrial applied.

Table 3 Flue gas data and stream compositions after SO_2 separation

Streams	FGIN	FGOUT	SO_2 feed	FGIN	FGOUT	SO_2 feed
	Mole Flow (kmol/h)			Mole Frac		
H_2O	0	23.365	57.236	0	0.015	0.761
N_2	1197	1197	0	0.798	0.787	0
O_2	300	300	0	0.2	0.197	0
SO_2	3	0	0.001	0.002	0	0
HSO_3^-	0	0	2.887	0	0	0.038
SO_3^{2-}	0	0	0.112	0	0	0.001
Na^+	0	0	10.56	0	0	0.14
$C_6H_8O_7$	0	0	0.02	0	0	0
CA-ION_1	0	0	1.429	0	0	0.019
CA-ION_2	0	0	2.834	0	0	0.038
CA-ION_3	0	0	0.118	0	0	0.002

3.3. Liquid Claus reaction

In this section, Claus reaction occurs in CSTR reactor and Table 4 shows the result of reactants and products composition. The stream *S-OUT* is the solid product of the react after separated and the stream *CAOUT* is the liquid phase after react and separated. Such composition shows that the total yield of sulfur reaches up to 91.16 %. The yield of sulfur in Li (2008) is about 90 %; the sulfur recovery after optimization is about 96.32% in Ghahraloud (2017); the conversion rate is about 65 % in Cao et al. (2014); and the total yield of sulfur is about 99.32 % in Zhang et al.(2016). Although those normal Claus processes can give superior yield such as 99.32 %, such result is obtained by multiple reactors in series and the reaction conditions are rigor. Compared with traditional Claus process, the yield can be raised from 65 % to 91.16 % after one reactor in this process. Thus, this process are superior to traditional ones both on performance and conversion condition aspects. From comparison in Tables 1-4, the H_2S and SO_2 components in shale gas and flue gas are concentrated and then converted to sulfur product. Furthermore, as all absorption, desorption and reaction occurs in liquid phase, the separation and conversion are high efficient. We can see that the whole process is simple, reaction and separation conditions are mild and the absorbents are recoverable. Thus, the process design and results are not only meaningful theoretically but also remarkable to industrial application.

Table 4 Reaction and separation products composition

Streams		FGIN	SGIN	S-OUT	CAOUT
Mole Flow (kmol/h)	H_2S	0	6.2	0	0.028
	HS^-	0	0	0.002	0.549
	S^{2-}	0	0	0	0
	SO_2	3	0	0	0
	HSO_3^-	0	0	0	0.003
	SO_3^{2-}	0	0	0.001	0.201
	S	0	0	8.387	0

4. Conclusions

This paper presents a novel wet Claus sulfur recovery process design and performs its simulation based on existing kinetic and reactive equations by ASPEN Plus. Compared with other processes of sulfur recovery, the new process has the advantages of simple process design, mild reaction condition, high safety and recyclable absorbents. Furthermore, it can not only deal with the sulfide in the flue gas and the shale gas to reduce environmental pollution but also obtain high value-added sulfur product. It can be expected good environmental and application prospects.

5. Acknowledgement

Financial support from the National Natural Science Foundation of China under Grant 21506169 is gratefully acknowledged.

References

H. Cao, S. Huang, F. Wang, 2014, Simulation of Claus Sulfur Recovery Process using HYSYS Software[J], Technology & Development of Chemical Industry. In Chinese

M. Chardonneaua, S. Ibrahim, A. K. Gupta, 2015, Role of Toluene and Carbon Dioxide on Sulfur Recovery Efficiency in a Claus Process[J], Energy Procedia, 75(24):3071-3075

G. Chen, 2003, Review and prospect for desulfurization and decarbonization by oxyamine process[J]. Chemical Engineering of Oil & Gas, 32(3):134-133. In Chinese

G. Chen, 2007, Progress of applying Claus sulfur recovery technology[J], Petroleum Processing & Petrochemicals. In Chinese

N. I. Dowling, J. B. Hyne, D.M.Brown, 1990, Kinetics of the reaction between hydrogen and sulfur under high-temperature Claus furnace conditions[J], Industrial & Engineering Chemistry Research, 29(12):2327-2332

R. El-Bishtawi, N. Haimour, 2004, Claus recycle with double combustion process[J], Fuel Processing Technology, 86(3):245-260

H. Ghahraloud, M. Farsi, M. R. Rahimpour, 2017, Modeling and optimization of an industrial Claus process: Thermal and catalytic section[J]. Journal of the Taiwan Institute of Chemical Engineers, 76:1-9

S. Ibrahim, A. A. Shoaibi, A. K. Gupta, 2013, Role of toluene in hydrogen sulfide combustion under Claus condition[J], Applied Energy, 112(112):60-66

J. Li, J. Sun, D. Wang, et al. 2006, Desulfurization process for refinary sour dry gas[J]. Petrochemical Technology & Application. In Chinese

R. Li, 2008, Comparison between Sulfur Recovery Processing Process[J]. Fuel & Chemical Processes. In Chinese

L. Lin, 2013, Simulation and analysis for selective removing H_2S with MDEA, Guangdong Chemical Industry. In Chinese

M. Sassi, A. K. Gupta, 2008, Sulfur recovery from acid gas using the Claus process and High Temperature Air Combustion (HiTAC) technology[J], American Journal of Environmental Sciences, 4(5)

G. Zhang, X. Ma, J. Gao, 2016, Process simulation calculation and analysis of Super Claus[J], Chemical Engineering(CHINA), Vol. 44 No. 8. In Chinese

Y. Zhang, Z. Zheng, C. Zhang, 2003, A study on eeaction kinetics of H_2S and SO_2 in liquid phase[J], Journal of Chemical Engineering of Chinese Universities. In Chinese

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50231-7

CFD-assisted Modeling and Analysis on Residual Monomer Stripping Using Spinning Cone Column

Seongwoong Bae[a], Shinhyuk Kim[a], Jay H. Lee[a*]

[a]*Department of Chemical and Biomolecular Enginnering, KAIST, Daejeon, 34141, Korea*
jayhlee@kaist.ac.kr

Abstract

Spinning Cone Column (SCC) showing higher removal efficiency under a mild operation condition can be an attractive option for the residual monomer removal process in polymer production. The effect of the rotation speed should be more carefully explored since the strong shear force and impingement on the wall, caused by a high rotational force, can induce operational stability issues. In this work, transport phenomena inside SCC are studied from a different angle. The Eulerian-Lagrangian approach is introduced in the CFD modeling to simulate the dynamics of the liquid droplets formation and movements as well as the complex motion of the fluid. Full 3D CFD transient simulations are conducted under various operation and design conditions. Key features for the enhancement in the residual monomer removal rate of SCC will be discussed based on the results.

Keywords: Spinning Cone Column, CFD modeling, Monomer stripping.

1. Introduction

Spinning Cone Column (SCC) is a stripping column that contains repeated units of a rotating cone attached to the rotating shaft at the central axis of the column and a stationary cone that is attached to the inner wall (Figure 1). It has been utilized as an equipment for stripping, especially in the food and cosmetics industries given its higher stripping efficiency with the shorter residence time that enable a mild operation as required by products like wine and perfume (Casimir, 1974).

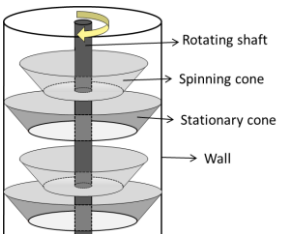


Figure 1. Simplified scheme of Spinning Cone Column

Due to its capability to operate in mild conditions with a satisfactory efficiency, our industrial partner has been exploring SCC (Joo et al., 2012) for the removal of residual monomers in polymer latex. A stripping process to remove residual monomers in polymer latex, which is a resultant mixture of emulsion polymerization, is essential as the residual monomers can negatively affect health and polymer quality.

Several studies have been performed to model and analyze the fluid dynamics inside SCC. Through a series of investigations into the flow regime followed by CFD simulations of SCC, a mass transfer analysis was performed (Makarytchev, et al., 2004).

Although comprehensive analyses have already been carried out on SCC, still there are necessities for additional investigations related to the phenomenon of monomer removal. First, to apply SCC for the different target mixture that is polymer latex in this study, separation should occur in milder condition with the adjusted design specification. Second, there are advantages of selecting different flow models in CFD modeling of SCC. Specifically, necessary data for analysis can be easily achieved through careful selection of flow model.

In this study, transient three-dimensional CFD modeling of a lab-scale SCC is performed. A novel approach, which is a combination of Discrete Phase Model (DPM) and Eulerian Wall Film (EWF) for liquid phase including droplets formation and movements, coupled with the Eulerian framework for vapor phase, is applied to express the complex flow fields in a more realistic way. Critical effects of the rotation on monomer removal in SCC will be discussed using several simulations at different conditions.

2. CFD modeling

CFD modeling is done in full three-dimension with ANSYS Fluent 18.0 to consider the effect of strong swirling motions and turbulences caused by rotation.

2.1. Geometry and mesh

The number of cones in SCC can be varied depending on the desired degree of separation and the properties of the mixture. In this work, not an entire, but a part of lab-scale SCC is modeled to reduce the computational cost of CFD. Total three stationary cones and two spinning cones are carefully selected so as to minimize the direct distortion of the by the adjacent inlet/outlet boundary conditions (Figure 2).

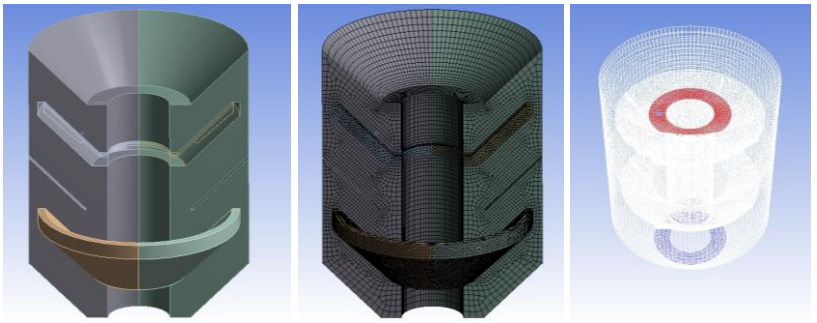


Figure 2. Details of CFD modeling: Geometry, meshes, and boundaries from the left

Geometry is drawn by referring to the design specification provided by the industrial partner. Meshes are finely set at the adjacent region of the cones where the complex flows of both liquid and vapor are expected. Mesh motion, which is a direct rotation of designated meshes, is utilized to express the rotation of the cone. Particularly fine meshes are used for the rotating meshes (The second plot in Figure 2). The total number of the mesh is 505,691. Hexagonal meshes are used in most regions except the rotating meshes. The rotating meshes are set as the tetrahedral type. Average skewness of the meshes is 0.21 and the maximum is 0.95, which sets the acceptable mesh quality.

2.2. Boundary conditions

For the inlet of a liquid phase, an arbitrary sized small surface is drawn near the end of the top stationary cone (a blue square in the red region shown in the third diagram of Figure 2). Liquid outlet is also designated at the bottom surface indicated in the blue donut shape, and it is also set as a vapor inlet so as to express counter-current flow inside SCC. With this configuration, injected liquid coming from the blue rectangle falls onto the spinning cone and is spread to the wall by the rotational force. Then, it flows down along the wall to the stationary cone by gravity. Vapor injected at the bottom goes upward and makes counter-current flow or crossflow.

With the above configuration of geometry and meshes, water is used as a pure component so as to focus on the fluid dynamics only without unduly increasing the computational cost. Water is injected in the form of uniformly distributed liquid droplets. Regardless of the setting of the diameter of the injected water, all droplets are converted to liquid film after touching the upper surface of the spinning cone, and the CFD results are unchanged. So, the diameter of injected water is set to an arbitrary value of 1mm.

2.3. Model description

This study involves novel selection of a model to reproduce the different flow types in SCC. The selected model is arranged in the following table.

Table 1. Summary of selected model

Phenomena		Selected model
Flow type	Vapor	Eulerian Equation
	Liquid film	Eulerian Wall Film
	Liquid droplet	Discrete Phase Model
Viscous model		Standard $k-\epsilon$ (scalable wall function)

The combination EWF and DPM enable the realistic tracking of liquid droplets movements as well as their generation and extinction. Information on the liquid film including film velocity and thickness can be obtained with a EWF model. At the edge of the cones, the film is torn off and droplets are generated according to the Foucart's criterion. Liquid droplets are affected by the vapor flow field until they impinge to the wall. When the droplets touch the cones and the wall, their fates are determined by Stanton-Rutland model which is the criteria deciding whether the droplets are attached, rebound, spread or splashed based on their kinetic energy (Stanton and Rutland, 1996).

Due to the spatial limitation, detail equations for each model are omitted. Related equations would be easily found from the references or through the Fluent theory guide.

3. Results

Effect of rotation on each flow and its relevance on the enhancement of residual monomer removal is the main focus of this study. Due to the space limit, some key results are written in flat sentences without figures.

3.1. Vapor

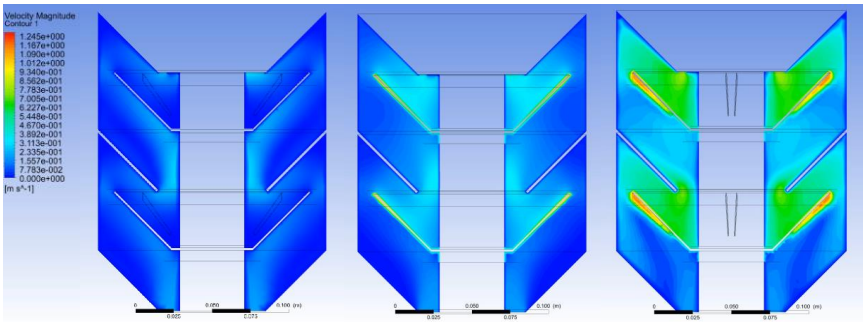


Figure 3. Vapor velocity (left: 0rpm, center: no fin/170rpm, right: fin/170rpm)

Figure 3 shows the overall characteristics of vapor flow inside SCC. All contour plots are expressed in the same range of velocity magnitude. These are the cases when the different operating and design conditions are applied, but the same amount of vapor injected. The first one is the case of no rotation (0rpm), the second is the case of no fin with the rotational speed of 170rpm, and the third one is the case of fins with 170rpm.

The effect of rotation is clearly observed by comparing the first with the second and third plots. Without rotation, vapor goes upward with the lower velocity along the shortest path. The maximum velocity occurs in the inner throat region that is a narrow space between the rotating shaft and the edge of the stationary cone.

When cones rotate, vapor flows faster, especially upon the spinning cone. An area-averaged velocity magnitude upon the spinning cone shows about 2.1 to 5.9 times higher value compared to other regions. This velocity is intensified with the existence of fin attached beneath the spinning cones. Main velocity component is the circumferential velocity, and this trend is also more distinctive on the spinning cone.

In addition, a remarkable observation is the decrease of axial velocity components heading upward at the inner throat region. When the cones rotate (170rpm), the average axial velocity at the inner throat region is reduced about 60% compared to the case of 0rpm. This indicates that advantages of rotation are the effective utilization of stripping agent as well as enhanced vapor velocity. The advantage also comes from its unique structure which causes bottleneck near the rotating shaft. In short, through the rotation, injected stripping agent sweeps the entire area inside SCC at a higher speed and a longer residence time by rotation. This advantage enhances the removal rate of monomer significantly since it is usually known that the contact of steam at the liquid interface enhances stripping in a polymer de-volatilization process.

3.2. Liquid film

Another continuous flow in SCC is the liquid film, which is modeled with a EWF model. Calculated data of liquid film velocity are extracted through the line drawn along the flow direction upon the spinning cone, the wall, and the stationary cone. Since the liquid film velocity shows significant fluctuations at each time step, time-averaged data are analyzed.

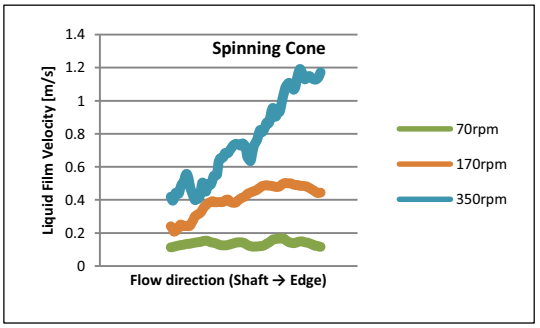


Figure 4. Liquid film velocity on the spinning cone with different rotation speed

CFD simulations are performed at different rotation speeds. Figure 4 shows the liquid film velocity enhancement on the spinning cone where the rotational force is directly applied to the liquid. Average values of above result are 1.36E-01, 4.01E-01, and 7.85E-01 [m/s] for 70, 170, and 350 [rpm], respectively.

To examine the effect of the liquid film velocity increase on the residual monomer removal, mass transfer coefficients are calculated based on the film theory using the CFD results. A mass transfer correlation of Eq.(1), developed by Vivian and Peaceman (1956), is applied.

$$Sh_L = \frac{k_L z}{D_{AB}} = 0.433(Sc_L)^{\frac{1}{2}}\left(\frac{\rho_L^2 g z^3}{\mu_L}\right)^{\frac{1}{6}}(Re_L)^{0.4} \quad (1)$$

D_{AB} indicates diffusivity of residual monomer in the liquid. z is computation domain. ρ_L and μ_L are density and viscosity of the liquid. g is gravitational acceleration. Re_L and Sc_L are Reynolds number and Schmidt number for the liquid phase.

Averaged values of liquid side mass transfer coefficient (k_L) for 70, 170, and 350 rpm are calculated as 4.05E-05, 7.03E-05, and 1.00E-4, respectively. It shows the clear increasing trend as rotating speed increased. By adjusting the rotation speed from 70 rpm to 350 rpm, stripping rate can be locally enhanced about 2.5 times. This implies that rotation of cone contributes to monomer removal by a direct increase of the liquid film velocity on the spinning cone.

3.3. Liquid droplets

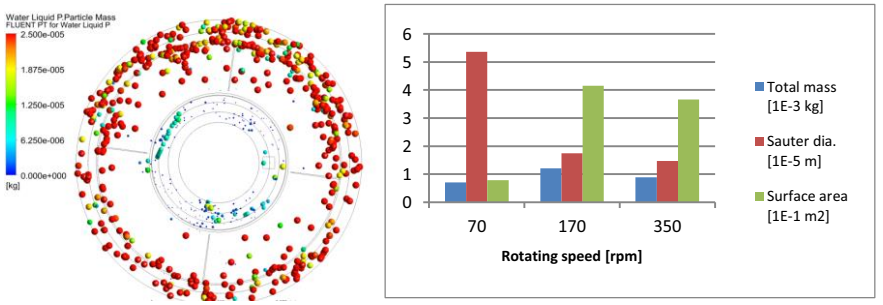


Figure 5. Simulation results on liquid droplets (left-overall configuration, right-summary)

Liquid droplets are generated usually at the edge of the cones by tearing of the liquid film. This phenomenon is reproduced by the combination of EWF and DPM. Film velocity at the edge is related to the flowrate of droplets, and their movements are affected changed by drag force from the vapor flows. Droplets are tracked in the Lagrangian frame of reference using DPM so that droplets information such as the location and the size of droplets can be easily effectively visualized. In addition, it is possible to obtain the quantitative information of overall liquid droplets with this approach.

The left picture in Figure 5 indicates the liquid droplets distribution in SCC viewed from z+ direction (looking down from the top). Heavy droplets are observed near the edge of the spinning cone while fine droplets exist around the rotating shaft.

The plot on the right side of Figure 5 summarizes the overall information on the liquid droplets existing in SCC at different rotating speed. A noticeable fact related to this summary is that the estimated values of the surface area of droplets are from 7.87E-02 to 4.15E-01 m^2. These correspond to 56 % or 298 % increases compared to the maximum area where the film can cover in SCC. In short, the residual monomer removal rate may be greatly enhanced by the significant increase in the mass transfer area due to the formation of droplets.

4. Conclusions

CFD modeling of SCC was performed with the combination of EWF and DPM. Several observations for each flow from the CFD simulation results conducted in different design and operating conditions were presented. Key features that can affect the monomer removal rate were analysed.

Vapor flows faster but with a longer residence time by rotation of the cones and the fluid dynamics of vapor is much intensified by the fins. Rotation also contributes to an increase in the liquid film velocity. The mass transfer coefficient for the monomer removal rate can be enhanced up to 2.5 times by changing the rotating speed from 70 to 350 rpm, according to the film theory. In addition, the surface area of liquid is significantly enlarged by the generation of liquid droplets. A 56 % to 298 % increase in the surface area is expected by the rotation.

References

D. J. Casimir, 1974, Technological aspects of the production of concentrates of passionfruit—Passiflora edulis, Ph.D. Thesis, University of New South Wales, Sydney, NSW, Australia.

D. W. Stanton and C. J. Rutland, 1996, Modeling Fuel Film Formation and Wall Interaction in Diesel Engines, SAE Paper, 960628.

E. J. Joo, D. Y. Shin and J. K. Lee, 2012, SCC with improved mobility of material to be removed, US9162210 B2

S. V. Makarytchev, T. A. G. Langrish and D. F. Fletcher, 2004, Mass transfer analysis of spinning cone columns using CFD, Trans IChemE, Part A, Chem. Eng. Res. Des., 92(A6), 752-761

J. E. Vivian and D. W. Peaceman, 1956, Liquid-side resistance in gas absorption, AIChE Journal, Vol. 2, Issue 4, 437-443

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50232-9

Dynamic Modelling and Simulation of Solid Oxide Fuel Cell Based Auxiliary Power Unit System

Daewook Kim, Sohyun Jeong, Jay H. Lee*

Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon, 34141, Republic of Korea
jayhlee@kaist.ac.kr

Abstract

Solid oxide fuel cell (SOFC) based auxiliary power unit (APU) system using diesel fuel is expected to be an environmentally-friendly and highly-efficient alternative to supply non-propulsion power for long-haul trucks, replacing engine idling or diesel generator based APUs. The main components that govern the overall system performance are the auto-thermal reformer (ATR), which converts diesel fuel into hydrogen-rich gas, and the SOFC stack, which produces power by electrochemical reaction of the reformed fuel and oxygen. In this study, a dynamic ATR model including a reaction kinetic model with experimental validation and a dynamic SOFC stack model with induced theoretically derived electrochemical model, are constructed. Then, in order to investigate the effect of operating variables of the ATR on the SOFC stack, the two models are integrated by connecting the outlet of the ATR model to the anode inlet of the SOFC stack model. Finally, using the developed dynamic model, dynamic responses of the system to step changes in the oxygen-to-carbon ratio (OCR) and steam-to-carbon ratio (SCR), which are the main operating variables of the ATR, are examined.

Keywords: SOFC based APU system, dynamic modelling and system dynamics

1. Introduction

Recently, the solid oxide fuel cell (SOFC) based auxiliary power unit (APU) system has emerged as a new alternative to provide non-propulsion power to long-haul trucks or recreational vehicles, given its high efficiency, environment friendliness, and quiet operation, to replace engine idling or diesel generator based APU. It is important to understand the dynamics of the SOFC based APU system to develop control strategies during load changes and rapid start-ups, since the system is known to exhibit frequent load changes as well as "on- offs"s given the nature of its application. Despite this, most of the studies regarding this system have focused on steady-state behavior. A few dynamic studies on the system considered a system with CH_4 or natural gas as fuel, unlike in the actual heavy duty truck applications, which use the diesel as fuel (Andersson et al. (2011), Sorrentino and Pianese (2009)).

Therefore, in this study, we develop a dynamic model that can properly simulate the dynamics of the system using diesel as the fuel, in order to address the common practical situation. In addition, instead of steam reforming which requires a heat exchanger of considerable size, auto-thermal reforming (ATR), which is better suited to the application due to its thermally independent characteristic, is selected to reform diesel to hydrogen-rich gas.

The dynamic ATR model is constructed by including a reaction kinetic model of diesel reforming, which has been validated with experimental data. On the other hand, in the case of the SOFC model, the developed model could not be validated due to a lack of experimental data. Instead, the dynamic model is developed by introducing a rigorous electrochemical model obtained from the literature. The two models are integrated into a system model by connecting the outlet of the ATR model to the anode inlet of the SOFC stack model.

This paper focuses on the propagated effect of inlet variables of the ATR on the SOFC stack. As such, using the developed dynamic model, transient responses of the system to step changes in OCR and SCR, which are the main operating variables of the ATR, are analyzed and discussed.

2. Model development

2.1. Dynamic ATR model

Heavy-hydrocarbons such as diesel have difficulty in reaching the equilibrium of the reforming reaction unlike methane and natural gas. In addition, the reformed gas may contain light-hydrocarbons such as methane due to incomplete cracking. Thus, a reaction kinetic model for the catalyst used is required for dynamic simulation of ATR using diesel. In addition, since light-hydrocarbons such as methane can be reformed in the SOFC stack and affect the performance of the stack, the kinetic model that can account for light-hydrocarbons is needed to more accurately predict the system performance. Accordingly, a dynamic model of the ATR is built by introducing the reaction kinetic model of diesel reforming, which can account for methane. The model has been validated with experiments using a nickel-based catalyst. The reaction set considered in the kinetic model can be found in Table 1. A detailed description of the kinetic model can be found in Jeong et al. (2018).

In this study, a monolith type reactor coated with the nickel-based catalyst with high mechanical superiority is considered as the reactor type of ATR given its popularity in mobile applications. Also, the dynamic ATR model is constructed with the assumption that one single channel inside the monolith can represent the entire ATR reactor. The balance equations that make up the model are shown in Table 2 below. Assumptions applied to the dynamic ATR model construction as well as a more detailed description are given in Jeong et al. (2018).

Table 1. The reaction set considered in the reaction kinetic model of the ATR using diesel fuel (C_nH_m: diesel surrogate)

Total oxidation (TOX)	$C_nH_m + \left(n + \frac{m}{4}\right)O_2 \rightarrow nCO_2 + \frac{m}{2}H_2O$
Pre-reforming (PR)	$C_nH_m + \left(n - \frac{m}{4}\right)H_2O \rightarrow \left(\frac{n}{2} - \frac{m}{8}\right)CO_2 + \left(\frac{n}{2} + \frac{m}{8}\right)CH_4$
Steam reforming (SR)	$C_nH_m + nH_2O \rightarrow nCO + \left(n + \frac{m}{2}\right)H_2$
Water-gas shift (WGS)	$CO + H_2O \leftrightarrow CO_2 + H_2$

Table 2. The balance equations of the dynamic ATR model

Mass balance for gas phase $\frac{\partial(\rho_g x_i)}{\partial t} = -\frac{\partial(ux_i)}{\partial z} + MW_i \sum_j v_{i,j} r_j$
Energy balance for gas phase $\rho_g c_{p,g} \frac{\partial T_g}{\partial t} = -u_g c_{p,g} \frac{\partial T_g}{\partial z} - \hat{a}_s u_{overall} \left(T_g - T_s\right) + \sum_j \left(-\Delta H_j\right) r_j$
Energy balance for solid phase $\rho_s c_{p,s} \frac{\partial T_s}{\partial t} = \hat{a}_s u_{overall} \left(T_g - T_s\right) - \hat{a}_{amb} h_{amb} \left(T_s - T_{amb}\right)$

2.2. Dynamic SOFC stack model

The dynamic SOFC stack model is composed of two mass balances for the fuel and air channels, four energy balances for the fuel and air channels, PEN (Positive electrode / Electrolyte / Negative electrode) structure and interconnect, and the electrochemical model, which relates the composition and temperature to the electrochemical variables. In fact, in the case of SOFC, carbon monoxide or light-hydrocarbons, such as methane, can be used as fuel as well as hydrogen due to the characteristic of high operating temperature. In this paper, however, a theoretically derived electrochemical model is applied, which is based on the assumption that only hydrogen has electrochemical activity. Nevertheless, methane and carbon monoxide can be considered to be used as fuel because methane and carbon monoxide are converted into hydrogen through methane steam reforming and water gas-shift reactions, which are assumed to take place at the fuel channel as shown in Table 3. A detailed description of the electrochemical model can be referenced in Chan et al. (2001).

Similar to the monolith type of the dynamic ATR model, the dynamic SOFC stack model is built based on the assumption that one repeating unit inside the stack can be representative of the entire stack. In order to facilitate the analysis of the dynamic response of the stack to step changes in the main operating condition of the ATR, the model is established assuming a co-flow pattern, which minimizes the dimension of space, among several possible stack flow patterns. The balance equations that make up the model are summarized in Table 4. More detailed explanations of the dynamic SOFC stack model and further assumptions can be found in Aguiar et al. (2004).

2.3. Dynamic integrated system model

(a) of Figure 1 shows the layout of the system considered in this study. As can be seen from the layout, it is assumed that the fuel reformed from the ATR directly enters the anode inlet of the SOFC stack with no composition change or heat loss between them. The two heaters in the layout are respectively introduced to heat the water entering the ATR to 100 ℃ steam and to heat the room air to the cathode of the stack up to 650 ℃, and their dynamics were not considered.

Table 3. The reaction set considered in the dynamic SOFC stack model

Methane steam reforming (MSR)	$CH_4 + H_2O \rightarrow CO + 3H_2$
Water-gas shift (WGS)	$CO + H_2O \leftrightarrow CO_2 + H_2$
Electrochemical reaction (ER)	$H_2 + 1/2\,O_2 \rightarrow H_2O$

Table 4. The balance equations of the dynamic SOFC stack model

Mass balance for fuel channel

$$\frac{\partial C_{i,f}}{\partial t} = -u_f \frac{\partial C_{i,f}}{\partial z} + \sum_j \upsilon_{i,j} r_j \frac{1}{h_f}$$

Mass balance for air channel

$$\frac{\partial C_{i,a}}{\partial t} = -u_a \frac{\partial C_{i,a}}{\partial z} + \upsilon_{i,ER} r_{ER} \frac{1}{h_a}$$

Energy balance for fuel channel

$$\rho_f c_{p,f} \frac{\partial T_f}{\partial t} = -u_f \rho_f c_{p,f} \frac{\partial T_f}{\partial z} + k_{f,P}\left(T_P - T_f\right)\frac{1}{h_f} + k_{f,1}\left(T_I - T_f\right)\frac{1}{h_f} + \sum_{j \in MSR,WGS} \left(-\Delta H\right)_j r_j \frac{1}{h_f}$$

Energy balance for air channel

$$\rho_a c_{p,a} \frac{\partial T_a}{\partial t} = -u_a \rho_a c_{p,a} \frac{\partial T_a}{\partial z} + k_{a,P}\left(T_P - T_a\right)\frac{1}{h_a} + k_{a,I}\left(T_I - T_a\right)\frac{1}{h_a}$$

Energy balance for PEN structure

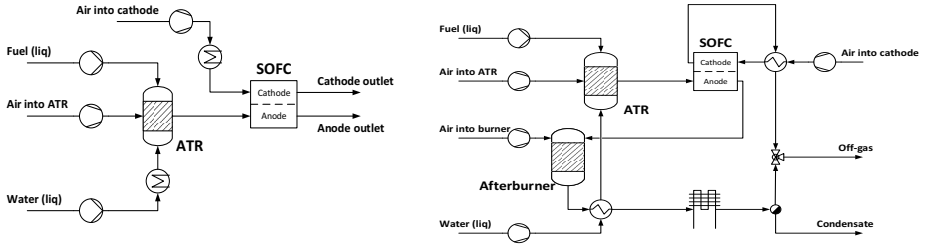
$$\rho_P c_{p,P} \frac{\partial T_P}{\partial t} = \lambda_P \frac{\partial^2 T_P}{\partial z^2} - k_{f,P}\left(T_P - T_f\right)\frac{1}{\tau_P} - k_{a,P}\left(T_P - T_a\right)\frac{1}{\tau_P} + \left[\left(-\Delta H\right)_{ER} r_{ER} - jU\right]\frac{1}{\tau_P}$$
$$+ \left[\frac{\sigma\left(T_I^{\,4} - T_P^{\,4}\right)}{1/\in_I + 1/\in_P - 1}\right]\frac{1}{\tau_P}$$

Energy balance for Interconnect

$$\rho_I c_{p,I} \frac{\partial T_I}{\partial t} = \lambda_I \frac{\partial^2 T_I}{\partial z^2} - k_{f,I}\left(T_I - T_f\right)\frac{1}{\tau_I} - k_{a,I}\left(T_I - T_a\right)\frac{1}{\tau_I} - \left[\frac{\sigma\left(T_I^{\,4} - T_P^{\,4}\right)}{1/\in_I + 1/\in_P - 1}\right]\frac{1}{\tau_I}$$

One of the important advantages of the SOFC system is that the outlet gas of the stack can be used as a high quality heat source due to its high temperature operating condition. Thus, in a practical system, such heat can be recovered by introducing an afterburner and heat exchangers, as shown in (b) of Figure 1.

However, in this study, the process simulation is carried out assuming the simplest layout shown in (a) of Figure 1, to place the priority on studying the effect of operating conditions of the ATR on the dynamic responses of the stack. This study can serve as a starting point, to investigate the dynamics of more complex and practical systems with interactions between the components of the system, including heat exchangers and recycling.



(a) The simplest system layout (b) exemplary practical system layout

Figure 1. System layouts

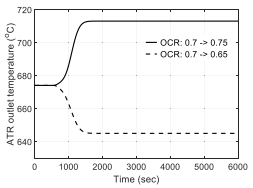
3. Simulation results and discussion

Using the developed dynamic model of the integrated system, step changes in the oxygen-to-carbon ratio (OCR) and steam-to-carbon ratio (SCR), which are the main operating variables of ATR, are introduced to see their eventual effect on the dynamic responses of the SOFC stack through the ATR. The simulation run of 6000 seconds interval starts from the steady state with OCR and SCR values of 0.7 and 2.5, respectively, and the step changes in the OCR (±0.05) and SCR (±0.3) are introduced to the system at 600 seconds into the run.

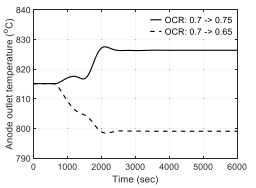
Although the dynamic responses of many variables such as compositions, flowrate, and electrochemical variables are investigated in this study, only the dynamic responses of the outlet temperatures of the ATR and the anode of the SOFC stack, which show the most remarkable behaviour, are discussed in here.

(a) and (c) of Figure 2 show the dynamics of the ATR outlet temperature in response to step changes in the OCR and SCR, respectively. A positive step change in the OCR means an increase of oxygen entering the ATR, which increases the reaction extent of the strongly exothermic total oxidation. Therefore, the ATR outlet temperature gradually increases with time as shown in the solid line of (a) and vice versa. In the case of SCR, although the response direction of the ATR's outlet temperature is reversed, it can be interpreted in similar frames. A positive step change in the SCR increases steam entering the ATR, increasing the reaction extent of the endothermic steam reforming. As a result, the ATR outlet temperature decreases as shown by the solid line of (c).

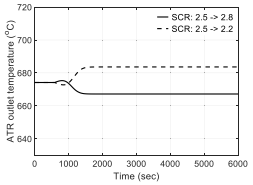
(b) of Figure 2 shows the dynamics of the anode outlet temperature, which is affected by the dynamics of the ATR due to the step change in the OCR. An oscillation is observed with time, and the influences of positive and negative step changes are asymmetric. The dynamics are results of complex interplay among various phenomena, but two factors that are most responsible for such response characteristics are the change in the anode inlet temperature and the change in the rate of methane steam reforming reaction near the anode inlet. In the case of the positive step change, which induces a more oscillatory behavior, the anode inlet temperature is exactly equal to the ATR outlet temperature, so the temperature gradually increases with time as shown by the solid line of (a). At the same time, however, this increase in the inlet temperature causes a cooling effect by



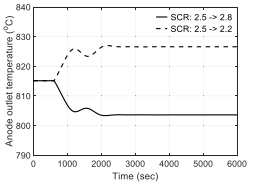
(a) ATR outlet temperature according to step change of OCR



(b) Anode outlet temperature according to step change of OCR



(c) ATR outlet temperature according to step change of SCR



(d) Anode outlet temperature according to step change of SCR

Figure 2. Dynamic responses of ATR outlet temperature and anode outlet temperature according to step changes of OCR and SCR

increasing the rate of the endothermic methane steam reforming reaction near the inlet of the anode. Eventually, the two conflicting behaviors work together to determine the temperature of the anode outlet. In the case of the negative step change, however, the above two factors do not conflict. Even though the rate of the methane steam reforming reaction is slowed down due to the decrease in the temperature of the inlet, this reaction is an endothermic reaction, so only the cooling effect is reduced. Thus, the asymmetric response characteristic shown in (b) is observed.
In (d) of Figure 2, which shows the dynamics of the anode outlet temperature to the step change of SCR, the responses to the positive and negative step changes are almost symmetric unlike the case of OCR. This result can be explained by the temperature change in the anode inlet and the corresponding voltage change. If the temperature of the anode inlet decreases due to the positive step change of SCR, the temperature inside the stack will gradually decrease and the voltage will also drop. However, such voltage drop increases the heat generation due to the loss of electrical energy, so the two effects are in conflict, and eventually, the response of the anode outlet temperature oscillates. On the other hand, in the case of the negative step change, the situation is completely reversed. Thus, the dynamic responses of the anode outlet temperature to the positive and negative step changes are symmetrical as shown in (d).

4. Conclusion

In this study, in order to investigate the dynamics of the SOFC based APU system, an integrated dynamic model of the system was developed, simply by connecting the outlet of the dynamic ATR model to the anode inlet of the dynamic SOFC stack model. The ATR model included an experimentally validated reaction kinetic model of diesel reforming, whereas the SOFC stack model was based on a theoretically derived electrochemical model. Then, by using the developed model, it was investigated how step changes in the OCR and SCR, which are the main operating variables of the ATR, affect dynamic responses of the SOFC stack through the system connection. This investigation was done using the simplest system layout. In the future, the developed dynamic model will be applied to more complex and practical system layouts like (b) of Figure 1 and control issues that can occur during such load changes or start-up will be identified and addressed.

References

D. Andersson, E. Aberg, J. Eborn, 2011, Dynamic modeling of a solid oxide fuel cell system in Modelica, Proceedings 8th Modelica Conference

M. Sorrentino, C. Pianese, 2009, Control oriented modelling of solid oxide fuel cell auxiliary power unit for transportation applications, Journal of Fuel Cell Science and Technology, 6, 041011

P. Aguiar, C.S. Adjiman, N.P. Brandon, 2004, Anode-supported intermediate temperature direct internal reforming solid oxide fuel cell. 1: model-based steady-state performance, Journal of Power Sources, 138, 120-136

S.H. Chan, K.A. Khor, Z.T. Xia, 2001, A complete polarization model of a solid oxide fuel cell and its sensitivity to the change of cell component thickness, Journal of Power Sources, 93, 130-140

S. Jeong, D. Kim, J.H. Lee, 2018, Modeling and simulation of autothermal reforming reactor of diesel over TNR catalyst in solid oxide fuel cell based auxiliaty power unit system, Proceedings of the 13th International Symposium on Process System Engineering, (submitted)

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50233-0

A multi-component mass transfer rate based model for simulation of non-equilibrium crystal growth

Yidan Shu[a], Yang Li[a], Yang Zhang[a], Jing J. Liu[b] and Xue Z. Wang[a,b*]

[a]*School of Chemistry and Chemical Engineering,South China University of Technology, Guangzhou 510640, China*
[b] *School of Chemical and Process Engineering, University of Leeds, Leeds LS2 9JT, U.K.*
xuezhongwang@scut.edu.cn

Abstract

To overcome the short-coming in the previous tools on the simulation of composition of crystals during growth, we proposed a model based on multi-component mass transfer for modelling the growth behavior of crystals. The core of the model is a phase non-equilibrium model of a single crystal that treats the solute components equally and describes them by a set of mass transfer equations. The size, shape and composition of a crystal at a given time during the crystallization process, i.e. the size in each face direction and the area of each face, depend on the amount of each material deposited on each surface. This is unlike conventional facet growth kinetic model in which the growth rate is correlated directly to supersaturation.

Such models can be applied to describe co-crystallization processes, estimate the impurity content in product and enables the prediction of stability of crystalline pharmaceuticals. The modelling method is illustrated by reference to a detailed case study of $NaNO_3$ crystal growth in water in the presence of KNO_3. The dynamic evolution of crystal composition, shape, and size is predicted in simulation runs under different initial KNO_3 concentrations.

Keywords: Crystal growth; multi-component mass transfer; non-equilibrium; crystal shape; crystal composition; co-crystallization

1. Introduction

Considering that solution crystallization is regarded as a separation process for product purification in the chemicals, pharmaceuticals and food processing industries, its modeling and simulation should surely be able to estimate the composition of the multiple components in the crystal solids in the crystals at any time during the crystallization process. The modeling for crystallization, i.e. population balance (PB) modelling (Ramkrishna and Singh 2014) of crystallization however has focused on modeling crystal size and shape distributions, rather than the composition of multiple components in crystals. It seems illogical for the most important modeling tool for crystallization processes not to be able to estimate the composition of components in the crystals.

As a matter of fact, some attention was paid to the study of crystal growth behavior in an impure media via simulation, but the focus was still on the size and shape of crystals, or more accurately, on how an impurity, often in a much smaller concentration in comparison with the material to be purified, affects the size and shape change of the

crystal. There is limited simulation work on estimation of impurity content in the crystal solids which is often based on a partition coefficient which is obtained based on the assumption of solid-liquid phase equilibrium. Some researchers (Kubota and Mullin 1995, Kubota 2001, Kubota, Sasaki et al. 2004) considered adsorption models for impurities to improve the crystal growth rate estimation. To consider the incorporation rate of impurities, Borsos et al. assumed a concentration equilibrium between the incorporating impurities and the impurities in the solution (Borsos, Majumder et al. 2016). However, the model was based on the thermodynamic equilibrium while the kinetics of mass transfer was not taken into consideration. Another observation motivated the current work is the way crystal growth rate in population balance models is described. The kinetic equation of crystal face specific growth rate is often directly correlated to supersaturation. However, when the compositions are of similar order of magnitude (for example, co-crystallization), when talking about the solubility and supersaturation which material they refer to? In the work of (Holan, Ridvan et al. 2015) about co-crystallization, solubility and supersaturation was about the mixture at a fixed composition of the major materials in the co-crystal. But the composition of the major materials in the crystals clearly changes from batch to batch, so the solubility and supersaturation for one composition does not apply to crystals of a different composition.

In this work, a multi-component mass transfer rate based model for simulation of non-equilibrium growth of crystals is introduced. In the model, the multiple solute components are described by a set of diffusion, adsorption and integration equations to form a multi-component mass transfer system. Thus both the crystal growth in terms of size and shape and change of composition of all components in the crystal can be simulated simultaneously by solving the mass transfer equations.
The remaining part of this paper is organized as follows: Section 2 gives a detailed description of the proposed multi-component mass transfer models for modeling the non-equilibrium growth of crystals. Then in Section 3 a detailed case study on $NaNO_3$ crystallization in impure media is presented to illustrate the proposed modeling approach. Finally in Section 4 conclusions will be drawn and pointers to future research will be provided. .

2. Multi-component mass transfer rate based model for single crystal growth

2.1. Stage model

The multi-component mass transfer rate based model takes inspirations from the non-equilibrium stage model for the distillation column simulation (Krishnamurthy and Taylor 1985). In our model, each “stage” stands for a discretized moment during simulation of the crystal growth. At each stage, mass transfer occurs between the solution and the crystal and leads to the growth of the crystal. The left part of Figure 1 shows the details of a stage in the model for crystal growth. The simulation of the crystal growth is discretized in time, $j = 1 \ldots J$. In total there are C components, $i = 1 \ldots C$. Stage j refers to the status of the crystal and solution at time jth moment of the simulation. At stage j, the liquid phase and the solid phase were considered separately.

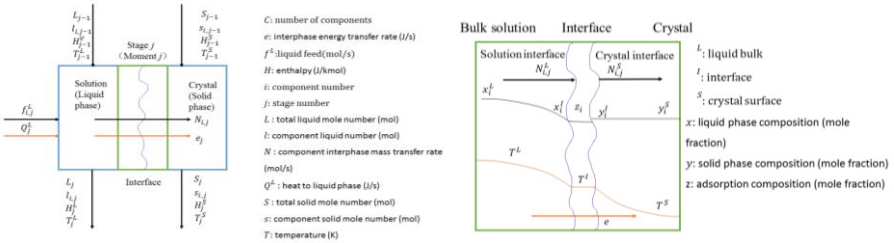


Figure 1 Schematic diagram of non-equilibrium stage model for solution crystallization: stage model (left) and interface model (right)

On stage *j,* the two phases are not at equilibrium because the process is moved to the next stage *j+1* before they reach equilibrium. There are mass transfers for each component between the liquid and solid phases in both directions. $N_{i,j}$ is the net mass transfer rate of component *i* from the liquid phase to the solid phase, for stage *j*, in mol/s. Q^L, Energy transfer between the two phases, and f^L, the liquid feed can also be modeled in the framework of our model but in the present work they are ignored. Then the stage model can be described with the following mass balance equations (*i* =1,2,…,*C; j= 2,3, ...,J; k=1,2,..F)*:

$$M_{i,j}^{L} \equiv l_{i,j-1} - l_{i,j} - \sum_{k=1}^{F} N_{i,j,k}^{L} \Delta t = 0 \tag{1}$$

$$M_{i,j,k}^{S} \equiv s_{i,j-1,k} - s_{i,j,k} + N_{i,j,k}^{S} \Delta t = 0 \tag{2}$$

$$M_{i,j,k}^{I} \equiv N_{i,j,k}^{L} \Delta t - N_{i,j,k}^{S} \Delta t = 0 \tag{3}$$

2.2. Interface model

The net mass transfer rate can be modeled with the interfacial model as the right part of Figure 1shows. In the solution, liquid mass transfer occurs due to the concentration difference between the bulk and the liquid interface. On the interface, adsorption equilibrium prevails between the liquid interfacial concentration and the surface layer concentration. In the crystal, solid mass transfer occurs in the form of molecule integration from the surface layer to the crystal surface.

It should be noted that Figure 1 only represent a "stage" on one crystal face. For different crystal faces, different model parameters should be considered. The total number of faces is *F*.

The mass transfer rate in the liquid phase and solid phase are described by equation (4) and (5) separately:

$$R_{i,j,k}^{L} \equiv N_{i,j,k}^{L} - k_{i,k}^{d} A_k (x_{i,j}^{L} - x_{i,j,k}^{I}) = 0 (i = 1,...,C-1) \tag{4}$$

$$R_{i,j,k}^{S} \equiv N_{i,j,k}^{S} - k_{i,k}^{r} A_k (z_{i,j,k} - K_{i,k}^{r} y_{i,j,k}^{I})^{r} = 0 (i = 1,...,C) \tag{5}$$

The composition in the adsorption phase is described with the Langmuir adsorption isotherms (*i* =1,2,…,*C*-1):

$$Q_{i,j,k}^{ads} \equiv z_{i,j,k} - \frac{K_{i,k} x_{i,j,k}^{I}}{1 + \sum_{m=1}^{C-1} K_{m,k} x_{m,j,k}^{I}} = 0 \tag{6}$$

The solid composition related to the mass transfer is equal to the composition of the crystal layer that formed in the last moment (i =1,2,…,C).

$$Q^{S}_{i,j,k} \equiv y^{I}_{i,j,k} - y^{S}_{i,j,k} = 0 \tag{7}$$

2.3. Summation relations

The summations components of the liquid phase, solid phase and adsorption phase should be 1. The solid phase summation equations can be obtained from the summation of $Q^{S}_{i,j,k}$ so they are not independent equations and won't be listed here (j=1,2,…,F).

$$S^{L}_{j,k} \equiv \sum_{i=1}^{C} x^{I}_{i,j,k} - 1 = 0 \tag{8}$$

$$S^{ads}_{j,k} \equiv \sum_{i=1}^{C} z_{i,j,k} - 1 = 0 \tag{9}$$

2.4. Numbers of equations and variables

There are $C+6FC$ unknown quantities for each stage j. These are the component liquid mole number ($l_{i,j}$: C in number), the solid mole number ($s_{i,j,k}$: CF), the liquid composition at the interface ($x^{I}_{i,j,k}$: CF), the solid composition at the interface ($y^{I}_{i,j,k}$: CF), the adsorption phase composition at the interface ($z_{i,j,k}$: CF), the mass transfer rate ($N^{L}_{i,j,k}$: CF and $N^{S}_{i,j,k}$: CF). It should be noted that $x^{L}_{i,j}$ and $y^{S}_{i,j,k}$ are not independent because they can be obtained from $l_{i,j}$ and $s_{i,j,k}$ separately. There are $C+6CF$ independent equations which are listed above (Equations (1) to (9)). Therefore, the equations can be solved.

2.5. Simulation of crystal shape and composition

Neither the face growth rate nor the normal distance straightforward appears in our model. However, based on the calculation of solid mole number of all components, the normal distance change can be obtained by summing up the solid mole number of each component divided by its density and molecular mass. On the other hand, based on the solid mass of each component, both the overall composition of the whole crystal or on a specific face normal direction, and the composition distribution along a specific face normal direction, can be obtained.

3. Case study

Sodium nitrate (NaNO3) crystallization is selected for the case study of our proposed model. Fig. 3 is the equilibrium morphology of a NaNO3 crystal which was predicted by Benages-Vilau (Benages-Vilau 2013) and also observed by Wu et al. (Wu, Ma et al. 2016) A NaNO3 crystal is rhombohedral with six faces. These faces can be classified into three groups of crystallographically equivalent faces, i.e. two {104}, two $\{1\bar{1}4\}$ and two $\{01\bar{4}\}$. Wu et al. also measured the crystal face specific growth kinetics of all the three faces.

In the case study, we will use our proposed model to simulate the shape and purity evolution of a single $NaNO_3$ with or without the presence of KNO_3 as the impurity. A pure $NaNO_3$ crystal is loaded in the crystallizer as a crystal seed and the crystallization temperature is kept constant. According to Benages-Vilau's work, KNO_3 is assumed to be adsorbed on {104} face only.

The mass transfer coefficients from the solution to the crystal surfaces can be predicted using a correlation given by Geankoplis(Geankoplis 1978). The adsorption constant $K_{i,k}$ and the equilibrium ratio $K^r_{i,k}$ are related with the solubility data of $NaNO_3$(Xu and Pruess 2001) and KNO_3(Rolfs, Lacmann et al. 1997). The other model parameters corresponding to $NaNO_3$ are regressed with the experimental data of Wu et al 19 while the other model parameters corresponding toKNO_3 are assumed.

Initially a pure $NaNO_3$ crystal is loaded in the crystallizer as a crystal seed and the crystallization temperature is kept constant. KNO_3 is assumed to be adsorbed on {104} face only. The crystallization temperature is kept 218.15K. The initial the molar fractions of NaNO3 is 0.1689. The initial face normal distances of {104}, $\{\bar{1}14\}$ and $\{0\bar{1}4\}$ are 100.06μm, 216.83μm and 253.28μm respectively. KNO_3 is added with $NaNO_3/KNO_3$ molar ratio of 10^{-3} and 10^{-5}. The normal distance evolution of face {104} and shape evolution during crystallization are shown in Figure 2. Because of the parameter setting of KNO_3 on $\{\bar{1}14\}$ and $\{0\bar{1}4\}$, the normal distance evolutions of faces $\{\bar{1}14\}$ and $\{0\bar{1}4\}$ show no differences from simulation in the absence of KNO_3.

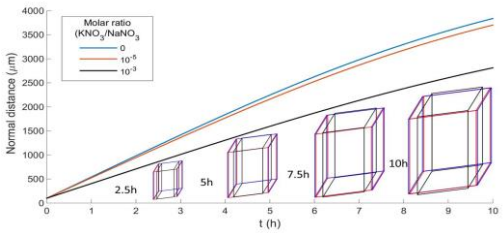


Figure 2 Evolution normal distance of face {104} and crystal shape during NaNO3 crystal growth with different KNO3/NaNO3 molar ratio

The overall molar fraction of KNO_3 of the growing $NaNO_3$ can also be predicted with the proposed model as is shown in Figure 3(a). On the other hand, due to the change of solution phenomenon, the composition of crystal in the normal direction of {104} is different at different locations and the distribution of KNO_3 molar fraction in the normal direction of {104} can also be simulated with the proposed model, as is shown in Figure 3(b). The left of the distribution curves is missing because it represents the initial KNO_3 molar fraction of 0 and cannot be shown in logarithmic coordinate.

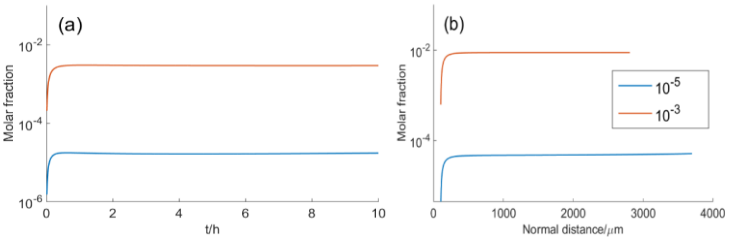


Figure 3 Crystal with different KNO3/NaNO3 molar ratio. (a) Evolution of KNO3 molar fraction in the whole crystal (b) KNO3 molar fraction distribution in the normal direction of face {104}

4. Conclusion

In this work, a multi-component mass transfer model is proposed for modelling the crystal growth of multi-component systems. In the model, all the solutes are equally

described by a set of diffusion, adsorption and integration equations. Through solving the equations, evolution of both the crystal shape and composition can be predicted.
A case study of NaNO3 crystal growth in the presence of KNO3 is conducted. The shape and composition evolutions of NaNO3 crystals of crystallization with different initial KNO3 concentrations are modelled.

The main contribution of this work is the proposal of a novel model for the prediction of composition evolution of multi-component crystallization processes. The model can be expected to work on other multi-component crystallization processes like co-crystallization. In future work, the proposed model will be validated with multi-component crystallization experiments.

References

R. Benages-Vilau, 2013, Growth, Morphology and Solid State Miscibility of Alkali Nitrates. PhD, Universitat de Barcelona.

A. Borsos, A. Majumder and Z.K. Nagy, 2016, Multi-Impurity Adsorption Model for Modeling Crystal Purity and Shape Evolution during Crystallization Processes in Impure Media, Crystal Growth & Design 16(2), 555-568.

C.J. Geankoplis, 1978, Transport processes and unit operations, Allyn & Bacon.

J. Holan, L. Ridvan, P. Billot and F. Stepanek, 2015, Design of co-crystallization processes with regard to particle size distribution, Chemical Engineering Science 128, 36-43.

R. Krishnamurthy and R. Taylor, 1985, A nonequilibrium stage model of multicomponent separation processes. Part I: Model description and method of solution, 31(3), 449- 456.

N. Kubota, 2001, Effect of impurities on the growth kinetics of crystals, Crystal Research and Technology 36(8-10), 749-769.

N. Kubota and J.W. Mullin, 1995, A kinetic model for crystal growth from aqueous solution in the presence of impurity, Journal of Crystal Growth 152(3), 203-208.

N. Kubota, S. Sasaki, N. Doki, N. Minamikawa and M. Yokota, 2004, Adsorption of an Al (III) impurity onto the (100) face of a growing KDP crystal in supersaturated solution, Crystal Growth & Design 4(3), 533-537.

D. Ramkrishna and M.R. Singh, 2014, Population Balance Modeling: Current Status and Future Prospects. Annual Review of Chemical and Biomolecular Engineering, Vol 5. J. M. Prausnitz, M. F. Doherty and R. A. Segalman. Palo Alto, Annual Reviews. 5, 123-146.

J. Rolfs, R. Lacmann and S. Kipp, 1997, Crystallization of potassium nitrate (KNO3) in aqueous solution .1. Growth kinetics of the pure system, Journal of Crystal Growth 171(1-2), 174-182.

K. Wu, C.Y. Ma, J.J. Liu, Y. Zhang and X.Z. Wang, 2016, Measurement of Crystal Face Specific Growth Kinetics, Crystal Growth & Design 16(9), 4855-4868.

T. Xu and K. Pruess, 2001, Thermophysical properties of sodium nitrate and sodium chloride solutions and their effects on fluid flow in unsaturated media, Office of Scientific & Technical Information Technical Reports.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50234-2

Two-phase Dynamic Model for PEM Electrolyzer

Mariana Corengia[a], Ana I. Torres[a*]

[a]*Instituto de Ingeniería Química, Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay*
aitorres@fing.edu.uy

Abstract

This work develops a dynamic model for hydrogen production in Proton exchange membrane electrolyzer cells (PEMECs). The novelty of the model lies in considering gaseous and liquid phases for the species in both anode and cathode, and linking both phases using rigorous mass transfer equations. With these considerations the model properly accounts for the operating conditions usually encountered in PEMECs. The model thus derived defines an ODE system whose steady state solution is used as an input to compute the current-voltage operation curve of the cell. Validation of the model is performed comparing the steady state solution against experimental operation curves. Simulations show that for a constant current supply the anode reaches steady state faster and in a smother way, than the cathode.

Keywords: Energy storage, PEMEC dynamics, hydrogen production.

1. Introduction

Modern renewable energy sources such as wind or sunlight cannot be programmed to be available when needed, a feature that represents a barrier for their widespread usage. Conversion to chemicals is a promising alternative for storage of excess electrical energy. In particular, hydrogen production has been considered as an energy storage vector, either for reconversion to electricity (Gabrielli et al., 2017) or for conventional fuels substitution (Pereira and Samsatli, 2017).

Proton exchange membrane electrolyzer cells (PEMECs) are devices that produce hydrogen from water; hydrogen is produced in the cathode, and separated from the oxygen producer anode by a proton conductive polymeric membrane. PEMECs are regarded as the replacement of the current low temperature hydrogen production technology which is based on alkaline electrolyzers.

During operation, PEMECs will be integrated into the electrical grid, possibly as one of many technologies for renewable energy use and storage. Therefore, their operation will be inherently intermittent and mathematical models that capture their dynamics are required even at the design stage.

Dynamic models for PEMECs are already available in the literature. For instance, (Görgün, 2006) presented one in which the electrolyzer is divided in 4 subsystems: anode, cathode, membrane and an auxiliary component that models the relation between voltage and electric current. Although this model has been used by many later contributions, its development stems from models originally derived for proton exchange membrane fuel cells (PEMFCs), which received much more attention during the last decade (Carmo et al., 2013). Therefore, some of the assumptions included in current PEMEC models are inherited from the ones originally valid for PEMFCs, but do not represent the operational

conditions in PEMECs. In this direction, a recent review (Olivier et al., 2017) points out the need for alternative dynamic models.

In this work, we take the modular division proposed by (Görgün, 2006) as the basis to develop a dynamic two-phase (gas-liquid) model for both cathode and anode. This dynamic model allows for water to accumulate as a liquid in both cathode and anode. In each electrode, rigorous mass transfer equations are considered to link the concentration of the species among the two phases. A CSTR-like behaviour and incompressible biphasic flow is assumed to compute the outlet flow of the species for each electrode. The accuracy of the incompressibility simplification is checked by calculating the Mach number corresponding to each electrode outlet.

2. Model description

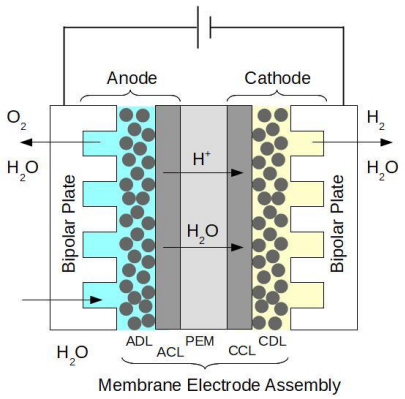


Figure 1. PEMEC diagram

Figure 1 shows a typical construction of a PEMEC cell. PEMECs consist of conductive solid bipolar plates that enclose the contents of the anode, the cathode and the membrane. The membrane is a polymer electrolyte that allows for selective diffusion of protons and water. The anode (in blue in the picture) is the section between the bipolar plate and the PEM itself and consists of a gas/liquid chamber, a diffusion layer (ADL) and a catalyst layer (ACL). The diffusion layer is a porous layer that allows both electrical conductivity and diffusion of the reactants. Electrochemical reactions take place in the catalytic layer, which has to be placed next to the polymeric membrane so that protons can migrate. The cathode (in beige in the picture) is the section enclosed by the bipolar plate and the other side of the PEM and as the anode is divided in three sub-sections: CCL, CDL and gas-liquid chamber, that have similar characteristics to the anode´s. An external power source connected to the circuit provides the required energy.

In the derivations that follow, the volume of each electrode (anode, cathode) is assumed to be constant. In contrast to previous dynamic models, we consider that two separate phases may be present in each electrode. The volume of the liquid phase (V_L) is calculated from the mass of water in the electrodes and assuming the electrolyzer temperature for calculation of density. The volume of the gas phase is calculated substracting (V_L) from the total volume of the electrode.

2.1. Mathematical model for the anode

Water is continuously fed into the anode (flowrate H_2O_{in}) and oxidizes to form gaseous oxygen and protons according to the electrochemical reaction: $2H_2O_{(l)} \rightarrow O_{2(g)} + 4H^+_{(ac)} + 4e^-$. Water consumed by this reaction is regarded as H_2O_{con}. Protons generate and immediately diffuse through the membrane, then do not accumulate in the anode. However, when diffusing they drag an amount of water H_2O_m through the membrane (electro-osmotic drag, see Section 2.3). O_2 generates at a rate $O_{2,gen}$ as dry bubbles within the aqueous phase, hence an amount of water vaporizes and transfers to these bubbles with rate $k_{MT}(p_{sat} - p_{H_2O,a})$.

Flow out of the anode occurs when the total pressure in the anode P_a exceeds the back-pressure (P_{a0}). This total pressure is computed as the sum of the partial pressure of each gas component (O_2, H_2O_{ag}): $P_a = p_{O_2} + p_{H_2O,a}$. When $P_a > P_{a0}$, a valve opens and both gas phase species as well as liquid water exit the anode. The exit flowrate of species X is represented as $X_{a,out}$. Considering the discussion above, the following three mass balances represent the operation of the anode:

$$\frac{dN_{O_2}}{dt} = O_{2,gen} - O_{2,a,out} \tag{1}$$

$$\frac{dN_{H_2O_{ag}}}{dt} = k_{MT}\left(p_{sat} - p_{H_2O,a}\right) - H_2O_{g,a,out} \tag{2}$$

$$\frac{dN_{H_2O_{al}}}{dt} = H_2O_{in} - H_2O_{con} - k_{MT}\left(p_{sat} - p_{H_2O,a}\right) - H_2O_{la,out} - H_2O_m \tag{3}$$

In here, partial pressure of each gas component (p_{O_2}, $p_{H_2O,a}$) is calculated by the ideal gas law; $O_{2,gen}$ and H_2O_{con}, are expressed as a function of the electric current in the electrolyzer (i_{el}) and the number of cells (n_{el}) by Faraday's law: $2O_{2\,gen} = H_2O_{con} = i_{el}n_{el}/2F$. Outlet flowrates for each species ($O_{2,a,out}$, $H_2O_{g,a,out}$ and $H_2O_{l,a,out}$) are calculated from the total exit flow (EF_a^*) assuming the outlet flowrates and the electrode have the same species distribution (CSTR-like behaviour). EF_a^* depends on the pressure difference between the pressure in the anode and the pressure outside and is computed with Bernoulli's equation:

$$EF_a^* = A_{a,out}\sqrt{2(P_a - P_{a0})/\rho_a} \tag{4}$$

In Eq. 4, $A_{a,out}$ is the area of the exit orifice and ρ_a is the average density in the anode. A discharge coefficient (C_D), is included to account for pressure losses and calculate a more realistic exit flow (EF_a): $EF_a = C_D EF_a^*$. It should be noted that Eq. 4 is valid only if there is not a significant variation of the density of the biphasic flow (i.e. incompressible flow). This assumption is valid if the Mach number is under 0.3, a limit that is checked in all simulations.

Finally, temperature of the electrolyser (T_{el}) is assumed to be constant, as it is common practice to feed an excess of water to the make this happen (Medina and Santarelli, 2010), hence energy balances are not considered.

2.2. Mathematical model for the cathode

There are three species present in the cathode: H_2, H_2O (both in liquid and vapor form) and N_2. Gaseous hydrogen is produced by the electrochemical reaction: $2H^+_{(ac)} + 2e^- \rightarrow H_{2(g)}$ at a rate given by Faraday´s law $H_{2gen} = i_{el}n_{el}/2F$; liquid water, is transported through the membrane at a rate H_2O_m as a result of the electro-osmotic drag; as in the anode part of the liquid water vaporizes at a rate given by $k_{MT}\left(p_{sat} - p_{H_2O,c}\right)$; N_2 is, for safety reasons, used to purge the cathode when the cell is not in operation, but it is not continuously fed. The four mass balances that represent the operation in the cathode are:

$$\frac{dN_{H_2}}{dt} = H_{2,gen} - H_{2c,out} \tag{5}$$

$$\frac{dN_{H_2O_{cg}}}{dt} = k_{MT}\left(p_{sat} - p_{H_2O,a}\right) - H_2O_{gc,out} \tag{6}$$

$$\frac{dN_{N_2}}{dt} = -N_{2c,out} \tag{7}$$

$$\frac{dN_{H_2O_{cl}}}{dt} = H_2O_m - k_{MT}\left(p_{sat} - p_{H_2O,a}\right) - H_2O_{lc,out} \tag{8}$$

As in the anode, partial pressure of every gas component is calculated by the ideal gas law, and total pressure in the cathode (P_c) as their sum: $P_c = p_{H_2} + p_{N_2} + p_{H_2O,cg}$. When P_c exceeds the back-pressure of the cathode, there is an exit flow (EF_c^*), calculated with an equation like Eq. 4. Same considerations are taken in both electrodes to calculate the exit rate of each component in each phase, and the validity of Eq. 4.

2.3. Mathematical model for the membrane

As in (Görgün, 2006) accumulation of water species in the membrane is neglected, then the flux of water from the anode to the membrane is identical to the flux of water from the membrane to the cathode and is referenced as H_2O_m. As we are considering that both electrodes operate at the same pressure, only two water-transport mechanisms are considered: electro-osmotic drag and diffusion. In short H_2O_m is computed as:

$$H_2O_m = \frac{n_d P M_{H_2O} i_{el}}{F} + \frac{D_W \rho_{m,d} P M_{H_2O} A(\lambda_a - \lambda_c)}{t_m M_{md}} \tag{9}$$

Electro-osmotic drag (first term) depends on the drag coefficient n_d (average number of water moles "dragged" through the membrane with every H^+ mole) and the current i_{el} applied to the electrolyzer. Diffusion (second term) is computed with Fick´s law and depends on the water content in the membrane (λ) which is estimated from the water activity at each side using the equations in (Görgün, 2006). Other parameters in this term include: water diffusion coefficient D_W, electrode area A, membrane water-free density $\rho_{m,d}$, membrane thickness t_m and the molar mass of the membrane compound M_{md}, all obtainable from vendor data or common correlations.

2.4. Current - Voltage Correlations

Once mass balances are solved, the operating voltage for each cell (V_{el}) is computed by summation of the open circuit potential (E) and the activation and ohmic overpotentials,

$$(\eta_{act}, \eta_{ohm}): V_{el} = (E + \eta_{act} + \eta_{ohm}) \tag{10}$$

Nernst equation is used to calculate the open circuit potential $E = E_0 + RT_{el}/(2F) * log\left(p_{H_2} {p_{O_2}}^{1/2}\right)$. Water activity in the anode is omitted from this equation as it can be assumed to be close to one. Net activation overpotential is the sum of the contribution of each electrode and calculated from Tafel's equation: $\eta_{act} = RT_{el}/(\alpha F) * log(j_{el}/j_0)$. The current density $j_{el} = i_{el}/A$ is the same for both electrodes; the charge transfer coefficient α is estimated as 0.5 for the cathodic reaction and 2 for the anodic one (Marangio et. al., 2009). j_0 , the exchange current density, is a characteristic of the catalyst employed in the catalyst layer, and also depends on the amount used. Ionic transport trough the membrane is the main contributor to ohmic overpotential: $\eta_{ohm} = t_m i_{el}/(A\sigma_m)$. In here σ_m is the membrane proton conductivity, which is obtained from correlations (Marangio et. al., 2009).

3. Numerical Results

The model in Eqs. 1-10 was simulated in Matlab© considering a single cell electrolyzer operating at a temperature of 80 °C and a backpressure of 101 kPa in each electrode.

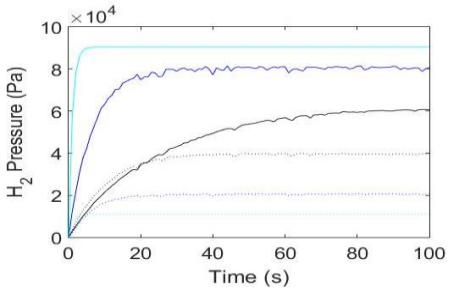


Figure 2. Partial pressures in the cathode. H_2 solid line. H_2O dotted line. $j_{el} = [0.02,0.1,1]$ Acm^{-2} (darker to lighter).

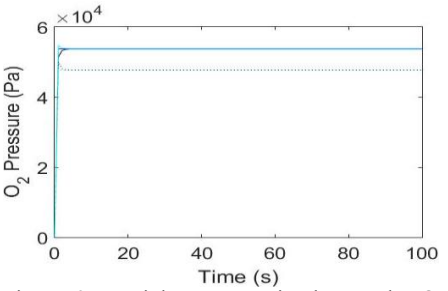


Figure 3. Partial pressures in the anode. O_2 solid line. H_2O dotted line. $j_{el} = [0.02,0.1,1]$ Acm^{-2} (darker to lighter).

According to available information for small commercial PEMECs, the area of the exit orifice was set to be 1% of the area of the electrodes. Physical and construction properties for the membrane were taken from Dupont© Nafion© 115 series. Mass transfer coefficients k_{MT}, were estimated from correlations in (Cussler, 2008). For contact areas, bubbles with a mean size diameter of 50μm were assumed in the anode, whereas in the cathode we assumed an area equal to the area of the membrane, representative of a water film covering the cathode diffusion layer. The assumed incompressibility of the flow was checked in all simulations; confirming a maximum Mach number below 0.01.

Figures 2 and 3 show the partial pressures in the cathode and anode respectively for three operation current densities: 0.02, 0.1 and 1 Acm^{-2}; which were selected following the usual ranges reported in the literature (Yigit and Salomet, 2016). It can be seen that while in the anode partial pressures build up fast and reach a current-independent steady state, the cathode shows a slower response, whose steady state also depends on the current applied to the cell. As the current increases, H_2 production rate increases proportionally, hence, N_2 leaves the chamber sooner, and sooner the pressure inside the chamber is only due to the presence of just H_2 and $H_2O(g)$. In addition, when j_{el} increases, as the characteristic time for water transfer to the gas phase is independent of it, pressure builds up having a larger contribution of H_2 with respect to $H_2O(g)$. Therefore H_2 exiting the cathode contains a lower amount of water in it.

Figure 4 shows the fraction of each electrode occupied by the liquid phase. Notice that although the anode reaches a steady state pressure independent of the current density, the number of moles in the gas phase are not independent of it. This happens because the model accounts for the possibility of changes in the volume of each phase. In particular in the anode, as the current increases, the fraction of liquid largely decreases, from 40% at $j_{el} = 0.1$ Acm^{-2} to 6% at $j_{el} = 1$ Acm^{-2}.

Figure 5 shows the operation curve for the cell $(V_{el}\text{-}i_{el})$ and the overpotentials as described in Section 2.4, for a current density in the range [0.02:0.02:1]. Each point then represents a steady-state solution of the system such as the one shown in Figs. 2 and 3. For Fig. 5 an exchange current density j_0=1E-7 Acm^{-2} was selected for the anode and 1E-3 Acm^{-2} for the cathode. The plot shows that the steady state reached with the proposed model is in agreement with the operation curves and overpotential values reported in literature, and obtained from experimental measurements (Awasthi et al, 2011).

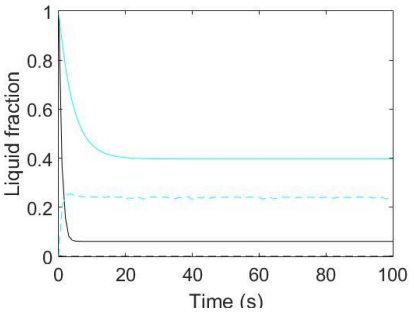


Figure 4. Liquid fraction of the electrodes: anode solid line; cathode (*10E4) dashed line. j_{el} =0.1 Acm^{-2} (lighter) j_{el} =1 Acm^{-2} (dark)

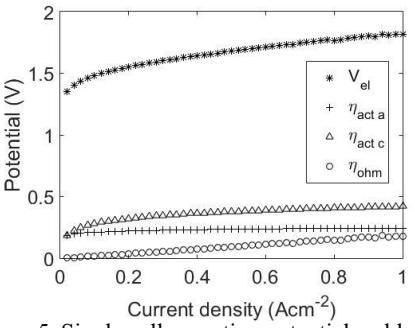


Figure 5. Single cell operating potential and losses. *: Operation voltage, +: anodic activation overpotential, Δ: cathodic activation overpotential, o: ohmic overpotential.

4. Conclusions

In here, a two-phase dynamic model for PEMEC was developed. The model results in steady states whose operational curves are in good agreement with previously reported experimental data and simulations. The dynamics of the individual electrodes show that steady state is not instantaneous but can be reached in the order of minutes, with the cathode being the slower electrode. Water mass transfer limitations in the cathode play a key role in this. In any case, these minutes-scale dynamics reasonably allow for matching with intermittent renewable energy sources.

References

A. Awasthi, K. Scott, S. Basu, 2011, Dynamic modeling and simulation of a proton exchange membrane electrolyzer for hydrogen production. International Journal of Hydrogen Energy 36 (22), 14779 – 14786.

M. Carmo, D. L. Fritz, J. Mergel, D. Stolten, 2013, A comprehensive review on PEM water electrolysis. International Journal of Hydrogen Energy 38 (12), 4901 – 4934.

E. L. Cussler, 2009, Diffusion: Mass Transfer in Fluid Systems, Cambridge University Press.

P. Gabrielli, M. Gazzani, E. Martelli, M. Mazzotti, 2017, A MILP model for the design of multi-energy systems with long-term energy storage. Computer Aided Chemical Engineering 40, 2437-2442.

H. Görgün, 2006, Dynamic modelling of a proton exchange membrane (PEM) electrolyzer. International Journal of Hydrogen Energy 31 (1), 29 – 38.

F. Marangio, M. Santarelli, M. Cal, 2009, Theoretical model and experimental analysis of a high pressure PEM water electrolyser for hydrogen production. International Journal of Hydrogen Energy 34 (3), 1143 – 1158.

P. Medina, M. Santarelli, 2010, Analysis of water transport in a high pressure PEM electrolyzer. International Journal of Hydrogen Energy 35 (11), 5173 – 5186.

P. Olivier, C. Bourasseau, P. B. Bouamama, 2017, Low-temperature electrolysis system modelling: A review. Renewable and Sustainable Energy Reviews 78 (Supplement C), 280 – 300.

A. Pereira, S. Samsatli, 2017, Optimal design and operation of heat networks utilising hydrogen as an energy carrier. Computer Aided Chemical Engineering 40, 2527-2532.

T. Yigit, O. F. Selamet, 2016, Mathematical modeling and dynamic simulink simulation of high-pressure PEM electrolyzer system. International Journal of Hydrogen Energy 41 (32), 13901 – 13914.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50235-4

Study on the Distillation Sequence with Dividing Wall Column for Five-component Separations

Yunlu Zhang[a], Jingde Wang[a], Xiuyu Liu[b],Min Zhang[c], Milan Zhang[c], Wei Sun[a*]

[a]*Beijing Key Lab of Membrane Science and Technology, College of Chemical Engineering, Beijing University of Chemical Technology, 100029 Beijing, China*
[b]*Key Laboratory for Green Chemical Technology of Ministry of Education, R&D Center for Petrochemical Technology, Tianjin University, 300072 Tianjin, China*
[c]*Advanced Manufacturing Technology Center China Academy of Machinery Science & Technology, 100044 Beijing, China*
sunwei@mail.buct.edu.cn

Abstract

Distillation is one of the primary separation technique widely adopted in process industry. For separating a multicomponent mixture by distillation, a sequence of columns will be required. In this work, distillation sequences with both simple column and dividing wall column (DWC) are contained under the sharp split assumption for five-component separations. A two-level optimization is conducted with the objective of minimum total annual cost. The optimal distillation sequence with DWC can be efficiently identified based on dynamic programming method. In this work, alcohol system is studied to compare with a hydrocarbon system. Based on the optimal distillation sequence obtained, the impact of intermolecular forces is discussed, which could be instructive to reduce optimization computation load of distillation sequence and narrow the searching scope of the optimal position for multicomponent separation DWC. The optimal sequences with DWC included at the last step in five-component hydrocarbon system with various mixture contents are also investigated.

Keywords: DWC, distillation sequence, position

1. Introduction

Distillation is a well-known separation technique. For separating multicomponent mixtures, distillation operations are conducted sequentially. The energy-saving has always been the focus of researchers. Dividing wall column (DWC) proposed by Wright (1949) is an atypical distillation column with an internal partition wall for separating multicomponent mixture into three output streams. It is reported that about 30 % cost can be saved by the application of DWC (Triantafyllou and Smith, 1992). When including DWC, the number of columns in each sequence can be reduced, while the number of feasible distillation sequences will increase rapidly (Zhang et al., 2017), which makes it computationally challenging for finding the optimal sequence. The number of all possible sequences is obtained by simple enumeration of any possible subproblems, or columns for adjacent components, without considering the inherent characteristics of separation problem itself. The aim of this work is to find the possibility to remove some enumerated sequences by considering related physical-chemical property of materials in the separation system.

In this work, five-component separations in alcohol system and hydrocarbon system have been studied as examples. Based on the simulation and optimization results of all

subproblems achieved by Aspen Plus software, the correspondingly optimal distillation sequences are found by dynamic programming method. Then intermolecular forces are discussed in terms of their impact on energy consumption, as well as on the optimal separation sequence.

2. Methodology

2.1. Column level optimization

The column models are based on the equilibrium stage concept (Mueller and Kenig, 2007). Short-cut method implemented in Aspen Plus is used to get design parameters first. The parameters obtained were used to initialize the stage-to-stage rigorous method RadFrac for simple column. The simulation of DWC is achieved by functionally equivalent configuration Petlyuk (Petlyuk et al., 1965). Total annual cost is estimated based on optimization results obtained. To evaluate the operating cost, which contains recycling water cost and low pressure steam cost, 8000 hour operation per year is assumed. According to the diameter, height, temperature, pressure and other data of column, equipment costs can be indirectly calculated based on engineering calculation software SW6. The investment payback period is assumed as 6 years. The cost difference of dynamic control between simple column and DWC has not been considered in this work.

2.2. Distillation sequence level optimization

Separation sequences are different combinations of subproblems realized by specific columns, which have been optimized in previous section. The optimal sequence of separation system in this research is obtained through multi-stage decision-making by the dynamic programming method proposed by American mathematician Bellman in 1957, i.e., in such a problem, a sequence for a subproblem has to be optimized if it exists in the optimal sequence for the whole problem.

The objective function of multi-stage decision defined by Howard (1966) can be written as follow:

$$\Phi = min \sum_{i=0}^{N-1} C(X_k, U_k) \tag{1}$$

where X_k refers to the end state of k stage decision or the start state of k+1 stage decision; U_k represents the control or decision of k+1 stage; C represents the cost function of k+1 stage, which is the function of X_k and U_k.

Let the $V_j(X_i)$ refers to the minimum value of the objective function since the X_i state decision transfer to the end state. Recursive formula based on dynamic programming method can be shown as follow ($V_0(X_N) = 0$):

$$V_{N-k}(X_k) = \min_{U_k \in U} \{C(X_k, U_k) + V_{N-(k+1)}(X_{k+1})\} \quad (\mathrm{k} = 0,1,\ldots,\mathrm{N}-1) \tag{2}$$

2.3. Intermolecular force calculation

Considering the process of distillation, Wayner mentioned in 1999 that part of the energy consumption comes from overcoming intermolecular forces in micro perspective. Intermolecular force mainly includes: the interactions between ions, charged groups, dipoles and induced dipoles; hydrogen bonds; the interactions of hydrophobic groups and non-bond electronic repulsion (Duan and Zhou, 1989). By using Materials Studio

software, the intermolecular forces can be calculated and analyzed to show its impact on the final distillation sequence.

3. Case study

3.1. Cases

Hydrocarbon and alcohol are common separation systems in chemical industry and also typical systems for the DWC relevant studies. Feed fractions for all three cases considered in this work are given in Table.1. Product specifications are 0.990 (mol frac).

In most industrial process, pure product is more preferred. If a DWC is used to separate

Table.1 Feed Conditions

Cases	**Feed Fraction (mol frac)**				
Case 1	isobutane	pentane	hexane	heptane	octane
	0.200	0.200	0.200	0.200	0.200
	0.177	0.235	0.177	0.235	0.177
Case 2	hexane	heptane	octane	2-methyloctane	n-nonane
	0.200	0.200	0.200	0.200	0.200
Case 3	methanol	n-propanol	n-butanol	n-pentanol	n-hexanol
	0.200	0.200	0.200	0.200	0.200

three components into three pure products, it will appear at the last separation step in a sequence. In Case 1, light hydrocarbon separation system is studied with two sets of mixture contents. The number of distillation subproblems U and distillation sequences S_R can be calculated by Eq.(3) and Eq.(4), which are 23 and 20 respectively according to our previous work (Zhang et al., 2017).

$$U = \frac{R(R-1)(R+1)}{6} + (R-2) \tag{3}$$

$$S_R = \sum_{m=0}^{\frac{R}{2}} \frac{binomial(R-2m+1,m)\cdot binomial(2R-4m,R-2m)}{R+1-2m} \tag{4}$$

where R is the number of components. (R=1,2,3,…,n)

Table.2 Boiling points (BP) of five components in Case 2 and Case 3

	Component1	**Component2**	**Component3**	**Component 4**	**Component5**
Case 2	hexane	heptane	octane	2-methyloctane	n-nonane
BP (℃)	68.7	98.3	125.7	143.4	150.8
Case 3	methanol	n-propanol	n-butanol	n-pentanol	n-hexanol
BP (℃)	64.7	97.2	117.7	137.8	157

In practical separation process, the use of a dividing wall column in multicomponent separation is also a choice. In case 2 and case 3, DWC is not limited to the last step in the distillation sequence. It can be seen in Table.2 that boiling points of corresponding components are quite close in both case 2 and case 3.

The number of distillation subproblems U and distillation sequences S_R in case 2 and case 3 can be calculated by Eq.(5) and Eq.(6), which are 35 and 38 respectively according to our previous work (Zhang et al., 2017).

$$U = \frac{(R-1)R(R+1)(R+2)}{24} \tag{5}$$

$$S_{Q+1} = \sum_{k=0}^{\lfloor \frac{Q}{2} \rfloor} \frac{C_{2Q-k}^{Q+k} C_{Q+k}^{k}}{Q+1} \ (R = Q+1) \tag{6}$$

Peng-Robinson thermo model is used to calculate the vapor-liquid equilibrium for all cases. The optimization of single column and the distillation sequence is achieved as described in section 2.1 and 2.2. Number in optimal sequence represents carbon number.

3.2. The optimal sequence of Case 1

Through four-stage decision-making, each node in the dynamic programming tree has an optimal choice. Then the optimal sequence (4,5,6,7,8)→4/5678→5/678→6/7/8 is achieved from 20 possible sequences with two kinds of mixture contents. The optimal sequence with simple column only is (4,5,6,7,8)→4/5678→567/8→56/7 →5/6. The specific steps are shown in Figure 1. "*C*" refers to the cost function *C* in formula (1).

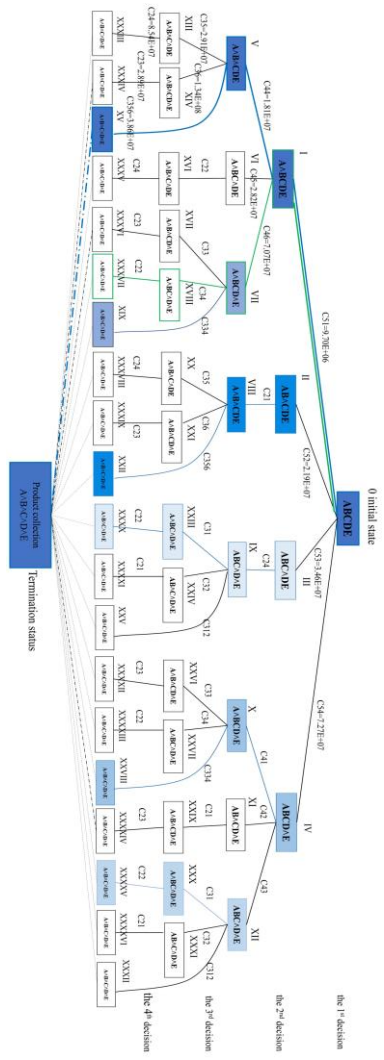


Fig. 1 Dynamic programming sequence tree for Case 1

3.3. The optimal sequence for Case 2

The optimal distillation sequence with DWC in new hydrocarbon system is (6,7,8,9a,9b) → 67/8/9a9b → 6/7+9a/9b. Besides, the best sequences with simple column only is (6,7,8,9a,9b) → 67/89a9b → 6/7+89a/9b → 8/9a, in which 9a and 9b represent 2-methyloctane and n-nonane.

3.4. The optimal sequence for Case 3

(1,3,4,5,6)→1/3/456→4/5/6 as the optimal distillation sequence with DWC in alcohol system, achieved from 38 possible sequences. The best sequences with simple column only is (1,3,4,5,6)→13/456→1/3+4/56→5/6.

Table.3 Cost comparisons

Optimal sequence	Hydrocarbon system(Case 2)		Alcohol system(Case 3)	
	Operating cost	Total cost	Operating cost	Total cost
DWC+Simple column	9.75E+07	9.91E+07	4.69E+07	4.74E+07
Simple column	1.14E+08	1.15E+08	6.89E+07	6.92E+07
Cost saving	14.3%	13.8%	31.9%	31.5%

The operating cost and total annual cost comparison between the two optimal distillation sequences are given in Table.3. The energy consumption of the sequence with dividing wall column is less that without it in both cases.

4. Results and discussion

4.1. The impact of intermolecular force on the optimal separation sequence.

It can be seen that DWC is adopted at the first step of optimal sequence in both case 2 and case 3, but the output components of DWC in the optimized sequences are different which should be determined by the inherent characteristic of each system. As mentioned above, the difference of boiling points between adjacent components are quite similar in case 2, case 3. Among all the factors, intermolecular forces cannot be neglected, which is considered as one of energy barrier for the separation of different components. By using Materials Studio software, the intermolecular forces can be simulated for single component, two to four adjacent components.

Table.4 The sum of intermolecular forces

Alcohol system	The sum of intermolecular forces /eV	Hydrocarbon system	The sum of intermolecular forces /eV
1/3/456	1.4114	6/7/89a9b	2.2267
1/34/56	3.2808	6/78/9a9b	4.1930
1/345/6	3.6290	6/789a/9b	4.1572
13/4/56	1.7300	67/8/9a9b	2.7604
13/45/6	3.8118	67/89a/9b	4.0601
134/5/6	1.6437	678/9a/9b	1.8159

*The numbers in the first and third columns in the table refer to the carbon numbers, and 9a and 9b represent 2-Methyloctane and n-nonane respectively.

The sum of intermolecular forces through the complete sequence is assumed positively related to the total energy for separating each adjacent components. As shown in Table.4, the sum of intermolecular forces can be clustered into two groups, and that of the optimal distillation sequence is found in the lower cluster, as marking with shadows in the table,

i.e. first step cutting form like A/B/CDE, AB/C/DE and ABC/D/E, are possibly candidates for optimal sequence. It can be concluded that those within the higher group shouldn't be considered during the searching for the optimal sequence, which could significantly reduce the computation load, especially at the stage of the optimization of single column. Besides, it is also found that the cutting point of the first step in optimal distillation sequence with simple columns only is one of the two cutting points of DWC in the optimal distillation sequence with both simple column and DWC.

Based on the discussion above, it can be seen that the intermolecular force does affect the total energy consumption of a distillation sequence. With the knowledge of it, the computation load can be significantly reduced in the optimization of distillation sequence.

4.2. The subproblem separated by DWC in the optimal sequence

It can be observed from the result of case 1 that the optimal sequences obtained are totally different when DWC is included, however, the DWC is always applied for separating the three most heaviest components once it is included. Same result can also be found in case 3. It can be explained that the benefit on energy saving of DWC is more significant when heavier mixture is processed.

5. Conclusions

The optimal distillation sequences with DWC for five-component separation in alcohol system and hydrocarbon system have been identified by dynamic programming method based on the results of simulation and optimization of single columns. Intermolecular forces are calculated, their impact on the optimal sequence has been discussed based on existing results, which could be a reference for reducing the computation load in the optimization of distillation sequence in other separation system by eliminate sequences with significant higher total intermolecular forces.

Acknowledgements

The authors gratefully acknowledge the following institutions for their support: the National Natural Science Foundation of China (Grant No.21576015); the National Basic Research Program of China, 973 program (Grant No. 2013CB733600).

References

Duan Lianyun, Zhou Gongdu, 1989, Intermolecular force-an important factor that determines the property of materials, College chemistry, 4(2):1-7.

Howard, R. A., 1966, Dynamic programming. Management Science, 12(5), 317-348.

Mueller I., Kenig EY., 2007, Reactive distillation in a dividing wall column: rate-based modeling and simulation, Industrial & Engineering Chemistry Research, 46, 3709-3719.

Petlyuk F. B.; Platonoy V. M.; Slavinskii D. M., 1965, Thermodynamically optimal method for separating multicomponent mixtures. Znt. Chem. Eng. 5 (31, 555-561.

Triantafyllou C., Smith R., 1992, The design and optimisation of fully thermally coupled distillation columns: Process design, Chemical Engineering Research & Design, 70, 118-132.

Wayner, P. C., 1999, Intermolecular forces in phase‐change heat transfer: 1998 kern award review, Aiche Journal, 45(10), 2055-2068.

Wright R O., 1949, Fractionation apparatus, United States patent, US 2471134.

Zhang Yunlu, Han Guangyue, Sun Wei, Estimation of the number of distillation sequences with dividing wall column for multi-component separation. Chemical Engineering Transactions, 61, 343-348.DOI:10.3303/CET1761055.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50236-6

High-Pressure Polymerization of Ethylene in Tubular Reactors: Prediction of the Bivariate Distributions of Molecular Weight-Branches with a Rigorous Reactor Model

Maira L. Dietrich[a], Claudia Sarmoria[a,b], Adriana Brandolin[a,b], Mariano Asteasuain[a,b*]

[a]*Planta Piloto de Ingeniería Química (Universidad Nacional del Sur-CONICET), Camino La Carrindanga km 7, Bahía Blanca 8000, Argentina.*
[b]*Dpto. de Ingeniería Química, Universidad Nacional del Sur (UNS), Avenida Alem 1253, Bahía Blanca 8000, Argentina.*
masteasuain@plapiqui.edu.ar

Abstract

Low-density polyethylene (LDPE) is a branched polymer whose processing and end-use properties are mainly determined by the molecular weight distribution (MWD), the long chain branching distribution (LCBD) and short chain branching distribution (SCBD). In general, a suitable characterization of LDPE needs a simultaneous knowledge of these three distributions. This work presents a rigorous model of the high-pressure ethylene polymerization in tubular reactors, capable of predicting the joint bivariate distributions MWD-LCBD and MWD-SCBD, as well as monomer conversion and average molecular properties. The bivariate distributions are modeled using the 2D probability-generating function (2D pgf) technique. This deterministic technique allows a simple and efficient calculation of the distributions, in spite of the complex mathematical description of the process. The reactor model is used to analyze the effect of different process and design variables on the above mentioned joint distributions and on other properties of the product.

Keywords: Low-density polyethylene, High-pressure, Tubular reactors, Mathematical model, Branching distributions.

1. Introduction

LDPE is one of the most widely used commodity polymers, with a production rounding 19 Mt per year. It is used in a broad variety of applications, such as food packaging, general-purpose containers, bags, toys, films, etc. It is suitable for different processing techniques, such as extrusion, blowing or injection. The molecular properties of the polymer, and therefore its processing and end-use properties, are strongly determined by the reactor operating conditions. For this reason, a predictive tool able to simulate the reactor output for a given operating point is very useful to gain insight into the process. The processing and end-use properties of LDPE are mainly determined by the MWD, the LCBD, and the SCBD. It is well established that in highly branched polymers, long chain branches have a great impact on the rheological behaviour, affecting the flow properties of the material. Additionally, they affect some of its solid state properties (Pladis and

Kiparissides, 1998). On the other hand, short chain branches have an important influence on the polymer density. In general, a complete characterization of LDPE requires knowing simultaneously all three distributions. A high-fidelity reactor model capable of calculating these three distributions is very desirable.

Several mathematical models have been reported for the production of LDPE in tubular reactors, but few of them deal with the joint distributions of molecular weight and branches. Meimaroglou and Kiparissides (2010) presented a stochastic algorithm that provides information about topological characteristics of LDPE, including both MWD-LCBD and MWD-SCBD. This model was later extended to include the prediction of the viscoelastic behaviour of the polymer (Kiparissides et al., 2010; Meimaroglou et al., 2011; Pladis et al., 2015).

Although stochastic methods provide very detailed information on the molecular structure, they may have high computational demands. Furthermore, they may have difficulties with the quantification of the high molecular weight tail of MWDs, where chain concentrations are very low (Meimaroglou and Kiparissides, 2010). In the present work, a novel deterministic model of the high-pressure ethylene polymerization in tubular reactors is developed. The model can predict joint bivariate distributions MWD-LCBD and MWD-SCBD, as well as average molecular properties and conversion.

2. Reactor model

The mathematical model is an extension of a previous work by the authors (Asteasuain and Brandolin, 2008). The former model was rigorous, considered a realistic reactor configuration and included detailed correlations for the physical and transport properties. It could predict average molecular properties, conversion, and the univariate MWD. The updates to that model presented here consist of the prediction of the bivariate MWD-LCBD and MWD-SCBD. The bivariate distributions are calculated applying the 2D pgf method (Asteasuain and Brandolin, 2010; Brandolin and Asteasuain, 2013; Brandolin et al., 2016).

Deterministic models of polymer processes are built based on population balance equations (PBEs) derived from a kinetic mechanism of the process. These equations describe the evolution of the population of polymer molecules characterized by a set of internal coordinates, the distributed properties of interest (i.e. molecular weight, number of short or long branches). Usually, direct solution of the PBEs is not feasible because they form a very large (infinite in theory) system of coupled equations. The 2D pgf method works by transforming the infinite system of bivariate PBEs (i.e. PBEs characterized by two internal coordinates) into the pgf domain, obtaining a finite system of equations in which the dependent variable is the pgf transform of the distribution. Pgf values obtained from the solution of the pgf equations are fed to an inversion formula to recover the desired bivariate distributions. Then, the final mathematical model is composed of the finite set of pgf equations and the algebraic pgf inversion equations. This method is a general modelling tool that does not assume any shape of the distribution and can be applied to different systems even those with complex kinetic mechanisms. In addition to the bivariate distributions, the extended model computes conversion, the number- and weight-average molecular weights ($\overline{M_n}$,$\overline{M_w}$), long-chain branches (LCB/1000 C) and short chain branches (SCB/1000 C).

The mathematical model is built upon two parallel set of equations for the prediction of the MWD-LCBD and the MWD-SCBD, respectively. Table 1 shows the corresponding

kinetic mechanisms. Here $R_{n,m}$ and $P_{n,m}$ represent living radical and dead polymer chains with chain length m and n-1 long chain branches (i.e subscript n computes the main chain+branches). Similarly, $R_{h,\,m}$ and $P_{h,\,m}$ are living radicals and dead polymer chains with chain length m and h short chain branches. Long chain branches are considered to be formed by chain transfer to a dead polymer chain (Eq. 17). In its turn, an intramolecular transfer reaction (Eq. 22) is considered to be responsible for the formation of short chain branches.

Table 1. Kinetic mechanism for chain branching prediction.

Step	Long chain branching prediction		Short chain branching prediction	
Peroxide initiation	$I_k \xrightarrow{f_k\, k_k} 2R_{1,0}$	(1)	$I_k \xrightarrow{f_k\, k_k} 2R_{0,0}$	(2)
Monomer thermal initiation	$3M \xrightarrow{k_m} R_{1,1} + R_{1,2}$	(3)	$3M \xrightarrow{k_m} R_{0,1} + R_{0,2}$	(4)
Oxygen initiation	$O_2 + M \xrightarrow{k_o} 2R_{1,1}$	(5)	$O_2 + M \xrightarrow{k_o} 2R_{0,1}$	(6)
Generation of inert	$O_2 + R_{\mathrm{n,m}} \xrightarrow{f_o k_o} X$	(7)	$O_2 + R_{\mathrm{h,m}} \xrightarrow{f_o k_o} X$	(8)
Propagation	$R_{\mathrm{n,m}} + \mathrm{M} \xrightarrow{k_p} R_{\mathrm{n,m+1}}$	(9)	$R_{\mathrm{h,m}} + \mathrm{M} \xrightarrow{k_p} R_{\mathrm{h,m+1}}$	(10)
Termination by combination	$R_{\mathrm{n,m}} + \mathrm{R}_{\mathrm{y,j}} \xrightarrow{k_{tc}} P_{\mathrm{n+y-1,m+j}}$	(11)	$R_{\mathrm{h,m}} + \mathrm{R}_{\mathrm{y,j}} \xrightarrow{k_{tc}} P_{h+\mathrm{y,m+j}}$	(12)
Thermal degradation	$R_{\mathrm{n,m+1}} \xrightarrow{k_{tdt}} P_{\mathrm{n,m}} + R_{1,1}$	(13)	$R_{\mathrm{h,m+1}} \xrightarrow{k_{tdt}} P_{\mathrm{h,m}} + R_{0,1}$	(14)
Chain transfer to monomer	$R_{\mathrm{n,m}} + \mathrm{M} \xrightarrow{k_{trm}} P_{\mathrm{n,m}} + \mathrm{R}_{1,1}$	(15)	$R_{\mathrm{h,m}} + \mathrm{M} \xrightarrow{k_{trm}} P_{\mathrm{h,m}} + \mathrm{R}_{0,1}$	(16)
Chain transfer to polymer	$R_{\mathrm{n,m}} + \mathrm{P}_{\mathrm{j,y}} \xrightarrow{k_{trp}} P_{\mathrm{n,m}} + R_{\mathrm{j+1,y}}$	(17)	$R_{\mathrm{h,m}} + \mathrm{P}_{\mathrm{j,y}} \xrightarrow{k_{trp}} P_{\mathrm{h,m}} + R_{\mathrm{j,y}}$	(18)
Chain transfer to transfer agent	$R_{\mathrm{n,m}} + \mathrm{S} \xrightarrow{k_{trs}} P_{\mathrm{n,m}} + R_{\mathrm{1,0}}$	(19)	$R_{\mathrm{h,m}} + \mathrm{S} \xrightarrow{k_{trs}} P_{\mathrm{h,m}} + R_{0,0}$	(20)
Backbiting	$R_{\mathrm{n,m}} \xrightarrow{k_{bb}} R_{\mathrm{n,m}}$	(21)	$R_{\mathrm{h,m}} \xrightarrow{k_{bb}} R_{\mathrm{h+1,m}}$	(22)
β-Scission of secondary radical	$R_{\mathrm{n,m+1}} \xrightarrow{k_{\beta 1}} P_{\mathrm{n,m}} + R_{\mathrm{1,1}}$	(23)	$R_{\mathrm{h,m+1}} \xrightarrow{k_{\beta 1}} P_{\mathrm{h,m}} + R_{0,1}$	(24)
β-Scission of tertiary radical	$R_{\mathrm{n,m+1}} \xrightarrow{k_{\beta}} P_{\mathrm{n,m}} + R_{\mathrm{1,1}}$	(25)	$R_{\mathrm{h,m+1}} \xrightarrow{k_{\beta}} P_{\mathrm{h,m}} + R_{0,1}$	(26)

3. Results and discussion

The reactor model was used to study the influence of the operating conditions on the bivariate molecular weight-branching distributions. As an example, Table 2 summarizes the operating conditions of two of the cases analysed. Operating the reactor under Case I conditions produces a polymer with a unimodal MWD, while operation under Case II conditions gives rise to a bimodal MWD with a low molecular weight tail. In Case II, average molecular weights and dispersity index are larger than in Case I, while average branching content is lower. More detailed information about the MWDs obtained and the reactor configuration can be found in Asteasuain and Brandolin (2008).

Figure 1 shows contour plots of the MWD-LCBD at the reactor exit for the two cases. For both of them, a high fraction of the polymer contains between 0 and 2 LCBs per molecule (excluding the main chain). Furthermore, as the number of LCBs increases the molecular weight distributions at a given branching content become narrower and the weight fractions of the polymer chains decrease. At the same time, distributions shift towards higher molecular weights. That is expected because branching formation implies reactivating a dead polymer chain that starts growing again. Note that the MWD-LCBD of Case II is shifted towards higher molecular weights compared to Case I, which means that the same branching content is present in chains that are longer than those in Case I.

Table 2. Case I and II operating conditions (Asteasuain and Brandolin, 2008).

Operating Conditions	Case I	Case II
Inlet temperature (° C)	77	85
Inlet pressure (bar)	2300	2173
Oxygen main feed (kg/s)	6.9×10^{-5}	8.9×10^{-5}
Transfer agent main feed (kg/s)	0.00762	0.064
Transfer agent first lateral injection (kg/s)	0	0
Transfer agent second lateral injection (kg/s)	0	0.61
Peroxide first lateral injection (kg/s)	0.00102	8.5×10^{-5}
Peroxide second lateral injection (kg/s)	1.57×10^{-4}	6.9×10^{-5}
Monomer main feed (kg/s)	11	10.2
Monomer first lateral injection (kg/s)	0	0
Monomer second lateral injection (kg/s)	0	0.8
Location of first lateral injection (z/L)	0.12	0.03
Location of second lateral injection (z/L)	0.63	0.62
Jacket temperature zones 1-8 (° C)	170-225-170-170-170-170-170-170	270-150-270-266-150-150-159-270
Conversion (%)	25.8	20.5
$\overline{M_n}$	21602	25670
$\overline{M_w}$	162958	224149
SCB/1000C	24.73	14.79
LCB/1000C	2.14	1.37

Similarly, Figure 2 shows the contour plots of the MWD-SCBD. It can be seen that most chains have between 0 and 40 short chain branches. At low SCB content, the molecular weight distributions at a given branching content shift towards higher molecular weights as the number of branches increases. This occurs until a certain value of SCB (about 40), from which the peaks of the distributions do not change significantly. Similarly to the MWD-LCBD, there is also a shift towards higher molecular weights of the MWD-SCBD, but more subtle in this case.

Figure 3 depicts the branching density, expressed as average number of branches/1000 C, as a function of molecular weight. It may be noted that for Case I the short and long branching density are higher in almost all ranges of molecular weight. This is consistent

with the data in Figures 1 and 2, which show that at constant branching content the chains are longer in Case II than in Case I. The effect is much more noticeable for the LCB. Figure 3 also shows that the change in the operating conditions has a stronger effect on the LCB density distribution than in the SCB density distribution, in the form of an important shift towards high molecular weights.

4. Conclusions

In the present work, a deterministic model is presented that can predict the bivariate molecular weight-long and short chain branching distributions of LDPE produced in high-pressure tubular reactors. To the best of our knowledge, this is the first deterministic model for this process that includes these features. To demonstrate the predictive capabilities of the proposed approach, two operating conditions were simulated, obtaining, as a result, detailed information on the LCBD-MWD, SCBD-MWD and average number of short and long chain branches.

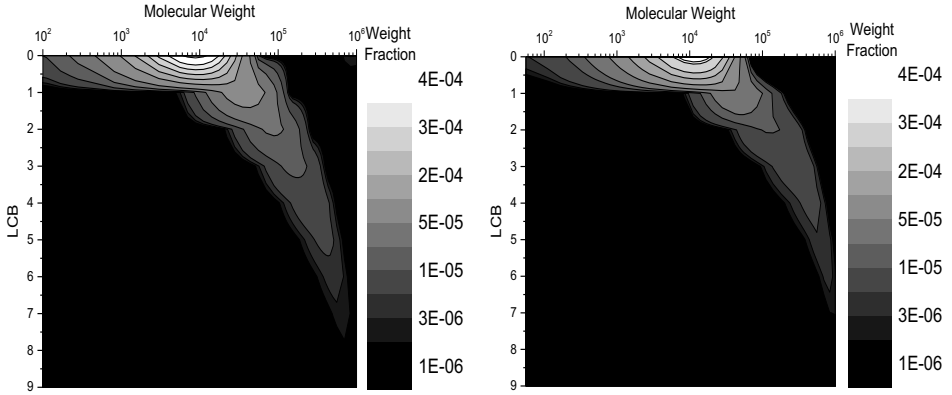


Figure 1. MWD-LCBD contour plots for Case I (left) and Case II (right).

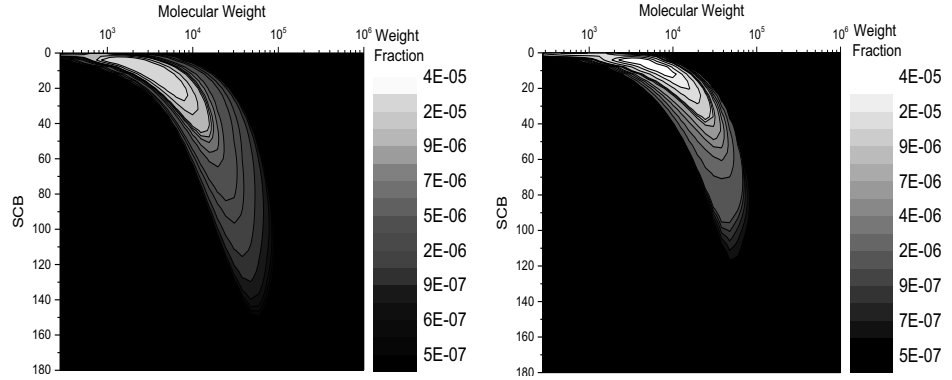


Figure 2. MWD-SCBD contour plots for Case I (left) and Case II (right).

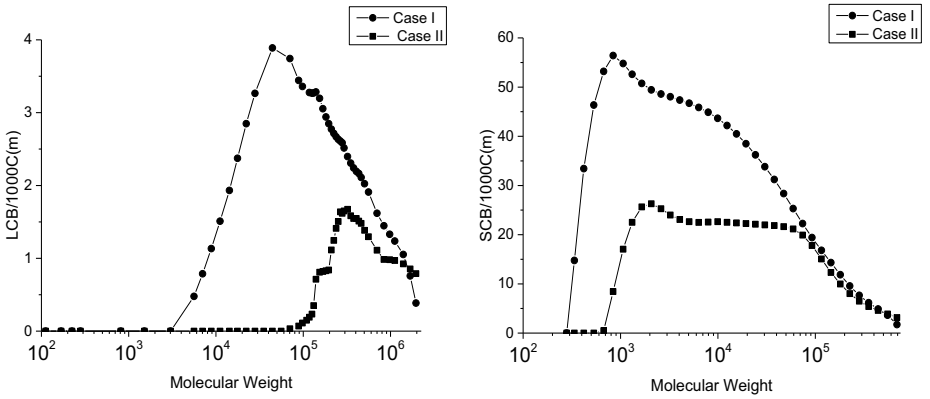


Figure 3. LCB/1000 C (left) and SCB/1000 C (right) for Case I and II.

References

Asteasuain, M. and A. Brandolin (2008). "Modeling and Optimization of a High-Pressure Ethylene Polymerization Reactor using gPROMS." Computers & Chemical Engineering **32**: 396-408.

Asteasuain, M. and A. Brandolin (2010). "Mathematical modeling of bivariate polymer property distributions using 2D probability generating functions, 1 - Numerical inversion methods." Macromolecular Theory and Simulations **19**(6): 342-359.

Brandolin, A., A. Assini Balbueno, et al. (2016). "Improved numerical inversion methods for the recovery of bivariate distributions of polymer properties from 2D probability generating function domains." Computers & Chemical Engineering **94**(Supplement C): 272-286.

Brandolin, A. and M. Asteasuain (2013). "Mathematical modeling of bivariate distributions of polymer properties using 2D probability generating functions. part II: Transformation of population mass balances of polymer processes." Macromolecular Theory and Simulations **22**(5): 273-308.

Kiparissides, C., A. Krallis, et al. (2010). "From Molecular to Plant-Scale Modeling of Polymerization Processes: A Digital High-Pressure Low-Density Polyethylene Production Paradigm." Chemical Engineering & Technology **33**(11): 1754-1766.

Meimaroglou, D. and C. Kiparissides (2010). "A novel stochastic approach for the prediction of the exact topological characteristics and rheological properties of highly-branched polymer chains." Macromolecules **43**(13): 5820-5832.

Meimaroglou, D., P. Pladis, et al. (2011). "Prediction of the molecular and polymer solution properties of LDPE in a high-pressure tubular reactor using a novel Monte Carlo approach." Chemical Engineering Science **66**(8): 1685-1696.

Pladis, P. and C. Kiparissides (1998). "A comprehensive model for the calculation of molecular weight-long-chain branching distribution in free-radical polymerizations." Chemical Engineering Science **53**(18): 3315-3333.

Pladis, P., D. Meimaroglou, et al. (2015). "Prediction of the Viscoelastic Behavior of Low-Density Polyethylene Produced in High-Pressure Tubular Reactors." Macromolecular Reaction Engineering **9**(3): 271-284.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50237-8

High Performance Agent-Based Modeling to Simulate Mammalian Cell Culture Bioreactor

Robert Jackson[a], Elif Seyma Bayrak[b], Tony Wang[b], Myra Coufal[c], Cenk Undey[b], Ali Cinar[a*]

[a]*Illinois Institute of Technology Department of Chemical and Biological Engineering, 10 W. 33rd Street, Chicago, IL 60616, USA*
[b]*Amgen Inc. Process Development, One Amgen Center Drive, Thousand Oaks, CA 91320, USA*
[c]*Amgen Inc. Process Development, 360 Binney St, Cambridge, MA 02141, USA*
cinar@iit.edu

Abstract

Agent-based modeling (ABM) is a novel modeling approach to address the complexity of systems that comprise heterogeneous interacting individuals. ABM is naturally hybrid for its ability to integrate quantitative and qualitative knowledge, deals with multiple levels of actions, and performs best when combined with conventional modeling approaches. In this study, a hybrid agent-based platform is developed using high performance computing to simulate a mammalian cell culture bioreactor. This platform enables communication of the agent-based model with the first principle models to account for quantitative changes in nutrient and metabolite concentrations and their distribution in the bioreactor. The model can predict viable cell density, and the cell cycle distributions along with the important nutrients and metabolites such as glucose and lactate. Integrating this cell culture agent-based model with high-performance computing leverages parallel processing to allow the ABM program to run faster, more efficiently, and with a higher capacity for the number of cells that can be modeled. The model is validated using bench-scale bioreactor experiments and showed good agreement with experimental data.

Keywords: Agent-Based Modeling, Mammalian Cell Culture, Cell Cycle, High-Performance Computing

1. Introduction

In the production of therapeutic proteins, cell cycle distribution, viable cell density and viability provide critical information on the health of the cell culture within the bioreactor. It is also useful to know the metabolic flux rates of glucose consumption and lactate production/consumption. This information can give insights into optimal feeding strategies and harvest decisions as well as the optimal process conditions for higher productivity.

Cell culture bioreactors have relied heavily upon experimental data for the purposes of scale-up, optimization, monitoring, and process control. Various types of computational models, including population balances, have been developed to predict cell physiology and behaviour (Mantzaris et al. 2001). Knowledge of cell cycle phase behaviour in response to environmental conditions and inputs can be used to optimize the production of therapeutic proteins because different proteins are optimally expressed in different

phases of the cell cycle. Studies showed it may be beneficial to arrest the cells in either the G1 or S phase through a temperature shift (Al-Rubeai and Emery, 1990; Leelavatcharamas et al. 1999).

Agent-based modeling is a powerful programming paradigm where autonomous computational entities, the agents, act in accordance with stochastic and deterministic rules. These rules describe the behaviour of agents within an environment or context. An environment can be a continuous or discrete space such as the substrate of a bioreactor or any type of network, such as a social or opinion network. The individual behaviour of heterogeneous agents on a microscopic level will combine to give the macroscopic behaviour of the entire system being modeled. The ABM approach easily integrates first principles models, heuristics, and probabilistic variation. These powerful capabilities make ABM a great candidate to model and understand complex cell behaviour in mammalian cell bioreactors. There are numerous software packages that can be used to construct ABMs including Repast (Recursive Porous Agent Simulation Toolkit) Simphony and Repast HPC (High Performance Computing) (Collier and North, 2013; North *et al.*, 2013). As the number of cells grows exponentially high performance computing becomes necessary to parallelize the code and speed up simulations. An ABM of CHO cell culture was developed in Repast Simphony (Bayrak *et al.*, 2016). Due to inherent computational limitations of desktop computer simulation the number of cells simulated using this model could not reach to realistic scales. In this work, it has been upgraded to Repast HPC to speed up the simulation time and increase the capacity of agents that can be handled through Message Passing Interface (MPI) used for parallel processing (Collier and North, 2013).

The overall goal of this work is to improve upon a model of CHO cell behaviour that was published in Bayrak et al. (2016), and develop this model using Repast HPC. The model can serve as an in-silico monitor to aid in process control and optimization of cell culture processes. This model intends to reduce the level of experimentation involved in cell culture process monitoring and optimization. It can potentially provide better insight into CHO cell behavior in response to their environment within bioreactors in order to reduce the need for black box parameter optimization and suggest promising regions for experimental validation and eliminate regions not worthy of testing.

2. Methods

The ABM rule-base used in this paper, based on Bayrak et al. (2016), is shown in Figure 1 below. Any set of conditions with high sodium or lactate is considered severe to cells. High oxygen is considered at least favourable or highly favourable depending on whether the glucose level is low or high, respectively. Based on these sets of environmental conditions, probabilities have been assigned to give the autonomously acting CHO cell agents actions to stochastically take as they progress through the cell cycle while sensing their environment. In the favourable conditions, CHO cells progress through the cell cycle and proliferate the most efficiently. Metabolic shift to lactate metabolism is also considered when glucose concentration is low and lactate production is high. Under mild and mild to severe conditions, cells have the option of shifting to G0 or a slight probability of entering apoptosis if they have not shifted to consuming lactate, however, G0 is experimentally indistinguishable from G1, and if they have shifted to consuming lactate then they will progress through the cell cycle inefficiently. As observed in the literature (Hwang *et al.*, 2011), pH outside a cell-line dependent

specified range within pH 7 and/or a shift to a temperature lower than 37 °C results in G1 or G2M arrest, but cells in S phase continue to replicate their DNA and progress through the cell cycle. Under severe conditions, cells have a high probability of progressing through the stages of apoptosis until cell death. A CHO cell metabolic flux model for glucose consumption and lactate production/consumption was also introduced into this ABM to predict glucose and lactate levels in the cell culture process.

The variables such as shear stress (τ), energy dissipation rate (ϵ), and the Kolmogorov micro-scale of turbulence (λ) are also considered in the rule base. According to Motobu et al. (1998), at a shear stress of 0.082 N/m^2, the percentage of confluent cells in S phase is much lower at 20% compared to 30% for 0 shear and a slight majority, close to 50%, of cells are in G2M compared to 40% for 0 shear. This observation supports the rule that shear stress contributes to G1 and G2M arrest. According to Vickroy et al. (2007), when the local Kolmogorov micro-scale of turbulence in a bioreactor is between 50% and 100% of the size of the cell's diameter, then the observed growth rate is reduced to 66% of its maximum rate. This observation supports the new rule that when a cell is exposed to an appropriately sized Kolmogorov eddy, its doubling time will increase by 50% and it will have a 33% probability of entering early apoptosis. According to Sieck et al. (2013), the energy dissipation rate seems to only have an impact on monoclonal antibody productivity, reducing it to 75% of its normal rate, and most of today's suspension CHO cell lines can thrive at energy dissipation rates several orders of magnitude higher than typically needed for plant-scale production, up to 1e5 W/kg compared to 1e0 W/kg. Since hydrodynamic conditions affect the viability of cells, it was decided in the rule base to check the Kolmogorov micro-scale of turbulence and shear stress first before checking the concentrations of the soluble media components.

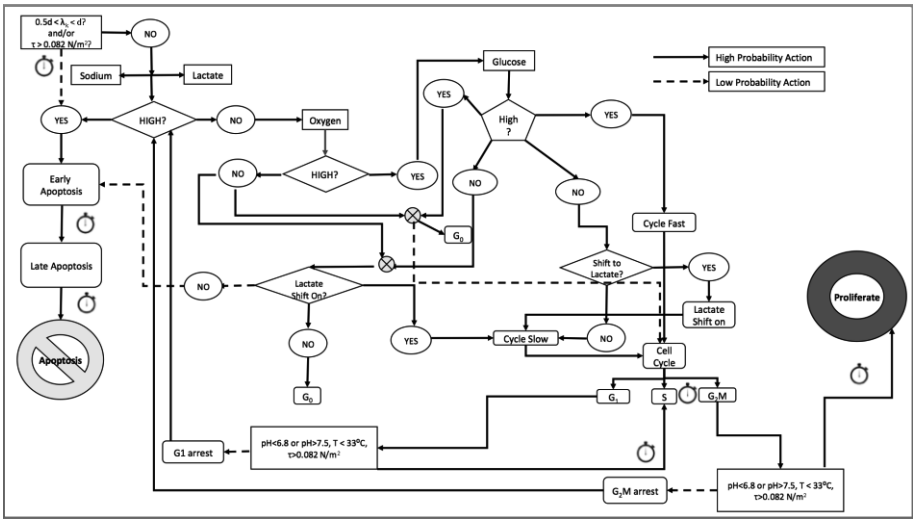


Figure 1. ABM rule-base.

Additionally, in order to represent heterogeneity of the culture, the ages of the initially inserted CHO cell agents were randomized. In other words, the amount of time they have previously spent in their initially assigned phase at the start of the simulation was randomized and could be non-zero.

The ABM was initialized with the first experimental data point and the CHO cell agents are subjected to the same pH, temperature, glucose, lactate, and sodium as the CHO cells in the corresponding experiment. However, the CHO Cell agents take actions autonomously based on the ABM rule-base.

A Red Hat (Maipo release 7.1) Linux server with an Intel(R) Xeon(R) 3.10GHz CPU and 62.5 GB RAM was used to build and run this model. The Agent-Based Modelling software used was Repast HPC version 2.2.0. Repast HPC consists of software libraries for a C++ compiler.

3. Results

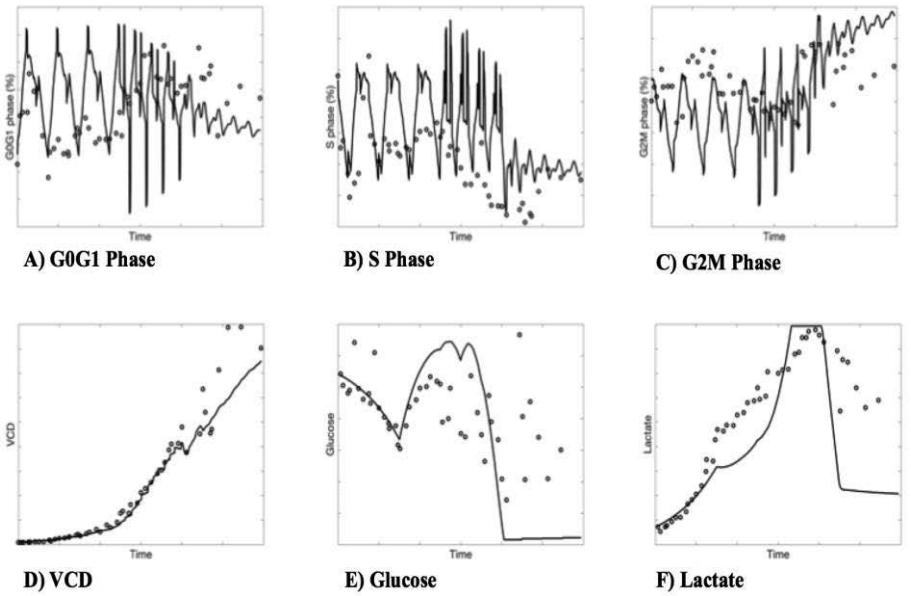


Figure 2. HP-ABM Experimental Validation for a cell culture process

The ABM predictions are compared with cell culture experiments without taking into account the hydrodynamic variables in the rule base, so the bioreactor was assumed to be homogeneous. Figure 2 shows that the ABM predictions for cell cycle distribution, viable cell density (VCD), glucose concentration, and lactate concentration for an Amgen proprietary cell culture process agree well with the experimental data. The simulation started initially with 9,600 agents to proportionally represent the proprietary initial VCD, and ended with 1,024,041 agents and took on the order of 10 minutes on the hardware mentioned in the methods section. The model prediction showed higher cyclic behavior compared to experimental data for the cell cycle predictions. The oscillations near the end of the experiment dampen because of cell cycle arrest. As expected, lactate production increases for the initial duration of the experiment, and then decreases as the cells shift to consuming lactate due to decreasing glucose and high availability of lactate to use as an energy and carbon source. The glucose agreement is good the duration of the experiment before a process change made that was not

accounted for in the model. The agreement for viable cell density to experimental data is higher. Different phases observed in the data correspond to Amgen proprietary cell culture process changes.

We have also conducted high performance bench-marking tests to measure the computational efficiency. The results of weak and strong scaling tests are shown in Figure 3 below. In the weak-scaling test where all processes have the same number of agents (10,000), the trend is near linear growth with slight increase for up to 32 processes. This might be due to less efficient parts in the algorithm. In the strong-scaling test where the same number of agents is divided among an increasing number of processes, significant decrease in computational time is observed as expected.

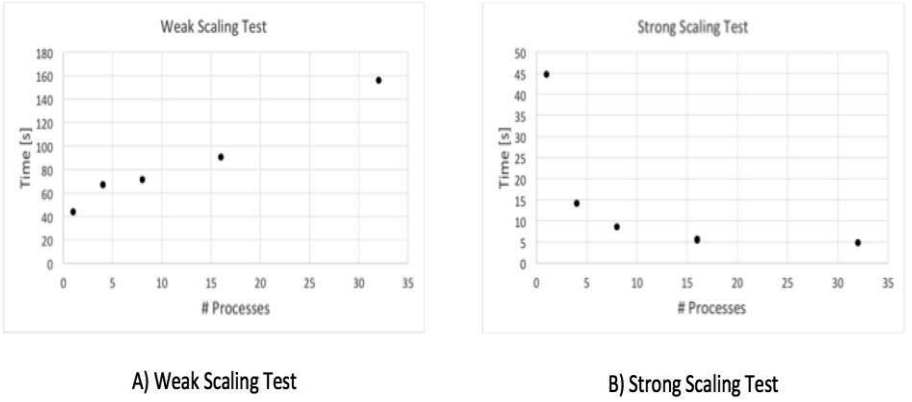


Figure 3. Weak and Strong Scaling Tests

4. Conclusions

The HP-ABM has been demonstrated to reasonably predict cell cycle phase distribution and more accurately predict viable cell density, glucose, and lactate. The computational time was reduced 100-fold using Repast HPC when compared to the desktop version of the model published in Bayrak et al. (2016). Repast HPC uses C++, which naturally compiles and runs faster than ABM code written in JAVA for, the language of the previous model done with Repast Simphony. Furthermore, the Repast HPC code has the capability of running on multiple CPU cores simultaneously. The model developed can be potentially used to monitor and control cell culture processes towards optimal productivity.

Acknowledgments: This study is based on research supported in part by the National Science Foundation (IIS- 1125412) and Amgen, Inc.

References

Al-Rubeai, M. and Emery, A. N. (1990) 'Mechanisms and kinetics of monoclonal antibody synthesis and secretion in synchronous and asynchronous hybridoma cell cultures', *Journal of Biotechnology*. Elsevier, 16(1–2), pp. 67–85. doi: 10.1016/0168-1656(90)90066-K.

Bayrak, E. S., Wang, T., Jerums, M., Coufal, M., Goudar, C., Cinar, A. and Undey, C. (2016) 'In Silico Cell Cycle Predictor for Mammalian Cell Culture Bioreactor Using Agent-Based Modeling Approach', *IFAC-PapersOnLine*, 49(7), pp. 200–205. doi: 10.1016/j.ifacol.2016.07.249.

Collier, N. and North, M. (2013) 'Parallel agent-based simulation with Repast for High Performance Computing', *SIMULATION*. SAGE PublicationsSage UK: London, England, 89(10), pp. 1215–1235. doi: 10.1177/0037549712462620.

Hwang, S. J., Yoon, S. K., Koh, G. Y. and Lee, G. M. (2011) 'Effects of culture temperature and pH on flag-tagged COMP angiopoietin-1 (FCA1) production from recombinant CHO cells: FCA1 aggregation', *Applied Microbiology and Biotechnology*, 91(2), pp. 305–315. doi: 10.1007/s00253-011-3266-7.

Leelavatcharamas, V., Emery, A. N. and Al-Rubeai, M. (1999) 'Use of cell cycle analysis to characterise growth and interferon-γ production in perfusion culture of CHO cells', *Cytotechnology*. Kluwer Academic Publishers, 30(1/3), pp. 59–69. doi: 10.1023/A:1008055904642.

Mantzaris, N. V., Daoutidis, P. and Srienc, F. (2001) 'Numerical solution of multi-variable cell population balance models: I. Finite difference methods', *Computers & Chemical Engineering*. Pergamon, 25(11–12), pp. 1411–1440. doi: 10.1016/S0098-1354(01)00709-8.

Motobu, M., Wang, P.-C. and Matsumura, M. (1998) 'Effect of Shear Stress on Recombinant Chinese Hamster Ovary Cells', *Journal of Fermentation and Bioengineering*, 85(2), pp. 190–195.

North, M. J., Collier, N. T., Ozik, J., Tatara, E. R., Macal, C. M., Bragen, M. and Sydelko, P. (2013) 'Complex adaptive systems modeling with Repast Simphony', *Complex Adaptive Systems Modeling*. Springer Berlin Heidelberg, 1(1), p. 3. doi: 10.1186/2194-3206-1-3.

Sieck, J. B., Cordes, T., Budach, W. E., Rhiel, M. H., Suemeghy, Z., Leist, C., Villiger, T. K., Morbidelli, M. and Soos, M. (2013) 'Development of a Scale-Down Model of hydrodynamic stress to study the performance of an industrial CHO cell line under simulated production scale bioreactor conditions', *Journal of Biotechnology*. doi: 10.1016/j.jbiotec.2012.11.012.

Vickroy, B., Lorenz, K. and Kelly, W. (2007) 'Modeling shear damage to suspended CHO cells during cross-flow filtration', in *Biotechnology Progress*. doi: 10.1021/bp060183e.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50238-X

HPC enabled parallel, multi-scale & mechanistic model for high shear granulation using a coupled DEM-PBM framework

Chaitanya Sampat[a], Yukteshwar Baranwal[a], Ioannis Paraskevakos[b], Shantenu Jha[b], Marianthi Ierapetritou[a] and Rohit Ramachandran[a,*]

[a]*Department of Chemical and Biochemical Engineering, Rutgers – The State University of New Jersey, Piscataway, New Jersey 08854, USA*
[b] *Electrical and Computer Engineering, Rutgers – The State University of New Jersey, Piscataway, New Jersey 08854, USA*
rohitrr@soe.rutgers.edu

Abstract

A multiscale model combines the computational efficiency of a macro-scale model and the accuracy of a micro-scale model. With the current cyberinfrastructure resources available, using more computationally intensive and concurrent multiscale models are more feasible. This study proposes to use Discrete Element Method (DEM) and a Population Balance Model (PBM) in a simultaneous manner to model the granulation process of a pharmaceutical product inside a high shear granulator. The DEM provides the collision data while the PBM helps in predicting the macroscale phenomena like aggregation and breakage. The execution of each of the components is governed by a multilevel job scheduler which allocates resources. This method of using shorter bursts of each simulation led to faster simulation times as well as a more accurate model of the high shear granulator. The Quality by Design (QbD) approach is addressed using such a modelling framework and it also helps us understand the granulation process in a quantitative as well as in a mechanistic manner.

Keywords: Multi-scale Model, Population Balance Model, Discrete Element Method, High-performance computing, MPI.

1. Introduction

Particulate materials are products or intermediates of around 60% of the chemical industry (Ingram and Cameron, 2004). The modelling of these particulate processes pose a challenge when compared to modelling if uniform liquid systems. These solid systems usually comprise of solids which co-exist with various sizes and shape configuration, which have a large effect on the final composition and the performance of the product (Ingram and Cameron, 2005). This variation in the configuration of the product is unacceptable by the pharmaceutical industry. It could lead to differences in the dissolution rates and bioavailability of the product thus, affecting the quality of the product. The unit process that helps control the size of a granule of the powder in the pharmaceutical product is granulation. During this unit operation, fine powders are converted to the larger particles by the process of aggregation. This occurs due to addition of a liquid binder, making the particles agglomerate into larger granules. This process is also referred to as wet granulation as a liquid binder is added. Due to high restrictions set by the Food and Drug Administration (FDA) these products undergo

various batch rejections or have high recycle ratios (Sen *et al.*, 2014). One of the major aims of the industry currently is to reduce this waste, thus creating a need for better modelling of this process.

The most popular way to model the granulation process is to use a population balance model (PBM). This model tracks the number of particles with certain set of characteristics such as size, porosity and liquid content. It also takes into account the changes that may occur in these properties due to growth, breakage, consolidation, and other phenomena. This model is discussed in more detail in section 2.1. PBM fails to account for the particle level data, thus being inaccurate at times. This particle level data needed to improve the PBM can be obtained from discrete element methods (DEM). DEM helps obtain the velocity of each particle and the number of times it collides with other particles by solving the Newton's laws of motion. This information is vital to the PBM as it helps develop a more accurate model with a higher physical significance. As these methods are complementary to each other, they are usually coupled while building a model for the granulation process. These processes can be coupled serially like in Sampat *et al.*, 2017 or in parallel as discussed in this work. DEM is always executed at initiation and the particle level data is then transferred to the PBM. The DEM simulation is computationally very high as it needs to solve large amounts of simultaneous equations at every time step. The parallel execution of this type of coupling requires various instances of simulation of each of these models to be initiated, requiring high computational power. The most efficient way to make these simulations run faster is to take advantage of large number of CPU cores present on modern day high-performance computing systems (HPCs / supercomputers).

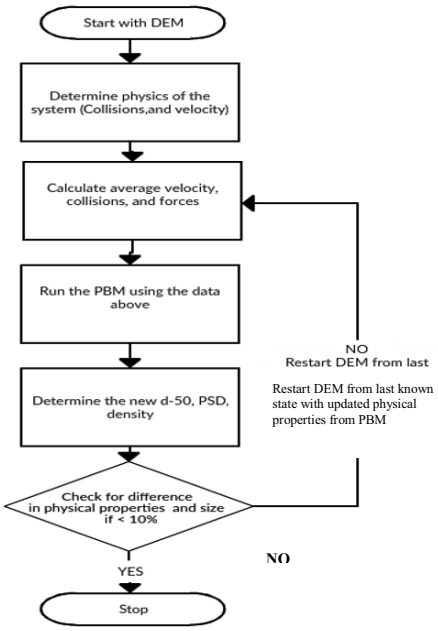


Figure 1: Schematic of the bi-directional approach used to model the high shear granulator.

This work focuses on the development of physically accurate model for a high shear granulator. Section 1 discusses the model equation used and the method of parallelization implemented on the PBM and DEM to enable them to run on various cores of HPCs. The resulting particle properties from the coupling and the speed improvements achieved when the number of cores used for the simulation are changed, is discussed in section 3.

2. Methods

Figure 1 represents the workflow that was used to model the high shear granulator. A two-dimensional PBM was developed for the solids present while a lumped model was used for the liquid and gas content. The PBM tracked particle flow, aggregation, breakage, consolidation and liquid addition. This PBM used the particle velocities and

collision data from the DEM simulations making it mechanistic in nature. The DEM simulation was used to obtain the particle residence time, velocities, inter-particle as well as particle-geometry collisions. These simulations were monitored using python scripts to switch in between the two simulations. The details of each of these components are discussed in this section.

2.1. Population Balance Model (PBM)

A detailed, mechanistic two-dimensional PBM was developed in C++, which tracked the particles as they undergo the aggregation, breakage, consolidation and liquid addition. The governing population balance model is shown in Equation 1.

$$\frac{\delta}{\delta t}F(s_1, s_2, x, t) = \Re_{agg}(s_1, s_2, x, t) + \Re_{br}(s_1, s_2, x, t) + \dot{F}_{in}(s_1, s_2, x, t) - \dot{F}_{out}(s_1, s_2, x, t) \quad (1)$$

Where, $\dot{F}(s_1, s_2, x)$ is the number of particles with an active pharmaceutical ingredients (API) volume of s_1 and an excipient volume of s_2 in the spatial compartment x. The rate of change of number of particles with time in different size classes depend on the rate of aggregation, $\Re_{agg}(s_1, s_2, x)$ and the rate of breakage, $\Re_{br}(s_1, s_2, x)$. Also, the rate of particles coming into, $\dot{F}_{in}(s_1, s_2, x)$ and going out, $\dot{F}_{out}(s_1, s_2, x)$ of the spatial compartment due to particle transfer affect the number of particles in different size classes.

The aggregation and breakage kernel as defined in Equations 2 and 3 respectively. These mechanistic kernels used the DEM inter-collision data as well as the impacts with the geometry to determine the collision and impact rate. The velocity of these particles were used to create a velocity distribution which decided whether the particle agglomerated with other particles or had to undergo breakage due to the impact on the geometry. The further details about the aggregation and breakage kernels can be found in Barrasso and Ramachandran, 2015. Details about the lumped model for the gas and liquid present in the system can be found in Sampat *et al.*, 2017.

$$\Re_{agg}(s_1, s_2, x) = \int_0^{s_1}\int_0^{s_2}\beta(s_1', s_2', s_1 - s_1', s_2 - s_2', x)F(s_1', s_2', x)F(s_1 - s_1', s_2 - s_2', x)ds_1'ds_2 - F(s_1', s_2', x)\int_0^{s_{max_1}-s_1}\int_0^{s_{max_2}-s_2}\beta(s_1, s_2, s_1', s_2', x)F(s_1', s_2', x)ds_1'ds_2' \quad (2)$$

Where, the aggregation kernel, $\beta(s_1, s_2, s_1', s_2', x)$ is expressed as a function of collision rate coefficient and probability that collision results in agglomeration.

$$\Re_{br}(s_1, s_2, x) = \int_0^{s_{max_1}}\int_0^{s_{max_2}} K_{br}(s_1', s_2', x)F(s_1', s_2', x)ds_1'ds_2' - K_{br}(s_1, s_2, x)F(s_1, s_2, x) \quad (3)$$

Where, the breakage kernel $K_{br}(s_1, s_2, x)$.

The number of solid bins selected for this simulation were 16 for each type of solid. The granulator was divided in 4 axial compartments to help with the parallelization of the *C++* code. Since each compartment needed to perform identical calculations which were independent each other, they could be sent different CPU cores without any decrease in the speed as there will be minimum communications between these

processes for a given time step. Thus, each of compartment was sent to a single process (core) using message parsing interface (MPI). The PBM code was made to dump the d50 and number of particles at an interval of *0.2s* such that the controller could monitor the execution.

2.2. Discrete Element Method (DEM) setup

The DEM simulation was performed on an open-source DEM software known as LIGGGHTS (Kloss *et al.*, 2012). The DEM simulation has to be initiated using a input file which was generated using a python script which was developed. The python script also helped generate the multiple restart files which would be required to restart the DEM simulation during the parallel multi-scale simulation. LIGGGHTS natively supports parallelization using MPI, which helps divide the geometry into various smaller simulation boxes thus, making them run faster compared to a serial code.

The particles used in the DEM simulation were of diameter 2mm and were considered to be granular in nature. A non-cohesive elastic contact model was used to model the collisions between the particles. The particles were introduced at the entry of the geometry with a constant mass flow rate of 15 kg/hr. The impeller of the granulator was set to rotate at 2000 rotations per minute. This was kept so high since this would lead higher number of particles colliding with the walls of granulator leading to breakage of the particles. Physical properties of the particles required for the simulation were taken from Sampat *et al.*, 2017. The collision and velocity data of the particles from the DEM simulation was printed at pre-decided intervals such that it could be monitored.

2.3. Controller Development

The controller scripts to monitor the DEM and PBM simulations were written in python. Each of these scripts were executed along with its respective simulation. The DEM controller script monitored the collision and velocity data being dumped by the DEM simulation. The data was first stored and the average of the velocities for each type of particles and the total number of collisions and impacts were determined. If there was a change in either of these properties of more than 15%, an exit status was sent which led to killing the execution of the DEM and indicated to start the PBM simulation. The DEM controller script also printed out various data files required for the PBM execution as well as for the restart of the DEM simulation if needed. In the case, the properties did not show a variation of more than 15% over the span of 5s of DEM simulation time, it was considered to be in steady state and that all the simulations were halted. The PBM controller read the d50 and number of particle files which were printed by the PBM after a constant time interval. The change in each of these properties in each compartment was monitored and an exit signal was sent if they varied by more than 15%. The halting of the PBM was accompanied by the printing of files required to restart DEM with new diameters and data for restarting the PBM if needed after the DEM simulation.

2.4. RADICAL – Pilot (RP)

Running several simulations that are dependent on High Performance Computing (HPC) resources can be extremely time consuming, since each simulation must be submitted in the resource independently, which can lead to varying queuing times. RADICAL-Pilot (RP) is a framework that implements the Pilot abstraction (Luckow *et al.*, 2012), where a container job is submitted to the HPC resource, the Pilot, and once

the necessary computing resources are being acquired they can be used to execute multiple simulations by reusing the same resources. By utilizing RADICAL-Pilot, we implemented an execution framework that allows to utilize the components described above. In addition, RP provides the necessary instructions to terminate or restart the execution of DEM or PBM simulations based on decisions made from the various controlling components.

3. Results and Discussions

This coupled DEM-PBM simulation were tested on the Stampede2 supercomputer with the Knights Landing configuration. Each of the compute node of the cluster consists of Intel Xeon Phi 7250 "Knights Landing" which has 68 cores on a single socket clocked at 1.4 GHz, with 96 GB of DDR4 RAM. Each of the cores consists of 4 hardware threads. Simulations were run on various configuration of the cores used for each of the component of the system. The number of cores used for the DEM simulation were 64, while the PBM used 4 MPI processes. The DEM and PBM controller scripts used 1 process each for their respective executions.

The coupled system took 10 DEM and 9 PBM simulations to reach steady state. Total simulation time of DEM execution was 3.575 seconds and PBM was 60 seconds.

3.1. Improved System Dynamics

The advantage of using a multiscale model in parallel is that the physical properties of the particles can be updated to the new values while moving among the simulations. This not only helps achieve better results but also reduces the time required for simulating the system. Granulation is used to control the size of the granules of a powder mixture. The median diameter of this powder mixture is usually used to characterise the behaviour of the system. For better control, median diameter of this system should increase gradually such that the time of granulation can be capped.

Figure 2 shows the evolution of median diameter of the simulated wet granulation process inside a high shear granulator with time. A slow but steady increase initially in the diameter is observed which is followed by breakage of granules into smaller ones due to the high amounts of shear present as the impeller rotates at high RPMs. This observation shows us that the model is better at predicting the characteristics of the process. The observed trend is better than previous works which only took into account only a serial coupling i.e. using the DEM collision information into the PBM for a single iteration.

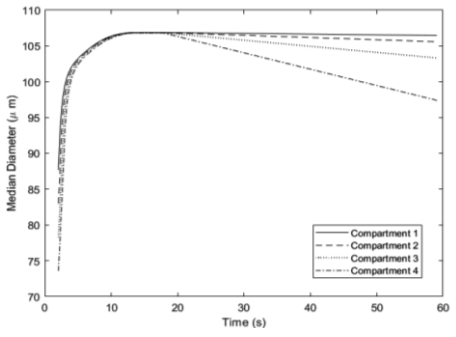


Figure 2: Median diameter of the system as a function of time. The value decreases in all compartments with time predicting high amounts of breakage as observed in a high

3.2. Simulation Speedup

When DEM and PBM are coupled for a single iteration, the steady state DEM data is passed to the PBM. DEM studies performed for this system showed that the system took

about 6 simulation seconds to reach steady state, which resulted in a wall time of 13 hours using 128 cores. The PBM was run on 4 MPI process which tool about 15 minutes to simulate. In this study, DEM studies ran for a total of 3.575 simulation seconds which took 9.76 hours. The PBM was simulated for a total of 60 simulation seconds consisting of 10 seconds of mixing and 50 seconds of liquid addition which took 49.27 minutes. Thus, it can be seen that the coupled simulation which has more components compared to a serial simulation finishes faster with better results. Since the data is updated after every iteration of each simulation and passed into the other it leads to the system reaching steady state quicker than a standalone simulation.

4. Conclusions

Multiscale models have recently been gaining popularity in pharmaceutical process modelling and some of its advantages are conveyed using this study. The coupled model is physically more accurate and has speed advantages over a serial model. The study also presents a model that utilizes HPCs for a quicker solution compared to traditional desktop computers and can also help design a control strategy for the granulation process. Currently, the challenge in implementation a control strategy lies in the online measurement of the particle size distribution (PSD) inside the equipment. The latest measurement techniques are not able to measure PSDs on the fly yet. Once the live data is available, a model can be initiated from the current state of the equipment and help reduce the computation required for the prediction. This work is a continued effort in the implementing a process control model to the granulator to help reduce pharmaceutical product waste.

References

Barrasso, D. and Ramachandran, R. (2015) 'Multi-scale modeling of granulation processes: bi-directional coupling of {PBM} with DEM via collision frequencies', *Chemical Engineering Research and Design*. Elsevier, 93, pp. 304–317.

Ingram, G. D. and Cameron, I. T. (2004) 'Challenges in multiscale modelling and its application to granulation systems', *Asia-Pacific Journal of Chemical Engineering*. Wiley Online Library, 12(3–4), pp. 293–308.

Ingram, G. D. and Cameron, I. T. (2005) 'Formulation and comparison of alternative multiscale models for drum granulation', *Computer Aided Chemical Engineering*. Elsevier, 20, pp. 481–486.

Kloss, C., Goniva, C., Hager, A., Amberger, S. and Pirker, S. (2012) 'Models, algorithms and validation for opensource DEM and CFD-DEM', *Progress in Computational Fluid Dynamics, an International Journal*. Inderscience Publishers, 12(2–3), pp. 140–152.

Luckow, A., Santcroos, M., Merzky, A., Weidner, O., Mantha, P. and Jha, S. (2012) 'P*: A model of pilot-abstractions', in *E-science, 2012 IEEE 8th International Conference on*, pp. 1–10.

Sampat, C., Bettencourt, F., Baranwal, Y., Paraskevakos, I., Chaturbedi, A., Karkala, S., Jha, S., Ramachandran, R. and Ierapetritou, M. (2017) *A parallel unidirectional coupled DEM-PBM model for the efficient solution and simulation of computationally intensive systems. (Unpublished Manuscript).*

Sen, M., Barrasso, D., Singh, R. and Ramachandran, R. (2014) 'A multi-scale hybrid CFD-DEM-PBM description of a fluid-bed granulation process', *Processes*. Multidisciplinary Digital Publishing Institute, 2(1), pp. 89–111.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50239-1

Evaluating the Boosting Approach to Machine Learning for Formation Lithology Classification

Vikrant A. Dev, Mario R. Eden*

Department of Chemical Engineering, Auburn University, Auburn, AL 36849, USA
edenmar@auburn.edu

Abstract

For petroleum exploration and engineering, the classification of underground formation lithology from well log data is a task of considerable import. This is because, lithology classification is the basis of reservoir parameter calculations and geological research studies. With the rising prowess of cheap computational devices, large amounts of data can now be analyzed with much higher accuracy and efficiency. Hence, there have recently been increased efforts to automate lithology classification. In one such effort, Xie et al. (2018) recently evaluated five machine learning methods to classify formation lithology by using data from the Daniudui gas field (DGF) and Hangjinqi gas field (HGF). Although their tree-based ensemble models performed well, there is still scope for improvement in the predictive ability. Motivated by the encouraging results obtained from the boosting approach to machine learning in their work, we applied this approach in our work with the aim of improving upon the predictive ability of their ensemble models. Specifically, we applied the AdaBoost and LogitBoost meta-algorithms using decision stumps and random trees as base learners. We evaluated the boosted trees by calculating metrics such as precision, recall, F1-score and PRC area after implementing 5-fold and 10-fold stratified cross-validation. In our analysis, amongst the applied meta-algorithms, we found that the implementation of the LogitBoost meta-algorithm, with random tree as a base learner, possessed the highest performance metrics.

Keywords: Lithology classification, ensemble machine learning, boosting, AdaBoost, LogitBoost

1. Introduction

Lithology identification involves the classification of rock formations based on their physical characteristics. In petroleum exploration and engineering, lithology identification forms one of the core tasks. It underlies the calculation of reservoir parameters. It also is the basis of geological research studies. Traditionally, the formation lithology has been identified in two ways (Salehi and Honarvar, 2014). First is through inference from cut pieces obtained during drilling. The second method of inference has been through analysis of core samples gathered from underground formations. In practice, however, both methods of inference can be unreliable. This may occur because geologists may introduce their bias while assessing rock samples. Consequently, different geologists may interpret rock samples differently (Akinyokun et al., 2009). Due to this concern, attention has shifted towards utilizing various computational techniques for assessing well log data for formation lithology classification. Well log data has proved useful not only for calculation of petrophysical parameters but also for sedimentologists and reservoir engineers. If computational models to identify different grain sizes of clastic rock with better accuracy are made

available, geologists can build better quantitative reservoir evaluation models (Xie et al., 2018). Recently, a variety of approaches have been introduced for lithology identification including those employing kernels (Sebtosheikh et al., 2015) and ensemble machine learning techniques like random forests. However, as Xie et al. (2018) note in their work, less work has been conducted to compare the performance of various machine learning methods for assessment of well log data. In their work, to address this paucity, they compared five machine learning methods. These were artificial neural networks, random forests, naïve bayes, support vector machines and gradient boosted trees. They analyzed data obtained from DGF and HGF located in China and found that ensemble methods, gradient boosted trees and random forests, overall, provided a good level of performance. Although the gradient boosted tree performs well on their dataset, there is still scope for improvement with regards to the classification accuracy metrics used by them. For example, for the dataset pertaining to DGF, they reported the average precision, recall and F1-scores to be 0.827, 0.824 and 0.822 respectively. Motivated by the encouraging performance of the gradient boosted tree on their datasets, we investigated the use of other boosting based machine learning approaches for classification in our work. Specifically, we implemented the AdaBoost and LogitBoost meta-algorithms in our work. To the best of our knowledge a comparative analysis of the AdaBoost and LogitBoost meta-algorithms has not been conducted on the dataset of Xie et al. (2018).

2. Methodology

Two diverse and labelled datasets consisting of well log data from DGF and HGF respectively were obtained from the work of Xie et al. (2018). The DGF dataset consists of 915 well log readings from five wells while the HGF dataset consists of 1238 log readings from seven wells. In both datasets, carbonate rock, coal, pebbly sandstone, coarse sandstone, medium sandstone, fine sandstone, siltstone and mudstone are the eight target lithology classes that need to be identified by the classification methods. Additionally, seven log parameters are used to identify the lithology. The seven log parameters consist of gamma ray log, acoustic log, density log, compensated neutron log, deep latero log, shallow latero log and caliper log. More details on the dataset and the features used can be found in the work of Xie et al. (2018). It is worth noting that the DGF and HGF datasets are imbalanced. Imbalanced datasets are those in which instances corresponding to certain classes are significantly lower than others. The imbalance is especially sharp in the HGF dataset where only 14 instances of the coal class exist. By comparison, 370 instances of pebbly sandstone exist in the HGF dataset. Also, unlike the DGF dataset, the HGF dataset has no instances of the carbonate rock class. In our work, like Xie et al. (2018), we use the reported seven features to identify the lithology classes. Also, we rely solely on the boosting approach to ensemble machine learning in our work. In our work, all ensemble boosting approaches were evaluated using the WEKA software (Hall et al., 2009).

2.1. Ensemble Machine Learning

As the name suggests, ensemble machine learning is performed by generating an ensemble of models, which can involve training constituent models on resampled datasets. These models are then concatenated in some form (for example, by averaging) to then provide predictions of response variables of interest (Dev et al., 2017). Ensemble methods have been applied successfully for tasks as disparate as image classification and compound classification. Contrary to expectations, ensemble learning methods may

generate constituent models that can predict barely better than what can be achieved by chance. The reason for employing such models for ensemble construction is that they are computationally inexpensive to generate. Such constituent models are termed 'weak learners' or 'base learners'. Two of the most common approaches to ensemble construction are randomization and boosting.

2.1.1. Randomization-Based Approaches

Randomization-based approaches (RBAs) produce distinct models from a single initial training set by introducing random perturbations into the learning procedure (Dev et al., 2017). Thus, ensemble construction in these approaches can be parallelized. A commonly utilized approach is bagging or bootstrap aggregation. Randomization-based approaches at the very least achieve a reduction in the error due to variance. Additional details and resources for consultation on RBAs can be found in Rokach (2016).

2.1.2. Boosting

Boosting involves choosing training sets for base learners such that they are forced to infer something new about the data each time they are called. Unlike RBAs, boosting constructs ensembles sequentially. Boosting works by assigning a large weight to training samples incorrectly classified by the current classifier. Next, another classifier is trained after accounting for the weights. This process is repeated until a certain stopping criterion is achieved. Essentially, at each iteration, a new base-learner is trained with respect to the error of the whole ensemble learnt so far. Unlike RBAs, the construction of each base learner is strongly dependent on the previously generated base learners. The Boosting approach to develop a model *f(x)* is visualized in Fig. 1, which has been adapted from Sugiyama (2016).

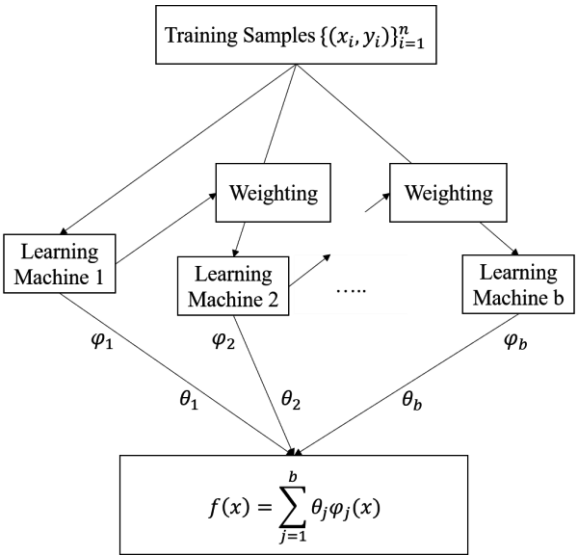


Figure 1: Boosting

In Fig. 1, φ_j is the j^{th} base learner and θ_j is the j^{th} weight associated with the j^{th} learner. Probably the most popular boosting approach is AdaBoost (Rokach, 2016). AdaBoost is short for adaptive boosting, and the algorithm is characterized by the adaptive manner in which it produces and concatenates weak learners. Friedman et al. (2000) observed that Adaboost is in effect approximating a stagewise additive logistic regression model by

optimizing an exponential criterion. Consequently, new variants of AdaBoost were developed that fit additive models directly. One such variant is LogitBoost (Friedman et al., 2000), which employs Newton-like steps to optimize the loss criterion. In our work, we have focused our attention on the AdaBoost and LogitBoost meta-learners. In our work, the meta-learners employed decision stumps (DSs) and random trees (RTs). A DS is a decision tree containing only the root node and the leaves, i.e., a stump contains no internal nodes. RTs are trees grown on a random subset of descriptors (Baskin et al., 2017). Thus, in total four methods were evaluated for classification in our work.

3. Results and Discussion

The average and weighted average performance metrics for classification, i.e., precision (Pr), recall (Re), F1-scores (F1) and the precision recall curve (PRC) area are reported in Tables 1-4. Table 1 and Table 2 contain metrics corresponding to 10-fold stratified cross-validation (SCV) and 5-fold SCV respectively, based on the DGF dataset. Similarly, Table 3 and Table 4 contain metrics corresponding to 10-fold and 5-fold SCV respectively, based on the HGF dataset. SCV is more appropriate for our datasets than cross validation (CV), which was implemented by Xie et al. (2018) in their work. This is because in the case of imbalanced datasets, it accounts for the distribution of the classes of the instances while creating the folds and offers additional variance reduction. Also, it should be noted that Xie et al. (2018) have not reported the weighted average values of the performance metrics despite class imbalance. Unlike Xie et al. (2018), we have also chosen to report PRC area values. They can especially be informative for imbalanced datasets (Saito and Rehmsmeier, 2015). Xie et al. (2018) have also restricted their analysis to 5-fold CV. However, it should be noted that, unlike our work, they have tuned their hyperparameters. In contrast, we chose the defaults offered by the WEKA software (Hall et al., 2009) for model evaluation. As can be noticed from the results, when AdaBoost employs DS the metrics are very low. LogitBoost with DS, however, improves the performance metrics. In general, LogitBoost with RT provides the best performance metrics, both for the DGF and the HGF datasets.

Table 1: Metrics for 10-fold SCV on DGF dataset

Method	Pr	Re	F1	PRC area	
AdaBoost w/DS	0.086	0.222	0.124	0.197	Average
	0.129	0.341	0.186	0.228	Weighted Average
AdaBoost w/RT	0.777	0.769	0.772	0.830	Average
	0.762	0.764	0.762	0.829	Weighted Average
LogitBoost w/DS	0.692	0.659	0.669	0.732	Average
	0.676	0.677	0.673	0.729	Weighted Average
LogitBoost w/RT	0.821	0.818	0.820	0.845	Average
	0.809	0.807	0.807	0.840	Weighted Average

Table 2: Metrics for 5-fold SCV on DGF dataset

Method	Pr	Re	F1	PRC area	
AdaBoost w/DS	0.124	0.223	0.148	0.200	Average
	0.163	0.336	0.207	0.232	Weighted Average
AdaBoost w/RT	0.774	0.755	0.765	0.820	Average
	0.764	0.764	0.763	0.818	Weighted Average
LogitBoost w/DS	0.686	0.652	0.665	0.724	Average
	0.666	0.666	0.661	0.718	Weighted Average
LogitBoost w/RT	0.801	0.791	0.795	0.830	Average
	0.778	0.777	0.777	0.818	Weighted Average

Table 3: Metrics for 10-fold SCV on HGF dataset

Method	Pr	Re	F1	PRC area	
AdaBoost w/DS	0.149	0.248	0.179	0.253	Average
	0.245	0.441	0.304	0.298	Weighted Average
AdaBoost w/RT	0.779	0.893	0.778	0.640	Average
	0.781	0.781	0.781	0.653	Weighted Average
LogitBoost w/DS	0.677	0.547	0.591	0.638	Average
	0.599	0.597	0.588	0.627	Weighted Average
LogitBoost w/RT	0.832	0.808	0.819	0.879	Average
	0.843	0.842	0.842	0.908	Weighted Average

Table 4: Metrics for 5-fold SCV on HGF dataset

Method	Pr	Re	F1	PRC area	
AdaBoost w/DS	0.150	0.248	0.180	0.213	Average
	0.247	0.442	0.305	0.301	Weighted Average
AdaBoost w/RT	0.765	0.741	0.752	0.604	Average
	0.766	0.766	0.766	0.635	Weighted Average
LogitBoost w/DS	0.633	0.531	0.565	0.638	Average
	0.581	0.581	0.572	0.624	Weighted Average
LogitBoost w/RT	0.845	0.797	0.819	0.879	Average
	0.829	0.829	0.828	0.899	Weighted Average

4. Conclusions

With the aim of developing classifiers with high accuracy and to improve upon the models trained in the work of Xie et al. (2018), in our work we investigated boosting based approaches to machine learning for lithology classification. Specifically, we applied AdaBoost and LogitBoost meta-learners to decision stumps and random trees. In our work, amongst the applied meta-algorithms, we found that AdaBoost when combined with decision stumps, performed poorly as compared to other combinations. We found that by using random trees as base learners instead of decision stumps much better accuracy was achieved. LogitBoost in combination with random trees performed best in our study of both the DGF and HGF datasets. This evaluation was based on 5-fold and 10-fold SCV. Despite not tuning the hyperparameters, we found our results to be comparable to those reported by Xie et al. (2018). We also made sure that the dataset imbalance was taken into consideration while reporting the performance metrics. This will help reservoir engineers and/or geologists in making a more informed decision. Our future efforts would likely be devoted to expanding the scope of our study to simultaneously analyse well log data from different countries. We would also like to overcome class imbalance by studying the impact of sampling techniques such as under-sampling, over-sampling and their hybrid variants. Additionally, we would like to optimize the hyperparameters and evaluate if it leads to better results.

References

O.C. Akinyokun, P.A. Enikanselu, A.B. Adeyemo, A. Adesida, 2009, Well log interpretation model for the determination of lithology and fluid contents, Pac. J. Sci. Technol, 10, 507-517

I. I. Baskin, G. Marcou, D. Horvath, A. Varnek, 2017, Random Subspaces and Random Forest, Tutorials in Chemoinformatics, 263-269

V. A. Dev, S. Datta, N. G. Chemmangattuvalappil, M. R. Eden, 2017, Comparison of Tree Based Ensemble Machine Learning Methods for Prediction of Rate Constant of Diels-Alder Reaction, Computer Aided Chemical Engineering, 40, 997-1002

J. H. Friedman, T. Hastie, R. Tibshirani, 2000, Additive logistic regression: A statistical view of boosting, The Annals of Statistics, 28, 337-374

M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, 2009, The WEKA Data Mining Software: An Update, SIGKDD Explorations, 11, 1

L. Rokach, 2016, Decision forest: Twenty years of research, Information Fusion, 27, 111-125

T. Saito and M. Rehmsmeier, 2015, The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets, PLoS ONE, 10, 3, e011843

S.M. Salehi, B. Honarvar, 2014, Automatic identification of formation lithology from well log data: a machine learning approach, J. Petrol. Sci. Res., 3, 2, 73-82

M. Sebtosheikh, R. Motafakkerfard, M. Riahi, S. Moradi, N. Sabety, 2015, Support vector machine method, a new technique for lithology prediction in an Iranian heterogeneous carbonate reservoir using petrophysical well logs, Carbonates Evaporites, 30, 59-68

M. Sugiyama, 2016, Ensemble Learning, Introduction to Statistical Machine Learning, 343-354

Y. Xie, C. Zhu, W. Zhou, Z. Li, X. Liu, M. Tu, 2018, Evaluation of machine learning methods for formation lithology identification: A comparison of tuning processes and model performances, Journal of Petroleum Science and Engineering, 160, 182-193

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50240-8

Modelling and analysis of oil shale refinery process with the indirectly heated moving bed

Huairong Zhou [a], Shuai Zeng [a], Lei Zhang [a], Guangwen Xu [b], Yu Qian [a*]

[a] *School of Chemical Engineering, South China University of Technology, Guangzhou, 510640, PR China*
[b] *Shenyang University of Chemical Technology, Shenyang, 110142, PR China*
ceyuqian@scut.edu.cn

Abstract

China is rich of oil shale as one of alternative fossil energy resources. The oil shale processing in China is mainly based on Fushun type technology. However, low utilization rate of raw material, low oil yield, and high content of heavy components in the oil are the main problems. An indirectly heated moving bed is developed by the Institute of Process Engineering of Chinese Academy of Sciences to de-bottleneck Fushun type technology. The new technology has shown good potential in its pilot operation. The oil shale refinery process with the indirectly heated moving bed is modeled and simulated in this work. Based on the simulation, a techno-economic analysis is performed and compared with the conventional Fushun oil shale refinery process. Results show that: for a shale refinery of 3.0 Mt/y scale, (1) 375 t/h oil shale retorting requires 149.6 MW of heat, in which 60% of the heat is produced by combustion of the oil shale pyrolysis gas, while remaining 40% from the coal gasified gas. (2) In consideration of investment and operation, the cost of product shale oil of the new process is 2636 CNY/t, which is 12% lower than that of the conventional Fushun refinery process. This benefit comes from higher utilization of raw material and oil yield of the new process. (3) The competitive crude oil price of the indirectly heated moving bed refinery process is about 51 $/bbl, while that of the Fushun refinery process is 58 $/bbl. During 2017 year, the crude oil price fluctuates at around 50 $/bbl, thus the new indirectly heated moving bed refinery process shows better economic competitiveness.
Keywords: indirectly heated moving bed, shale refinery, modeling, analysis

1. Introduction

The shale oil-equivalent reserves in the world is 6902×10^8 tons, three times the world's crude oil reserves, as shown in Figure 1 (Zuckerman, 2014). In 2014, the shale oil production reached 2.1×10^8 tons, one-third of conventional petroleum production in the U.S. This has completely changed the pattem of international oil and gas supply, known as the shale revolution (Zuckerman, 2014). In China, the amount of oil shale reserves is 476×10^8 tons oil-equivalent, twice China's crude oil reserves (Mulchandani and Brandt, 2011). Chinese oil shale mining is mainly concerned with surface mining of shallow shale seams stored 100-400 below the ground. By 2014, China produced 1 million tons of shale oil through oil shale refinery (Qian et al., 2014). The capacity of shale oil is expected to exceed 3 million tons by 2020 (Li et al., 2015). The exploitation and utilization of oil shale are strategically important in alleviating the shortage of oil and gas resources in China.

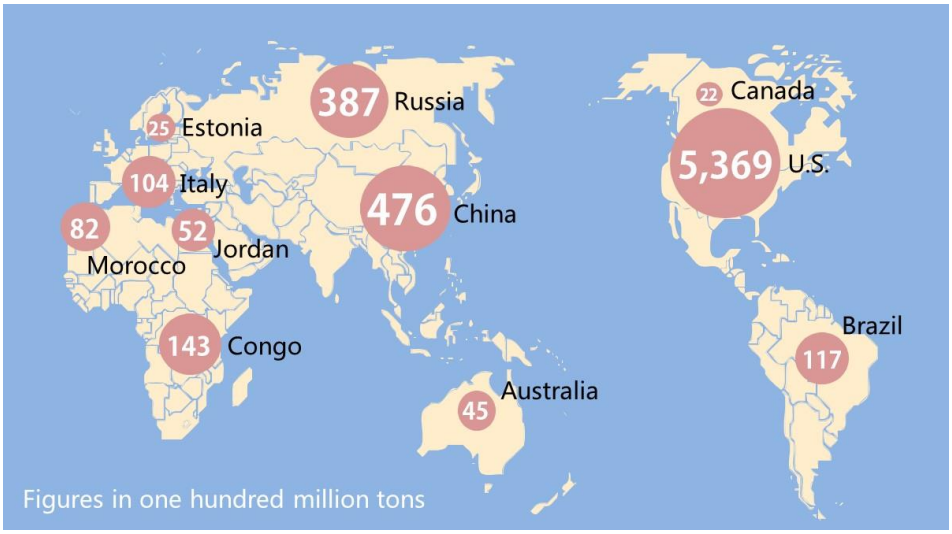
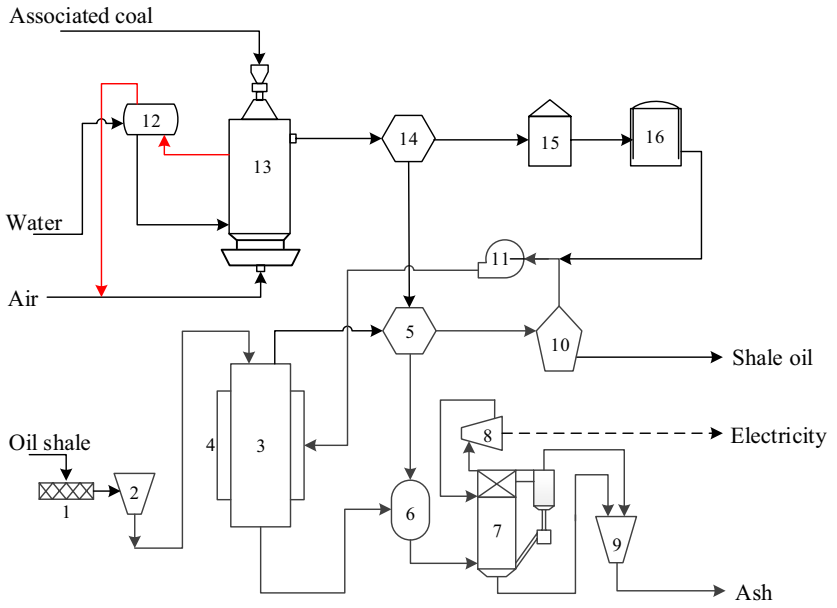


Figure 1. Shale oil-equivalent reserves in the world

There are two types of oil shale refinery technologies: in-situ and ex-situ (Wang et al., 2014). In China, only ex-situ represented by Fushun (FS) type technology has been applied (Qian et al., 2014). However, the Fushun technology suffers from low utilization of raw material (80%), low oil yield (65%), and low content of light components in the shale oil (20%) (Qian et al., 2014). The Chinese Academy of Science has developed an indirectly heated moving bed (IHMB) with the scale of 1000 tons/y oil shale (Lin et al., 2016). Compared with the Fushun technology, the utilization of raw material of this new technology is increased by 20%, the oil yield by 25%, and the content of light components in the oil by 46% (Lin et al., 2016).

The IHMB oil shale refinery process is shown in Figure 2. The process includes an oil shale pyrolysis unit, a coal gasification unit, an oil-gas separation unit, and a semi-coke for power generation unit. The oil shale is first milled and screened, and the shale particles below 10 mm in diameter enter the retort to pyrolysis. Part of the heat required for shale pyrolysis comes from the whole cycle of pyrolysis gas, while the remaining heat is provided by additional fuel gas. The syngas from associated coal gasification during the oil shale mining is used as fuel gas to complement the additional heat. The oil-gas mixture produced by pyrolysis is released from the top of the retort, and then fed into the cyclone separator to remove dust. After that, the oil-gas mixture enters the oil-gas separator to obtain retorting gas and shale oil. The retorting gas and coal gasification syngas are mixed and cycled to combust to supply heat for shale pyrolysis. In addition, the semi-coke produced by pyrolysis is discharged from the bottom of the retort. After drying, the semi-coke enters the circulating fluidized bed to burn to produce high-pressure steam, which is used to generate electricity through the steam turbine.



1 grinding mill; 2 shale hopper; 3 retort; 4 gas furnace; 5 flying ash separator; 6 mixture hopper; 7 CFB furnace; 8 steam turbine; 9 ash hopper; 10 oil-gas separator; 11 compressor; 12 steam pocket; 13 gasifier; 14 cyclone separator; 15 intercooler; 16 gas holder

Figure 2. Flow diagram of the IHMB refinery process

Followed in this paper, the key units of the IHMB refinery process are modeled and simulated in Section 2. Techno-economic analysis of the new process will be conducted in Section 3, and the results are compared with the conventional Fushun refinery process to point out the advantages and disadvantages of the new process.

2. Modelling and simulation of oil shale refinery process

In this work, the Aspen Plus (Version 7.2) process modeling software is used to simulate the IHMB refinery process, in which the oil shale pyrolysis unit is modeled and simulated in details. The processing scale of the new process is 375 t/h oil shale. The raw materials adopt Huadian oil shale and associated coal, corresponding proximate and elemental analyses are shown in Table 1 (Qian et al., 2014; Lin et al., 2016).

Table 1 Proximate and elemental analyses of oil shale and coal

	Proximate (wt. %, ar)				Elemental (wt. %, ar)				
	M	FC	V	A	C	H	O	N	S
Oil shale	3.43	1.77	22.83	71.97	50.16	6.37	38.12	1.99	3.35
Coal	18.42	15.64	32.21	33.73	61.42	4.93	15.30	0.86	0.29

Figure 3 shows the Aspen Plus simulation calculation procedure of the IHMB. According to the actual process, the retort is divided into a drying stage and a retorting stage. A RYield block is utilized to model the drying stage. Its temperature is 200 °C and pressure is 0.1 MPa. In addition, a RCSTR block is used for modeling of the

retorting stage. The temperature and pressure of the RCSTR reactor are set to 550 °C and 0.1 MPa.

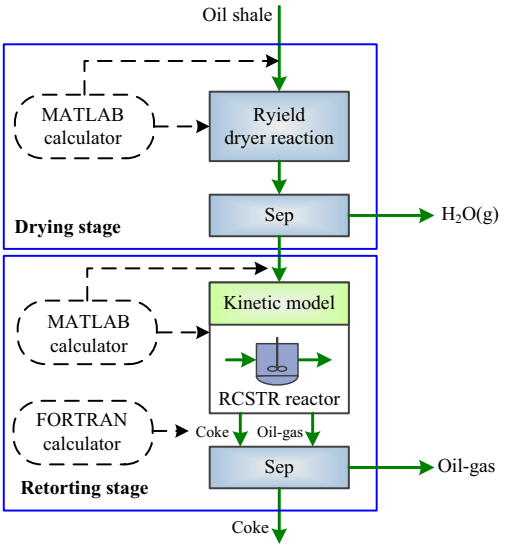


Figure 3. Aspen Plus simulation calculation procedure

The reaction in the drying stage is the removal of surface and internal moisture of the shale particles, while that in the retorting stage is kerogen pyrolysis, as follows:

$$\text{Kerogen} \rightarrow H_2 + H_2O + H_2S + NH_3 + CO + CO_2 + CH_4 + C_2H_4 + C_2H_6 + C_3H_8 + C_4H_{10}O + \text{Semi-coke} \quad (1)$$

The kinetic equation of the pyrolysis reaction is as follows (Syed et al., 2011):

$$\frac{\mathrm{d}x}{\mathrm{d}t} = k(1-x)^n \quad (2)$$

where x is the extent of conversion, t is the time, k is the specific rate constant, and n is the order of reaction. According to the Arrhenius equation, the reaction constant k is formulated as (Goldfarb et al., 2013):

$$k = A\mathrm{e}^{-Ea/RT} \quad (3)$$

where A is the apparent frequency factor, Ea is the apparent activation energy, and R is the gas constant. According to Qian et al. (2014), Ea is set to 158 KJ/mol, A to 2.03×10^9, and n to 1.0. The kinetic equation of the retorting reaction is as follows:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = k(1-x)^n = A\exp\left(-\frac{E_a}{RT}\right)(1-x)^n = 2.03\times10^9\,\mathrm{e}^{-19004/T}(1-x) \quad (4)$$

Based on the kinetic equation Eq. (4), a piece of external FORTRAN subroutines are programmed and utilized in Aspen Plus to calculate the pyrolysis yields of oil shale. The simulation data of yields of the oil shale pyrolysis products are shown in Table 2. The results are validated by comparing with industrial data (Lin et al., 2016).

Table 2 Comparison between simulation and industrial data of oil shale pyrolysis

Component	Mass fraction (%)	
	Simulation data	Industrial data
Semi-coke	81.44	81.46
CO	1.16	1.14
CO_2	3.82	3.78
CH_4	0.93	0.91
C_2H_4	0.07	0.06
C_2H_6	0.16	0.18
C_3H_6	0.11	0.13
C_3H_8	0.29	0.25
H_2	0.28	0.27
NH_3	0.08	0.07
H_2S	0.18	0.17
H_2O	3.36	3.40
$C_{14}H_{10}O$	8.12	8.18

3. Techno-economic analysis

3.1. Thermodynamic performance

In this paper, we take the retort as a benchmark to conduct thermal balance calculation. The heat carried by oil shale, Q_{shale} and the heat required for pyrolysis, Q_{ret} are equal to the sum of the shale pyrolysis reaction heat, Q_{pyr} and the heat carried by pyrolysis products. The heat balance equation is shown in equation (5):

$$Q_{\text{shale}} + Q_{\text{ret}} = Q_{\text{pyr}} + \sum\left(Q_{\text{oil}} + Q_{\text{gas}} + Q_{\text{semi-coke}} + Q_{\text{vap}}\right) \tag{5}$$

where Q_{oil}, Q_{gas}, $Q_{\text{semi-coke}}$, and Q_{vap} represent the heat carried by shale oil, retorting gas, semi-coke, and evaporated moisture, respectively. According to equation (5), the heat required for pyrolysis is 149.6 MW. The heat required for oil shale pyrolysis is provided by pyrolysis retorting gas and coal gasification syngas. The amount of coal gasification syngas is calculated by equation (6):

$$V_{\text{syn}} = Q_{\text{supl}} / \sum\left(\Delta h_{\text{L},k} \cdot vol_k\right) \tag{6}$$

where V_{syn} represents the volume flow rate of syngas, and vol_k represents the volume fraction of each component in syngas. After calculation, $V_{\text{syn}} = 5.3\times10^4$ Nm^3/h, and the amount of coal consumption is 16.9 t/h.

3.2. Economic analysis

The economic analysis is performed in terms of production cost (PC). PC mainly consists of raw material cost C_{r}, utilities cost C_{u}, operating and maintenance cost C_{om}, depreciation cost C_{d}, plant overhead cost C_{poc}, administrative cost C_{ac}, distribution and selling cost C_{dsc}, as shown in equation (7):

$$\text{PC} = C_{\text{r}} + C_{\text{u}} + C_{\text{om}} + C_{\text{d}} + C_{\text{poc}} + C_{\text{ac}} + C_{\text{dsc}} \tag{7}$$

The raw materials and utilities cost are calculated based on simulation results. The labor cost is referred to our previous work, which occupies 10% of the total PC. A straight-line method is adopted to perform the depreciation cost under a consumption of 20 years lifetime and 4% salvage value. Based on equation (7), production costs the IHMB and

Fushun (FS) refinery processes calculated and results show that: (1) their production cost of unit product shale oil are 2636 CNY/t and 3024 CNY/t. As high utilization of raw material and oil yield, the production cost of the IHMB refinery process is 12% lower comparing to the Fushun refinery process with the same processing scale. (2) The competitive crude oil price of the IHMB refinery process is 51 $/bbl, while that of the Fushun refinery process is 58 $/bbl. The IHMB refinery process shows better competitiveness, while the Fushun refinery process has no economic competitiveness in 2017 year with the crude oil price of 50 $/bbl. However, when the crude oil price reaches to a high level of 100 $/bbl, both the IHMB and Fushun refinery processes show more competitiveness.

4. Conclusions

In order to de-bottleneck the Fushun type technology, the Chinese Academy of Sciences developed an indirect heating moving bed (IHMB). The heat required for shale pyrolysis of the IHMB is partly provided by the whole cycled pyrolysis gas, while the remaining comes from gasification syngas from associated coal during oil shale mining. In this paper, the IHMB refinery process is modeled and analyzed, and the following conclusions are obtained. (1) Thermodynamic analysis shows that the heat required for 375 t/h shale pyrolysis is 149.6 MW, in which 89.6 MW is provided by pyrolysis gas and the rest 60 MW is supplied by coal gasification syngas. (2) Economic analysis shows that the production cost for product shale oil of the IHMB refinery process is 2636 CNY/t, which is 12% lower than that of the conventional Fushun refinery process. (3) The IHMB refinery process's competitive crude oil price is 51 $/bbl, which shows better economic advantages comparing with the Fushun refinery process's competitive oil price, 58 $/bbl.

Acknowledgments

The authors are grateful for the National Basic Research Program (2014CB744306) and financial support from the China NSF projects (21676101 and 21736004).

References

G. Zuckerman, 2013, The frackers: The outrageous inside story of the new billionaire wildcatters

H. Mulchandani, A.R. Brandt, 2011, Oil shale as an energy resource in a CO2 constrained: the concept of electricity production with in situ carbon capture, Energy Fuels, 25(4), 1633-1641.

Y. Qian, Q.C. Yang, J. Zhang, H.R. Zhou, S.Y. Yang, 2014, Development of an integrated oil shale retorting process with coal gasification for hydrogen production, Industrial & Engineering Chemistry Research, 53 (51), 19970-19978.

X.X. Li, H.R. Zhou, Y.J. Wang, Y. Qian, S.Y. Yang, 2015, Thermoeconomic analysis of oil shale process with gas or solid heat carrier, Energy, 113, 639-647.

S. Wang, X.M. Jiang, X.X. Han, J.H. Tong, 2014, Effect of retorting temperature on product yield and characteristics of non-condensable gases and shale oil obtained by retorting Huadian oil shales, Fuel Process. Technol., 121, 9-15.

L.X. Lin, D.G. Lai, E. Guo, C. Zhang, G.W. Xu, 2016, Oil shale pyrolysis in indirectly heated fixed bed with metallic plates of heating enhancement, Fuel, 163, 48-55.

S. Syed, R. Qudaih, I. Talab, I. Janajreh, 2011, Kinetics of pyrolysis and combustion of oil shale sample from thermo gravimetric data, Fuel, 90, 1631-1637.

J.L. Goldfarb, A. D'Amico, C. Culin, E.M. Suuberg, I. Kulaots, 2013, Oxidation kinetics of oil shale semi-cokes: reactivity as a function of pyrolysis temperature and shale origin, Energy & Fuels, 27, 666-672.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50241-X

Dimensionality reduction in feasibility analysis by latent variable modeling

Gabriele Bano[a], Zilong Wang[b], Pierantonio Facco[a], Fabrizio Bezzo[a], Massimiliano Barolo[a], Marianthi Ierapetritou[b*]

[a]*CAPE-Lab, Department of Industrial Engineering, University of Padova, Italy*
[b]*Rutgers University, 98 Brett Road, Piscataway, NJ, 08854, USA*
marianth@soemail.rutgers.edu

Abstract

We propose a systematic methodology to exploit partial least-squares (PLS) regression modeling to reduce the dimensionality of a feasibility analysis problem. PLS is used to project the original multidimensional space of input factors onto a lower dimensional latent space. We then apply a radial basis function (RBF) adaptive sampling feasibility analysis on this lower dimensional space to identify the feasible region of the process. A simple low-dimensional representation of the feasible region is thus obtained with this combined PLS-RBF approach. The performance of the methodology is tested on a mathematical example involving six inputs. We show the ability of this PLS-RBF approach to reduce the computational burden of the feasibility analysis while maintaining an accurate and robust identification of the feasible region.

Keywords: feasibility analysis, partial least-squares regression, radial basis function, adaptive sampling, model reduction

1. Introduction

The aim of feasibility analysis (Halemane and Grossman, 1983) is to determine the multivariate region of the input domain within which a process is considered to be feasible. When the original model is computationally expensive or the computation of model derivatives is difficult, surrogate-based approaches can be used (Boukovala *et al*., 2010). These methods build a surrogate as a computationally cheap and reliable approximation of the original model, and use this surrogate to identify the feasible region. Recently, Wang and Ierapetritou (2017) proposed a radial basis function (RBF) adaptive sampling approach that outperforms other surrogate-based approaches for low-dimensional test problems. Adaptive sampling is used to simultaneously explore regions of the original input domain that are close to the boundary of the feasible region and/or less explored regions, taking full advantage of the sampling budget.

The main drawback of feasibility analysis is that it suffers from the curse of dimensionality. The computational cost can increase very significantly when a large number of input factors is considered. Moreover, having a simple and compact graphical representation of the feasible region is impossible when more than three inputs are involved. In most common situations, not all the combinations of the original input factors have a strong effect on the model responses. A class of statistical models that can identify the dominant input combinations that mostly affect the model outputs are latent variable models (LVMs). In particular, partial least-squares (PLS) regression (Geladi and

Kowalski, 1986) is a multivariate latent variable (LV) technique that can be used to capture the variability of the input and output spaces by means of few meaningful variables, called LVs, that are usually much smaller in number than the original input factors. Therefore, PLS can be used to obtain a compact representation of the original multidimensional input space without losing significant information on the original input and output correlation structure.

The aim of this study is to overcome the curse of dimensionality problem of feasibility analysis by exploiting the ability of PLS to reduce the input space dimensionality. We use a PLS model to perform a linear transformation from the original multidimensional input space to a lower dimensional latent space. We then perform a RBF-based adaptive sampling feasibility analysis on this latent space and we assess the accuracy and robustness of the results with three appropriate metrics. We test the performance of the methodology on a 6-dimensional mathematical test problem, and we prove the ability of the methodology to give accurate and robust results by simultaneously reducing the computational burden for the feasibility analysis problem.

2. Methods

Let $\mathbf{y} = \mathbf{g}(\mathbf{d}, \mathbf{x})$ be the model of the process under consideration, where $\mathbf{y}$ is the vector collecting all the model outputs, $\mathbf{d}$ is the vector collecting the design variables (which are fixed since the process has already been designed) and $\mathbf{x}$ is the vector collecting all the input variables. Let $f_j(\mathbf{d}, \mathbf{x}),\ j = 1, \dots, J$ be a set of J inequality constraints for the process outputs $\mathbf{y}$. Two mathematical techniques are joint together in the proposed framework: PLS and RBF-based adaptive sampling feasibility analysis.
PLS (Geladi and Kowalski, 1986) is a data-driven technique that projects a historical dataset of inputs $\mathbf{X}$ $[I \times Q]$ and responses $\mathbf{Y}$ $[I \times M]$ onto a common latent space of A LVs, according to the following model structure:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E}_{\mathbf{X}} \tag{1}$$

$$\mathbf{Y} = \mathbf{T}\mathbf{Q}^{\mathrm{T}} + \mathbf{E}_{\mathbf{Y}} \tag{2}$$

$$\mathbf{T} = \mathbf{X}\mathbf{W} \tag{3}$$

where $\mathbf{T}$ $[I \times A]$ is the score matrix, $\mathbf{P}$ $[Q \times A]$ and $\mathbf{Q}$ $[M \times A]$ are the $\mathbf{X}$ and $\mathbf{Y}$ loading matrices, $\mathbf{E}_{\mathbf{X}}$ and $\mathbf{E}_{\mathbf{Y}}$ the residuals, $\mathbf{W}[Q \times A]$ is the weight matrix.

The number A of LVs can be chosen in such a way as to explain a significant fraction of the variability of both the input and the output data. If the input collinearity is strong and/or the model outputs are only affected by a reduced set of input combinations, a significant reduction of the problem dimensionality can be obtained (i.e. $A \ll Q$). Eqs. (1)-(3) can be interpreted as a linear transformation L_{PLS} between the original set of input factors $(x_1, x_2, \dots, x_Q)$ and the new set of A LVs, identified by the model scores $(t_1, t_2, \dots, t_A)$, according to Eq. (1).

Feasibility analysis (Halemane and Grossman, 1983) is a mathematical technique that can be used to find the set of input combinations within which the process under investigation is considered to be feasible. This is obtained by determining the set of input combinations

for which the maximum value of all the process constraints is less or equal to zero. The maximum value of the process constraints is defined as the feasibility function $\psi(\mathbf{d}, \mathbf{x})$:

$$\psi(\mathbf{d}, \mathbf{x}) = \max_{j \in J} f_j(\mathbf{d}, \mathbf{x}) \tag{4}$$

The set of input combinations that satisfy $\psi(\mathbf{d}, \mathbf{x}) \leq 0$ represent the feasible region of the process under consideration. These combinations can be found using a RBF-based adaptive sampling approach. Mathematical details on this approach can be found in Wang and Ierapetritou (2017).

2.1. Proposed methodology

The proposed methodology can be described as a 7-step procedure.

1. Step #1: collection of simulated data from the model. A dataset that collects the original input factors and the corresponding values of the feasibility function is obtained using the original process model. The input space is spanned using the Kennard-Stones's sampling algorithm (KSA; Kennard and Stone, 1969). This algorithm ensures a uniform mapping of the input data, with the advantage that samples can be sequentially increased until a satisfactory coverage is obtained (Daszykowski et al., 2002). For each selected sample $\mathbf{x}^*$, the corresponding value of the feasibility function (4) is computed using the process model. Repeating this procedure for all the selected samples gives a dataset $[\mathbf{X}; \boldsymbol{\psi}]$, where $\mathbf{X}$ is the matrix that collects the selected input combinations, and $\boldsymbol{\psi}$ is the vector that collects the corresponding values of the feasibility function.
2. Step #2: PLS model calibration. A PLS model (Eqs.(1)-(3)) is calibrated with the dataset generated in step #1. The PLS model is used to obtain a linear transformation L_{PLS} between the original set of input factors $(x_1, x_2, \dots, x_Q)$ and the new set of A LVs, identified by the PLS model scores $(t_1, t_2, \dots, t_A)$, according to Eq. (1). In most situations, 2 or 3 LV are used to describe the original multidimensional input space.
3. Step #3: PLS model diagnostics. The values of the cumulative **X**- and **y**-variability ($R^2_{X,cum}$ and $R^2_{y,cum}$) explained by the PLS model for the given number of LVs are computed. If the value of $R^2_{y,cum}$ is high (e.g., > 90%), the PLS model can be directly used to determine the feasible region of the process, avoiding the feasibility analysis step. If the value of $R^2_{y,cum}$ is low (< 70%) and $R^2_{X,cum}$ is sufficiently high (> 90%), the PLS-RBF approach can be used to identify the feasible region with the chosen number of LVs. If $R^2_{X,cum}$ is smaller than 90%, the user should consider the possibility to increase the number of LVs of the PLS model to increase the projection accuracy. Intermediate situations can be handled according to user's risk adversity.
4. Step #4: feasibility analysis. The RBF-based adaptive sampling feasibility analysis is performed using the reduced set of latent variables $[t_1, t_2, \dots, t_A]$, while maintaining the original structure of the process model.
5. Step #5: results validation on the latent space. The accuracy and the robustness of the results obtained in the feasibility step are assessed on the latent space using three metrics (Wang and Ierapetritou, 2017):
   $CF(\%)$: percentage of the projection of the feasible region on the latent space that has been correctly identified by the PLS-RBF feasibility analysis;

$CIF(\%)$: percentage of the projection of the infeasible region on the latent space that has been correctly discovered by the PLS-RBF feasibility analysis;
$NC(\%)$: percentage of the projection of the feasible region on the latent space that has been overestimated by the PLS-RBF feasibility analysis.
The closer the values of $CF(\%)$ and $CIF(\%)$ to 100, the higher the accuracy of the results on the latent space. The closer the value of $NC(\%)$ to 0, the greater the conservativeness of the PLS-RBF predictions.

6. Step #6: PLS model inversion. If the values of the above metrics on the latent space is satisfactory, the results on the latent space are expressed in terms of the original input factors $[x_1, x_2, \dots, x_Q]$ using the inverse of the linear transformation L_{PLS}, according to the model structure (1). Every feasible and infeasible point on the latent space, identified in terms of the PLS model scores $[t_1, t_2, \dots, t_A]$, is projected back onto its original representation $[x_1, x_2, \dots, x_Q]$.
7. Step #7: results validation on the original input space. The three metrics are computed with respect to the original input space. They represent the accuracy and conservativeness of the results with respect to the true multidimensional feasible region of the process, and should thus be considered as the final quantitative assessment of the quality of the results obtained. The closer the values of the first two metrics to 100, the greater the accuracy of the estimation of the true feasible region of the process. The closer the value of the third metric to 0, the greater the robustness of the results obtained.

The methodology presented above was implemented in MATLAB® 2015, and the simulations were performed on an InteL® Core™ I7-5600U CPU@2.60GHz processor with 16.0 GB RAM.

3. Results and discussion

We considered a case study of a process composed by two units in series and a total number of 6 input factors $[x_1, x_2, x_3, x_4, x_5, u_1]$, as schematically shown in Figure 1.

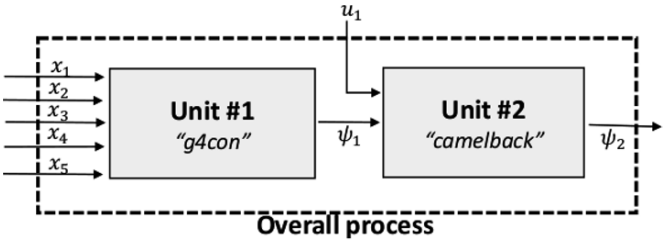


Figure 1. Block flow diagram of the two-unit case study considered.

The constraints for the first unit ("g4con") are described by the set of equations (Koziel and Michalewicz, 1999):

$$0 \leq 85.334407 + 0.0056858x_2x_5 + 0.0006262x_1x_4 - 0.0022053x_3x_5 \leq 92 \quad (5)$$

$$90 \leq 80.51249 + 0.0071317x_2x_5 + 0.0029955x_1x_2 + 0.0021813x_3^2 \leq 110 \quad (6)$$

$$20 \leq 9.300961 + 0.0047026x_3x_5 + 0.0012547x_1x_3 + 0.0019085x_3 \leq 25 \quad (7)$$

The maximum value of the constraints for the first unit is the feasibility function ψ_1. The constraint for second unit ("camelback") is described by:

$$\left(4 - 2.1\psi_1^2 + \frac{\psi_1^4}{3}\right)\psi_1^2 + \psi_1 u_1 + (-4 + 4u_1^2)u_1^2 \leq 0 \tag{8}$$

$$-3 \leq \psi_1 \leq 3;\ -2 \leq u_1 \leq 2 \tag{9}$$

The feasibility function for the second unit is ψ_2. Application of the RBF-based feasibility analysis on the original 6-dimensional problem was not computationally viable, due both to the high total computational time (within 24 h, with 64 initial points and 253 adaptive samples, the values of the three metrics were $CF = 54.1\ \%, CIF(\%) = 63.8\ \%, NC(\%)$=10.2%) and to the low accuracy of the results that could be obtained with a fixed sampling budget (accuracy less than 45% with a sampling budget of 198 (64 initial points +134 adaptive samples)). Therefore, we decided to apply the PLS-RBF methodology to overcome the curse of dimensionality. Following the step-by-step procedure, we first generated a dataset of 49 samples using the process model (5)-(10) and we calibrated a PLS model using this dataset (step #1 and step #2). The PLS diagnostics (step #3) is reported in Table 1. As can be noticed, 88.3% of the input variability is captured using two LVs only. We then used a PLS model with 2 LVs to obtain a linear transformation of the original 6-dimensional input space to a 2-dimensional latent space. The projection of the true feasible region of the process onto the latent space is represented by the grey-shaded area of Figure 2a. We applied the RBF-based adaptive sampling feasibility analysis to this latent space (step #4). We used 49 initial samples chosen according to a rectangular-grid sampling strategy to build the initial RBF surrogate, and a maximum number of 100 iterations for the adaptive sampling strategy. The results obtained are shown in Figure 2b. As can be noticed, after 100 iterations, most of the adaptive sampling points (circles in Figure 2b) are closely located near the boundary of the feasible region. The values of the three metrics $CF(\%), CIF(\%), NC(\%)$ were 95.47, 97.15 and 2.37 respectively, meaning that more than 95% and 97% of the feasible and infeasible regions (respectively) were correctly identified on the latent space (step #5). Since these values were satisfactory, we moved forward with the methodology by expressing the results in terms of the original 6-dimensional combinations of the input factors (step #6). The values of the three metrics $CF(\%), CIF(\%), NC(\%)$ on the original 6-dimensional space were 93.19, 96.85 and 2.89 respectively, meaning than more of 93% of the true 6-dimensional region was correctly identified, and more than 96% of the true 6-dimensional infeasible region. The total computational time for the simulation was 32 min. The results proved to be extremely satisfactory with respect to the original RBF approach.

Table 1. Diagnostics of the PLS model.

LV #	$R_X^2(\%)$	$R_{X,cum}^2(\%)$	R_y^2 (%)	$R_{y,cum}^2$ (%)
1	49.20	49.20	21.95	21.95
2	39.12	88.32	16.02	37.97
3	4.18	92.50	9.87	47.84
4	3.25	95.75	4.11	51.95
5	2.11	97.86	3.12	55.07
6	2.14	100.00	4.87	59.94

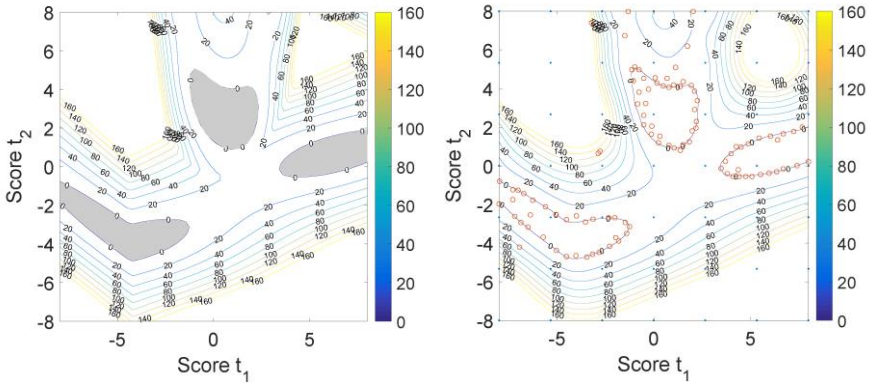


Figure 2. (a) Projection of the original function onto the 2-dimensional latent space. The grey-shaded area is the feasible region. (b) RBF model after 100 iterations. Dots represent the initial samples. Circles represent the adaptive sampling points after 100 iterations.

4. Conclusions

We proposed a methodology to identify the feasible region of a multiple-unit process by applying a combined PLS-RBF adaptive sampling feasibility analysis. We first used a PLS model to obtain a linear transformation between the original set of input variables and a reduced set of latent variables. We then applied RBF-based adaptive sampling feasibility analysis on the lower-dimensional latent space and we assessed the quality and robustness of the results with three metrics, that describe how well the feasible and infeasible regions are identified by the RBF model, and how large is the portion of overestimated feasible region. We tested the methodology on a mathematical test example involving two units and a total of 6 input factors. We proved the ability of the proposed approach to reduce the overall computational burden while maintaining a good accuracy and robustness of the results, with a correct estimation of the feasible region greater than 93%. In our future work, we will focus on the assessment of the performance of the methodology with real experimental data coupled with the original process model.

References

Boukouvala, F., Muzzio, F.J., Ierapetritou, M. G., 2010, Design space of pharmaceutical processes using data-driven-based methods. J. Pharm. Inn., 5, 119-137.

Daszykowski, M., Walczak, B., Massart D.L., 2002, Representative subset selection, Anal. Chim. Acta, 468, 91-103.

Geladi, P., Kowalski B.R., 1986, Partial Least-Squares Regression: A Tutorial. Anal. Chim. Acta, 185, 1–17.

Halemane, K. P., Grossmann, I. E., 1983, Optimal process design under uncertainty. AIChE J., 29, 425-433.

Kennard R.W, Stone L.A., 1969, Computer Aided Design of Experiments, Technometrics, 11, 137-148.

Koziel, S., Michalewicz, Z., 1999, Evolutionary algorithms, homomorphous mappings, and constrained parameter optimization. Evol. Comp., **7**, 19-44.

Wang, Z., Ierapetritou, M., 2017, A novel feasibility analysis method for black- box processes using a radial basis function adaptive sampling approach. AIChE J., 63, 532-550.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50242-1

Markov Chain MINLP Model for Reliability Optimization of System Design and Maintenance

Yixin Ye[a], Ignacio E. Grossmann[a*], Jose M. Pinto[b], Sivaraman Ramaswamy[b]

[a]*Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213*
[b]*Business and Supply Chain Optimization, Praxair, Inc., Danbury, CT 06810*
grossmann@cmu.edu

Abstract

The need for optimization tools for reliability design considering operational factors is motivated by the significance of availability of process systems and the lack of systematic and rigorous tools. In response to this need, this paper introduces a systematic approach to model the stochastic process of system failures and repairs as a continuous-time Markov chain, in which the impact of maintenance is incorporated in order to find the optimal selection of parallel units. An illustrative example is shown.

Keywords: Reliability, Availability, Maintenance, MINLP, Markov Chain

1. Introduction

Plant availability has been a critical consideration for chemical processes, as it represents the fraction of normal operating time, which directly impacts profitability. In practice, discrete-event simulation tools are used to examine the availability of a few selected designs of different redundancy levels under various maintenance policies (Sharda and Bury, 2008). However, the best plan selected through simulation is usually suboptimal because the list of candidates is not exhaustive. Thus, there is a strong motivation for systematic optimization tools for redundancy design considering operational factors.

Literature review of research in reliability engineering can be found in Ye et al. (2017) where a general mixed-integer framework is proposed to select standby units to maximize availability and minimize cost. However, for a more comprehensive optimization, it is also important to consider the impacts of operational factors such as maintenance. In particular, preventive maintenance (PM) is a major strategy to improve the availability of units (Ding and Kamaruddin, 2015). Pistikopoulos et al. (2001) and Goel et al. (2003) formulate an MILP model for the selection of units and production and maintenance planning for a fixed system configuration. Markov chain is a powerful mathematical tool being extensively used to capture the stochastic process of systems transitioning among different states. Bloch-Mercier (2002) models the deterioration process of a system as continuous-time Markov chain to optimize inspection intervals. Lin et al. (2012) model a simple utility system using Markov chain and carry out RAM (reliability, availability & maintainability) analysis iteratively to decide the optimal reliability design.

Given the aforementioned research gaps and knowledge basis, this work extends our recent mixed-integer framework Ye et al. (2017) by introducing a systematic approach to model the stochastic process of system failures and repairs as a continuous-time Markov chain, for which the impact of maintenance is incorporated in order to find the optimal selection of parallel units.

2. Problem statement

We define a general modelling framework for production systems with underlying serial structures for availability evaluation as shown in Figure 1. For each stage k, a set of potential parallel units J_k are available for selection at the design phase. The goal is to determine which one or several potential parallel units to install, as well as the length of inspection intervals t_k^i, in order to maximize the system availability (i.e. probability that the system performs without failures), while minimizing the total cost of the system.

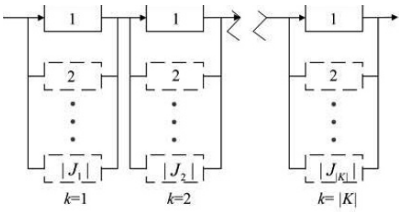


Figure 1 - A serial system

Each stage k has a set of potential units $j \in J_k$, for which the following is given:

- Availability parameters, i.e. Failure rate $\lambda_{k,j}$ and repair rate $\mu_{k,j}$.
- Operating priority within stage k (indicated by the order of j). A unit becomes active if and only if all the selected units with higher priorities have failed.
- Cost rates, including installation, inspection, maintenance and repair.

For each stage k, inspections are scheduled for active equipment at a certain time period to be determined, t_k^i, called inspection interval. If the inspection indicates that the equipment has a deterioration, a predictive maintenance task will be carried out in time. In that case, there will be enough time to order the spare parts and hence reduce the shipping costs, with which a maintenance task is cheaper than the repair upon failure. A deterioration can be detected by scheduled inspections in a certain period before it happens, called delay time (Christer, 1999), or PF-interval (Moubray, 1997) of length T_k^d.

Based on the parameters provided above, the availability of stage k will depend on the selection of parallel units $y_{k,j}$ and the inspection intervals of each processing stage t_k^i.

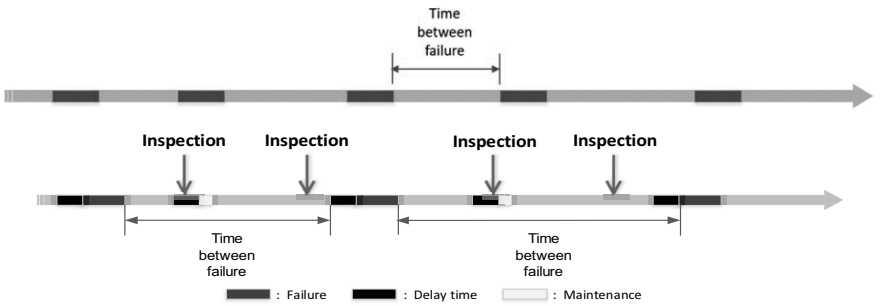


Figure 2 – Timeline of inspection/maintenance/repair

3. Modeling

3.1. Availability calculation

A continuous-time Markov chain is characterized by its transition rate matrix, the "**Q** Matrix", which can be used to solve for the stationary probability vector $\boldsymbol{\pi}$ through the linear equations, $\boldsymbol{\pi}^{\mathrm{T}}\mathbf{Q} = \mathbf{0}$ and $\boldsymbol{\pi}^{\mathrm{T}}\mathbf{1} = 1$. For the failure-repair system that we are considering, the availability can be obtained by adding up the probabilities of the non-failure states. Next, we show how to construct the "**Q** Matrix" of the system, and model the impact of the selection of units and inspection intervals on the "**Q** Matrix", and therefore on $\boldsymbol{\pi}$ and system availability.

First, for single stages, as we are making selections among the potential parallel units within stage *k*, there are several exclusive designs indexed by *h*, and each of them has a corresponding "**Q** Matrix" constructed as follows:

- A state space $T_{k,h}$ is enumerated where the transition from state *i* to state *j* ($i, j \in T_{k,h}$) is due to the happening of a failure or a repair. For example, the design decision of installing two units has the following 4 states:

Table 1 - State space example

Unit 1	Active	Active	Being repaired	Being repaired
Unit 2	Standby	Being repaired	Active	Being repaired

- Then the **Q** Matrix of design h is a 4×4 matrix where the element in row *i* and column *j* is equal to the failure rate or repair rate of the transition from state *i* to state *j*.
- And the element on the diagonal is equal to the opposite number of the sum of the other elements in the same row.

After the **Q** Matrices for the designs in stage *k* are formulated, they are put together to form a block-diagonal matrix $\mathbf{QM}_k$ called the "pseudo **Q** matrix" of stage *k,* which include all possible states for stage *k*. Following from that, the "pseudo **Q** matrix" of the system, $\mathbf{WM}_k$ is calculated using the following formula, where $\otimes$ is the Kronecker product and satisfies the associative law of addition.

$$WM = I_{n_{|K|}n_{|K|-1}\cdots n_2} \otimes QM_1 + I_{n_{|K|}n_{|K|-1}\cdots n_3} \otimes QM_2 \otimes I_{n_1} + I_{n_{|K|}n_{|K|-1}\ldots n_4} \otimes QM_3 \otimes I_{n_2 n_1} + \cdots$$
$$+ I_{n_{|K|}} \otimes QM_{|K|-1} \otimes I_{n_{|K|-2}n_{|K|-3}\cdots n_1} + QM_{|K|} \otimes I_{n_{|K|-1}n_{|K|-2}\cdots n_1}$$

The following constraints are used to manipulate the elements in $\mathbf{WM}_k$. The values of binary variables $y_{k,j}$ decide which columns and rows are eliminated from $\mathbf{WM}_k$ to find the actual $\mathbf{Q}$ matrix through equations (1)-(11). The variables t_k^i will affect the failure rates of single units, and therefore the values of the elements in $\mathbf{WM}_k$ ((12)-(18)).

$$instCost = \sum_{k\in K}\sum_{j\in J_k} y_{k,j} c\_inst_{k,j} \tag{1}$$

$$\sum_{j\in J_k} y_{k,j} \geq MIN_K, \qquad \forall k \in K \tag{2}$$

$$Z_{k,h} \Leftrightarrow \bigwedge_{\alpha_{j,k,h}=1} Y_{k,j} \bigwedge_{\alpha_{j,k,h}=0} \neg Y_{k,j}, \qquad \forall k \in K, h \in H_k \tag{3}$$

$$\bigvee_{h\in H_k} Z_{k,h}, \qquad \forall k \in K \tag{4}$$

$$ZZ_{k,s} \Leftrightarrow Z_{k,h}, \qquad \forall k \in K \tag{5}$$

$$\bar{Z}_{\bar{h}} \Leftrightarrow \bigwedge_{k\in K, h\in H_k, |h|=hc_{k,\bar{h}}} Z_{k,h}, \qquad \forall \bar{h} \in \bar{H} \tag{6}$$

$$\overline{ZZ}_{\bar{s}} \Leftrightarrow \bar{Z}_{\bar{h}}, \qquad \forall \bar{h} \in \bar{H}, \bar{s} \in \bar{T}_{\bar{h}} \tag{7}$$

$$\begin{bmatrix} \overline{ZZ}_{\bar{s}} \\ 0 \leq \pi_{\bar{s}} \leq 1 \end{bmatrix} \vee \begin{bmatrix} \neg\overline{ZZ}_{\bar{s}} \\ \pi_{\bar{s}} = 0 \end{bmatrix}, \qquad \forall \bar{s} \in \bar{S} \tag{8}$$

$$\left[\sum_{\bar{s}} \pi_{\bar{s}} WM(\bar{s},\bar{r}) = 0\right] \vee \left[\sum_{\bar{s}} \pi_{\bar{s}} WM(\bar{s},\bar{r}) < \infty\right], \qquad \forall \bar{r} \in \bar{R} \tag{9}$$

$$\sum_{\bar{s}} \pi_{\bar{s}} = 1 \tag{10}$$

$$A = 1 - \sum_{\bar{s}\in\bar{S}^f} \pi_{\bar{s}} \tag{11}$$

$$\lambda_{k,j}^0 - \lambda_{k,j} = \frac{e^{-\lambda_{k,j}^0 t_k^i} - e^{-\lambda_{k,j}^0 \left(t_k^i + T_k^d\right)}}{t_k^i}, \qquad \forall k \in K, j \in J_k \tag{12}$$

$$\sum_{l\in L} x_{k,l} = 1, \qquad \forall k \in K \tag{13}$$

$$\sum_{l\in L} x_{k,l} T_l^i = t_k^i, \qquad \forall k \in K \tag{14}$$

$$inspCost \geq \sum_{k\in K} \frac{T}{t_k^i} c\_insp_k \tag{15}$$

$$repaCost \geq T \sum_{\bar{s}\in\bar{S}^f} -WM(\bar{s},\bar{s})\pi_{\bar{s}} \sum_{k\in K_{\bar{s}}^f} c\_repa_k \tag{15}$$

$$prevTimes_k \geq \frac{y_{k,j}\left(\lambda_{k,j}^0 - \lambda_{k,j}\right)}{\lambda_{k,j}}, \qquad \forall j \in J_k \tag{16}$$

$$prevCost \geq T \sum_{\bar{s}\in\bar{S}^f} -WM(\bar{s},\bar{s})\pi_{\bar{s}} \sum_{k\in K_{\bar{s}}^f} prevTimes_k c\_prev_k \tag{17}$$

$$prevTime \geq T \sum_{\bar{s}\in\bar{S}^f} -WM(\bar{s},\bar{s})\pi_{\bar{s}} \sum_{k\in K_{\bar{s}}^f} prevTimes_k T\_prev_k \tag{18}$$

$$A^{net} = A - \frac{prevTime}{T} \tag{19}$$

3.2. Economic dependence

The net profit is the sum of the revenue, penalty and bonus minus the sum of costs. The total revenue is considered proportional to the availability of the system. Generally, in the contract between the plant and the customer, two reference bounds are set for the availability of the plant. If the actual availability of the plant does not meet the lower bound, the plant will be charged proportional to the difference. Conversely, if the actual availability exceeds the upper bound, the customer reward the plant with a bonus that is also proportional to the difference.

$$\max NP = RV - PN + BN - instCost - repaCost - inspCost - prevCost \quad (20)$$

$$RV = rvA^{net} \quad (21)$$

$$W_1 \veebar W_2 \veebar W_3 \quad (22)$$

$$\begin{bmatrix} W_1 \\ A^{net} \le A\_lo \\ PN = (A\_lo - A^{net})pn \\ BN = 0 \end{bmatrix} \vee \begin{bmatrix} W_2 \\ A\_lo \le A^{net} \le A\_up \\ PN = 0 \\ BN = 0 \end{bmatrix} \vee \begin{bmatrix} W_1 \\ A^{net} \ge A\_up \\ PN = 0 \\ BN = (A^{net} - A\_up)bn \end{bmatrix} \quad (23)$$

4. Example

In this section, a small example is shown featuring the system in Figure 3 with 2 serial stages, where stage 1 has 3 potential units and stage 2 has 2 potential units. Both of the two stages require at least 1 unit to function properly. Table 2 displays the parameters for the case study. A 10-year time horizon is considered.

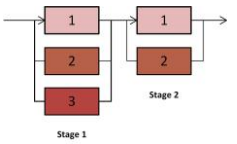


Figure 3 - Two stage-two unit system

Table 2 - Parameters

Stage	Unit	MTBF (day)	MTTR (day)	Fixed cost (k$)	Repair cost rate (k$ per time)	Inspection cost rate (k$ per time)	Maintenance cost rate (k$ per time)	Maintenance time (day)	Delay time (day)
1	1	50	7	200	12	0.1	0.6	1	10
	2	45.5	7.7	160					
	3	41.7	8.3	120					
2	1	66.7	2.6	240	10	0.1	0.5	1	12
	2	50	2.8	200					

Revenue rate (k$)	Penalty rate (k$)	Bonus rate (k$)	Availability lower bound	Availability upper bound
700	1000	10000	0.988	0.998

The MINLP model without maintenance is solved with BARON in 91.26 CPUs ((1)-(9) and (17) -(19)). There are 44921 equations and 44737 variables with 40 binary variables. The optimal design is to have all potential units installed. The expected system availability is 0.999, and the net profit is $6039.6 k, with a revenue of $6990.1 k and a bonus of $5.83 k. $920 k is spent on unit investment, and $36.3 k is spent on repair. The entire MINLP model with maintenance ((1)-(19)) is solved with BARON in 14924.15 CPUs, which has 44938 equations, 44458 variables with 49 binary variables. The optimal design is to install the last two units for stage 1 and both two units for stage 2. The expected system availability is 0.994. The expected net profit is $6131.6 k, with a revenue of $6959.9 k

and no bonus or penalty. \$720 k is spent on unit investment, \$52.1 k on inspections, \$2.7 k on maintenance, and \$53.4 k on repair. Other results are shown in Table 3.

Table 3 - Optimization results

Stage	Inspection interval (day)	Equivalent MTBF (day)
1	14	Unit2: 98.0, Unit 3: 86.2
2	14	Unit1: 181.8, Unit 2: 117.6

From the above results we can see that when maintenance is considered, the model suggest additional costs on inspection and maintenance, while less is to spend on the unit investment, which leads to an overall higher net profit.

5. Conclusion

In this paper, we have proposed a general modeling framework to represent the failure-repair process of a multi-unit system as a continuous-time Markov Chain, and to incorporate the design decision of selecting one or more standby units, as well as the operation decision of determining inspection intervals. The resulting MINLP model was implemented and solved for an illustrative example.

6. References

E. Balas,. (1985). Disjunctive programming and a hierarchy of relaxations for discrete optimization problems. SIAM Journal on Algebraic Discrete Methods, 6(3):466-486.

S. Bloch-Mercier, (2002). A preventive maintenance policy with sequential checking procedure for a Markov deteriorating system. European Journal of Operational Research, 142(3):548-576.

A. Christer, (1999). Developments in delay time analysis for modelling plant maintenance. Journal of the Operational Research Society, pages 1120-1137.

S. H. Ding and S. Kamaruddin, (2015). Maintenance policy optimization literature review and directions. The International Journal of Advanced Manufacturing Technology, 76(5-8):1263-1283.

H. D. Goel, J. Grievink, and M. P. Weijnen, (2003). Integrated optimal reliable design, production, and maintenance planning for multipurpose process plants. Computers & Chemical Engineering, 27(11):1543-1555.

Z. Lin, Z. Zheng, R. Smith, and Q. Yin, (2012). Reliability issues in the design and optimization of process utility systems. Theoretical Foundations of Chemical Engineering, 46(6):747-754.

J. Moubray, (1997). Reliability-centered maintenance. New York: Industrial Press, 2nd ed edition. "RCM II"-Cover.

M. F. Neuts, (1981). Matrix-geometric solutions in stochastic models: an algorithmic approach. Courier Corporation.

E. N. Pistikopoulos, C.G. Vassiliadis, J. Arvela, and L. G. Papageorgiou, (2001). Interactions of maintenance and production planning for multipurpose process plants a system effectiveness approach. Industrial & Engineering Chemistry Research, 40(14):3195-3207.

B. Sericola, (2013). Markov Chains: Theory and Applications. John Wiley & Sons.

B. Sharda and S. J. Bury, (2008). A discrete event simulation model for reliability modeling of a chemical plant. In Proceedings of the 40th Conference on Winter Simulation, pages 1736-1740. Winter Simulation Conference.

Y. Ye, I. E. Grossmann, and J. M.Pinto, (2017). Mixed-integer nonlinear programming models for optimal design of reliable chemical plants. To appear in Computers & Chemical Engineering.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50243-3

From property uncertainties to process simulation uncertainties – Monte Carlo methods in SimSci PRO/II process simulator

Jérôme Frutiger[a], Mark Jones[a], Nevin Gerek Ince[b], Gürkan Sin[a*]

[a]*Department of Chemical and Biochemical Engineering, Technical University of Denmark (DTU), Søltofts Plads, 2800 Lyngby, Denmark*
[b]*SimSci by Schneider Electric, Houston 77001, TX, United States*
gsi@kt.dtu.dk

Abstract

This study presents a methodology to apply Monte Carlo methods for property uncertainty propagation in the process simulation software *SimSci PRO/II*. The aim of this work is to integrate advanced uncertainty and sensitivity analysis tools into commercial process simulators. The property uncertainty and sensitivity analysis tools were applied to a heat pump system with cyclopentane as a working fluid. Monte Carlo sampling technique was used to generate property samples of the SRK equation of state parameters critical temperature, critical pressure and acentric factor. The samples were subsequently evaluated in the heat pump flowsheet built in *SimSci PRO/II*. This allowed describing the process model output uncertainty in a distribution and with the 95% confidence interval. Furthermore, Monte Carlo based standard regression could be used to analyse the sensitivity of the respective fluid properties. The results showed that property uncertainty propagation strongly depends on the correlation between the property parameters. The sensitivity analysis showed that the acentric factor is the most sensitive SRK parameter. This works demonstrates that Monte Carlo methods are a simple and useful tool, which can be used in commercial process simulators by industrial users.

Keywords: Monte Carlo method, process simulator, uncertainty

1. Introduction

Process design and simulation heavily rely on property data. Property data is subject to varying uncertainties depending on the experimental data or property models used. The propagation of property uncertainties through the process simulation on to the process model output gives important information on the reliability and sensitivity of the property model and influences the design decisions to be taken (Sin et al., 2009).

There have been significant efforts on the quantification of uncertainties in experimental (Dong et al., 2005) and predicted property data (Frutiger et al., 2016). Furthermore, a variety of studies have analysed and emphasized the importance of property uncertainties of the uncertainties in physical-property models on process design over the past decades (Macchietto et al., 1986; Whiting, 1996; Hajipour and Satyro, 2011; Frutiger et al., 2017). However, property uncertainty analysis is still not applied in chemical engineering industrial practice. One possible reason is that methods suggested by researchers are currently not available in commercial process simulators, such as

SimSci PRO\II or Aspen Plus. This paper aims at showing a methodology that integrates property uncertainty analysis and commercial process simulators.

Recently Mathias et al. (2014; 2017) performed several studies performing property uncertainty analysis in Aspen Plus software. The authors quantified the property uncertainty propagation of mixtures through different unit operations using Margules-based perturbation scheme for the activity coefficients. They propagated the uncertainties through process flowsheets in Aspen Plus. However, Mathias and co-workers did not take into account correlations between the input property parameters. Considering large property databases such as NIST TDE (Diky et al., 2013; Frenkel et al., 2005) or AIChE DIPPR database (Rowley et al., 2014), the property parameters of one compound are not completely independent. If correlation is neglected the uncertainty propagation can lead to an overestimation of the output uncertainty (Helton and Davis, 2003). Furthermore, Mathias et al. did not analyse the global sensitivity of the properties with respect to the model output.

Monte Carlo methods have been used in a variety of studies (Sin et al., 2009). Recently it has also been used for property uncertainty analysis (Hajipour et al., 2014; Frutiger et al., 2017). However, none of the authors implemented Monte Carlo method in a commercial process simulator.

In the current study we will present a simple methodology based on Monte Carlo sampling and evaluation techniques in order to propagate and analyse property uncertainties through the commercial process simulation software SimSci PRO\II (PRO/II, 2017). The methodology is applied to a case study of a heat pump system for the recovery of heat in food industries. Monte Carlo uncertainty propagation is used to analyse the output uncertainty and Monte Carlo standard regression is applied to study the global sensitivity of the properties.

2. Monte Carlo based strategy to address property uncertainties in SimSci PRO/II process simulator

The strategy to apply Monte Carlo method in *SimSci PRO/II* is based on the methodology of Frutiger et al. (2017) for property uncertainty propagation in process models:

1) A process model is built in the commercial process simulations software *SimSci PRO/II*. Process constants and variables are specified. The property models and parameters for the uncertainty analysis need to be selected.

2) The property uncertainty data is collected from databases (e.g. NIST TDE (Frenkel et al., 2005), AIChE DIPPR (Rowley et al., 2014)) or literature studies.

3) Monte Carlo sampling technique is used to sample property values within its corresponding uncertainty range i.e. 95%-confidence interval using the software MATLAB (2015b). Latin Hypercube Sampling method can be utilized for probabilistic sampling of the fluid property input space of a compound. The probability of uncertainty is assumed to follow normal distribution A total of 100-250 samples is obtained, each sample containing one value for each input parameter. The rank-based method for correlation control of Iman and Conover (1982) allows taking into account correlations between the property parameters. This is necessary, when parameters are not completely independent, as often the case for property models.

4) The Monte Carlo samples are evaluated in the process model in SimSci PRO/II. To this extend the PRO/II COM Server can be used, which allows to read and write access to property information in PRO/II through Python software (Python, 2017).

5) Uncertainty analysis: The process model output uncertainty is quantified and analyzed (Frutiger et al., 2017). The Monte Carlo results provide a distribution function for desired process model output of PRO/II. This can be analysed using mean and percentile calculations. Hence, the 95%-confidence interval of the PRO/II output with respect to the corresponding input property values can be obtained.

6) Sensitivity analysis: By constructing linear regression models on the outputs obtained from the Monte Carlo procedure (Sin et al., 2009), it is possible to obtain the standardized regression coefficient (SRC) β_j of parameter set θ_j given by

$$\frac{W_{NET} - \mu_{WNET}}{\sigma_{WNET}} = \sum_{j=1}^{M} \beta_j \cdot \frac{\theta_j - \mu_{\theta j}}{\sigma_{\theta j}} + \varepsilon_j \tag{1}$$

W_{NET} is the simulator output, ε_j is the error of the regression model, $\mu_{\theta j}/\sigma_{\theta j}$ and μ_{NET}/σ_{NET} are the corresponding mean and standard deviations of the parameters and the model output. The SRC is a measure of how strongly the parameter contributes to the model output (Frutiger, 2016).

3. Case study

The methodology for Monte Carlo methods for property uncertainty analysis is demonstrated through a case study for a heat pump waste heat recovery. The heat pump system is used to recover heat from exhaust gas from spray drying in food industries (heat source) to preheat the air before the spray dryer (heat sink), utilizing secondary cycles with pressurized water (Zühlsdorf et al., 2017). Cyclopentane is selected as a working fluid. The process model output is the coefficient of performance (COP), which is the enthalpy provided by the heat source divided by the compressor work: $COP=\Delta H_{Source}/W_{Compressor}$ The system has been implemented in PRO/II (see Figure 1). The mass flows and specifications can be found in the work of Zühlsdorf et al. (2017).

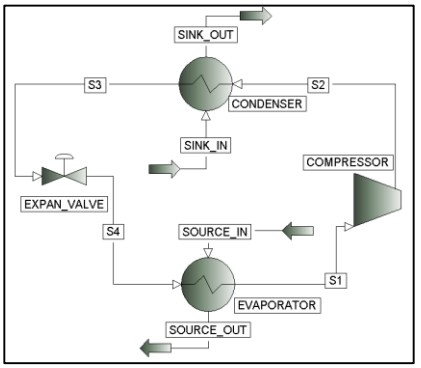


Figure 1: Pro/II flowsheet of a heat pump.

Table 1: Property parameter correlation

Correlation matrix			
	T_c	P_c	ω
T_c	1.00		
P_c	-0.89	1.00	
ω	0.80	-0.48	1.00

As a property model for the system Soave-Redlich-Kwong equation of state (SRK) is selected. The SRK parameters are the critical temperature T_c, the critical pressure P_c and the acentric factor ω. The goal of the case study is to propagate the uncertainty of T_c, P_c and ω through the SRK property model and process model in PRO/II following the methodology described in section 2. The uncertainty range and the correlation matrix of the SRK parameters are collected from the work of Frutiger et al. (2017), who obtained the correlation matrix using bootstrap method (Efron, 1979) (Table 1).

Furthermore, two different sampling strategies are applied. 1) Monte Carlo sampling without correlation, and 2) Monte Carlo sampling with correlation control according to Iman and Conover (1982). This allows comparing the influence of property parameter correlation on the process model output uncertainty. Figure 2 visualizes the two sampling procedures. Here, only the fluid properties of the working fluid cyclopentane are sampled and propagated. Property data of water stay constant.

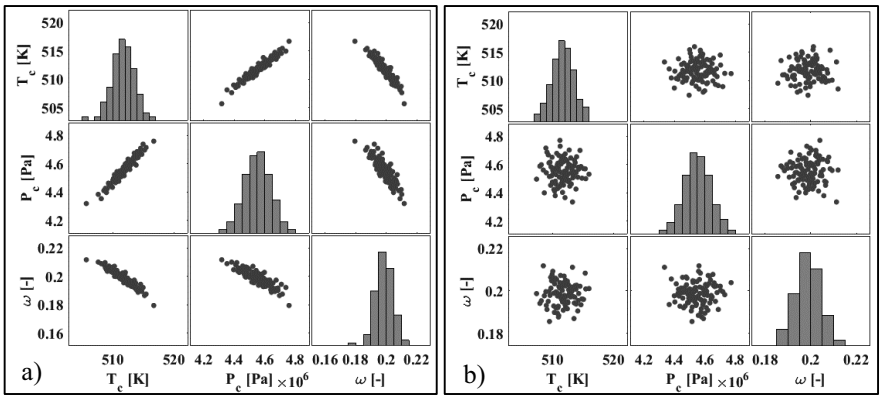


Figure 2: Illustration of samples generated by Latin Hypercube Sampling method with Iman and Conover correlation control (subplot a) and without correlation control (b).

4. Results and discussion

4.1. Uncertainty analysis

The Monte Carlo simulation results can be illustrated by distribution plots of the COP for both sample sets (see Figure 3). The larger the width of the COP output, the higher the output uncertainty. From the output distributions the 2.5% and 97.5% percentile defining the lower and upper bound of the uncertainty range (95%-confidence interval) for the process simulator output (here the COP) can be obtained (Table 2).

Table 2: Lower and upper bound of 95% confidence interval for COP for the correlated property samples (a) and the uncorrelated property samples

	COP mean values [-]	COP 95% confidence itnerval values [-]	
		Lower bound	Upper bound
Correlated samples	2.87	2.41	2.94
Uncorrelated samples	2.83	2.10	2.94

As it can be seen in Figure 3 and Table 2, the uncertainty range of the COP is larger for the results based on uncorrelated Monte Carlo sampling. In particular the lower bound of the uncertainty range is different. The uncorrelated samples also give COP values below 2.4 compared to the results of the correlated samples. This shows that, correlation between the property parameter samples can have a strong influence on the Monte Carlo uncertainty propagation procedures. If no or a wrong correlation is imposed on the property parameter samples, the process output uncertainty can be overestimated. This implies that parameter correlation should be considered carefully in the context of property uncertainty propagation through process simulators.

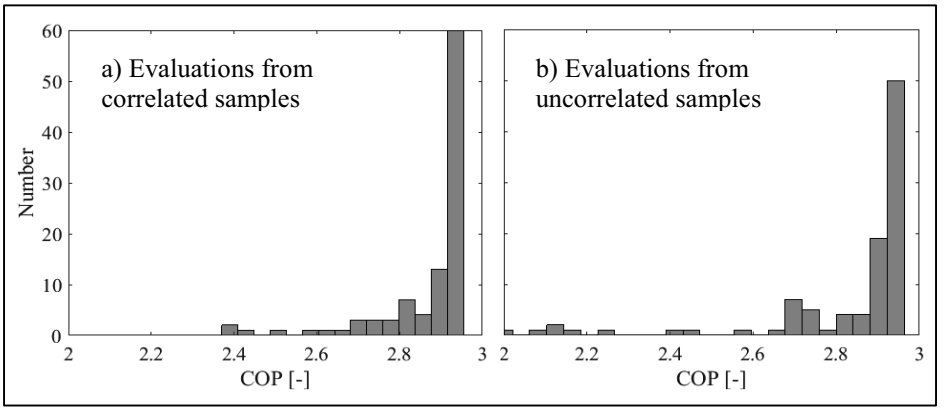


Figure 3: COP output distribution from the evaluation of the correlated property samples (a) and the uncorrelated property samples (b).

4.2. Sensitivity analysis

The regression coefficients – SRCs – of the fluid parameters were obtained from linear least square regression using Eq. 1 (see Figure 4). Correlated samples were used.

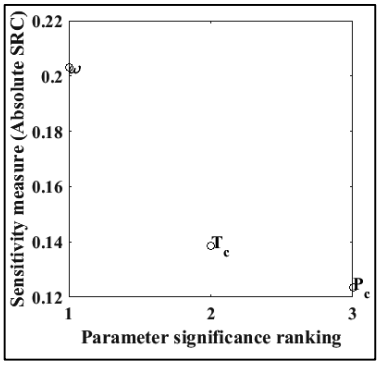


Figure 4: Standardized Regression Coefficients (SRCs) and ranking of SRK equations of state property parameters (using correlated samples).

The SRCs were ranked according to their absolute value. According to the analysis the acentric factor ω has by far the highest influence compared to the other parameters. This is in agreement with previous studies (Frutiger et al., 2016). Using a linear model between the inputs and outputs may mask some of the important properties. However, this can be analyzed through the Pearson correlation coefficient R^2 (>0.7), which shows to what

extend the model is linearizable. In case that it is not linearizable other statistical methods (such as polynomial chaos expansion need to be applied (Sin et al., 2009).

5. Conclusions

The study presented a Monte Carlo based methodology for property uncertainty and sensitivity analysis in a commercial process simulator. A heat pump system with cyclopentane as a working fluid was analyzed. The uncertainties of critical temperature, critical pressure and acentric factor were propagated through the SRK equations of state and the process model implemented in Pro\II. The results showed that considering correlation control in the sampling procedure provides influences the model output uncertainty. The most sensitive property parameter was the acentric factor. The study also showed that it is possible to use Monte Carlo techniques with commercial process simulators, which is currently not state-of-the-art in industrial practice.

References

Diky, V., Chirico, R.D., Muzny, C.D., Kazakov, A.F., Kroenlein, K., Magee, J.W., Abdulagatov, I., Won Kang, J., Gani, R., Frenkel, M., 2013. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. 8. Properties of Material Streams and Solvent Design. Chem. Inf. Model. 53, 249–266.

Dong, Q., Chirico, R.D., Yan, X., Hong, X., Frenkel, M., 2005. Uncertainty Reporting for Experimental Thermodynamic Properties. J. Chem. Eng. Data 50, 546–550.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. Ann. Stat. 7, 1–26.

Frenkel, M., Chirico, R.D., Diky, V., Yan, X., Dong, Q., Muzny, C., 2005. ThermoData Engine (TDE): Software Implementation of the Dynamic Data Evaluation Concept. J. Chem. Inf. Model. 45, 816–838.

Frutiger, J., Bell, I., O'Connell, J.P., Kroenlein, K., Abildskov, J., Sin, G., 2017. Uncertainty assessment of equations of state with application to an organic Rankine cycle. Mol. Phys. 115, 1225–1244.

Frutiger, J., Marcarie, C., Abildskov, J., Sin, G., 2016. A comprehensive methodology for development, parameter estimation, and uncertainty analysis of group contribution based property models – an application to heat of combustion. J. Chem. Eng. Data 61, 602–613.

Hajipour, S., Satyro, M.A., 2011. Uncertainty analysis applied to thermodynamic models and process design – 1. Pure components. Fluid Phase Equilib. 307, 78–94.

Hajipour, S., Satyro, M.A., Foley, M.W., 2014. Uncertainty analysis applied to thermodynamic models and fuel properties - Natural gas dew points and gasoline reid vapor pressures. Energy and Fuels 28, 1569–1578.

Helton, J.C., Davis, F.J., 2003. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliab. Eng. Syst. Saf. 81, 23–69.

Iman, R., Conover, W., 1982. A distribution-free approach to inducing rank correlation among input variables. Commun. Stat. Part B-simulation Comput. 11, 311–334.

Macchietto, S., Maduabeuke, G., Szcepanski, R., 1986. Exact determination of process sensitivity to physical properties. Fluid Phase Equilib. 29, 59–67.

Mathias, P.M., 2014. Sensitivity of process design to phase equilibrium-A new perturbation method based upon the Margules equation. J. Chem. Eng. Data 59, 1006–1015. doi:10.1021/je400748p

Mathias, P.M., Kister, H.Z., 2017. Effect of Phase-Equilibrium Uncertainties on Ethyl Acetate Purification. J. Chem. Eng. Data 62, 2872–2883.

MATLAB 2015b, The MathWorks, Inc., Natick, Massachusetts, United States, n.d.

Python 3.5, Python Software Foundation, 2017, n.d.

Rowley, R.L., Wilding, W. V., Oscarson, L., J., Knotts, T.A., Giles, N.F., 2014. DIPPR® Data Compilation of Pure Chemical Properties, Design Institute for Physical Properties, AIChE, New York, NY.

SimSci PRO/II - Process engineering comprehensive Process simulation, v.9.3, by Schneider electric, 2017., n.d.

Sin, G., Gernaey, K. V., Eliasson Lantz, A., 2009. Good Modeling Practice for PAT Applications: Propagation of Input Uncertainty and Sensitivity Analysis. Biotechnol. Prog. 25, 1043–1053.

Whiting, W.B., 1996. Effects of Uncertainties in Thermodynamic Data and Models on Process Calculations †. J. Chem. Eng. Data 41, 935–941. doi:10.1021/je9600764

Zühlsdorf, B., Bühler, F., Mancini, R., Cignitti, S., Brian Elmegaard, 2017. High Temperature Heat Pump Integration using Zeotropic Working Fluids for Spray Drying Facilities. 12th IEA Heat pump Conf.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50244-5

Strategic design and analysis of large-scale shale gas monetizing projects under mixed uncertainty

Hua Liu, Qingsong Xie, Ming Pan, Chang He*

School of Chemical Engineering and Technology, Guangdong Engineering Center for Petrochemical Energy Conservation, Sun Yat-sen University, Guangzhou 510275, China. hechang6@mail.sysu.edu.cn

Abstract

This article developed a systematic monetization framework that incorporates stochastic modeling and probabilistic analysis of a large-scale shale gas monetization process under mixed uncertainties including feedstock composition, estimated ultimate recovery, and economic parameters. The major products investigated were polymers including HDPE and PP. To accommodate the reality of uncertainty, a new stochastic data processing strategy integrating the SAA, KSM, and DSAC methods was proposed to generate appropriate scenarios and to determine the corresponding values of each uncertain variable. These output variabilities were passed to the process simulator to implement the detailed process design and techno-economic analysis. A distributed-centralized conversion network employing two monetization pathways of shale gas to polymers, namely NGLs pathway and hybrid pathway, was presented. Uncertain economic parameters were nested in this stochastic model to identify the probabilistic distributions and to analyze the MLE of techno-economic objectives. The proposed strategy and models are illustrated in four case studies with different plant locations and pathway designs.

Keywords: stochastic modeling, probabilistic analysis, shale gas, mixed uncertainty

1. Introduction

Uncertainty is an inherent characteristic in process industries, which leads to significant challenges in handling variability in parameter values. Especially, increasing scientific evidence is showing that the shale gas monetizing process is inevitably subject to mixed and multi-dimensional uncertainties from gas acquisition to end-use production, significantly affecting the profitability and environmental impacts of the whole project. First, raw shale gas acquisition (the initial step of shale gas projects) is most influenced by the endogenous uncertainties of gas compositions. (Gao and You, 2017; Li et al., 2011) In fact, most of the shale gas regions produce raw shale gases with high compositional variabilities due to the differences in geological and petrophysical characteristics in these regions and wells. (Gong and You, 2017) Another challenge involves the supply of shale gas including special production decline and well-refracturing operation. Meanwhile, the variability associated with shale gas acquisition refers to the estimated ultimate recovery (EUR), which is affected by the special decline in production rate and well-refracturing operation. (Tavassoli et al., 2013; Cafaro et al., 2016) In addition, exogenous uncertainties normally occur throughout the entire project caused by market variabilities such as product prices, drilling costs, utility prices, and significantly increases the complexity of uncertainty in practice. (Grossmann et al., 2016) The resulting mixed uncertainties must be explicitly figured out at the design and development stage. Given the complex

uncertain variability, a comprehensive systems design and analysis framework is required to decipher the true economic risk of the shale gas monetization to the end-use products process under changing feedstock conditions and market fluctuations. To date, a number of research challenges in this field remain unsolved. These challenges are summarized as follows (a) how to simultaneously account for the mixed uncertainties, (b) how to incorporate field data into the stochastic modeling, (c) how to develop efficient shale gas monetization pathways, (d) how to apply statistical techniques to interpret the risks.

This article focuses on addressing these research challenges and develops a systematic monetization framework that incorporates stochastic modeling and likelihood analysis of a large-scale shale gas to polymers process with consideration of endogenous and exogenous uncertainties. To capture the feedstock variabilities, for the first step it proposes a new data processing strategy to generate appropriate scenarios from real-world data related to shale gas compositions and EUR. Next, this work employs the detailed process design and techno-economic modeling, by which the raw shale gas in each scenario is processed in multi-site distributed processing facilities, and then either NGLs, or the mixture of NGLs and methane, can be pipelined to a centralized plant for polymer production. These generated scenarios are considered together with a set of uncertain economic parameters to identify their probabilistic distributions and techno-economic objectives based on likelihood analysis. Finally, the applicability of the proposed framework is illustrated through four case studies under different plant locations and monetization pathways.

2. Methodology

As shown in Figure 1, the proposed methodology involves a three-step statistical analysis including (II) feedstock variability quantification, (II) process modeling and synthesis, and (III) probabilistic analysis. Herein, due to page limitations we only summary the bullet points of this methodology. More details are reported by He et al. (2018).

Step I: it proposes a stochastic data processing strategy to characterize and quantify the endogenous uncertainties including raw gas composition and EUR associated with the feedstock acquisition. A finite size of scenarios is generated to accommodate the diverse nature of feedstock variability. Three rigorous statistical methods included in the data processing strategy are sample average approximation (SAA), Kriging-based surrogate model (KSM), and integrated decline-stimulate analysis curve (DSAC). First, we used an SAA method to generate a scenario set of initial C_1 production rate of well i ($P^{r_0}_{i,C_1}$) to approximate the original expected objective value by calculating the sample average. The number of scenarios is determined by the desired level of accuracy of the solution, which can be measured by the confidence interval of the optimal solution of a stochastic programming (SP) problem. Next, KSM method implemented by DACE (Design and Analysis of Computer Experiments) toolbox in MATLAB includes multiple sub-steps in series to determine the corresponding initial production rate of other components that corresponds to C_2($P^{r_0}_{i,C_2}$), C_3($P^{r_0}_{i,C_3}$), C_{4+}($P^{r_0}_{i,C_4}$), N_2($P^{r_0}_{i,C_5}$), CO_2($P^{r_0}_{i,C_6}$), and H_2S($P^{r_0}_{i,C_7}$). (He et al., 2018) In each sub-step, we fit a single response surface with respect to the original component distribution using field data reported by Zavala-Araiza et al. (2018) Finally, an integrated DSAC method integrating Arps decline formula and refracturing formula is developed to quantitatively estimate the EUR under uncertainty. In addition, this step classifies the shale wells into wet and dry ones by checking the high heating value of raw

gas quality (S_{hhv}) in each scenario. Only raw shale gas extracted from wet wells is considered in this work since it is suitable for the downstream monetizing technologies.

Step II: the feedstock variability is propagated through performing a detailed process modeling method on distributed-centralized conversion network systems. Herein, detailed distributed-centralized network systems of shale gas-to-polymers considering different monetization pathways are developed through rigorous process design, modeling and synthesis. Specially, either the hybrid methane and NGLs or solely the NGLs are converted to polymers in a centralized production plant, resulting in two monetization pathways termed hybrid pathway (H) and NGLs pathway (O). The first pathway includes five process units (mixture fractionation, H_2/olefins recovery, C_2-C_3 pyrolysis, C_1 pyrolysis, and CO_2 capture and storage) and a utility facility. While the last two process units, namely C_1 pyrolysis unit and CO_2 capture and storage unit, as well as the syngas separator in C_2-C_3 pyrolysis unit, need to be eliminated in the second pathway due to the absence of C_1 conversion.

Once all the scenario runs are completed, a set of model output variables of interest is mapped and collected. We obtain a stochastic model with a finite number of independent scenarios. For per-scenario, the model outputs including plant-level yields, consumptions, equipment sizes, and emissions are known perfectly for the next step.

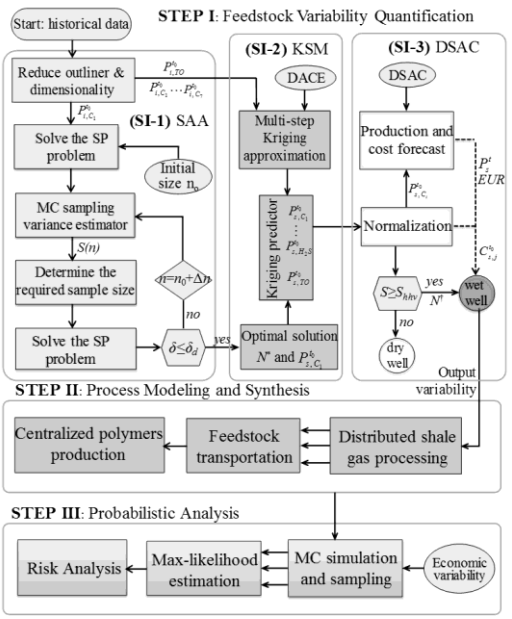


Figure 1. Flowchart of the proposed methodology

Step III: uncertain economic parameters extracted from the entire process are nested in the stochastic model to identify the probabilistic distributions and to perform a likelihood analysis of techno-economic objectives. In particular, each economic parameter can be assigned a certain type of probability distribution. As shown in Figure 1, Monte Carlo simulation method is used to perform the uncertainty analysis of economic parameters. The values of varying economic parameters are sampled randomly from the user-specified probabilistic distribution. In this way, for each scenario, a set of resulting outputs (sub-scenarios), with respect to the process performances, is recorded by

conducting a sufficient number of sets of samples. In this way, we can identify the probability distributions and then estimate the max-likelihood estimation (MLE) of levelized production cost of HDPE (LCOP, $/t):

$$\mathrm{LCOP}=\frac{TPC+\sum_{t\in T_{ls}}\left[\frac{AF+AO}{(1+DR)^t}-\frac{RV}{(1+DR)^t}-\frac{\sum_{bp}R_{bp}}{(1+DR)^t}\right]}{T_{ls}\times AP_{\mathrm{HDPE}}} \tag{1}$$

where AF, AO, and AP_{PE} are annualized parameters including the feedstock cost, operati ng cost, and production amount of PE, respectively; TPC, DR, RV, and R_{bp} are the total p roject cost, discount rate, residual value, and the revenue by selling by-product bp. Note that, in the calculation of LCOP, we consider that the raw shale gas is directly purchased from the raw shale gas producer at its real extraction cost instead of the market price.

3. Case Studies

The proposed shale gas monetization framework is applied to four distinct case studies including two sets of influencing factors as follows. (1) Plant location: feedstock is collected from two shale regions, namely the Appalachians (AP) and the Gulf Coast (GC); (2) monetization pathway: end-use products can be manufactured via either only a NGLs pathway design or a combination of C_1 pathway design and NGLs pathway design. Thus, these examples will be denoted by AP-O, AP-H, GC-O, and GC-H where AP and GC denote the plant located in the AP and GC regions. The distributed processing facility and centralized production plant are considered at 50 mmscfd of raw gas input and 300 kt per year of polymers production. Besides, allowable by-products consist of PG, butanes+, sulphur, and electricity.

4. Result and Discussion

Figure 2 shows the financial variability in terms of the LCOPs. In a vertical comparison, the histograms of LCOP distribution of the shale gas monetizing projects located in the same region but employing different production routes are presented. It is interesting that the case studies employing the NGLs pathway design has a much broader cost distribution. For example, as the monetization projects located in the AP region, the LCOP associated with the hybrid pathway ranges from $200/t to $2,250/t, while for the NGLs pathway design this value ranges from $-1,000/t to $4,350/t. This difference suggests that hybrid pathway design can effectively reduce the variability of production costs. It should be noted that a negative value of LCOP indicates that the monetizing project of interest is economically competitive even only considering the sale of by-products in some extreme scenarios. In a horizontal comparison, the LCOP distributions of the monetizing project located at different shale regions but considering the same production route are investigated. As an example, when the hybrid pathway design is taken into account, the projects located in the AP region had a slightly narrower distribution of LCOP compared with the GC region case. In the case of the NGLs pathway design, the difference in the distribution width of LCOPs becomes wider and more apparent for the two shale regions investigated.

To present the financial risk analysis, we consider that the LCOP cannot exceed a certain threshold, which is assumed to be $1,400/t of HDPE according to the high level of market

prices in the last three years. Once the LCOP exceeds this value, the distributing range is marked by a translucent dashed-box and the resulting cumulative probability distributions (Ω_R) are provided in Figure 2. As shown in this figure, the monetizing projects considering the hybrid pathway design have a much lower financial risk in relation to the Ω_R. For example, the financial risk can be remarkably mitigated from 29.6% for the AP-O case study to 17.2% for the AP-H case study.

Figure 3 compares the max-likelihood of LCOP results for different case studies with consideration of multi-refracturing treatments. Compared with the mean value shown in Figure 2, the MLE of LCOPs has a greater value by 8~12% for all case studies investigated. This increment clearly indicates that the true production cost is underestimated using the conventional statistical method under uncertainty which, in turn, results in an over-optimistic forecast of financial risk. For each case study, first the refracturing treatment leads to a decrease in the MLE of LCOP, but when the number of well-refracturing continues to grow, the LCOP's MLE remarkably increases. The single refracturing treatment case (M=1) has the minimum MLE of LCOP, which is \$975/t for the AP-H case study, \$1,038/t for the AP-O case study, \$994/t for the GC-H case study, and \$1,153/t for the GC-O case study. This result suggests that a single refracturing treatment should be performed over the lifespan of the wells, which is sufficient to guarantee a minimum LCOP. Note that the optimal number of well-refractures comes from the EUR growth and nominal operation expense for an additional treatment.

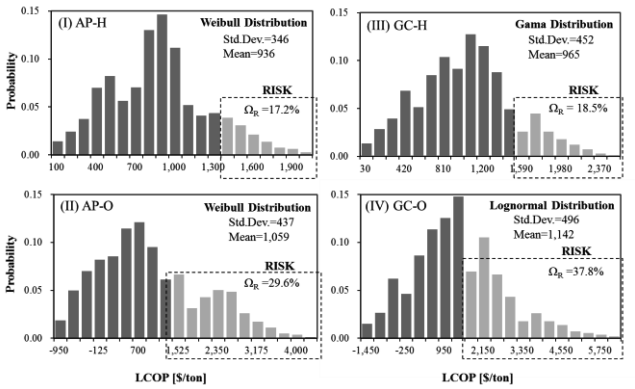


Figure 2. Comparison of the LCOP distributions for the HDPE production.

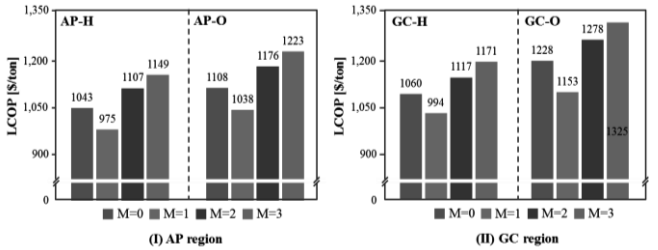


Figure 3. Comparison of the MLE of LCOPs.

As has been concluded above, the hybrid pathway design provided an effective solution to reduce the financial risk of shale gas monetizing projects. The advantage is not limited to this aspect. For the GC region case shown in Figure 3, the hybrid pathway design can

reduce the MLE of LCOPs by 10~15%, although this reduction significantly declines to about 5% for the AP region. This reflects that the hybrid pathway design is more amenable to reducing the risk of projects fed by raw shale gas that have high NGLs content. In other words, wet shale gas used as feedstock provides an opportunity to perform the process integration and equipment sharing considered in the hybrid pathway design, as NGLs are comparable to the methane in terms of the feed amount; otherwise, the improvement on the real production cost is not as much as anticipated.

5. Conclusions

This work proposed a systematic framework for monetizing raw shale gas to polymers and other by-products with consideration of complex mixed uncertainties. The mixed uncertainty considered in the monetization framework not only included endogenous variabilities of gas composition and EUR, but also the exogenous ones of the entire project, often corresponding to market variabilities. To accommodate the reality of uncertainty, a new stochastic data processing strategy integrating the SAA, KSM, and DSAC methods was proposed to generate appropriate scenarios and to determine the corresponding values of each uncertain variable. These output variabilities were passed to the process simulator to implement the detailed process design and techno-economic analysis. Uncertain economic parameters were nested in this stochastic model to identify the probabilistic distributions and to analyze the MLE of techno-economic objectives. The proposed framework was applied to four case studies based on different plant locations and the above-mentioned pathways. The results highlighted the benefits of the hybrid pathway as it reduced the MLE of LCOPs. In particular, the hybrid pathway design was more amenable to reducing the economic risk of the projects located in the GC shale region where the raw shale gas has appropriate NGLs content but suffers from expensive extraction expenses.

References

C. He, Ming Pan, B. Zhang, Q. Chen, et al., 2018, Monetizing Shale Gas to Polymers under Mixed Uncertainty: Stochastic modeling and likelihood analysis, AIChE J., DOI 10.1002/aic.16058.

D. C. Cafaro, M. G. Drouven, and I. E. Grossmann, 2016, Optimization models for planning shale gas well refracture treatments, AIChE J., 62, 12, 4297–4307.

D. Zavala-Araiza, D.T. Allen, M. Harrison, et al., 2015, Allocating Methane Emissions to Natural Gas and Oil Production from Shale Formations, ACS Sustainable Chem. Eng., 3, 3, 492-498.

J. Gao, F. You, 2017, Design and Optimization of Shale Gas Energy Systems: Overview, Research Challenges, and Future Directions, Comput. Chem. Eng., 106, 699–718

J. Gong, M. Yang, and F. You, 2017, A systematic simulation-based process intensification method for shale gas processing and NGLs recovery process systems under uncertain feedstock compositions, Comput. Chem. Eng., 105, 259–275.

I. E. Grossmann, R.M. Apap, B. A. Calfa, et al., 2016, Recent advances in mathematical programming techniques for the optimization of process systems under uncertainty, Comput. Chem. Eng., 91, 3–14.

X. Li, E. Armagan, A. Tomasgard, and P. I. Barton, 2011, Stochastic pooling problem for natural gas production network design and operation under uncertainty, AIChE J., 57, 8, 2120–2135.

S. Tavassoli, W. Yu, F. Javadpour, et al., 2013, Well Screen and Optimal Time of Refracturing: A Barnett Shale Well, J. Pet. Eng., 2013, 1–10.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50245-7

An Improved L-shaped Method for Two-stage Convex 0-1 Mixed Integer Nonlinear Stochastic Programs

Can Li[a,*], Ignacio E. Grossmann[a]

[a]*Center for Advanced Process Decision-Making, Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, United States*
canl1@andrew.cmu.edu

Abstract

We propose an improved L-shaped method to solve large-scale two-stage convex 0-1 mixed integer nonlinear stochastic programs with mixed-integer variables in both first and second stage decisions. Solving the deterministic equivalent of the two-stage stochastic programs with a large number of scenarios directly with commercial solvers is computationally expensive. To address the difficulties in solving large problems, we propose a Benders-like decomposition algorithm that includes both Benders cuts and Lagrangean cuts in the Benders master problem. The proposed algorithm is applied to solve a batch plant design problem under demand uncertainty. It is shown that the proposed algorithm outperforms the commercial solver DICOPT for the problems with a large number of scenarios. Also it is proved that the proposed algorithm can always yield a lower bound that is at least as tight as the one from Lagrangean decomposition algorithm.

Keywords: Stochastic programming, L-shaped method, Convex MINLP, Integer recourse

1. Introduction

Stochastic programming is an optimization framework that deals with decision-making under uncertainty. A special case is two-stage stochastic programming. Decomposition algorithms like Benders decomposition (Geoffrion, 1972) and Lagrangean decomposition (Guignard, 2003) have been used to solve stochastic programs with linear constraints and continuous recourse. However, for two-stage nonlinear 0-1 mixed-integer stochastic programs, relatively little work has been done. Li et al. (2011) propose a nonconvex generalized Benders decomposition (NGBD) algorithm for mixed-integer nonlinear stochastic programs with pure binary first stage variables. Ogbe (2016) proposes a joint decomposition method for mixed-integer nonlinear nonconvex programs.

To the best of our knowledge, there is no effective decomposition algorithm that can solve two-stage 0-1 mixed-integer convex nonlinear stochastic programs with mixed-integer variables in both first and second stage decisions. In this paper, we propose an improved L-shaped algorithm that has different types of valid inequalities in the Benders master problem.

2. An improved L-shaped method

In this paper, we consider two-stage stochastic programs with convex nonlinear constraints and 0-1 mixed-integer variables in both first and second stage. The deterministic equivalent of the two-stage stochastic program that we address in this paper is defined as (*P*) and shown in Eq.(1)-Eq.(5).

$$(P): \quad min \quad z = c^T x + \sum_{\omega \in \Omega} \tau_\omega d_\omega^T y_\omega \tag{1}$$

$$s.t. \quad A_0 x \geq b_0, \quad g_0(x) \leq 0 \tag{2}$$

$$A_{1,\omega} x + g_{1,\omega}(y_\omega) \leq b_{1,\omega} \quad \forall \omega \in \Omega \tag{3}$$

$$x \in X, \quad X = \left\{x : x_i \in \{0,1\}, \forall i \in I_1, \;\; 0 \leq x \leq x^{ub}\right\} \tag{4}$$

$$y_\omega \in Y \quad \forall \omega \in \Omega, \quad Y = \left\{y : y_j \in \{0,1\}, \forall j \in J_1, \;\; 0 \leq y \leq y^{ub}\right\} \tag{5}$$

Here, x represents the first stage decisions. y_ω represents the second stage decisions in scenario ω. $g_0, g_{1,\omega}, g_{2,\omega,}$ are convex functions. Both the first and the second stage decisions are 0-1 mixed-integer. Let *I* be the index set of all the first stage variables. $I_1 \subseteq I$ is the subset for indices of the binary first stage variables. Let *J* be the index set of all the second stage variables. $J_1 \subseteq J$ is the subset for the indices of the binary second stage variables. x^{ub} is a vector that represents the upper bound of all the first stage variables. y^{ub} is a vector that represents the upper bound of all the second stage variables. In this paper, we assume problem (*P*) has relatively complete recourse, i.e., any solution *x* that satisfies the first stage constraints has feasible recourse decisions in the second stage. Solving (*P*) with a large number of scenarios directly using MINLP solvers is prohibitive since the computational time grows exponentially with the number of scenarios. In order to solve problem (*P*) more efficiently, we propose an improved L-shaped method that is based on generalized Benders decomposition (Geoffrion, 1972). The basic idea of generalized Benders decomposition is to decompose (*P*) into a Benders master problem which only contains first stage decisions and Benders subproblems each of which only contains second stage decisions for a given scenario. In this paper, we add both Benders cuts and Lagrangean cuts to the Benders master problem to tighten the relaxation.

We first define the Benders master problem (MB^k), with both Lagrangean cuts Eq.(8), and Benders cuts Eq.(9) for a given set of iterations $k \in K$. z_{MB}^k is the objective of the Benders master problem at iteration k. $z_{SL,\omega}^{*k}$ and $z_{SB,\omega}^{*k}$ are the optimal values of the Lagrangean subproblem (SL_ω^k) and the Benders subproblem (SB_ω^k), which are defined next.

$$(MB^k): \quad min \quad z_{MB}^k = \sum_{\omega} \eta_\omega \tag{6}$$

$$s.t. \quad A_0 x \geq b_0, \quad g_0(x) \leq 0, \quad x \in X \tag{7}$$

$$\eta_\omega \geq z_{SL,\omega}^{*k} - \mu_\omega^k x \quad \forall \omega \in \Omega, k \in K \tag{8}$$

$$\eta_\omega \geq z_{SB,\omega}^{*k} + (\lambda_\omega^k)^T (x - \tilde{x}^k) + \tau_\omega c^T x \quad \forall \omega \in \Omega, k \in K \tag{9}$$

Next, we show how the valid inequalities Eq.(8) and Eq.(9) can be derived. The deterministic equivalent problem (P) can be reformulated by duplicating the first stage decisions x for each scenario ω and adding nonanticipativity constraints (*NACs*), $x_{\omega 1} = x_{\omega 2}, x_{\omega 1} = x_{\omega 3}, \cdots x_{\omega 1} = x_{\omega|\Omega|}$, to guarantee that all the first stage decisions made for all the scenarios are the same. The *NACs* can be dualized by multiplying π_ω to the constraints, $x_{\omega 1} = x_{\omega 2}$, $x_{\omega 1} = x_{\omega 3}, \cdots x_{\omega 1} = x_{\omega|\Omega|}, \omega = \omega_1, \omega_2, \cdots, \omega_{|\Omega|-1}$. Then the deterministic equivalent problem can be decomposed into $|\Omega|$ scenarios. Each subproblem is defined as a Lagrangean subproblem (SL_ω^k) for iteration k:

$$(SL_\omega^k): \quad min \quad z_{SL,\omega}^k = \tau_\omega (c^T x_\omega + d_\omega^T y_\omega) + \mu_\omega^k x_\omega \tag{10}$$

$$s.t. \quad A_0 x_\omega \geq b_0, \quad g_0(x_\omega) \leq 0 \tag{11}$$

$$A_{1,\omega} x_\omega + g_{1,\omega}(y_\omega) \leq b_{1,\omega} \tag{12}$$

$$x_\omega \in X, \quad y_\omega \in Y \tag{13}$$

where $\mu_{\omega 1}^k = \sum_{\omega=\omega_1}^{\omega_{|\Omega|-1}} \pi_\omega^k$, $\mu_{\omega+1}^k = -\pi_\omega^k$, $\omega = \omega_1, \omega_2, \cdots, \omega_{|\Omega|-1}$. Each Lagrangean subproblem (SL_ω^k) is solved to optimality. The optimal value of the subproblem ω at iteration k is defined as $z_{SL,\omega}^{*k}$. $\eta_\omega \geq z_{SL,\omega}^{*k} - \mu_\omega^k x$ is a valid inequality for the Benders master problem defined as a Lagrangean cut. The proof of the validity of the Lagrangean cuts is shown by Li and Grossmann (2017). After solving the Lagrangean subproblems at each iteration k, the multipliers of the Lagrangean subproblems can be updated using the subgradient method (Oliveira et al., 2013). After adding the Lagrangean cuts to the Benders master problem, the lower bound obtained by solving the Benders master problem is at least as tight as the lower bound that can be obtained by using Lagrangean decomposition (Li and Grossmann, 2017).

Assume the solution to the Benders master problem at iteration k is $\tilde{x}^k$, x is fixed at $\tilde{x}^k$ for the relaxation of deterministic equivalent problem (*RP*). (*RP*) can be decomposed into $|\Omega|$ subproblems. Each subproblem is defined as the Benders subproblem (SB_ω^k):

$$(SB_\omega^k): \quad min \quad z_{SB,\omega}^k = \tau_\omega d_\omega^T y_\omega \tag{14}$$

$$s.t. \quad x = \tilde{x}^k \tag{15}$$

$$A_{1,\omega} x + g_{1,\omega}(y_\omega) \leq 0 \tag{16}$$

$$0 \leq y_\omega \leq y^{ub} \tag{17}$$

According to Geoffrion (1972), a Benders cut can be generated at iteration k:

$$\eta_\omega \geq z_{SB,\omega}^{*k} + (\lambda_\omega^k)^T (x - \tilde{x}^k) + \tau_\omega c^T x \tag{18}$$

where λ_ω^k is the optimal dual multiplier for Eq.(15) at iteration k. $z_{SB,\omega}^{*k}$ is the optimal value of (SB_ω^k).

After the Benders master problem is solved at iteration k, x in problem (P) is fixed at $\tilde{x}^k$. The rest of the problem can be decomposed into $|\Omega|$ scenarios. A feasible solution to the deterministic equivalent problem can be obtained by solving the upper bound subproblems (UB_ω^k) defined as Eq.(19)-(20). Let $z_{UB,\omega}^{*k}$ be the optimal value of (UB_ω^k). An upper bound of (P) can be obtained by calculating $c^T \tilde{x}^k + \sum_{\omega\in\Omega} z_{UB,\omega}^{*k}$.

$$(UB_{\omega}^{k}): \quad min \quad z_{UB,\omega}^{k} = \tau_{\omega} d_{\omega}^{T} y_{\omega} \tag{19}$$

$$s.t. \quad g_{1,\omega}(y_{\omega}) \leq b_{1,\omega} - A_{1,\omega}\tilde{x}^{k} \tag{20}$$

The steps of the improved L-shaped method are outlined in Figure 1. The lower bound is initialized with $-\infty$. The upper bound is initialized with the optimal value obtained by solving the stochastic program with the first stage decisions fixed at the optimal solution of the expected value problem. The Lagrangean multipliers are initialized to 0. At each iteration, the Lagrangean subproblems, the Benders master problem, the Benders subproblems, and the upper bound subproblems are solved sequentially. The optimal solution of the Benders master problem is used to initialize the Benders subproblems and the upper bound subproblems. Valid inequalities for the Benders master problem can be derived by solving the Lagrangean subproblems and the Benders subproblems. The algorithm iterates until the relative optimality gap is within a given value ϵ.

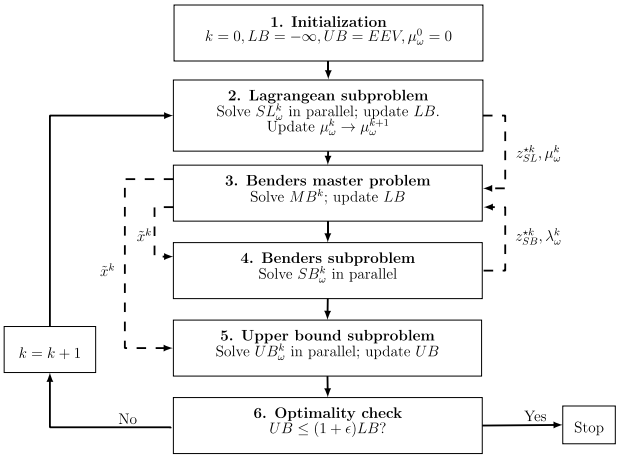


Figure 1 Algorithmic flow (solid line) and information flow (dotted line) of the improved L-shaped method

3. Optimal design of multi-product batch plant under demand uncertainty

We study a multi-product multi-stage batch plant (Figure 2) under demand uncertainty. Every product has to be processed through all the stages. It is assumed that in each stage there can be multiple processing units with equal volume. The number and the volume of the processing units purchased have to be decided before the demand for each product is known. After the demand of each product is realized, operating decisions including the number of processing units actually used and the batch size of each product have to be made. We assume the batch sizes for any given product are the same. Not all the processing units purchased have to be used, i.e., processing units can be idle when the demand is low. There will be a fixed cost for using one processing unit, so one should try to use as few processing units as possible. One should also try to complete all the processing tasks within a given time horizon and there is a penalty cost for exceeding the time horizon. This problem can be formulated as a two-stage stochastic program where the first stage decisions are the number and volume of the processing units. The second stage decisions are the batch size of each product, the number of processing units used in each stage, and the cycle time of each product. The problem can be formulated as a convex MINLP which is shown by Li and Grossmann (2017).

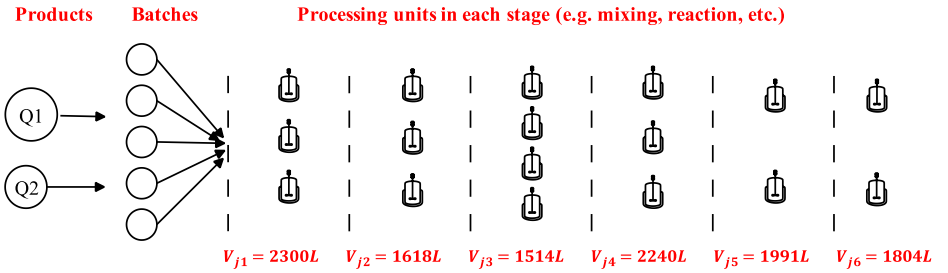


Figure 2. A multi-product multi-stage batch plant for two products with corresponding demands Q_1 and Q_2.

In the case study, we study a batch plant with 6 stages, 5 products. Each stage can have at most 4 processing units. We assume that there are 3 scenarios, where the demand of each product can be high, medium, or low with probability of 25%, 50%, 25%, respectively. The high and low demand are +10% and -10% away from the medium. The problem is first solved in deterministic equivalent form in GAMS using DICOPT on the 12 processors of an Intel Xeon (2.67GHz) machine with 64 GB RAM. The model has 96 binary variables, 64 continuous variables, 316 constraints. The optimal solution yields a minimum cost of 423.5(10^3\$). The NLP relaxation of this problem is 408.4(10^3\$). The wall time of solving the deterministic equivalent problem is 0.8 sec. The number and the volume of each processing units purchased in each stage is shown in Figure 2. The number of processing units that are operating in each stage in each scenario ($N_{j\omega}^S$) is shown in Table 1, where $\omega_1, \omega_2, \omega_3$, correspond to high, medium, and low demand, respectively. It is easy to see that the recourse decisions of different scenarios are different. When the demand is low, not all the processing units purchased are used. However, if we solve the expected value problem with the demand of all products fixed at their mean values, the number of units purchased at each stage is 3, 2, 4, 3, 2, 2, respectively. Compared to the stochastic solution, the number of units at stage *j2* decreases from 3 to 2. Therefore, with the expected value solution, the high demand cannot be satisfied without an extension of the time horizon. The expected value solution has an expected cost of 433.4(10^3\$). The value of stochastic solution (*VSS*) of this problem is 9.9(10^3\$).

Table 1: Optimal number of processing units in each stage in each scenario

Stage	*j1*	*j2*	*j3*	*j4*	*j5*	*j6*
$N_{j\omega_1}^S$	3	3	4	3	2	2
$N_{j\omega_2}^S$	3	2	4	3	2	2
$N_{j\omega_3}^S$	3	2	4	2	2	2

If the demands for the 5 products are independent and the demand for each of the products can be high, medium, or low as in the 3-scenario problem, the total number of scenarios is 243. The deterministic equivalent model has 5,856 binary variables, 4,144 continuous variables, 23,596 constraints. In this case, DICOPT fails to obtain a feasible solution for the deterministic equivalent problem within 50,000 seconds. The proposed decomposition algorithm is implemented to solve both the 3-scenario and the 243-scenario problem. The decomposition algorithm is allowed to run for 30 iterations

adding both Lagrangean cuts and Benders cuts to the Benders master problem in each iteration, or the relative optimality gap is less than 1%, whichever comes first. If the relative optimality gap after 30 iterations is greater than 1%, the algorithm is allowed to run for another 30 iterations only adding Benders cuts to the Benders master problem because it is observed that for most cases the Lagrangean multipliers are close to convergence after the first 30 iterations. As before, all the problems are solved using the 12 processors of an Intel Xeon (2.67GHz) machine with 64 GB RAM. All the MINLP problems are solved using DICOPT and all the NLP problems are solved using CONOPT. All the subproblems are solved in parallel using Grid GAMS. The wall time of solving the 3-scenario problem and the 243-scenario problem are 26 seconds and 604 seconds, respectively. The relative optimality gaps for the two cases are 0.4% and 0.7%, respectively. The upper bound of the 243-scenario problem is 413.3(10^3\$), while the lower bound is 410.5(10^3\$). The *VSS* of the 243-scenario problem is 4.9(10^3\$) based on the upper bound obtained by the proposed decomposition algorithm.

From the case study, we can see that although DICOPT can solve the small-scale deterministic equivalent problem quite fast, it fails to obtain a feasible solution for the large-scale deterministic equivalent problem. In comparison, the proposed decomposition algorithm scales better with the number of scenarios.

4. Conclusion

In this paper, we have proposed an improved L-shaped method that can effectively solve two-stage convex 0-1 mixed-integer nonlinear stochastic programs with mixed-integer variables in both first and second stage decisions. The decomposition algorithm is tested using the batch plant design problem under demand uncertainty. The algorithm can achieve a relative optimality gap of less than 1% for the problems tested. The proposed decomposition algorithm is faster than using DICOPT to solve the deterministic equivalent problem directly when applied to the problem with a large number of scenarios. However, the proposed algorithm is not guaranteed to close the optimality gap of the problems. A more rigorous algorithm that has finite convergence is currently being developed.

Reference

Li, X., Tomasgard, A., and Barton, P. I. (2011). Nonconvex generalized benders decomposition for stochastic separable mixed-integer nonlinear programs. Journal of Optimization Theory and Applications, 151(3):425.

Ogbe, E. (2016). New rigorous decomposition methods for mixed-integer linear and nonlinear programming. PhD thesis, Queen's University (Canada).

Guignard, M. (2003). Lagrangean relaxation. Top, 11(2):151-200.

Oliveira, F., Gupta, V., Hamacher, S., and Grossmann, I. E. (2013). A Lagrangean decomposition approach for oil supply chain investment planning under uncertainty with risk considerations. Computers & Chemical Engineering, 50:184-195.

Geoffrion, A. M. (1972). Generalized benders decomposition. Journal of Optimization Theory and Applications, 10(4):237-260.

Li, C., Grossmann, I.G. (2017). An improved L-shaped method for two-stage convex 0-1 mixed integer nonlinear stochastic programs. (2017). Submitted for publication.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50246-9

Optimal Strategies for Carbon Dioxide Enhanced Oil Recovery under Uncertainty

Demian J. Presser[a], Vanina G. Cafaro[a,b], Miguel Zamarripa[c], Diego C. Cafaro[a,b*]

[a]*FIQ-UNL, Santiago del Estero 2829, Santa Fe 3000, Argentina*
[b]*INTEC (UNL-CONICET),Güemes 3450, Santa Fe 3000, Argentina*
[c]*Oak Ridge Institute for Science and Education, National Energy Technology Laboratory, 626 Cochrans Mill Road, Pittsburgh, PA 15236-0940,United States*
dcafaro@fiq.unl.edu.ar

Abstract

This work presents a two-stage stochastic programming model to optimize the expected net present value (ENPV) of CO_2-EOR projects under uncertainty. The mathematical formulation relies on a multi-period planning approach aimed to find the optimal exploitation strategy for a mature oil reservoir. Given uncertain prices and productivity scenarios, the model sets the most convenient time to launch the CO_2-EOR project, and establishes efficient operating conditions over the planning horizon. It determines the number of production and injection wells to operate at every period, the CO_2 injection rate in every well, and the timing for maintenance and conversion tasks. The problem complexity grows rapidly with the number of wells and scenarios considered, resulting in a large-scale decision-making problem. Well productivity forecast functions are nonlinear (typically hyperbolic), yielding a mixed integer nonlinear (MINLP), nonconvex formulation. A moving horizon framework is adopted to take recourse actions when uncertain production parameters are revealed. The proposed approach helps operators to increase CO_2-EOR profitability by minimizing losses in low-price and productivity scenarios and maximizing the gain under more promising conditions.

Keywords: Carbon Dioxide Enhanced Oil Recovery, Stochastic Programming, Optimization, MINLP

1. Introduction

Carbon dioxide enhanced oil recovery (CO_2-EOR) and CO_2 sequestration in oil reservoirs is a relatively new technique with great potential to increase hydrocarbons production from mature reservoirs[1]. CO_2-EOR can increase the oil production from mature fields that have been primary and secondary exploited. While primary and secondary production stages typically yield between 30% and 50% of the original oil in place (OOIP, Melzer, 2012), CO_2-EOR techniques can increase the oil production by 15% or more (Advanced Resources International, 2011). Carbon dioxide injection in all its variants, specially the miscible option (combined with water), proves to be an effective option in different types of reservoirs. The main purpose is both to extend the reservoir productive lifespan and to increase the field productivity. Moreover, CO_2-EOR can reduce greenhouse gas emissions since large amounts of anthropogenic and/or natural CO_2 remain stored into the oilfield after performing the EOR, thus mitigating the environmental impacts of the oil and gas industry. Although many EOR projects take advantage of existing wells, data and facilities of the secondary recovery phase, several investments have to be made to achieve a reliable operation. The most relevant

investments are due to: (a) workover of existing wells (for production or injection); (b) CO_2 supply, compression and recycling facilities; (c) drill and completion of new wells; (d) field analysis to predict the reservoir response to CO_2 injection; and (d) new monitoring and operating equipment.

This work presents a two-stage stochastic programming model to determine the optimal investment and operation strategy for a CO_2-EOR project under uncertainty. To assess the performance of the EOR process, a nonlinear model is adopted, thus yielding a Mixed-Integer Nonlinear (MINLP) formulation. The oil production is forecasted by the Capacitance-Resistance-Model (CRM, Tao and Bryant, 2015, Eshraghi et al., 2016) combined with the Fractional Flow Method (FFM, Albertoni, 2002), based on hyperbolic decline functions of the oil composition. The main advantage of these approaches is their proven effectiveness to predict the oil production based on simple evaluations of historical data, regardless of geological simulations (Gentil, 2015). We embed the model on a moving horizon framework to account for changes in the problem conditions (Drouven et al., 2017).

2. General Problem Statement

We consider an oil reservoir that has completed its secondary recovery phase. We assume that: (a) Reservoir characterization data is available, given that the field has produced oil during a considerable amount of time. However, the well productivity after CO_2 injection is uncertain. We use a set of discrete scenarios $\{p \in P\}$ to account for the oil production rate variability. Once the EOR project is launched and the field starts to produce, production uncertainty is assumed to vanish. (c) CO_2 purchase and delivery cost, as well as future oil prices are uncertain parameters of the mathematical formulation. We consider a finite set of price forecast scenarios $\{f \in F\}$ to address this exogenous uncertainty. (d) Injection and production wells can be operated at different capacity levels. Higher operation rates imply larger facilities, but unit fluids handling costs reduce due to economies-of-scale.

Given the above assumptions, the proposed approach is aimed to: (1) decide whether or not to invest in the EOR-CO_2 project; (2) select new and/or existing wells to use for the EOR; (3) determine the most suitable operating mode (injection or production) for every well at every period; (4) set the most profitable well operation capacity level; and (5) find the right amount of carbon dioxide to inject at every period in every injection well; so as to maximize the expected net present value (ENPV) of the project. Furthermore, the model should find proper timings for maintenance and/or switching interventions, according to given prices and performance scenarios. In fact, the well's selection strategy should account for every future scenario (first stage, here-and-now decisions), while the CO_2 injection rates and the operation mode of each well over future periods will depend on the oil production response after the first interventions (second stage, wait-and-see decisions).

3. Mathematical Formulation

Besides production and price forecast scenarios, the optimization model is based on the following sets: time periods $t \in T$ (commonly quarters), new and existing available wells $i \in I$, production and injection capacities $cp \in CP$, $ci \in CI$. Furthermore, the MINLP model comprises the following 0-1 variables: V (= 1 if the project is launched), $Y_{i,t,ci}$ (= 1 if the well i is used as an injector during period t, at capacity level ci), $X_{i,t,cp}$ (=

1 if well *i* is set as producer during *t* at capacity level *cp*), $NW_{i,t}$ (indicating that the well *i* is inactive during period *t*), $MI_{i,t,ci}$ / $MP_{i,t,cp}$ (denoting that production/injection well *i* is under maintenance during *t*), $WPI_{i,t,ci}$ / $WPP_{i,t,cp}$ (denoting that the well *i* begins to inject/ produce at capacity *ci* / *cp* during period *t*), and $WRI_{i,t,ci}$ / $WRP_{i,t,cp}$ (= 1 if the production/injection well *i* turns to be an injector/producer at period *t*). Except for the variable *V* and the time period $t = 1$ (here-and-now decisions), all of the above 0-1 variables can take a different value for every production scenario *p* and price forecast *f* (wait-and-see decisions). The two-stage stochastic programming formulation seeks to:

$$Max\ ENPV = \sum_p \varphi_p \{ \sum_f \varphi_f [\sum_t CF_t^{(f,p)}/(1+r)^t] - Inv^{(f,p)} \} \quad (1)$$

where φ are the probabilities of every single scenario, *CF* are the cash flows, *r* is the discount rate and *Inv* is the initial investment. The main model constraints are presented below. For the sake of simplicity, indexes *p* and *f* (scenarios) are omitted.

$$\sum_{ci} Y_{i,t,ci} + \sum_{cp} X_{i,t,cp} + NW_{i,t} = V \qquad \forall i \in I, t \in T \quad (2)$$

$$WPP_{i,t,cp} \geq X_{i,t,cp} + NW_{i,(t-1)} - 1 \qquad \forall\, i \in I, t \in T, cp \in CP \quad (3)$$

$$WPI_{i,t,ci} \geq Y_{i,t,ci} + NW_{i,(t-1)} - 1 \qquad \forall\, i \in I, t \in T, ci \in CI \quad (4)$$

$$WRI_{i,t,ci} \geq X_{i,t,cp} + \sum_{ci} Y_{i,(t-1),ci} - 1 \quad \forall\, i \in I,\ t > 1,\ cp \in CP \quad (5)$$

$$WRP_{i,t,cp} \geq Y_{i,t,ci} + \sum_{cp} X_{i,(t-1),cp} - 1 \quad \forall\, i \in I,\ t > 1,\ ci \in CI \quad (6)$$

Eq.(2) to Eq.(6) determine the condition of every single well at every period. The injection/production rate is limited by Eq.(7) and Eq.(8). Parameters η and μ account for losses in the well productivity due to switching and maintenance interventions.

$$Q_{i,t,ci} \leq capi_{ci}\ (Y_{i,t,ci} - \eta i\ \ WRI_{i,t,ci} - \mu i\ \ MI_{i,t,ci}) \quad \forall i \in I, t \in T, ci \in CI \quad (7)$$

$$Q_{i,t,cp} \leq cap_{cp}\ (X_{i,t,cp} - \eta p\ \ WRP_{i,t,cp} - \mu p\ \ MP_{i,t,cp}) \quad \forall i \in I, t \in T, cp \in CP \quad (8)$$

Maintenance tasks are enforced to be performed periodically, depending on the operating conditions of the well (for the sake of shortness, the equations are omitted).

One of the most challenging issues is the relationship between inflows and outflows to/from injection and production wells. We adopt the Capacitance-Resistance-Model [3,4], omitting the transient term (initial phase) given that this period usually lasts few hours (negligible when compared to quarters). The CRM is given by Eq.(9) to Eq.(11).

$$QP_{i,j,t} \leq \psi_{i,j}\ \ QI_{i,t}\ (1 - \rho_t) \qquad \forall i \neq j, t \in T \quad (9)$$

$$\sum_{cp} Q_{i,t,cp} = \sum_{j \neq i} QP_{j,i,t} \quad ; \quad \sum_{ci} Q_{i,t,ci} = QI_{i,t} \qquad \forall i \in I, t \in T \quad (10)$$

$$\sum_j \sum_{cp} Q_{j,t,cp} = \sum_i \sum_{ci} Q_{i,t,ci}\ (1 - \rho_t) \qquad \forall\, t \in T \quad (11)$$

Every pair of wells is characterized by an inter-well connectivity factor (ψ_{ij}) where *i* and *j* are injection and production wells, respectively. Larger values of ψ_{ij} imply higher transmissibility. Parameter ρ is the retention factor of CO_2 into the reservoir, which usually decreases with time.

Another modelling challenge is to predict the oil composition in the flows produced by every well over the time horizon. In this case, we use the basis of the Fractional Flow Method[4,5] that estimates the oil yield of a single production well in a CO_2-EOR project. In our model, the oil fraction for every pair *i-j* declines hyperbolically with the cumulative amount of CO_2 already pumped through that link (*CGII*), in any direction, as expressed by Eq.(12) and Eq.(13). More specifically, the oil composition in the outflow of a production well can be forecasted through parameters a_j and b_j (specified for every single well, under every production scenario). In our model, similar to Drouven et al.[7], endogenous uncertainty is related to the production yield (a_j and b_j), which is assumed to be revealed as soon as the EOR project is started.

$$CGII_{i,j,t} \geq \sum_{t' \leq t} \psi_{i,j}\, QI_{i,t'} + \sum_{t' \leq t} \psi_{j,i}\, QI_{j,t'} \qquad \forall (i \neq j) \in I, t \in T \tag{12}$$

$$Qoil_{i,j,t} \left(1 + a_j \left(CGII_{i,j,t} \right)^{b_j} \right) \leq QP_{i,j,t} \qquad \forall (i \neq j), t \in T \tag{13}$$

Finally, the overall production from the reservoir during the whole time horizon should not exceed the fraction of the original oil in place expected to be recovered in each scenario, as imposed by constraint (14).

$$\sum_t \sum_{i \neq j} Qoil_{i,j,t} \leq u\ OOIP \tag{14}$$

To calculate revenues, investment and operating costs during every time period we account for: oil selling, CO_2 acquisition, CO_2 capture and storage (CCS), drill and completion of new wells, set up and maintenance of existing wells, operation mode switching, CO_2 / oil handling equipment, CO_2 recycling plant, production/injection equipment, among others (Wei et al., 2015). We consider exogenous uncertainty in the CO_2 purchase cost and the oil selling price. This uncertainty realizes regardless of what the decision-maker does. To react against this, a moving horizon framework is implemented, periodically updating production strategy.

4. Case Study Results

The MINLP formulation is applied to a representative case study comprising 15 existing wells that can be adapted to operate in a miscible CO_2-EOR project. The time horizon length is ten years, divided into quarters. The production strategy is revised every year, following a moving horizon framework. It is assumed that three price forecast scenarios are likely to occur (low, average and high) at every planning instance. As shown in Figure 1, EOR projects are the largest CO_2 consumers. As a result, we assume a direct correlation between oil prices and CO_2 purchasing cost. Figure 2 shows the oil price forecast scenarios. Each scenario is assumed to have the same occurrence probability.

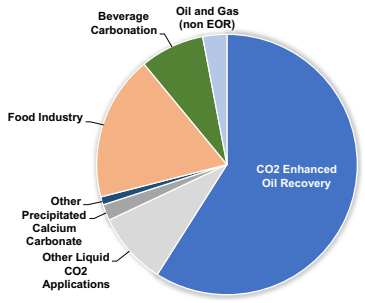


Figure 1: Global CO_2 market demand share (Source: CCS Institute, 2017)

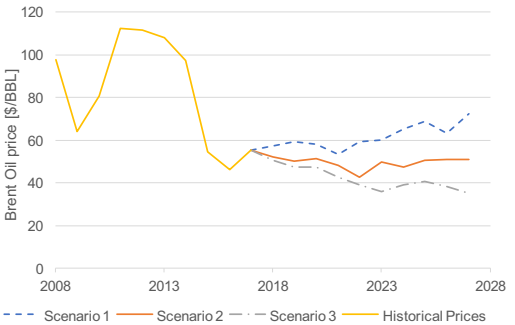


Figure 2: Oil price forecast scenarios

Similarly, two productivity scenarios (optimistic and pessimistic) are generated by introducing different sets of parameters a_j and b_j for every well. The optimistic scenario implies moderate declines of the well productivity (Scenario A), while pessimistic values mean a more rapid reduction of the oil fraction with the cumulative amount of CO_2 injected (Scenario B), as shown in Fig. 3 for a single production well.

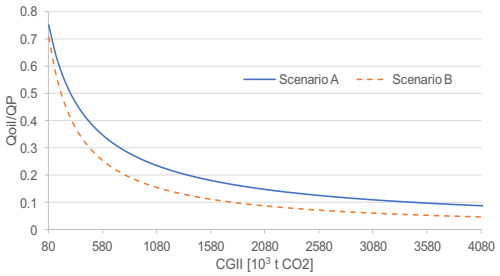


Figure 3: Oil fraction vs. Cumulative CO_2 reaching the well

Predictive maintenance tasks need to be performed at least every 2 and 3 years for production/injection wells, respectively. It is assumed a 10% reduction in the production/injection capacity of the well during the corresponding quarter. Analogously, the well capacity is reduced by half during the quarter in which the operation mode is switched.

From the hyperbolic decline function (13) the MINLP model is clearly nonconvex, thus requiring global solvers to guarantee optimality. However, the model dimensions largely exceed the maximum number of variables and equations to be solvable by commercial global solvers. For the case study, for instance, the model has more than 48 thousand variables (more than 10 thousand discrete) and 50 thousand constraints. As a result, an outer approximation solver (DICOPT) is adopted in this case.

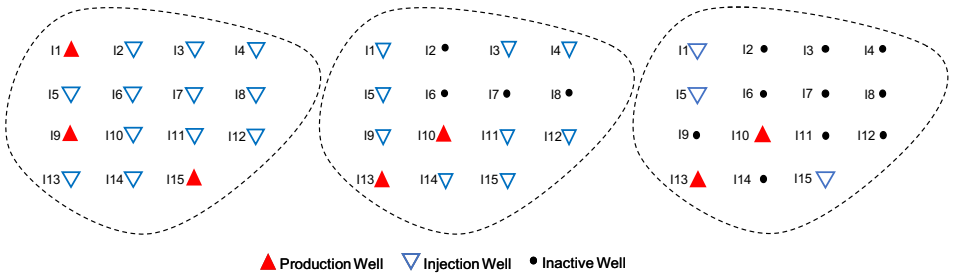


Figure 4: Well selection and operation modes for quarters T1, T25 and T40, respectively

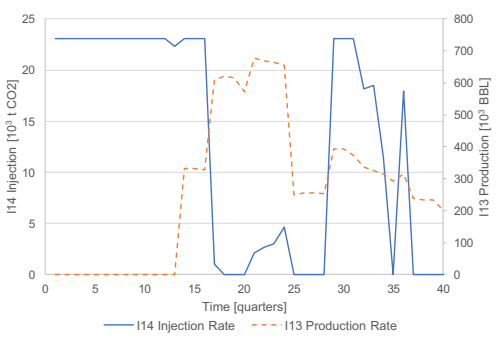


Figure 5: Production/Injection rates planned for wells I14 and I13, respectively

Using GAMS software on an Intel Xeon CPU, 8 GB RAM, (NLP solver: CONOPT4, MILP solver: CPLEX) it takes more than 4 hours and 5 major iterations to reach a local optimal solution. Some details on the solution found for the optimistic scenario are illustrated in Figs. 4 and 5. The most productive wells are I1, I5, I9, I10, I13, and I15. However, the model suggests a progressive development of the field and some changes in the wells operation strategy over time. As shown in Fig. 4, the recovery is more intense

in the first years, while several injection wells are gradually shut-in as oil production declines below threshold values. Under the optimistic scenario, the NPV amounts to 957.77 M$, 211 M$ better than the NPV of aggressive production strategies commonly applied in practice under promising scenarios. Maintenance tasks in production wells are planned at periods T8-T29 (I1); T14-T18-T26 (I5); T13-T26-T38 (I9); T13-T24-T37 (I10); T17-T26-T33 (I13); and T9-T13-T26-T38 (I15). Moreover, I1 is converted to an injection well at period T17; I5 switches to injection at period T25; I9 starts operating as an injection well at period T6; and I15 is converted to injection in quarter T13. Injection rates are properly managed, seeking to keep a quite stable oil production rate over time, maximizing the revenues when oil prices are high. To illustrate this, Fig. 5 represents injection and production rates of wells I14 and I13, respectively.

5. Conclusions

The two-stage stochastic approach presented in this work proves to be a valuable tool to aid EOR-CO_2 planning under uncertainty. The proposed MINLP model finds profitable solutions to a very complex combinatorial problem in reasonable CPU times. Although global optimality cannot be warranted since the model is nonconvex, the EOR development strategy is very efficient. By properly selecting the wells, managing injection rates and planning switching and maintenance tasks, project margins increase significantly (more than 28% with regards to EOR strategies typically implemented in practice). Future work will focus on reducing the computational burden by adding linear approximations of the decline functions and decomposition strategies. Reservoir representation can also improve by using more detailed well connectivity models.

Acknowledgement

Financial support from UNL and CONICET is fully appreciated. Miguel Zamarripa acknowledges partial support through Oak Ridge Institute for Science and Education.

References

L.S. Melzer, 2012, Carbon Dioxide Enhanced Oil Recovery (CO2-EOR): Factors Involved in Adding Carbon Capture, Utilization and Storage (CCUS) to Enhanced Oil Recovery.

Advanced Resources International (ARI), Improving Domestic Energy Security and Lowering CO2 Emissions with “Next Generation” CO2-EOR, June 2011, DOE/NETL-2011/1504.

Q. Tao, and S.L. Bryant, 2015. Optimizing carbon sequestration with the Capacitance/Resistance Model. SPE J. 20, 1094-1102.

S. Ehsan Eshraghi, M. Reza Rasaei, and S. Zendehboudi, 2016, Optimization of miscible CO2 EOR and storage using heuristic methods combined with capacitance/resistance and Gentil fractional flow models, Journal of Natural Gas Science & Engineering 32, 304-318.

A. Albertoni, 2002, Inferring interwell connectivity only from well-rate fluctuations in waterfloods, University of Texas, Austin (MSc Thesis).

P.H. Gentil, 2005, The use of multilinear regression models in patterned waterfloods: physical meaning of the regression coefficients, University of Texas, Austin (MSc Thesis).

M.G. Drouven, D.C. Cafaro, and I.E. Grossmann, 2017, Stochastic programming models for optimal shale well development and refracturing planning under uncertainty, AIChE J., 62, 4297-4307.

N. Wei, X. Li, R.T. Dahowski, C.L. Davidson, S. Liu, and Y. Zha, 2015, Economic evaluation on CO2-EOR of onshore oil fields in China, Int. J. of Greenhouse Gas Control, 37, 170-181.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50247-0

Optimal Production Scheduling of Industrial Gases under Uncertainty with Flexibility Constraints

M. Paz Ochoa[a*], Hao Jiang[a,b], Ajit Gopalakrishnan[c], Irene Lotero[c], Ignacio E. Grossmann[a]

[a]*Carnegie Mellon University, 5000 Forbes, Pittsburgh, PA 15213, U.S.A*
[b]*Tsinghua University, 30 Shuangqing Rd, Beijing 100084, P.R. China*
[c]*Air Liquide,200GBC Drive, Newark, DE 19702 U.S.A*
mochoa@andrew.cmu.edu

Abstract

In this work, we address the scheduling problem under uncertainties in electricity price and product demand in an air separation plant. The operation of the plant is represented by an efficient discrete-time MILP model as a process state transition network in order to deal with short-term production scheduling. On the one hand, uncertainties in electricity are addressed with stochastic programming techniques to find a schedule that minimizes expected cost over a proposed set of scenarios. On the other hand, uncertainties in product demand are tackled as flexibility constraints in order to ensure flexible operation over the entire range of variation of this uncertain parameter.

Keywords: Optimal schedule under uncertainty, power-intensive processes, two-stage stochastic programming, flexibility analysis.

1. Introduction

Demand side management is critical for maintaining profitability, especially in industrial power intensive processes where operating cost can be reduced by adjusting the production schedule to time-dependent electricity pricing schemes. However, uncertainty in these systems not only arises from electricity price but also from product demand.

Stochastic programming is the framework that models mathematical programs with uncertainty by optimizing the expected value over the possible realizations (Birge and Louveaux, 2011). The expected value is computed by integrating over the set of uncertain parameters. A simplification of this calculation involves a discretization of the uncertainty sets, where the realizations can be characterized with a finite number of scenarios. The most common stochastic programs that consider recourse is the Mixed-Integer Linear Programming (MILP) with continuous recourse in a second stage. The two-stage stochastic programming formulation considers two types of decisions: first stage, made before uncertainty reveals, and second stage, independent for each scenario.

Stochastic programming has found applications in process synthesis. For example, Halemane and Grossmann (1983) presented a two-stage approach for the design of flexible processes under uncertainty. The design problem is formulated to minimize a cost function and to guarantee feasible operation over a polyhedral uncertainty set.

The design of robust or flexible chemical processes under uncertainty has been dominated by the concept of flexibility analysis, which was first proposed and then further developed by Grossmann and co-workers, a description of its historical development can be found

in Grossmann *et al*. (2016). Recently, Zhang *et al*. (2016) presented insights on the strong connection between flexibility analysis and robust optimization.

The basic idea of flexibility analysis is to consider explicitly uncertainties in chemical processes. In general, the study of the flexibility can be performed at different stages. First at the design stage, where the optimization models for process design explicitly include flexibility constraints. Then with fixed design variables, to evaluate if certain design can tolerate the uncertainties in a specified range, known as the flexibility test problem. Finally, to quantify the range of uncertainties that the design can tolerate, called the flexibility index problem.

In this work, we propose a novel approach for the optimal production schedule with flexibility constraints, an analogous approach for the design of flexible processes under uncertainty proposed by Halemane and Grossmann (1983). We deal with uncertainties in electricity price and product demand in a production scheduling model of an air separation plant (Basán *et al*., 2018). Operation is represented by a discrete-time MILP model as a process state transition network, a scheme of the states can be seen in Fig. 1. We propose to explore the use of stochastic programming to account for uncertainties in electricity price in order to find the optimal production schedule that minimizes expected cost over a given set of scenarios. A two-stage stochastic programming problem is formulated considering the plant state binary variables as first-stage and the operation variables, such as power consumption, production and inventory levels, as second-stage.

Moreover, we address uncertainty in product demand to achieve robustness. We consider flexibility constraints to ensure flexible operation over the entire range of this parameter values. To obtain a flexible schedule with respect to product demand, the two-stage stochastic model evaluated at discrete electricity price scenarios with flexibility constraints need to satisfy the flexibility test. The test is evaluated with two different methodologies: affinely adjustable robust optimization and an iterative framework based on a dual-based flexibility analysis (Zhang *et al*., 2016). Once a flexible schedule is obtained, we calculate the flexibility index with both approaches to obtain a measure of the flexibility that can actually be achieved. We illustrate the application of the proposed models with a number of examples that show the effectiveness of the proposed approach for handling the uncertainties.

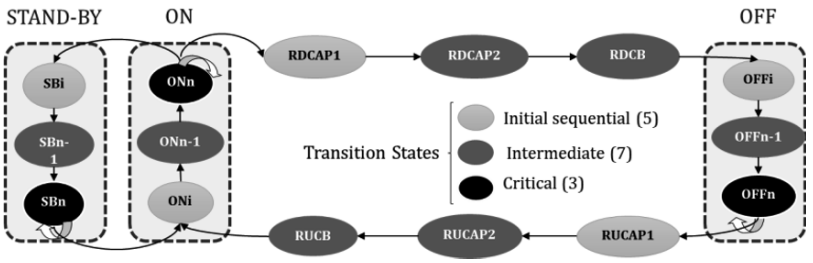


Figure 1. Process Transition Network (Basán *et al.,* 2018). Operation modes with minimum stay time: on (ON_i, ON_{n-1}, ON_n), stand-by (SB_i, SB_{n-1}, SB_n), and off (OFF_i, OFF_{n-1}, OFF_n). Start-up operation modes: ramp-up compressor A phase I (RUCAP1) and II (RUCAP2), and ramp up compressor B (RUCB). Shutdown modes: ramp down compressor A phase I (RDCAP1) and II (RDCAP2), and ramp down compressor B (RDCB).

2. Problem Statement and Methodology

The main objective is to minimize the expected value of the optimal cost function by selecting schedule variable *s* over the entire region *T*, which is the uncertainty set. First, the scheduling model corresponds to a MILP of the general form (see Basán *et al*., 2018):

$$\min c^T x + d^T y \tag{1}$$

$$s.t. \qquad Ax + By \leq d\,, x \leq 0, y \in \{0,1\} \tag{2}$$

In order to represent the feasible region of operation, the model is expressed as: $f(\mathrm{s}, z, \theta) \leq 0$, where *s* correspond to the 0-1 scheduling variables, z are the control variables and θ the uncertain parameters. The problem of optimal schedule under uncertainty corresponds to a two-stage programming problem, leading to an infinite programming problem because the feasibility constraint must be satisfied for the whole range of variation of θ.

$$\min_{S} \mathrm{E}_{\theta \in T} \left\{ \left(\min_{z} c(\mathrm{s}, z, \theta) \,\middle|\, f(\mathrm{s}, z, \theta \leq 0) \right) \right\} \tag{3}$$

$$s.t. \qquad \forall\, \theta \in T \{ \exists\, z \left(\forall\, j \in J \left[f_j(s, z, \theta) \leq 0 \right] \right) \} \tag{4}$$

The flexibility test problem states that for every point $\theta \in T$ there must exist at least one value of the vector control variable z that gives rise to non-positive values for all the individual constraint functions *j*. This means that irrespective of the actual values taken by the bounded parameters θ, the plant schedule *s* has the flexibility of operating to satisfy the specifications. A first simplification involves the discretization over the parameter space in order to approximate the expected cost by a weighted cost function. By using global maximum and minimum operators and inductive reasoning Halemane and Grossmann (1983), derived an equivalent multilevel optimization for the logical condition (4): $\max_{\theta \in T} \min_{z} \max_{j \in J} f_j(s, z, \theta)$

The general formulation of optimal schedule with flexibility analysis is then:

$$\min_{S, z_1, z_2, \ldots, z_n} \sum_{i=1}^{n} \varphi_i \cdot c(s, z_i, \theta_i) \tag{5}$$

$$s.t. \qquad f(s, z_i, \theta_i) \leq 0, i = 1, 2, \ldots, n \tag{6}$$

$$\max_{\theta \in T} \min_{z} \max_{j \in J} f_j(s, z, \theta) \leq 0 \tag{7}$$

where φ_i are discrete probabilities or weights for the selected finite number of parameter points θ_i. To solve the optimal schedule for the air separation plant with flexibility demand constraints we address two approaches: dual-based flexibility analysis (DFA) and adjustable robust optimization (AARO) following Zhang *et al*. (2016).

2.1. Schedule with Flexibility Constraints

Consider a set of *m* inequality constraints

$$f_j(s, z, \theta) = a_j s + b_j z + c_j \theta \leq 0 \;\; \forall\, j \in J \tag{8}$$

where $s \in \mathbb{R}^{n_s}$ are schedule variables, $z \in \mathbb{R}^{n_z}$ are control variables and $\theta \in \mathbb{R}^{n_\theta}$ are uncertain parameters, a_j, b_j and c_j are row vectors, and J is the set of constraints.

Table 1. DFA and AARO formulations for the scheduling problem under uncertainty and for flexibility index.

Schedule problem with Dual-based Formulation	Schedule problem with AARO Formulation
$\chi(S) = \max_{\lambda,\bar{\lambda},x} s^T A^T \lambda + \sum_{j\in J}\sum_{p\in\Theta} c_{jp}\left[\lambda_j\left(\theta_p^N - \Delta\theta_p^-\right) + \bar{\lambda}_{pj}\left(\Delta\theta_p^+ + \Delta\theta_p^-\right)\right]$	$\eta = \min_{d,\bar{z},p,Q,\mu,\upsilon} \hat{c}^T s + \sum_{i\epsilon I} \varphi_i \bar{c}^T \bar{z}_i$
s.t. $e^T\lambda = 1$	*s.t.* $As + B\bar{z}_i + C\bar{\theta}_i \le 0 \ \forall\, i \in I$
$B^T\lambda = 0$	$a_j s + b_j p +$
$\bar{\lambda}_{pj} \ge (\lambda_j - 1) + x_p \quad \forall\, p \in \Theta, j \in J$	$\left\{(\theta^U)^T\mu_j - (\theta^L)^T\upsilon_j\right\} \le u \ \forall\, j \in J$
$\bar{\lambda}_{pj} \le \lambda_j \quad \forall\, p \in \Theta, j \in J$	$\mu_j - \upsilon_j = \left(b_j Q + c_j\right)^T \forall\, j \in J$
$\bar{\lambda}_{pj} \le x_p \quad \forall\, p \in \Theta, j \in J$	
$\lambda \ge 0, \bar{\lambda} \ge 0, x \in \{0,1\}$	
Flexibility Index with DFA	**Flexibility Index with AARO**
$FI(s) = \min_{\lambda,\bar{\lambda},x} (-As - C\theta^N)^T \lambda$	$FI(s) = \max_{p,Q,\mu,\delta} \delta$
s.t. $B^T\lambda = 0$	
$\sum_{j\in J}\sum_{p\in\Theta} c_{jp}\left[\lambda_j \Delta\theta_p^- + \bar{\lambda}_{pj}\left(\Delta\theta_p^+ + \Delta\theta_p^-\right)\right] \ge 1$	*s.t.* $\delta \ge 0$
$\bar{\lambda}_{pj} \ge \left(\lambda_j + 1\right) + x_p \forall\, p \in \Theta, j \in J$	$a_j s + b_j p + c_j(\theta^N - \delta\Delta\theta^-) + e^T\mu_j \le 0 \quad \forall\, j \in J$
$\bar{\lambda}_{pj} \le \lambda_j \ \forall\, p \in \Theta, j \in J$	$\mu_j \ge \left(b_j Q + c_j \delta(\Delta\theta^- + \Delta\theta^+)\right)^T \quad \forall\, j \in J$
$\bar{\lambda}_{pj} \le x_p \ \forall\, p \in \Theta, j \in J$	
$\bar{\lambda}_{pj} \le 0, \lambda_j \le 0, x_p \in \{0,1\}$	

DFA is derived using LP duality theory, considering that solution must lie in the vertices, and then applying exact linearization to bilinear terms, leading to an MILP problem. Whereas, in the AARO approach the control actions are considered affine functions of θ, i.e. $z(\theta) = p + Q\theta$, which allow some degree of recourse, but leading to a restricted problem. In this reformulation, it is also assumed that the solution lies in the vertices, which is true when the constraints are jointly convex. To obtain the affinely adjustable robust counterpart, a constraint-wise worst-case approach and LP duality theory are applied. An advantage of this approach is that it does not add any binary variable. The final formulation of both approaches is described in Table 1. Details of the reformulation steps for both approaches can be found in Zhang *et al*. (2016).

The flexibility function $\chi(s)$ of schedule *s* with respect to the uncertainty set *T* represents the projection of the feasible region. If *T* is inscribed in the projection, then $\chi(s) \le 0$ meaning that a feasible schedule is obtained. When *T* is not completely constrained in the projection, then $\chi(s) > 0$.

The algorithm for the DFA is based on an iterative column-and-constraint generation approach, which relies on the fact that a schedule is feasible for all $\theta \in T$ if it is feasible for the worst-case realization of the uncertainty, which lies at one of the vertices of *T*. Vector θ is partitioned into two subsets, θ_{EP} for electricity prices and θ_D for demand uncertainty. It includes the following steps:

1) Set k=0. Choose an initial set $\widehat{T_0}$ of N_0 critical points
2) Solve:

$$\eta = \min_{S,\bar{z},\hat{z}} \hat{c}^T s \sum_{i=1}^{n} \varphi_i \cdot c^T \bar{z}_i \tag{9}$$

$$s.t: \quad As + B\bar{z}_i + C\bar{\theta}_{EP,i} \le 0, \forall\, i \in I \qquad \text{Electricity Price Scenarios} \tag{10}$$

$$As + B\hat{z}_t + C\hat{\theta}_{D,t} \leq 0, \forall\, t \in \hat{T}_k \qquad \text{Product Demand Vertices} \qquad (11)$$

where $\bar{z}_i$ and $\hat{z}_t$are the control variables that compensate electricity prices uncertainty, $\bar{\theta}_{EP,i}$, of the *i* scenarios, and demand uncertainty, $\hat{\theta}_{D,t}$, of the *t* vertices, respectively . To obtain s_k:

3) Solve Problem: Dual-based Flexibility Analysis with $s = s_k$, Obtain critical point θ_k^c
   If $\chi(s) \leq 0$, Stop. Else go to step 4.
4) Set $\hat{\theta}_{k+1} = \theta_k^c,\ N_{k+1} = N_k + 1,\ \hat{T}_{k+1} = \{1,2,\dots,N_{k+1}\},\ k = k+1$. Go to Step 2.

The algorithm converges in a finite number of iterations, since there is a finite number of vertices. Another difference between both approaches is that AARO does not require an iterative framework.

2.2. Flexibility Index

Consider the uncertainty set as $T(\delta) = \{\theta: \theta^N - \delta\Delta\theta^- \leq \theta \leq \theta^N + \delta\Delta\theta^+\}$, where δ is a positive scalar, $\Delta\theta^-$ and $\Delta\theta^+$ are negative and positive deviations from the nominal value. The flexibility index is a measure of how much flexibility can actually be achieved in the given schedule. The objective is to find the largest δ such that by proper adjustment of control variables *z*, the inequalities $f_j(s,z,\theta) \leq 0\ j \in J$ hold for all $\theta \in T(\delta)$. Following a similar reasoning as in the flexibility test, the flexibility index problem can be reformulated into the problems described in Table 1. If the value of *FI*>1, then schedule exceeds the flexibility target, whereas when *FI* < 1, it does not meet the target.

3. Numerical Results

The presented methodology is applied to scheduling model of an air separation plant with a time horizon of 168 hours, proposed by Basán *et al*. (2018). First, two-stage stochastic programming is applied considering the plant state (binary variables) as the first stage decision. Second stage decisions are production level, inventory and power consumption. The different electricity price scenarios profiles proposed are shown in Fig. 2, together with the nominal demand profile, whose range of variation is ±10%.

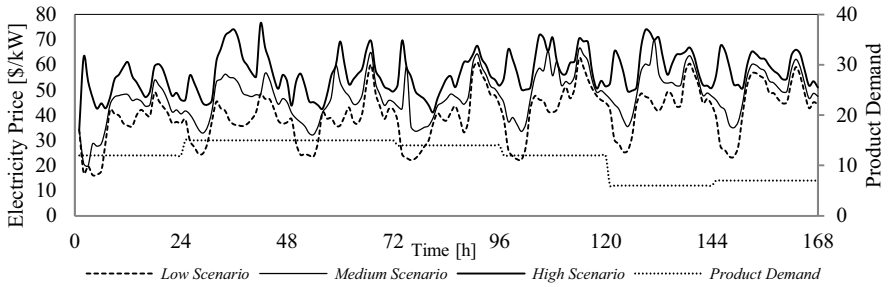


Figure 2. Proposed electricity price scenarios and nominal product demand profile.

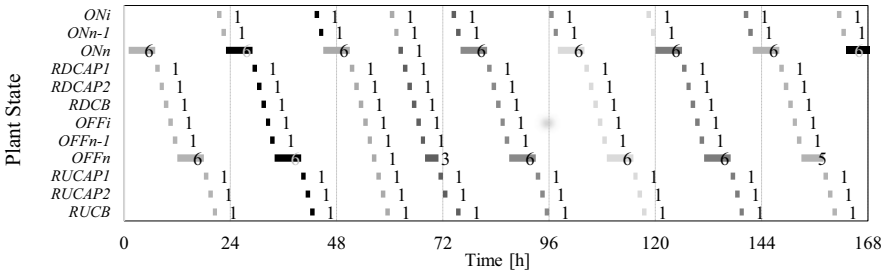


Figure 3. Optimal and flexible production schedule

Table 2. Scheduling under uncertainty with flexibility constraints and flexibility index results.

Problem	Approach	# of bin. variables	# of cont. variables	# of constraints	Result	Gap (%)	Solution Time (s)
Schedule under uncertainty	Dual-based	168	116,948	348,355	Cost 40814*	0	46
	AARO		247,218	116,779	40814		10.6
Flexibility Index	Dual-based	168	116,948	348,355	FI=1.083	0	10,739
	AARO		131,131	116,780	FI=1.083		135.6

(*) The DFA problem converges in the first iteration.

DFA and AARO formulations lead to the same production schedule, shown in Fig. 3. Additionally, both approaches provide the same flexibility index value for the uncertain demand. This implies that the same level of flexibility can be achieved even when considering a restricted control variable as in the case of AARO. The solution time required by the AARO LP models is less, as compared to the DFA, especially in the case of the flexibility index with two orders of magnitude of difference, as detailed in Table 2.

4. Conclusion

In this paper, we have addressed the scheduling problem under uncertainty. We proposed to deal with uncertainty in electricity price stochastically and in product demand in a robust way by considering it as the uncertain parameter set of the flexibility constraint. We solved the problem with the dual-based flexibility analysis and affinely adjustable robust optimization formulations.

The proposed approach has allowed the efficient calculation of the optimal schedule for the expected value of electricity price profiles and feasible operation for the range of variation of product demand over a time horizon of 168 hours. In addition, flexibility indices were computed successfully with both formulations, providing a measure of feasibility of the obtained schedule. Both formulations led to the same results for the different problems, with shorter solution times for the AARO approach, in line with the results obtained by Zhang *et al*. (2016).

References

N.P. Basán, I.E.Grossmann, A. Gopalakrishnan, I. Lotero, C.A. Méndez, 2018, A novel MILP scheduling model for power-intensive processes under time-sensitive electricity prices, Industrial & Engineering Chemistry Research, DOI: 10.1021/acs.iecr.7b04435.

J. Birge, FV Louveaux, 2011, Introduction to stochastic programming. 2nd ed. New York(NY): Springer.

I.E. Grossmann, R.M. Apap, B.A. Calfa, P. García-Herreros, Q. Zhang, 2016, Recent advances in mathematical programming techniques for the optimization of process systems under uncertainty, Computers & Chemical Engineering, 91,3-14.

K.P. Halemane, I.E. Grossmann, 1983, Optimal process design under uncertainty, AIChE Journal, 29 (3) 425–433.

Q. Zhang, R. Lima, I.E. Grossmann, 2016, On the relation between flexibility analysis and robust optimization for linear systems, AIChE Journal 62(9):3109-3123.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering–PSE 2018
July 1–5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50248-2

Raw material supply strategy for petrochemical process under market uncertainty

Byeonggil Lyu, Seokyoung Hong, Seunghyeon Oh, Il Moon*

Department of Chemical and Biomolecular Engineering, Yonsei University, 50 yonsei-ro, Seodaemun-gu, Seoul 03722, Korea
Ilmoon@yonsei.ac.kr

Abstract

In the petrochemical industry, raw material supply is a very important consideration that directly affects the operating condition and profit of the process. In particular, due to fluctuation in raw material price, there can be a huge difference in profit with the same amount of raw material procurement. However, due to various uncertainties regarding price fluctuation, it is very difficult for decision makers to arrive at an optimum supply strategy. In this study, we tried to establish an optimal supply strategy for petrochemical process considering the price uncertainty of naphtha and crude oil. To consider price uncertainty, we construct a system dynamics model to predict fluctuation of "crack," which is the price gap of naphtha and crude oil. Based on the system dynamics model, comparative analysis and optimization of naphtha purchase and production from crude oil were performed. Since the forecasting model focused on the rising and falling of the price gap rather than absolute numbers, it was more helpful in establishing short-term supply strategies. The developed supply strategy shows improved cost saving compared with the conventional supply model.

Keywords: supply strategy, system dynamics, forecasting, market uncertainty, decision making

1. Introduction

Raw material supply is a key issue directly affecting a plant's profit. However, it is very difficult to formulate material supply strategy because of market price uncertainty. Therefore, in response to this market uncertainty, supply decision makers developed several strategies. The raw material procurement problem is a topic of ongoing research. Deterministic and stochastic optimization models considering various cases of crude oil and refinery and petrochemical products are suggested (Al-Othman et al., 2008). Suggested model considered uncertainty in market prices and demand as a case study of each price, and market demand varied in the 20% range. With fluctuating raw material prices, the net present value framework of the raw material procurement model is suggested (Arnold et al., 2009). The crude oil procurement problem is also solved by considering sales, purchase, and the manufacturing cost (Chen et al., 2015). In this study, a stochastic multi-period price model is used for optimal refinery operation. The price uncertainty follows a Brownian motion price model. The problem of selecting procurement contracts is also solved (Calfa & Grossmann, 2015). In this study, contract selection is based on selling price optimization under supply and demand uncertainty.

Previous studies used various methods to solve the supply and procurement problems under uncertainty. However, several case studies of price fluctuation were used to find

solutions to price uncertainty. This is because of "the difficulty of forecasting." Accurate forecasting models that reduce market uncertainty would help in supply decision making. Figure 1 illustrates the novel decision-making framework suggested in this study.

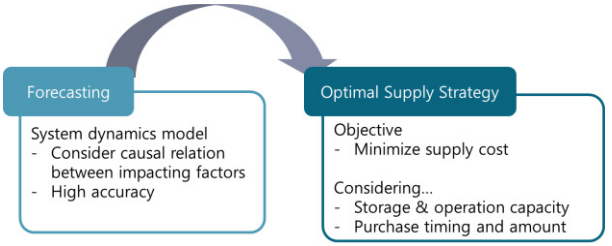


Figure 1. Framework for supply decision making

First, an accurate forecasting model that predicts raw material price at least three months in advance is developed using the system dynamics (SD) theory. Based on the forecasting result, a raw material supply strategy that minimizes the supply cost is suggested.

2. Problem Statement

The problem considered in this study is the supply of naphtha for naphtha cracker. Figure 2 shows that a cracker has two sources of naphtha: one is produced from condensate oil and the other is naphtha procured from the market. When condensate oil price falls, the naphtha produced is more competitively priced. On the contrary, naphtha procured at a decreased market price adds more competitiveness than naphtha produced from condensate oil. Because of the high price fluctuations of naphtha and condensate oil, the decision of whether to procure naphtha from the market or to produce it from oil is an issue. Therefore, our focus is to decide which, when, and how much to buy.

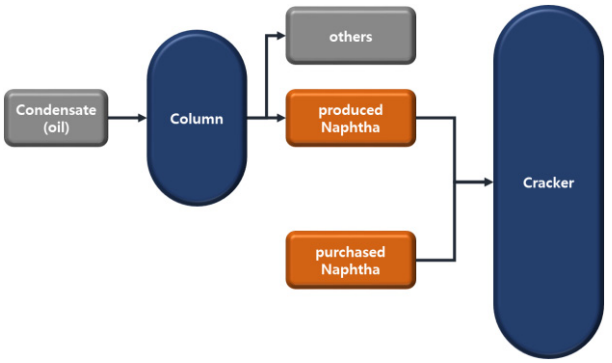


Figure 2. Process diagram for naphtha supply

Price forecasting of condensate oil and naphtha is necessary to solve the problem. However, to forecast the price of each material is very difficult and not very accurate. Therefore, we suggest a new forecasting term "crack," which is the price gap of naphtha and condensate oil. Since our objective is minimizing the supply cost, forecasting changes in "crack" can help lower the cost compared with forecasting price value.

3. Forecasting of Naphtha Crack

3.1. Forecasting methodology

System dynamics (SD) is a theory of problem solving based on the feedback control theory. SD uses relevant variables for the presented problem, defines the system, and models relationships among variables. It is based on the synthesis of many themes such as operating theory and system theory, control theory, information feedback theory, decision-making theory, mechanical system, and computer science. SD uses various control elements, such as feedback loops and delay times, to observe how the system reacts and responds to trends. The prediction procedure using the SD model is described in Figure 3. The naphtha crack forecasting model is suggested based on a causal loop diagram, a quantitative model, and an old data set.

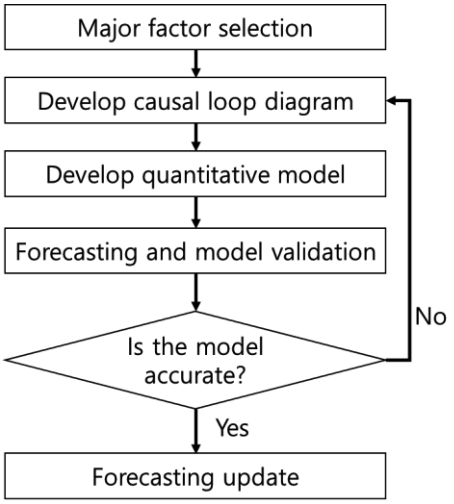


Figure 3. SD model development procedure

3.2. Forecasting model description

The forecasting model described in this section follows the model developed by our group (Lyu et al., 2017). A brief description as well as figures are introduced in this section. Following the model development procedure described in Figure 1, the forecasting model of naphtha crack is developed. First, major factors are selected by the heuristic of field engineers and the Pearson product-moment correlation coefficient (PPMCC) analysis.

$$p_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \tag{1}$$

PPMCC, as expressed in Eq.(1), is a measure of linear correlation between two variables, which give values between −1 and 1. When the PPMCC value is 1, it means that the two variables have a positive correlation, −1 means a negative correlation, and 0 means no correlation. Following the factor selection procedure, the causal loop diagram between the selected factors is developed. Vensim, a widely used SD software, is used for modeling the naphtha crack forecasting SD model. Vensim supports drawing of the causal loop diagram and editing of interaction formula and equations. Our model and forecasting result are described in Figures 4 and 5 and Table 1.

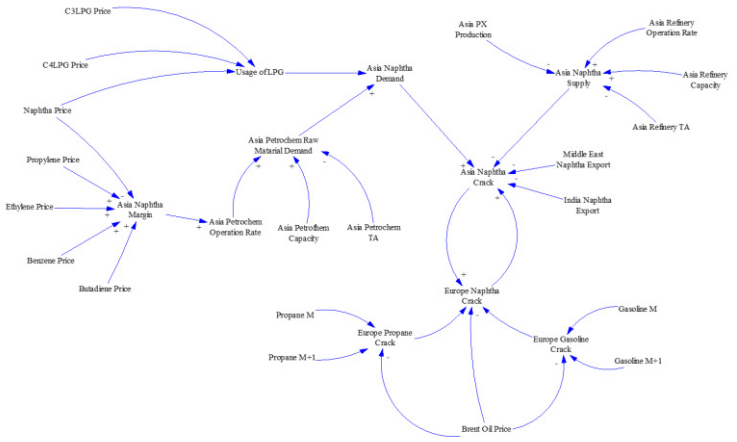


Figure 4. SD model of the naphtha crack (Lyu et al., 2017)

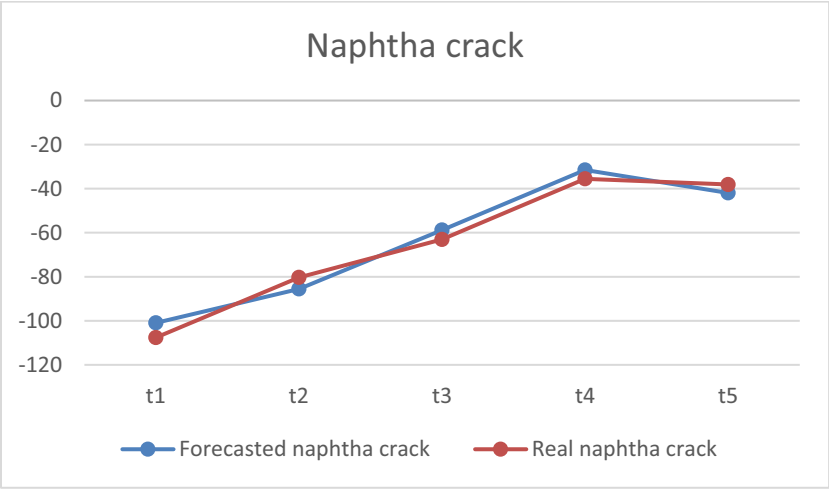


Figure 5. Naphtha crack forecasting result

4. Raw Material Supply Optimization

In the problem presented in section 2, a naphtha cracker has two naphtha sources: one is naphtha produced from condensate oil and the other is naphtha procured from the market. Because of high price uncertainty of naphtha and condensate oil, decision making regarding supply strategy is difficult. With the developed forecasting model, which helps to reduce uncertainty, an optimal raw material supply model under market uncertainty is suggested in this section.

4.1. Optimization model description

The objective is to minimize the total supply cost in Eq.(2). The cost which expressed in Eq.(3) and (4) is calculated as the product of the average price *Pa* and the feed amount *F* of each material.

$$total\_cost = \sum cost_m \ (m = naphtha\ and\ condensate) \tag{2}$$

$$cost_m = \sum F_{m,t} * Pa_{m,t} \tag{3}$$

Average price *Pa* is defined based on the amount of existing inventory, previous average price, newly added amount, and its trading price. The defined model has three storage tanks of condensate, naphtha from condensate, and naphtha from the market. The equations for average price and inventory are described as follows.

$$Pa_{m,t_{i+3}} = \frac{T_{m,t_0} * Pt_{m,t_0} + I_{m,t_{i+2}} * Pa_{m,t_{i+2}}}{T_{m,t_0} + I_{m,t_{i+2}}} \ (i = 0,1,2,\dots) \tag{4}$$

Eq.(4) shows the average price of each material. As the transportation time is three months, three-month-before trading amount and the price are used to define the average price.

$$I_{m,t_{i+3}} = T_{m,t_0} - F_{m,t_{i+3}} + I_{m,t_{i+2}} \tag{5}$$

Eq.(5) shows the inventory of each material. Material procured three months ago is added to the existing inventory and the feed amount is subtracted.

Trading price of each material changes in every time domain, but the future trading price information is unknown. Therefore, the forecasted naphtha crack is used to define the trading price of each material.

$$Pt_{m,t_{i+1}} = Pt_{n,t_i} \pm 0.5 * crack_{t_{i+1}} \tag{6}$$

We applied the forecasted crack equally to each material. The forecasted crack value is divided into two, added to the previous trading price of naphtha, and subtracted from the previous trading price of condensate.

The optimization model assumes a storage capacity of 800,000 tons and a maximum operating capacity of 300,000 tons per month.

4.2. Optimization result

The conventional supply cost and optimized supply costs for five months are expressed in Table 1 and 2.

Table 1. Conventional supply cost

(million USD)

	Conventional		
	Naphtha	Condensate	Total
t1	186.540	213.437	399.978
t2	208.487	228.560	437.047
t3	234.950	250.709	485.660
t4	242.614	251.500	494.114
t5	238.600	248.132	486.732
Ave.	222.238	238.467	460.706

Table 2. Optimized supply cost

(million USD)

	Optimized		
	Naphtha	Condensate	Total
t1	198.792	213.437	412.230
t2	242.162	228.560	470.722
t3	276.029	61.997	338.027
t4	285.566	37.501	323.068
t5	280.947	31.980	312.927
Ave.	256.699	114.695	371.395

The conventional supply strategy does not adjust the purchase amount in each time period because of high market uncertainty. However, the optimized model with reduced uncertainty changes the purchase amount in each time period. The forecasted price helps to decide when to buy large amounts. In the optimization model based on the forecasting result, the average cost of naphtha increased but the condensate cost decreased considerably. Finally, the optimized supply strategy saves US$89 million in the five-month period.

5. Conclusions

Because of dynamic market conditions, market uncertainty is vital yet challenging for the development of a raw material supply strategy. To deal with this problem, a novel framework is suggested in this study. First, a price forecasting model is suggested using the SD theory. The forecasting model forecasts the price gap between two raw materials, not the price value. This model helps build a strategy for when, what, and how much to buy. Subsequently, an optimization model that minimizes supply cost is developed with the forecasting result information. The optimization model considers the average price of raw materials, trading amount, and inventory capacity. With the forecasted crack (price gap between naphtha and condensate oil), the trading prices of naphtha and condensate are defined. This price information helps make a decision regarding the purchase amount of each raw material. As a result, the suggested supply model shows a cost saving of US$ 89 million compared with the conventional supply strategy.

References

W.B. Al-Othman, H.M. Lababidi, I.M. Alatiqi, and K. Al-Shayji, 2008. Supply chain optimization of petroleum organization under uncertainty in market demands and prices. European Journal of Operational Research, 189, 822-840.

J. Arnold, S. Minner, and B. Eidam, 2009. Raw material procurement with fluctuating prices. International Journal of Production Economics, 121, 353-364.

B.A. Calfa, and I.E. Grossmann, 2015. Optimal procurement contract selection with price optimization under uncertainty for process networks. Computers & Chemical Engineering, 82, 330-343.

R. Chen, T. Deng, S. Huang, and R. Qin, 2015. Optimal crude oil procurement under fluctuating price in an oil refinery. European Journal of Operational Research, 245, 438-445.

B. Lyu, H. Kwon, and I. Moon, 2017. A novel system dynamics model for forecasting naphtha price. Korean Journal of Chemical Engineering.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50249-4

Inventory Pinch Based Algorithm for Gasoline Blend Planning with Uncertainty in Components Qualities

Mahir Jalanko, Vladimir Mahalec*

Dept. of Chemical Engineering, McMaster University, 1280 Main Street West, Hamilton L8S 4L7, Canada
mahalec@mcmaster.ca

Abstract

In order to meet gasoline specifications, qualities of the blend components are analysed either in a lab or via on-line analysers. Deterministic gasoline blend planning assumes a quality of each blend component, even though there is an uncertainty associated with these qualities. In this work we present a chance constrained based methodology which enables a blend planner to decide what percent of the blends will meet specifications even if it means that some blends will give away qualities. This enables the planner to trade-off the cost of reblending the off-spec blends vs. the increase in the blend costs which guarantees that the quality specifications will be met. We employ a chance constrained formulation in (i) a two-level inventory-pinch based algorithm to optimize gasoline blend planning with uncertainty in components qualities; (ii) full space model for both linear and nonlinear blending rules. Inventory pinch based algorithm performs much better than the full-space formulation in terms of time required to reach the optimal solution, time to close the optimality gap to the required level, and robustness in achieving the optimal solution. Proposed algorithm enables construction of a trade-off curve of probability that the constraints will be satisfied vs. blend cost, which is not practical via full space formulation since the latter typically does not close the optimality gap or often it does not converge.

Keywords: Stochastic programming; Gasoline blending planning with uncertainty; Inventory pinch concept; Uncertainty in components qualities

1. Introduction

Refining plant section is a key element in the supply chain of the petroleum industry. About 60 % - 70 % of refinery revenue comes from selling gasoline products, therefore minimizing the cost of blending can greatly impact the profitability of the refinery. Blending components qualities fluctuate and are uncertain. Gasoline blend plan needs to compute blend recipes which are likely to be feasible and also minimize the overall costs. Fine adjustments to the blend recipes are via closed-loop blending where analyzers are used to correct the blend recipe based on the actual measured qualities. However, analysers have inherent measurement errors. Therefore, the blend plan and the closed loop blending need to consider uncertainties in the component qualities in order to guarantee on-spec products. Nonlinear nature of gasoline blending, and the multiperiod planning model make gasoline blend planning a mixed integer nonlinear problem.

One approach to solve the gasoline blend planning problem (MINLP) is to compute the blend recipes and the blended volume of gasoline products at each period simultaneously. Such approach is computationally expensive and might lead to suboptimal solution for

large scale problem. Therefore, this approach has been limited to use approximate linear formulation to simplify the model. (Li et al., 2009) developed a slot-based MILP formulation using blending indices and linear blending correlation to address nonlinear blending properties and solve the gasoline blend planning problem. Another approach which utilizes inventory pinch was introduced by (Castillo and Mahalec, 2014a) which decomposes the problem into two levels solve in sequence. At the top-level, number of periods are delimited by inventory pinch points and amount to blend at each of these periods are fixed. In this level, the solver computes the optimal blend recipes during each first level period. The lower-level typically has higher number of periods, and it computes amounts of volume blended for each product during each second level periods using the blend recipes computed at the top-level.

Stochastic optimization is mainly used when full knowledge about probability distribution of the uncertain parameters exist; this information is used to obtain a less conservative solution (Yang and Barton, 2016). (Monder, 2001) incorporated uncertainty in gasoline blend optimization problem using probabilistic programming. Single probability constraint was used to deal with uncertain parameters in multiple constraints. Monder designed a real-time optimization layer that deals with the disturbance (uncertainty) in the components qualities for linear and nonlinear blending rules. (Zhang et al., 2002) extended the previous results by dealing with uncertain parameters in multiple constraints as a joint probability constraint.

This work deals with gasoline planning (off-line blend recipe optimization over extended time horizon) under uncertainty in components blending qualities utilizing chance constrained formulation. It uses a two-level algorithm that utilizes inventory pinch concept to solve the blend planning problem with uncertainties. We compare this solution method to solving the problem as a single level full space model to evaluate the effectiveness of using inventory pinch algorithm versus the full space algorithm.

2. Problem Statement

The system studied has nine feedstocks components produced by the upstream processes and stored in individual component tanks. Components are sent to blenders where mixing occurs to produce four different products with different qualities specifications. Products are sent from blenders to eleven storage tanks before being shipped to satisfy the required demands. For the linear case, aromatic content (ARO), benzene content (BI), olefin content (OLF), research octane number (RON), rapid vapor pressure (RVP), Specific gravity (SGI), flammability limit (FI) are blended linearly on a volume basis, while oxygenates limit (OXI) and sulfur content (SI) are blended linearly on a weight basis. For the nonlinear case, research octane number (RON) is assumed to blend as described by a nonlinear model and the rest of the qualities are the same as the linear case. The objective of the model is to compute optimal blend recipes and volume blended for all products along the planning horizon when component qualities are uncertain. Optimal blend recipes correspond to the recipes that minimize the blending cost while meeting required probability of meeting products specifications. The model has tanks capacity constraints, products quality specifications constraints, and blending capacity constraints.

3. Solution Approach

Gasoline blend planning problem can be formulated as a full space MINLP model with uncertainty in components qualities for fixed discretized periods along the planning horizon. Such model faces difficulties in computing an optimal solution for large scale

cases due to the nonlinearities in the blend recipe calculation, some of them stemming from the nonlinear qualities blending while the others are due to the chance constraint formulation. The proposed approach to solve the gasoline blend planning problem with uncertainty utilizes inventory pinch concept, aggregation and disaggregation techniques to compute optimal blend recipes in shorter executions times by decomposing the problem into two smaller subproblems. The inventory pinch points (the boundaries of the first level periods) can be identified by constructing a cumulative total demand (CTD) curve. Then a cumulative total average production (CATP) curve is constructed as a straight line with a starting point representing initial inventories at the beginning of the planning horizon and extending to the point where the line is tangent to the CTD curve. The point where both curves intersect is called the inventory pinch point. At inventory pinch point, the slope of the CATP curve changes as shown in Figure 1. Hypothesis is that the optimal blend recipe for a time period between the inventory pinch points is constant. At the top-level chance constrained programming formulation is implemented to account for uncertainties in the component qualities. The deterministic quality constraints are presented by Eq. (1), where k, p, i, s, and S refers to the period, product, component, quality and uncertain quality sets respectively. The deterministic constraint is transformed to account for uncertainty in components qualities ($\widehat{Q_{i,s}^{comp}}$) for linear blending case is presented in Eq. (2) and its equivalent under normal distribution assumption is presented in Eq. (3). Where α is the probability of meeting the constraint and $\Phi(\alpha)$ is the α-quantile of the standard normal distribution.

$$\sum_i \left(V_{i,p,k}^{comp} \, Q_{i,s}^{comp} \right) \geq V_{p,k}^{blend} \, Q_{p,s}^{prod,min}, \ \forall p, s \epsilon US, k \tag{1}$$

$$\mathrm{P}\left\{ \sum_i \left(V_{i,p,k}^{comp} \, \widehat{Q_{i,s}^{comp}} \right) \geq V_{p,k}^{blend} \, Q_{p,s}^{prod,min} \right\} \geq \alpha \ \ \forall p, s \epsilon US, k \tag{2}$$

$$\sum_i \left(V_{i,p,k}^{comp} \widehat{Q_{i,s}^{comp}} \right) + \Phi(\alpha) \sqrt{\sum_i (V_{i,p,k}^{comp\,2} \sigma_{i,s}^{comp\,2})} \geq V_{p,k}^{blend} Q_{p,s}^{prod,min}, \forall p, s \epsilon US, k \tag{3}$$

The solution of this level gives optimal blend recipes for each grade of the gasoline products while minimizing the total cost of blending. At the bottom-level, an integer linear programming problem (MILP) is solved to using blending recipes calculated at top-level to determine amount produced of each product in each unit in each period. At this level the planner defined fixed length for the periods. If the blend recipes calculated at the top-level result in inventory infeasibilities at the bottom level, then the top-level period corresponding to the point of infeasibility is subdivided and the blend recipe at top level is re-optimized.

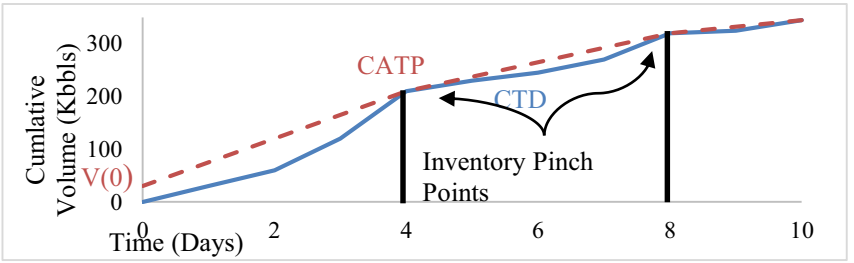


Figure 1: Example of inventory pinch points obtained from constructing CTD and CATP curves

In this paper, chance constrained programming is used to take uncertainty in qualities of the blending components into account. The chance constraints formulation in this paper assumes all uncertain parameters to follow normal distribution and the user knows the mean and standard deviation of all these parameters. The chance constrained formulation

seeks to satisfy the product qualities constraints at a predefined confidence interval chosen by the operator. Chance constrained can be introduced directly for the linear blending rules model. While, introducing chance constrained formulation into the nonlinear blending rule model for RON quality requires linearization of the constraints with respect to the uncertain parameters around their mean values.

4. Results & Discussion

In this work, test problems from (Li and Karimi, 2011) are used. The test problems were solved based on two models. The first model used assumed linear blending rules for all qualities utilizing blending index method. While, the second model assumed nonlinear blending rules for RON quality based on ethyl RT-70 equations. The two different models were solved using both full space MINLP and inventory pinch algorithm to compare the effectiveness of using the inventory pinch algorithm versus the traditional full space algorithm. (Li and Karimi, 2011) examples do not specify MON values for the components which are required to calculate products RON quality when ethyl RT-70 equations are used. Therefore, MON values for the blending components were assumed but MON quality was not included in the product quality specification constraints. Uncertainty in RON quality is assumed for linear blending rule case and uncertainty in both RON and MON are assumed for nonlinear blending rule case, while all other qualities are assumed to have fixed and known values. The standard deviation of all RON and MON are assumed to equal 1 % from their mean values.

All case studies were implemented in GAMS 24.4.1 software. CPLEX 12.6 and ANTIGONE 1.1 solvers were used to solve linear and nonlinear models. All problems were solved on HP Notebook 1703 (Intel ® Core™ i7-6500U CPU, 2.50 GHz, and 8.0 GB RAM) running Windows 10 Home. For linear blending rules cases, a 0.01 % optimality gap or maximum computational time of 3600 seconds is the stopping criteria, while 0.1 % optimality gap or maximum computational time of 10800 seconds is used as stopping criteria for nonlinear blending rule cases.

Table 1 shows the model size of linear and nonlinear blending rule cases for the full-space model and the inventory pinch top and bottom levels model. Figure 2 shows the relationship between blending cost and probability of meeting products specifications for linear and nonlinear blending rules. Higher probability of meeting products specifications results in higher blending cost. Nonlinear blending rules model gives a higher blending cost compared to linear blending rules model and the extra blending cost of incorporating uncertainty is higher for nonlinear blending because the nonlinear blending model is affected by uncertainty in RON and MON qualities, while the linear blending model is affected by RON uncertainty only.

Table 1: Full space and inventory pinch models size for linear and nonlinear blending rules

Example ID	Linear blending rules	Nonlinear blending rules
Full space (MINLP for uncertainty cases)		
# periods	4	4
# equations	3467	4175
# continuous variables	941	2041
# discrete variables	371	371
# nonlinear terms	576	1408
Top-level Pinch Point (NLP for uncertainty cases)		
# periods	2	2
# equations	488	574
# continuous variables	238	272
# discrete variables	0	0
# nonlinear terms	144	208

Bottom-level Pinch Point (MILP)		
# periods	4	4
# equations	2323	2323
# continuous variables	1399	1399
# discrete variables	371	371
# nonlinear terms	0	0

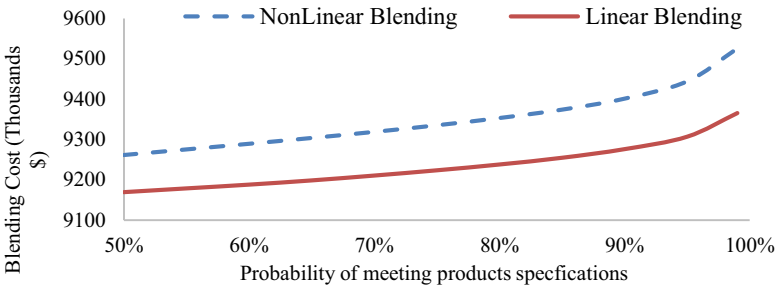


Figure 2: Blending cost versus probability of meeting products specs.

Table 2 shows the results obtained from solving linear and nonlinear blending rules models using both full space and inventory pinch algorithms.

Table 2: Inventory pinch and full space models results for linear and nonlinear blending rules

Nonlinear Blending Rules				Linear Blending Rules						
99	95	90	50	99	95	90	50	Probability of meeting products specifications		
9525.72	9444.28	9401.34	9261.63	9365.56	9307.23	9276.38	9169.91	Objective function (thousands$)		
2.85	1.97	1.51	-	2.13	1.50	1.16	-	Increase in blending cost (%)		
0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	GAP (%)		Pinch Point
21.8	16.7	16.5	1.28	0.17	0.24	0.25	0.37	Optimal solution	CPU time to reach (s)	Pinch Point
1	5	4	1	0.17	0.24	0.25	0.37	Final upper bound	CPU time to reach (s)	Pinch Point
0.1	0.1	10.3	0.93	1.66	1.10	0.92	0.01	GAP (%)		Full Space
8202	1864	10800	10800	3600	3600	3600	0.262	Optimal solution	CPU time to reach (s)	Full Space
257	227	268	1	5	5	5	0.26	Final upper bound	CPU time to reach (s)	Full Space

The nonlinear blending rules model are harder to solve due to the nonlinearity added for the RON quality computation. The results show that computation of optimal solution and closing the optimality gap can be done for all cases when inventory pinch is used, while full-space took longer time to compute the optimal solution and failed to close the optimality gap for some of the cases.

5. Conclusion

Gasoline blend planning (off-line optimization) problem under uncertainty in component qualities is solved using chance constrained formulation. Chance constrained formulation deals with uncertainty in components qualities by giving some qualities away to ensure products blended are on-specs products. By giving quality away, the planner is willing to operate at higher blending cost at the benefit of minimizing the risk of reblending off-specs products which can result in high additional cost. Incorporating chance constrained formulation into the gasoline blend planning model increase the nonlinearity terms making it more complex and harder to optimize. For the linear blending rule model, inventory pinch algorithm could compute the optimal solution in less than a second, while Full space model required at least twenty-fold that time to reach the optimal solution and failed to close the optimality gap after one hour. For the nonlinear blending rules, inventory pinch algorithm could compute the best solution and closed the optimality gap within seconds, while full space algorithm needed forty-fold the times to reach the optimal solution and failed to close the gap for some cases after three hours. Although results for large scale case are not included in the report, the results showed that full space algorithm might fail in computing a feasible solution after hours of running, while the inventory pinch algorithm can achieve the optimal solution and close the optimality gap within an hour. Results of this work indicate that the inventory pinch based approach is likely to be effective in solving the refinery planning problems with uncertainties in the crude qualities and the components qualities. Therefore, future work will investigate implementing similar approach to the refinery model.

References

Castillo, P. A. C., & Mahalec, V. (2014). Inventory pinch based, multiscale models for integrated planning and scheduling-part I: Gasoline blend planning. *AIChE Journal*, *60*(6), 2158–2178. https://doi.org/10.1002/aic.14423

Li, J., & Karimi, I. A. (2011). Scheduling Gasoline Blending Operations from Recipe Determination to Shipping Using Unit Slots. *Industrial & Engineering Chemistry Research*, *50*(15), 9156–9174. https://doi.org/10.1021/ie102321b

Li, J., Karimi, I. A., & Srinivasan, R. (2009). Recipe determination and scheduling of gasoline blending operations. *AIChE Journal*, *56*(2), NA-NA. https://doi.org/10.1002/aic.11970

Monder, D. S. (2001). Real-Tome Optirniration of Gasoline Blending with Uncertain Parameters. Retrieved from http://www.nlc-bnc.ca/obj/s4/f2/dsk3/ftp04/MQ60472.pdf

Yang, Y., & Barton, P. I. (2016). Integrated crude selection and refinery optimization under uncertainty. *AIChE Journal*, *62*(4), 1038–1053. https://doi.org/10.1002/aic.15075

Zhang, Y., Monder, D., & Fraser Forbes, J. (2002). Real-time optimization under parametric uncertainty: A probability constrained approach. *Journal of Process Control*, *12*(3), 373–389. https://doi.org/10.1016/S0959-1524(01)00047-6

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50250-0

Software for Creating Stochastic Scenarios for Optimization from Data

Andrea Staid[a], David L. Woodruff[b]

[a]*Discrete Math & Optimization, Sandia National Laboratories, Albuquerque, NM 87185-1326, USA*
[b] *Graduate School of Management, University of California, Davis, Davis, CA 95616-8609, USA*
dlwoodruff@ucdavis.edu

Abstract

Coefficients and constraint values for many optimization problems are best modeled as uncertain. Examples include reaction rates for chemical processes, demand for products in a supply chain, and the availability of renewable energy in a power grid. General purpose software is available for solving the resulting stochastic optimization problems given probabilistic scenarios as input [Hart et al., 2011, Hart et al., 2017, Watson et al., 2012]. In this research, we describe software that starts with data and produces these inputs. One of the motivating applications for the software is creating the stochastic inputs for the Institute for the Design of Advanced Energy Systems (IDAES). Other recent applications come from incorporation of renewables in the energy grid.

Here are the design goals and features of the software:

1. Allow for separation of the work of optimization modeling from the work of stochastic process and uncertainty modeling.
2. Support multi-stage optimization fully from the start.
3. Support "raw" scenario data that could be read by non-Python programs.
4. Provide a lot of flexibility.
5. Allow for easy use of data directly by deterministic models.

Keywords: Stochastic optimization; scenario creation; Uncertainty quantification.

1. Scenario Sets for Optimization

We present information and examples for the use of a software package called *daps*, which stands for 'Data to PySP.' This software is designed to use real, available data to create scenarios for use in stochastic optimization problems.

1.1. Optimization Under Uncertainty

We use x to represent the vector of variables whose values are set by the optimization process and use x^t to represent the part of the decision vector that corresponds to stage t. Only the first stage decisions can be implemented immediately, but of course additional stages must often be considered in order to make a good first-stage decision. We use $\vec{x}^t$ for $1 \leq t \leq T$ to represent the decisions for all stages up to, and including, stage t. We decompose the objective function by stages and for the first stage we write $f_1(x^1)$ and for subsequent stages $f_t\left(x^t; \vec{x}^{t-1}, \vec{\xi}^t\right)$, where ξ is the problem data and $\vec{\xi}^t$ is gives the values that will become known by time t.

This notation allows us to write the optimization problem as

$$\min_x f_1\left(x^1\right) + \mathbb{E}\sum_{t=2}^{T} f_t\left(x^t; \vec{x}^{t-1}, \vec{\xi}^t\right) \tag{1}$$

We use the notation $\mathbb{E}$ to indicate that an expectation will be taken over possible realization of ξ.

Some risk measures are easy to add directly. For example, the expected value of the $1-\alpha$ tail is known as CVaR (conditional value at risk) or sometimes TCE (tail conditional expectation). This can be implemented by augmenting this problem following the implementation given by Schultz and Tiedemann [Schultz and Tiedemann, 2005].

1.2. Scenarios

For practical reasons, we typically work with discrete scenarios. Define $\xi \equiv \{\xi(t)\}_{t=1}^{T}$ on a discrete probability space and use Ξ to represent the full set of scenarios. Each scenario, ξ, has probability π_ξ We organize ξ into a tree with the property that scenarios with the same realization up to stage t share a node at that stage. So, we note that $\vec{\xi}^t$ refers also to a node in the scenario tree. Let $\mathcal{G}_t$ be the set of all scenario tree nodes for stage t and let $\mathcal{G}_t(\xi)$ be the node at time t for a particular scenario, ξ. For a particular node $\mathcal{D}$ let $\mathcal{D}^{-1}$ be the set of scenarios that define the node.

This allows us to re-write the optimization problem in terms of the scenario tree as follows:

$$\min_{x,\hat{x}} \sum_{\xi\in\Xi} \pi_\xi \left[f_1(x^1(\xi)) + \sum_{t=2}^{T} f_t\left(x^t(\xi); \vec{x}^{t-1}, \vec{\xi}^t\right)\right] \tag{2}$$

$$\pi_\xi x^t(\xi) - \pi_\xi \hat{x}^t(\mathcal{D}) = 0, \quad t = 1,\ldots,T-1,\ \mathcal{D}\in\mathcal{G}_t,\ \xi\in\mathcal{D}^{-1} \tag{3}$$

Non-anticipativity must be enforced at each non-leaf node, which is done using $\hat{x}$ (with the same dimension as x) in this formulation.

Methods for obtaining scenarios include statistical models, perhaps obtained by data mining. Once a model has been fit, scenarios can be obtained by Monte Carlo sampling (see, e.g., [Löhndorf, 2016]) or moment matching (see, e.g.,[Høyland and Wallace, 2001]) Other ways include simulations (see, e.g.,[Watson et al., 2010]) and forecast error distributions when a forecast and a history of forecasts is available (see, e.g.,[Rios et al., 2015, Staid et al., 2017].)

1.3. Small Example

To illustrate some of the inputs, we will make use of the well-known "farmer" example from Birge and Louveaux [Birge and Louveaux, 1997]. The details of the model are not important to us here, but in order to understand the examples, one needs to know the following. Variables are vectors indexed by the three crops: Corn, Wheat and Sugar Beets. The variables are named DevotedAcreage, QuantitySubQuotaSold, QuantitySuperQuotaSold, and QuantityPurchased. Note that in Pyomo, an optimization model variable is called a `Var` in part to make a distinction with Python variables and random variables. The reserved word `Param` is used for some optimization model parameters. The uncertain optimization model parameter vector is a `Param` called Yield and it is also indexed by the three crops.

2. Using the Software

The software works with PySP, which is part of Pyomo [Hart et al., 2017]. Here, we present a list of items required to use the daps software.

- The Model: As one would expect, the modeler must provide a model file that is often called `ReferenceModel.py` and it must contain a cost `Expression` for each of the stages.
- Tree Template: The template provides the model-specific information about the scenario tree (as distinct from the scenario-specific data). Although there are other ways to specify a scenario tree template, this software presently makes use of two methods: 1) json and 2) AMPL [Fourer et al., 2003] format. The modeler must supply the stage names, the assignment of each `Var` to a stage and the names of the cost expressions for each stage. This can be done using a json file that contains a dictionary with outer keys that are PySP reserved words Stages, StageVariables, and StageCost. Wildcards are supported within the StageVariables declaration of each `Var` for each stage. Here is a sample json tree template file for farmer:

```
{
   "Stages": ["FirstStage", "SecondStage"],
   "StageVariables":
   {"FirstStage": ["DevotedAcreage[*]"],
     "SecondStage": [
 "QuantitySubQuotaSold[*]",
          "QuantitySuperQuotaSold[*]",
          "QuantityPurchased[*]"
     ]
   },
   "StageCost": {
"FirstStage": "FirstStageCost",
          "SecondStage": "SecondStageCost"
   }
}
```

- Scenario template data: This provides data needed by the model that does depend on the individual scenario. This is a flexible concept that might be a full scenario, might be root node data only, or might be empty. It is the starting point for every scenario and when the scenario is created, data for the scenario will be added to the template and/or replace template data. The template could be in a variety of formats, such as an AMPL format data file, a JSON file or a csv file, etc. The details of how it gets used depends on the format.
- Raw Scenario Tree Node Data: This stores the data for each scenario, and we use a tree structure to support multi-stage problems. For two-stage problems, nodes other than the root node are one-to-one with scenarios. These data need to find their way into a dictionary whose elements are either data or dictionaries. For abstract models using AMPL format input files, the dictionary names are intended to be `Param` names. For concrete models (and for use by abstract model build actions) they are intended to be whatever is needed by the modeler. If the elements are a dictionary, the indexes in the inner dictionary are intended to be the entire index string for a `Param` or else whatever they should be. When these data are supplied in a file, the goal is to have a lot of flexibility in the format and to be sure that packages written in other languages would have a shot

at using them if they wanted to. Furthermore, we want to be sure that for two-stage concrete models the files can be used directly as deterministic input if that is desired. Hence, we are coding the scenario tree in the file names:

NODE-name-PARENT-name-PROB-prob.ext

where the upper case tokens are literals and must be present and ext signals the type of data in the file (e.g., json). The special node name ROOT is assumed to exist unless a file is present with the special parent name NONE, which then defines the root node (it will be unusual for a raw root node data file to be created, but that is an option). Note that the numeric value `prob` is the conditional probability.

There is nothing stopping such a file from having a full scenario, but it is only responsible for being correct up to its node. When a scenario is assembled it is done in stage order.

3. Creating Node Data

Statistics and stochastic process modeling are needed in order to create node data. Exact details must always depend on the particular application, but the software presently provides some examples.

3.1. Sampling

We describe two examples now, both assume the Python program contains the following line:

```
import distr2pysp as dp
```

A sensible way to create scenarios in some settings is to fit a continuous probability distribution to data observations and then sample from that distribution. The first example shows how to use *daps* to read data given in a json file (datafiledict.json) to populate a scenario tree given in TreeTemplateFile.dat. It creates four scenarios and puts the resulting input files in dptest.

```
dp.indep_norms_from_data_2stage('concrete_farmer/dptest/datafiled
ict.json',
          'concrete_farmer/TreeTemplateFile.dat',
          4,
          'concrete_farmer/dptest',
          Seed = 7734)
```

But one might also fit the data using other software (e.g. R) and then provide the distribution information to *daps*, which uses *scipy* distributions described in distrdict.json. This time, only three scenarios are created, and unlike the previous example, the pseudo random number seed is not controlled.

```
dp.json_scipy_2stage('concrete_farmer/dptest/distrdict.json',
          'concrete_farmer/TreeTemplateFile.dat',
          3, 'concrete_farmer/dptest')
```

Here is `distrdict.json`

```
{"Yield":
  {"CORN": ["norm", {"loc" : 3, "scale" : 0.5}],
   "WHEAT": ["norm", {"loc" : 2.9, "scale" : 0.6}],
   "SUGAR_BEETS": ["norm", {"loc" : 24, "scale" : 4.5}]
  }
}
```

3.2. Importance Sampling

The expectation shown in Expression 1 is implemented by summing probabilities over a discrete probability space in Expression 2, which brings to mind ideas from numeric

integration such as importance sampling [Barrera et al., 2016, Infanger, 1992, Parpas et al., 2015]. The idea is that instead of drawing samples from the distribution of ξ as suggested in the previous subsection, one draws samples from a distribution where the objective function is likely to take extreme values. For the applications of interest to us, one can often anticipate which sorts of values those might be (e.g., large or small values, or both) based on engineering judgment. This enables a very simple form of importance sampling based on cutting points of a distribution that is sampled to obtain discrete values ξ. Previous sections used $f_t(\cdot)$ to refer to the objective function. In this subsection, $f(\cdot)$ refers to the probability density function (pdf) when ξ values could be directly sampled. $F(\cdot)$ is the corresponding cumulative distribution function (CDF). The symbol f is commonly used in these two contexts, although the functions $f_t(\cdot)$ and the function $f(\cdot)$ have nothing to do with each other.

We now describe the univariate version. We are given a list of C cutting points c_i. We assume here that $c_0 = 0$ and $c_C = 1$ and have in mind a cutting point list like (0, 0.1, 0.9, 1) or (0, 0.1, 0.9, 0.99, 1) or some other list that tends to emphasize the tails. We then obtain a list of region boundaries $v_i = F^{-1}(c_i)$. For long tailed distributions of ξ, the values of v_0 and v_C might need to be approximated to avoid excessively wide intervals (e.g. they might be placed five or six standard deviations from the mean instead of at zero and one respectively).

We select from among the regions with probability $1/C$ regardless of the width of the region; this is what enables emphasis of tails. Within the region we select a sample point uniformly; we refer to the pdf for this sampling distribution as $g(\cdot)$. For one of N samples, ξ, drawn from g, the corresponding probability with respect to expectations of f is

$$\frac{1}{N}\frac{f(\hat{\xi})}{g(\hat{\xi})}.$$

In practice, we rescale the probabilities so that they sum to exactly one. Once a region, i, is selected and a point is generated within the region according to a uniform distribution, the pdf for the point is

$$(C(v_i - v_{i-1}))^{-1}$$

4. Conclusions and Future Directions

We have described software that starts with data and produces inputs for optimization in the presence of uncertain data. We have described some of the downstream support to get the scenarios to PySP and some of the upstream support to create scenario data. Much of our present research centers on adding upstream support to create scenario data for a range of applications including parameter estimation and chemical process optimization.

An important area of further research involves using *daps* to facilitate scenario creation based on data from parameter estimation experiments, which explicitly recognizes that there is uncertainty around parameter values. This research has the potential to be interesting, since stochastic programming formulations can be used to obtain the parameter estimates [Word et al 2012] which are then intended for use in subsequent stochastic programs.

References

Barrera, J., Homem-de Mello, T., Moreno, E., Pagnoncelli, B. K., and Canessa, G. (2016). Chance-constrained problems and rare events: an importance sampling approach. Mathematical Programming, 157(1):153-189.

Birge, J. and Louveaux, F. (1997). Introduction to Stochastic Programming. Springer.

Fourer, R., Gay, D. M., and Kernighan, B. W. (2003). AMPL: A Modeling Language for Mathematical Programming, 2nd Ed. Brooks/Cole-Thomson Learning, Pacific Grove, CA.

Hart, W., Watson, J., and Woodruff, D. (2011). Pyomo: Modeling and solving mathematical programs in Python. Mathematical Programming Compuation, 3(3).

Hart, W. E., Laird, C. D., Watson, J.-P., Woodruff, D. L., Hackebeil, G. A., Nicholson, B. L., and Siirola, J. D. (2017). Pyomo. Springer.

Høyland, K. and Wallace, S. (2001). Generating scenario trees for multistage decision problems. Management Science, 47(2):295-307.

Infanger, G. (1992). Monte Carlo (importance) sampling within a benders decomposition algorithm for stochastic linear programs. Annals of Operations Research, 39:41{67.

Löhndorf, N. (2016). An empirical analysis of scenario generation methods for stochastic optimization. Technical report, WU Vienna University of Economics and Business.

Parpas, P., Ustun, B.,Webster, M., and Tran, Q. K. (2015). Importance sampling in stochastic programming: A Markov Chain Monte Carlo approach. INFORMS Journal on Computing, 27:358{377.

Rios, I., Wets, R. J.-B., and Woodruff, D. L. (2015). Multi-period forecasting and scenario generation with limited data. Computational Management Science, 12(2):267-295.

Schultz, R. and Tiedemann, S. (2005). Conditional Value-at-Risk in stochastic programs with mixed-integer recourse. Mathematical Programming, 105(2-3):365-386.

Staid, A., Watson, J.-P., Wets, R. J.-B., and Woodruff, D. L. (2017). Generating short-term probabilistic wind power scenarios via nonparametric forecast error density estimators. Wind Energy, 20(12):1911-1925.

Watson, J.-P., Hart, W. E., Woodruff, D. L., and Murray, R. (2010). Formulating and Analyzing Multi-Stage Sensor Placement Problems. ASCE.

Watson, J.-P., Woodruff, D., and Hart, W. (2012). PySP: Modeling and solving stochastic programs in Python. Mathematical Programming Computation, 4(2).

Word, D.P, Watson, J.-P., Woodruff, D.L., and Laird, C.D. (2012). A Progressive Hedging Approach for Parameter Estimation via Stochastic Nonlinear Programming, in Comp. Aided Chem. Eng. Editor(s): Iftekhar A. Karimi, Rajagopalan Srinivasan. Elsevier. 31:1507-1511

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50251-2

Scheduling of Seawater Reverse Osmosis Desalination under Two Kinds of Uncertain Factors

Jian Wang[a*], Chao Zheng[a], Wentao Shen[a], Chunhua Zhang[b], Haokun Wang[a], Aipeng Jiang[a]

[a]*School of Automation, Hangzhou Dianzi University, Xiasha Higher Education Zone, Hangzhou 310018, China*
[b]*School of Mechanical Engineering, Hangzhou Dianzi University, Xiasha Higher Education Zone, Hangzhou 310018, China*
wj@hdu.edu.cn

Abstract

This paper proposes a stochastic optimization framework for scheduling of a large-scale parallel-unit seawater reverse osmosis (SWRO) desalination plant. Two kinds of uncertainties, water consumption fluctuation and RO units malfunction, are considered. A combination scheduling frame including two routes, a robust pre-scheduling and a reactive rescheduling, is built to handle these two uncertainties. Both optimal scheduling problems are built as change constrained models, and the Monte Carlo method is used to solve the uncertainty problems. An improved genetic algorithm is constructed to compute the efficient solutions. The simulation results show that pre-scheduling can greatly overcome the effect of water consumption fluctuations, and the reactive rescheduling can improve the behaviours of dynamic operation.

Keywords: optimal scheduling, rescheduling, change constraint, Monte Carlo method

1. Introduction

The seawater reverse osmosis (SWRO) desalination is one of the most useful technology to alleviate the shortage of fresh water in coastal areas (Gao, 2016). A large-scale SWRO desalination plant is usually structured with a number of independent RO units in parallel (Figure 1, Wang, 2015). The optimal scheduling is necessary for a large-scale SWRO plant to reduce the operating cost (Jiang, 2015). However, there are many uncertainties in the desalination process. It is a challenge to make an excellent scheduling to deal with the uncertainties.

2. Pre-scheduling

Changing of freshwater demand and the stoppages of RO units are two typical uncertainties in the SWRO process. In order to cope with the uncertainty of freshwater demand, an optimal pre-scheduling is built.

2.1. Objective function

The objective function of pre-scheduling problem is the total cost of production, which includes energy consumption C_{en} and maintenance C_{mn}. The purpose of pre-scheduling is to minimize the objective function f.

$$\min\ f = C_{en} + C_{mn} \tag{1}$$

Where, the energy consumption C_{en} mainly considers power consumption (Eq.(2)), and the maintenance cost C_{mn} includes the maintaining expense when the machine runs and the cleaning and repairing expense when the machine stops (Eq.(3)).

$$C_{en} = \sum_{k=1}^{K} \left\{ Pe(k) \sum_{i=1}^{n} \left(\alpha_i(k) \times c_3 \times Q_i(k) \right) \right\} \tag{2}$$

$$C_{mn} = \sum_{k=1}^{K} \sum_{i=1}^{n} \left(c_1 \cdot \alpha_i(k) \cdot Q_i(k) + c_2 \cdot (1 - \alpha_i(k)) \right) \tag{3}$$

Where, K is the time period of scheduling, n is the number of RO units. $Pe(k)$ is electricity price at time k, and $\alpha_i(k)$ is the on/off status of unit i at time k, $\alpha_i(k)$=1 when the unit i is running; $\alpha_i(k)$=0 when it stops. $Q_i(k)$ is the amount of freshwater generated by unit i at time k. c_1 is the coefficient of maintaining expense which is proportional to $Q_i(k)$; and a constant c_2 during cleaning and repairing period, c_3 is the coefficient used to calculate the energy cost.

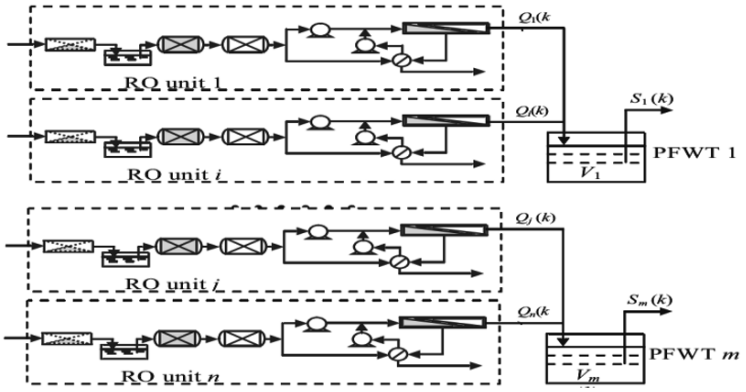


Fig. 1. Schematic diagram of a large-scale SWRO desalination plant

2.2. Constraints

The change constraint programing model is used to solve the uncertainty problem, and a confidence level is given to indicate the credibility of solution.

2.2.1 Freshwater supplement constraint

$$\Pr\left[\sum_{j=1}^{m} w_j(k) \geq D(k) \right] \geq \beta_1, \text{ where } D(k) = \xi_k, \quad \forall k \tag{4}$$

Where, $w_j(k)$ is the maximum water supply of produced freshwater tank (PFWT) j at time k; $D(k)$ is the consumption of freshwater at time k, which is a random variable ξ_k, m is the number of the PFWTs, β_1 is the confidence level of the constraint.

2.2.2 Capacity constraint of each PFWT

$$\Pr[V_j(k) \le V_j^U] \ge \beta_2 \quad j = 1, 2, \ldots, m; \forall k \tag{5}$$

$$V_j(k) = w_j(k) - S_j(k) + V_j^L \quad j = 1, 2, \ldots, m; \forall k \tag{6}$$

$$w_j(k) = V_j(k-1) + \sum_{i=1}^{n_j} \alpha_i(k) \cdot Q_i(k) - V_j^L \qquad j = 1, 2, \ldots, m; i = 1, 2, \ldots, n_i \tag{7}$$

$$S_j(k) = \frac{w_j(k)}{\sum_{i=1}^{m} w_i(k)} D(k) \quad j = 1, 2, \ldots, m; \forall k \tag{8}$$

Where, $V_j(k)$ is the capacity of PFWT j at the end of time k; V_j^U, V_j^L are the upper capacity limit and the lower capacity limit of PFWT j; $S_j(k)$ is the real amount of supplied freshwater by PFWT j at time k; n_j is the number of ROs at time k.

2.2.3 Constraints of produced freshwater

$$Q_i^L \le Q_i(k) \le Q_i^U, i = 1, 2, \ldots, n; \forall k \tag{9}$$

Where, Q_i^U and Q_i^L are the upper and the lower limit of the amount of RO unit i.

2.2.4 Robustness evaluation index

According to the change constraint programming, we define a robustness evaluation index γ_r to indicate the uncertainty of process, it is shown as Eq.(10).

$$\begin{aligned} &\Pr\{g_1(x^*, \xi, k) \wedge \ldots \wedge g_n(x^*, \xi, k) \le 0\} = \gamma_k \\ &\gamma_r = \min_k \gamma_k \ \forall k \end{aligned} \tag{10}$$

Where, γ_r is the robustness index of the scheme when the uncertain variable is ξ under the decision variable x^*.

3. Reactive scheduling

All the units will work according to the pre-scheduling plan. However, when unit malfunction happen, the pre-plan will not work well. A new reactive scheduling must be made to meet the changed situation.

3.1. Objective function

We use a bi-objective function in the reactive scheduling, shown by Eq.(11).

$$\min \quad F = \omega_1 \times F_{obj} + \omega_2 \times M \tag{11}$$

Where, F_{obj} is the economy goal of SWRO plant (shown by Eq.(12)) and M is the deviation between the pre-scheduling and the reactive one, given by Eq.(14), and ω_1 and ω_2 are the weights of F_{obj} and M respectively.

$$F_{obj} = C_{en} + C_{mn} + DF \tag{12}$$

Besides the energy cost C_{en} and the maintenance cost C_{mn}, a penalty value DF is added into the objective function of the reactive scheduling. That means that we do not hope to fail to meet the requirement of water consumption.

$$DF = p_w \cdot \sum_{k=1}^{K} \omega_{DF}(k) \max\left\{0, D(k) - \sum_{j=1}^{m} U(k)\right\} \tag{13}$$

Where, $\omega_{DF}(k)$ is the penalty factor, p_w is the average cost of pre ton production water.

$$M = \left|S_r - S_p\right| = p_w \cdot \sum_{k=k'}^{K}\left[\sum_{i=1}^{n}\left|\alpha_i'(k) \times Q_i'(k) - \alpha_i(k) \times Q_i(k)\right|\right] \tag{14}$$

Where, S_r is the rescheduling scheme which contains the new scheduling decision variables $\alpha'_i(k)$, $Q'_i(k)$; S_p is the pre-scheduling scheme which contains the pre-plan decision variables $\alpha_i(k)$, $Q_i(k)$; k' is the moment when the unit malfunction.

The constraints of reactive scheduling problem is similar to the pre-scheduling, to avoid repetition, it's not described here again.

4. Experimentation and discussion

The Monte Carlo simulation method is used to deal with the change constraints problem (Fan, 2015). An improved genetic algorithm (GA) is presented to calculate the results of change constraint. Two improvements are put forward to the GA: one is to keep the best 1/4 solutions into the next generation so that to keep the population on the optimal direction and maintain the diversity of species; another is to adjust the crossover rate and mutation rate from the adaptive function, which is calculated according to the individual fitness, the population fitness mean and the extreme value.

4.1. Simulation parameters

The actual water consumption at each hour is given by Eq.(15).

$$D_r(k) = D_p(k) + \Delta d,\ \Delta d \sim N(0, \sigma^2) \tag{15}$$

Where, $D_p(k)$ is the predictive value of water supply, Δd is the deviation between predicted and actual of water consumption, it is a random variables subject to normal distribution $N(0, \sigma^2)$.

The probability of unit malfunction is very small, so it is assumed that only one RO unit fails at the same time. The faulty unit is randomly selected from set [1, n]. The fault randomly occurs at the time m_1 and goes back to work at time m_2. The fault duration is uniformly distributed.

$$X = m_1 - m_2,\ X \sim U(a,b), a < b \tag{16}$$

Here, we take an example from "Wang, 2013" for numerical experimentation, which is a large-scale parallel-unit SWRO desalination plant with 8 RO units, 4 PFWTs and a

capacity of 100,000 m^3 freshwater a day. Different from "Wang, 2013", the maintaining plan of RO units is not considered in this paper.

4.2. Comparison of robust pre-scheduling and static scheduling

The parameters of the improved GA are set to: population size $N_{A,p}$=100, the maximum number of iterations $G_{A,m}$=100, mutation factor F_A =0.4, Crossover rate CR_A =0.8, the number of Monte Carlo simulation N_T = 1000, change constraint probability β_1=β_2=0.9, the standard deviation of fluctuations σ = 100.

Figure 2 shows the relation between production freshwater, freshwater demand and electricity price (underneath the Figure) under the robust pre-scheduling plan S_0. Figure 3 shows the total freshwater dynamic between the pre-scheduling plan S_0 and the static scheduling S_a (without considering the fluctuation of water consumption).

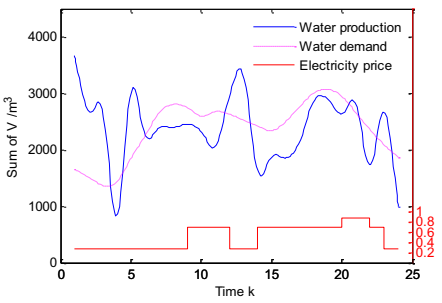


Figure 2: The relation between production water, water demand and electricity price under S_0 scheme

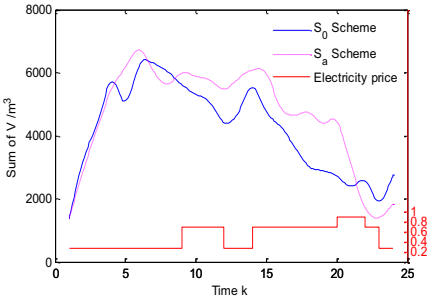


Figure 3: Comparison of total water volume with electricity price between S_0 and S_a

Figure 2 shows that the robust pre-scheduling has some predict ability to the electricity price. That is, when the price is low, the RO units work hard to produce much freshwater than the demand, and when the price is high, the production will go less than the demand, by supplying the pre storage of freshwater to meet the needs of freshwater. By this way, the electricity consumption reduces.

Figure 3 shows that the robust pre-scheduling has more sensitive to the electricity price and the fluctuation of water consumption than the static scheduling. That makes the former has less energy consumption than the latter.

4.3. simulation of reactive scheduling

Three cases of fault information are produced according to the probability of malfunction of the RO unit.

Table 1: Three sets of fault information

Case	Fault unit number	Failure time m_1	Recovery time m_2
1	4	18	20
2	6	7	12
3	5	13	17

Set $\omega_1=\omega_2=1$ are the parameters of multi-objective function, and reactive scheduling will be triggered when the malfunction happen and the maximum deviation of production freshwater demand is large than 10 % than pre-plan; penalty factor $\omega_{DF}(k)$ = 3.65; *Sp* means that the original scheduling scheme is still implemented after the failure happening; S_r means the reactive scheduling is triggered. The scheme comparison in each fault case is shown in Table 2.

Table 2: Comparison of S_p and S_r in 3 cases

Case	Cost of S_p	Cost of S_r	deviation degree M	Robust index γ_r
1	94552	93185	1088.4	0.80
2	99493	95279	1551.9	0.98
3	94066	92675	1386.2	0.98

Table 2 shows that reactive rescheduling reduces the total cost when the RO unit stoppage happens than the pre-scheduling plan, even considering the insufficient of water supply. The objective function of deviation degree *M* ensure the rescheduling plan changing in a small range from the pre-scheduling plan.

5. Conclusions

In this paper, a robust pre-schedule model with change constraint is proposed, which improves the stability of the system by sacrificing a certain amount of economy. A reactive rescheduling, taking into account both the deviations and the economy goal, is built to deal with the RO unit malfunction. An improved GA is used to compute the solution by using the Monte Carlo method. The simulation results show that the stochastic optimization framework proposed in this paper has high practicability.

Acknowledgments

This work was supported by the Natural Science Foundation of China (61374142) and the Natural Science Foundation of Zhejiang (LY16F030006).

References

A. Jiang, J. Wang, L. T. Biegler, W. Cheng, C. X. Xing, S. Jiangzhou, 2015, Operational cost optimization of a full-scale SWRO system under multi-parameter variable conditions. Desalination, 355, 124-140.

C. Gao, Y. Zhou, L. Liu, 2016, Present situation and Prospect of reverse osmosis seawater desalination technology, Journal of marine technology, 35(1), 1-14

J. Wang, X. Wang, 2013, Design of Monitoring and Scheduling System for a Large-scale Reverse Osmosis Desalination Plant, Proc. of 32nd CCC, 2480-2485

J. Wang, A. Jiang, L. Lian, G. Huang, 2015, A Rescheduling Approach for a Large-scale Reverse Osmosis Desalination Plant under Uncertain Fresh Water Demand, Computer Aided Chemical Engineering, 37:1991-1996

S. Fan, Q. Ai, X. He, 2015, Risk Analysis on dispatch of virtual power plant based on chance constrained programming, Proc. of the CSEE, 16, 4025-4034.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50252-4

Assessment of Accuracy and Computational Efficiency of Different Strategies for Estimation of Probability Distributions Applied to ODE/DAE Systems

Francesco Rossi[a,b*], Linas Mockus[a], Flavio Manenti[b], Gintaras Reklaitis[a]

[a]*Purdue University, Forney Hall of Chemical Engineering, 480 Stadium Mall Drive, West Lafayette, IN 47907-2100, United States*
[b]*Politecnico di Milano, Dipartimento Chimica, Materiali e Ingegneria Chimica "Giulio Natta", Piazza Leonardo da Vinci 32, Milano 20123, Italy*
frossi@purdue.edu

Abstract

This contribution employs a simulation-based approach to assessing the accuracy and computational efficiency of different strategies for estimation of probability distributions, applied to ODE/DAE systems. Specifically, two Bayesian Markov-chain Monte Carlo approaches are compared to a fast, new PDF estimation strategy, based on projection techniques. The two case studies analyzed confirm that this new PDF estimation strategy offers a very good trade-off between accuracy and computational efficiency, thus being excellent for time-critical PDF estimation tasks.

Keywords: Bayesian inference, PDF estimation, Reverse propagation of uncertainty.

1. Introduction

The application of statistical techniques to the quantification of model uncertainty is a new paradigm, which has recently emerged due to the growing interest of industry and of the PSE community in stochastic optimization frameworks, robust design strategies and quantitative risk assessment. Specifically, strategies for uncertainty quantification are commonly applied in areas such as robust process/product design (especially within the pharmaceutical sector) (Mockus et al., 2011), drug delivery (Laínez et al., 2011) and robust optimization/control of industrial processes (Rossi et al., 2016).

Typically, model uncertainty quantification comes down to the estimation of the joint probability distribution (PDF) of some key uncertain parameters of the model, which often is a system of differential equations (ODE) or differential-algebraic equations (DAE). Currently, there exist two principal types of strategies, applicable to the estimation of the PDF of the uncertain parameters of an ODE/DAE system: rigorous Bayesian inference coupled with random sampling approaches, e.g. Markov-chain Monte Carlo (Green and Worden, 2015), and approximate Bayesian inference exploiting optimization techniques, e.g. Variational Bayes (Beal, 2003) and frameworks based on the Laplace approximation (Dass et al., 2017). Although all of these techniques are well-established and commonly applied, they are usually very computationally demanding. Moreover, no systematic analyses have been conducted to assess their accuracy and computational efficiency, especially when ODE/DAE models must be dealt with.

Therefore, this contribution proposes a simulation-based comparison of two different PDF estimation strategies applied to ODE/DAE systems, namely, Bayesian Markov-chain Monte Carlo (BMCMC) and a new approach, named PDFE&U, which relies on a combination of fitting, back-projection techniques and maximum likelihood estimation. The level of accuracy and computational efficiency, attainable by these two methodologies, are evaluated by analyzing their outputs, especially their PDFs, and by comparing their computational times. The analysis of the PDFs is performed using both contour plots and well-known statistical indicators, i.e. expectation, variance, covariance, and quantiles. The specific ODE/DAE models used as benchmark systems for this analysis include a batch adaptation of the Tennessee Eastman Challenge problem and a pharmacokinetic/ pharmacodynamic model (PB/PK model).

The rest of the paper is organized as follows: first, we introduce the principal ideas, on which PDFE&U relies; then, we report the most interesting results of our analysis on the accuracy and computational performance insured by PDFE&U and BMCMC; finally, we discuss the most relevant consequences of such analyses.

2. Principal features of PDFE&U

PDFE&U is a strategy for estimating/updating the probability distribution of the uncertain parameters of a dynamic model using experimental data and, optionally, an existing PDF of the parameters (a prior). The rationale of this algorithm is summarized in Figure 1, where four distinct phases executed in series, namely, the initialization, Phase I, Phase II and Phase III, are indicated.

The initialization phase is the first task to be executed and serves to acquire user-supplied inputs and to set essential configuration options. The principal input data required by PDFE&U include one or more raw data sets containing measurement of the states, of the manipulated variables and of the disturbances of the target system, the system model formulated as a set of ODEs/DAEs, a parametric family of probability distributions used to approximate the joint PDF of the model parameters (a basis set), and, optionally, a prior (i.e. an existing PDF of the uncertain parameters).

After the initialization, the three steps included in Phase I are executed sequentially. First, we solve a conventional fitting problem and compute an initial estimate of the values of the uncertain parameters of the system model (Step A). Second, we scan the raw data sets for gross errors in the state measurements and eliminate them (Step B). The rationale used to identify gross errors is rather unconventional and relies on the sensitivity of the SSE (sum of squared errors) to this type of measurements. Finally, we project every state measurement, which has not been eliminated in Step B, onto the uncertainty space (Step C). This procedure is the core of Phase I, and is conceptually the inverse of the problem of propagation of a probability distribution through a system of ODEs/DAEs. It allows conversion of every state measurement into a corresponding set of parameters of the system model, i.e. a point belonging to the uncertainty space. Note that this back-projection procedure is the most computationally demanding task of PDFE&U, as it requires solution of a dynamic optimization problem for every state measurement, which must be projected. However, all these complex optimization problems are completely independent of each other, so parallel computing can be employed to drastically reduce the overall computational cost of their numerical solution.

The set of “pseudo-measurements” of the uncertain parameters, estimated in Step C of Phase I, can be easily converted into a probability distribution via maximum likelihood

estimation. This operation is performed in Phase II. Then, if no prior distribution has been supplied to PDFE&U in the initialization, Phase III is not executed (the algorithm is terminated at the end of Phase II). Conversely, if a prior distribution is available, we perform Phase III, which involves combination of the PDF, estimated in Phase II, with the prior, supplied by the user, to yield the posterior probability distribution of the parameters of the ODE/DAE model. In this last phase, we first discretize the two PDFs using an optimization-driven, incremental mapping approach, which relies on key level surfaces of the two distributions (Step A). Subsequently, we apply a modified maximum likelihood estimation strategy to combine the two discrete mappings and convert them into the posterior PDF of the parameters of the model (Step B). The relative weight, with which these two mappings are combined, is controlled by a bounded combination coefficient, which can alter the number of points in each mapping. This coefficient is treated as an additional optimization variable within the aforementioned, modified maximum likelihood estimation strategy.

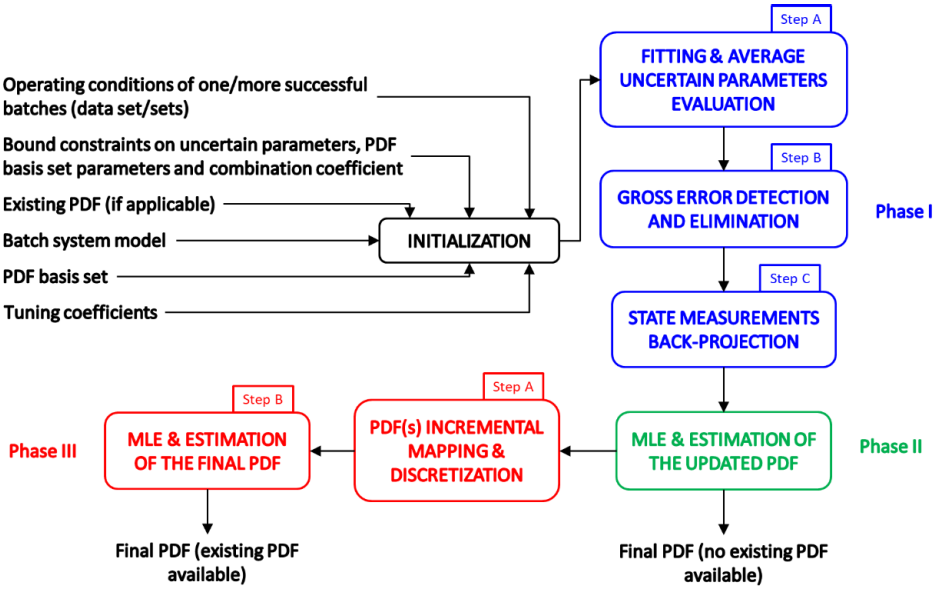


Figure 1: Architecture of the PDFE&U framework.

PDFE&U has been coded in C++ using a sequential, parallel implementation scheme, in which we make use of the BzzMath library (Buzzi-Ferraris and Manenti, 2012) to integrate the ODE/DAE model and solve the required optimization problems, and we exploit well-known multi-threading libraries to manage all the concurrent tasks.

Before moving on, note that, although brief, this outline of the rationale and implementation of PDFE&U should make it apparent that the proposed algorithm has little in common with conventional Bayesian inference strategies. In fact, it does not rely on Bayes theorem, thus does not require definition of a likelihood function. Moreover, it decomposes the estimation of the posterior PDF of the parameters of the system model into two stages, performed in series. We will see that it is this innovative decomposition strategy that yields a significant computational advantage over more conventional Bayesian inference approaches.

3. Performance and computational efficiency of PDFE&U and BMCMC

The computational performance and accuracy provided by PDFE&U and by two different tools for Bayesian Markov-chain Monte Carlo, i.e. BAT (Caldwell et al., 2009) and Stan (Carpenter et al., 2017), have been analyzed using two different case studies. The features of these two test problems forced the use of the two different BMCMC approaches.

The first test case involves estimation of the PDF of the uncertain parameters of the model of a fed-batch reaction process, which is a batch adaptation of the well-known Tennessee Eastman Challenge. The reactor model is comprised of 12 highly nonlinear ODEs, which include two uncertain parameters: the overall heat transfer coefficient (ν_U) and the activation energy of one of the main reactions (ν_{E1*}). The sets of measurements, used to estimate the joint PDF of ν_U and ν_{E1*} in several different scenarios, range from 170 to over 1000 data points.

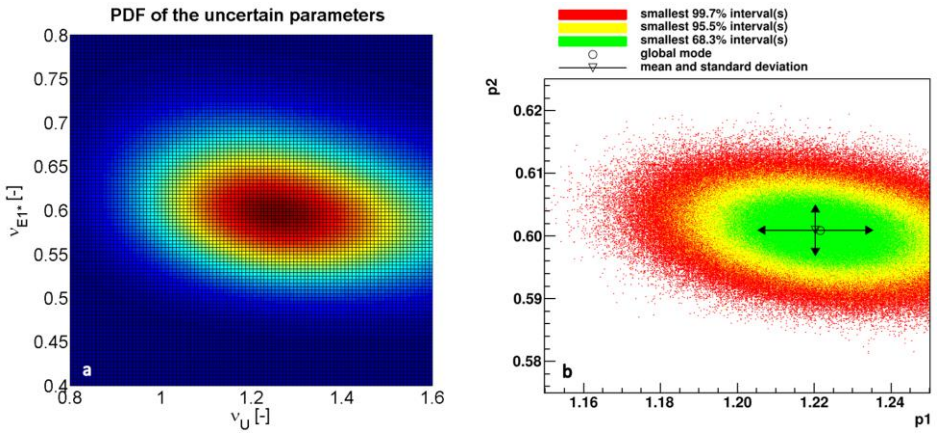


Figure 2: Contour plots of the PDFs yielded by PDFE&U (a) and by BAT (b) (the basis set used in PDFE&U is a bivariate lognormal distribution).

The results of one of these PDF estimation problems are reported in Figure 2, where we show the contour plots of the PDFs estimated by PDFE&U and BAT. Notice that PDFE&U tends to provide good estimates of central tendency measures (mode and expectation) and of the Pearson's correlation coefficient, but tends to overestimate variances. However, PDFE&U usually preserves variance ratios, thus can properly capture the qualitative shape of the probability distribution. Finally, note that a single PDFE&U run is approximately two orders of magnitude faster than a single run of BAT (Table 1). Both PDFE&U and BAT are C++ tools, so this is a fair comparison.

Table 1: Computational performance of PDFE&U, BAT and Stan.

Case study	Model size	Number of uncertain parameters	Time elapsed for solving a single PDF estimation problem [h]		
			PDFE&U	BAT	Stan
Batch reactor model	12 ODEs	2	0.05 – 0.4	6 – 24	-
PB/PK patient model	34 ODEs	8 – 10	0.3 – 0.6	-	≈ 120

The second case study involves the estimation of the joint probability distribution of the uncertain parameters of a PB/PK model, which simulates the response of a patient to the

administration of Gabapentin. The model is comprised of 34 nonlinear ODEs and includes either 8 or 10 uncertain parameters, depending on the characteristics of the patient. The data sets, used in these PDF estimation problems, range from 15 to 170 data points, and consist of measurements of plasma drug concentration.

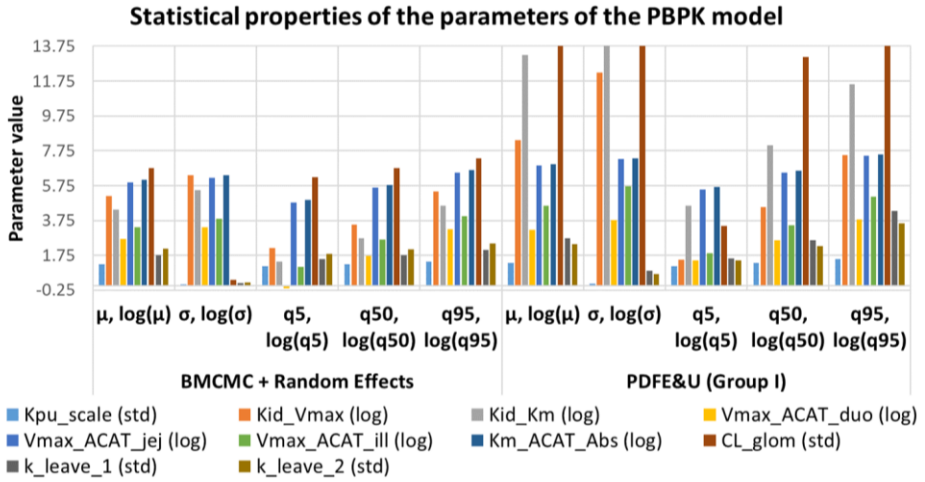


Figure 3: Statistical properties of the PDF of the parameters of the PB/PK model generated by PDFE&U and Stan (some quantities are reported in log-scale to improve the chart readability; μ, σ and qX stand for expectations, standard deviations and X % quantiles).

The results of one of the aforementioned PDF estimation problems are summarized in Figure 3 and Figure 4. In this scenario, the PB/PK model has 10 parameters, thus Figure 3 shows their most important statistical properties, computed by PDFE&U and Stan (BMCMC + Random Effects). In addition, Figure 4 shows the predicted enclosure of the temporal trajectories of drug concentration in a patient's blood stream, associated with the 95 % confidence region of the PDF of the model parameters.

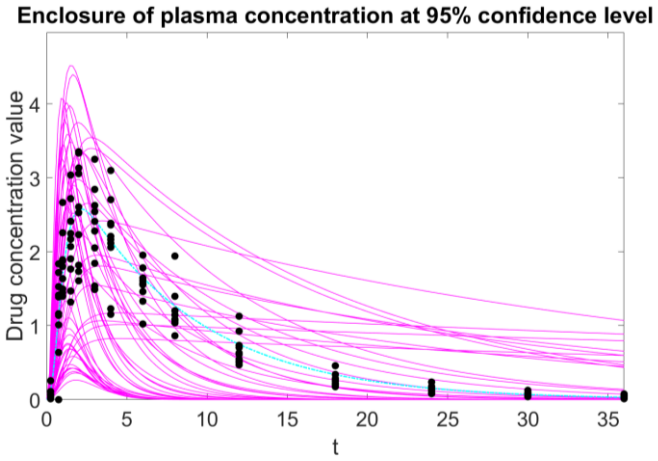


Figure 4: Predicted enclosure of the temporal trajectories of drug concentration in a patient's blood stream at 95% confidence level (the black dots are the experimental data; the light blue curve represents the profile of drug concentration in the blood stream of the average patient).

Both Figure 3 and Figure 4 confirm that PDFE&U tends to overestimate variances, but provides decent estimates of the central tendency measures and of the qualitative properties of the probability distribution, at a very attractive computational cost (Table 1). Finally, note that some of the largest deviations between the predictions of PDFE&U and those of Stan may be partially due to the small ratio of the number of data points to the number of uncertain parameters, which increases the probability of non-uniqueness of some of the projections of the state measurements (see the description of Step C of Phase I of PDFE&U).

4. Conclusions

The results of these analyses suggest that PDFE&U provides a favorable trade-off between accuracy and computational efficiency, by comparison to Bayesian Markov-chain Monte Carlo methods. This allows us to conclude that PDFE&U is more suitable for time-critical applications, where the PDF estimation task must be completed as soon as possible (e.g., in stochastic dynamic optimization and robust data reconciliation). Conversely, Bayesian Markov-chain Monte Carlo is more accurate and reliable, and is more suitable for offline, detailed PDF estimation studies.

In the near future, we are planning to extend these comparative studies by also analyzing the trade-off between computational performance and accuracy offered by Variational Bayes and Bayesian inference frameworks based on the Laplace approximation.

References

M.J. Beal, 2003, Variational algorithms for approximate Bayesian inference, London: University of London, UK.

B. Carpenter, A. Gelman, M.D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, A. Riddell, 2017, Stan: A probabilistic programming language, Journal of Statistical Software, 76, 1-32.

G. Buzzi-Ferraris, F. Manenti, 2012, BzzMath: library overview and recent advances in numerical methods, Computer Aided Chemical Engineering, 30, 1312-1316.

A. Caldwell, D. Kollar, K. Kröninger, 2009, BAT - The Bayesian Analysis Toolkit, Computer Physics Communications, 180, 2197-2209.

S.C. Dass, J. Lee, K. Lee, J. Park, 2017, Laplace based approximate posterior inference for differential equation models, Statistics and Computing, 27, 679-698.

P.L. Green, K. Worden, 2015, Bayesian and Markov chain Monte Carlo methods for identifying nonlinear systems in the presence of uncertainty, Philosophical Transactions A, 373, 20140405.

J.M. Laínez, G. Blau, L. Mockus, S. Orçun, G.V. Reklaitis, 2011, Pharmacokinetic based design of individualized dosage regimens using a Bayesian approach, Industrial and Engineering Chemistry Research, 50, 5114-5130.

L. Mockus, J.M. Laínez, G. Reklaitis, L. Kirsch, 2011, A bayesian approach to pharmaceutical product quality risk quantification, Informatica, 22, 537-558.

F. Rossi, G. Reklaitis, G. Buzzi-Ferraris, F. Manenti, 2016. Dynamic multi-scenario approach to robust and profitable online optimization & optimal control of batch processes, Computer Aided Chemical Engineering, 38, 955-960.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50253-6

Quantitative and Probabilistic approach for Underground Pipeline management Optimization

Keonhee Park*, Gunhak Lee, Chongyoung Nam, Wonbo Lee

Department of Chemical Engineering, Seoul National University, 615-8510 Seoul, KOREA
wblee@snu.ac.kr

Abstract

The risk of old underground pipelines installed in industrial complexes is increasing globally, Several methodologies for managing these equipment safely are being developed. Underground buried piping is damaged due to various causes such as corrosion and other construction, and it loses its function originally, which increases the risk of human life and environmental damage around facilities. QRA (Quantitative Risk Assessment) methods help to prioritize pipeline management with relative risk comparisons, but decisions on specific management strategies and methods are still in the hands of skilled experts. And it has limitation that relies heavily on experience. Also, it is difficult to fully consider the uncertain factors of piping in this deterministic approach.

In this study, a new piping management decision framework is proposed by combining risk assessment methodology for determining optimal piping life management strategy. The limit state function for each failure cause is defined and the accident probability is calculated by Monte Carlo simulation. The result is verified with the accident frequency using accident database of European Gas Incident Group (EGIG) and Department of Transportation USDOT.

In addition, pipeline accident consequence is defined by the Potential Impact Radius calculation method reflecting the characteristics of the Flammable and Toxic material. Finally, we can define objective function for management optimization using costing indexes for piping management. This optimal solution of this objective function helps us determining a cost-effective method for managing uncertain pipeline system. Based on the piping data embedded in the Ulsan National Industrial Complex in Korea, several case studies are performed and the expected results are presented in the proposed framework.

Keywords: Reliability, Quantitative Risk Assessment, Management Costing, Limit State Function, Probabilistic Approach, Optimization

1. Introduction

Pipeline systems, which is a typical fluid transport method, is managed by industrial companies. However, there have been fatalities in the world due to the explosion of gas pipeline in Kaohsiung, San Bruno and Ghislenghien pipeline leakage and pipe breakage accidents. Many accidents have been caused by defects in the welds, pipe breakage due to construction work, and the lack of proper piping management in accident reports. Most of the buried pipelines are not appropriately managed due to difficulties in management and investment of safety expenses. Various methods for efficiently

managing buried pipelines have been attempted. In order to detect and respond to faults causing major accidents, high-resolution devices such as ILI (Inline-inspection), Hydrostatic Testing, and External Corrosion Direct Assessment (ECDA) can be used. However, it is difficult for management companies to determine which methodology is cost-effective. In the world, aged pipelines are running in the industrial complex for more than 30-40 years, but there is no precise management system. Appropriate monitoring, inspection, and maintenance of piping can last for longer than the design life, which can be an optimization problem for the management company as there is a tradeoff between the uncertain damage caused by the accident and the costs involved in maintenance and management.

The Cost-Benefit Analysis problem has been a subject of continuous research and various methods have been tried. Several frameworks for diagnosing piping management by using a fuzzy logic model or introducing a risk based inspection concept has been proposed. However, these methodologies are only semi-quantitative and can only be used to prioritize the application of risk measures in piping, and it is difficult to quantitatively determine actual cost effects. In order to compensate for this, a methodology for evaluating pipeline integrity using a quantitative methodology based on the Structural Reliability Method can be used.Canadian Standards Association(CSA) summarizes the application of reliability based methods to the design and assessment of natural gas transmission pipelines. This method demonstrates the structural stability of a pipeline by making an explicit estimate of its reliability and comparing it to a specified reliability target. Structural reliability analysis are obtaining wider acceptance as a basis for evaluating pipeline integrity and these methods are ideally suited to managing metal corrosion damage as identified risk reduction strategies. These maintenance strategies considered included higher inspection frequencies and more higher accuracy test equipment usages. These results can explain whether the risk level of the current pipe is acceptable. This approach is to combine deterministic failure models with maintenance data and the pipeline attributes, experimental corrosion growth rate database, and the uncertainties inherent in this information. The calculated failure probability suggests the basis for informed decisions on which defects to repair, when to repair them and when to re-inspect or replace them.

This study provides an overview of pipeline management optimization solution using structural reliability method. Reliability based approach can be used to determine cost-effective safety measures. Consequence and probability calculation model was formulated to calculate the reliability using the limit state function model and the potential impact radius model referenced by Canadian standard association and British standards Some parameters of the literature model were adjusted based on piping information and management data in Ulsan and Yeosu, representative national industrial complexes located in Korea. The proposed decision making framework can help pipe management decision makers to establish an optimal maintenance strategy to time the investment decision; this strategy balances cost efficiency against stability in an uncertain future pipeline management system.

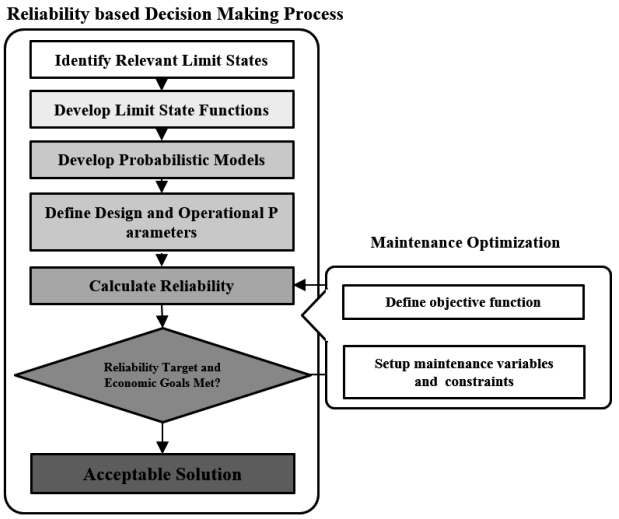


Figure 1. Proposed Decision making framework for underground pipeline management

2. Model Formulation

This paper is based on the probabilistic model and consequence model considering possible uncertainties of pipeline parameters. Failure mode can be defined to calculate the accident probability of the pipe.Failure modes are defined for small leaks, large leaks, and ruptures, and define a limit state function for each failure mode that is appropriate for the cause of the accident, such as corrosion or other construction. The model for calculating the failure probability of seven functions was established by citing the limit state function defined in CSA.The probabilistic distribution model of the piping variables included in each function was also referenced in the literature, and parameter tuning was performed based on the piping data in the Korean industrial complex and the experimental data from Korea Gas Safety Corporation. The calculation for this is based on the calculation setting of 100,000 times using Monte Carlo simulation. The results of the calculation of the accident probability are confirmed by the results of the PHMSA accident statistical data of the European Gas Incident Group EGIG (1970-2013), US Department of Transportation, and Pipeline & Hazardous Materials Safety Administration.

The Consequence model plays a role in predicting the magnitude of accidents occurring in each scenario and converting them into costs. The prediction model is classified into the Flammable gas consequence model and the Toxic gas consequence model depending on the type of pipeline transport material.Generally, the worst-case model is set as calculated in the risk assessment.The damage model assumes full rupture of the pipe and the type of accident to be modeled is defined as Jet fire and Toxic gas dispersion. In the Ulsan and Yeosu industrial complexes, fire and toxic hazardous materials are all included.

The distance of the accident influence in the Jet Fire accident type was calculated from the relation between the heat generation intensity and the distance to the maximum effect distance from death. The parameters used here were obtained by using Phast (DNV-GL) as a risk assessment program.

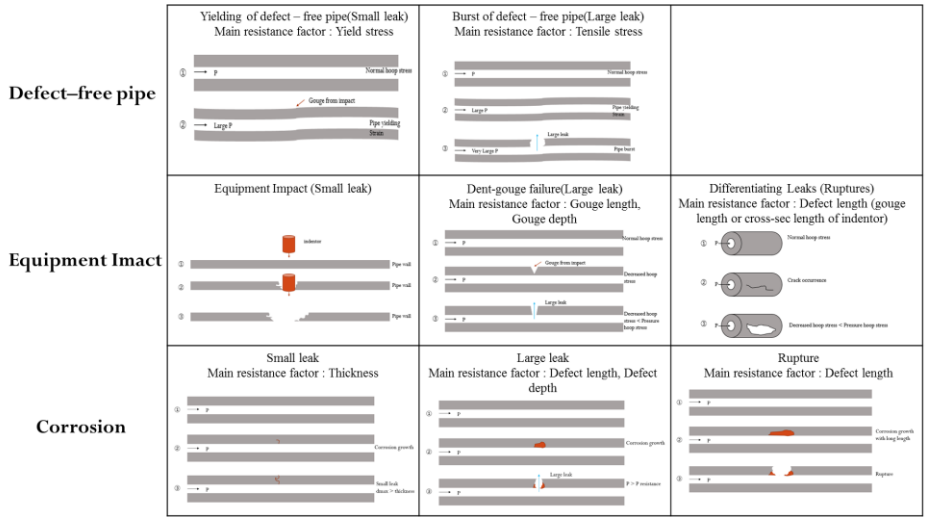


Figure 2. Description about main underground pipeline failure causes for limit state design

On the other hand, Probit probabilities of toxic substances were calculated by using Phast in the purple book, and the values of a, b, and n for each substance were substituted. The C-FER Technical Report shows that the distance that toxic gases are diffused and the harmful effects on the human body are related to the leakage amount and the exponential logarithm.The distance of influence was defined as the diffusion distance which corresponds to the concentration of 100% of the lethality.The amount of leakage was proportional to the square of the pressure and the diameter, and the influence distance (r) value was used.

Cost information was defined and applied to three risk reduction strategies: Inline Inspection, Hydrostatic Testing, and Direct Assessment. We define the objective function to approach the optimization problem of piping maintenance to derive the appropriate diagnosis time, cycle, and method for the pipe satisfying the acceptable risk level.The subsequent Constraint has the maximum allowable limit pipe wall thickness, reliability target, and so on.The allowable reliability target is set at 10-4 / km / yr, and the maximum allowable pipe wall thickness is calculated differently depending on the type of pipe, diameter, design pressure, etc.

Table 1. 8 Limit State functions and descriptions

Failure Mode	Failure Cause	Limit State function	Description
Small Leak	Yielding and burst of defect free pipe	$2\sigma_y t - PD$	σ_y : yield stress P : Operating pressure D : Diameter
Small Leak	Equipment Impact (1)Puncture	$r_a - q$	ra:estimated resistance q: normal impact force
Small Leak	Corrosion (1)Small leaks	$t - d_{max}$	t : wall Thickness dmax : maximum corrosion depth
Large Leak	Yielding and burst of defect free pipe	$1.906c\sigma_f t - PD$	c : model error σ_f : flow stress
Large Leak	Equipment Impact (2)Dent-gouge failure	$\sigma_c - \sigma_h$	σ_c : critical hoop stress σ_h : hoop stress

Large Leak	Corrosion (2) Large leak and ruptures	$r_a - P$	ra : estimated resistance (corrosion)
Rupture	Equipment Impact (3) Differentiating leaks and rupture	$S_{cr} - \sigma_h$	S_{cr} : critical resistance with unstable axial defect growth
Rupture	Corrosion (3) differentiating large leaks and ruptures	$\frac{1.8t\sigma_u}{m \times D} - P(SMYS \geq 241MPa)$ $\frac{2.3t\sigma_y}{m \times D} - P(SMYS \leq 241MPa)$	m : folias factor σ_u : tensile strength

3. Result

We will integrate the proposed methodology and data to define piping management optimization problem and interpret it using CPLEX 9.0 (GAMS). It is possible to analyze what is a single or multiple methodology with a lower cumulative management cost when viewed at 10, 20, and 30 years, while satisfying the Constraint.

Currently, it is possible to confirm the change in the probability of corrosion accident by the year when the pipe inspection cycle is different. In addition, the yearly changes in how the pipe design thickness affects the probability of an accident can be confirmed.
We will confirm the effectiveness of this reliability-based method by supplementing the optimized results.

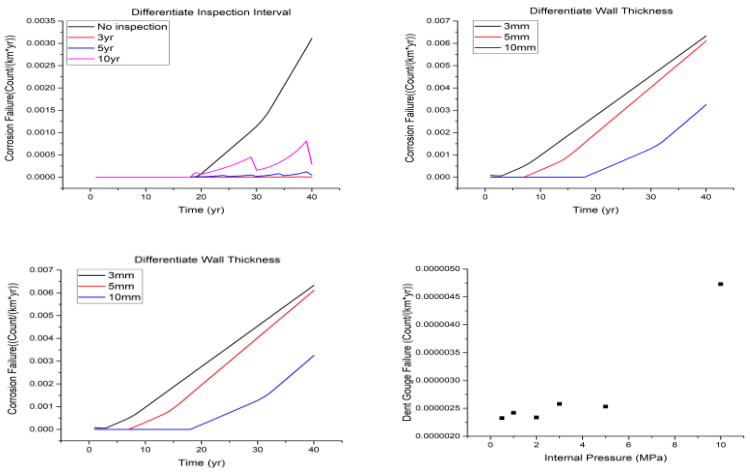


Figure 3. Case studies about propylene pipeline failure rate changes (Differentiate Inspection Interval and pieline parameters)

4. Conclusion

In order to solve the problem of setting the optimal piping management method, the framework was established by applying the reliability-based analysis method which is better than the existing semi-quantitative method. It is possible to approach the optimization problem by quantitatively reflecting the risk reduction effect of each risk reduction strategy by supplementing the uncertainty of the piping variables and forecasting the accident probability. In this study, only three risk reduction strategies have been considered, but if more diverse methodologies can be considered in the future,

optimal mathematical models and parameters can be supplemented to make optimal strategy analysis for more diverse methodologies. The proposed methodology can be used as a strategic tool for piping management decision makers who are concerned about cost and risk.

References

F.Caleyo, J.C. Velazquez, A. Valor, J.M. Hallen, 2009, Markov chain modelling of pitting corrosion in underground pipelines, Corrosion Science 51, pp2197-2207

H.P. Hong, 1999, Inspection and maintenance planning of pipeline under external corrosion considering generation of new defects, Structural Safety 21, pp203-222.

S.A. Timashev, M.G. Malyukova, L.V. Poluian, A.V, Bushiskaya, 2008, Markov description of corrosion defects growth and its application to reliability based inspection and maintenance of pipelines, in:Proc. 7th ASME Int. Pipeline Conf. IPC, Calgary, Canada, Paper IPC 2008-64546, September 26-29

Arup Dey, K.M. Mostafizur Rahman Sobuj, 2016, Limit state function formulation and time variant Relibility analysis, International Journal of Engineering Research and General Science Volume 4, Issue 2, March-April, 2016, ISSN 2091- 2730

Alexander Guzman, Behrooz Asgari, A cost-Benefit Analysis of Investing in Safety and Risk Engineering : The case of Oil & Gas Transportation Services by Pipelines, 2014 Proceedings of PICMET'14 : Infrastructure and Service Integration

Z662-07 Oil and gas Pipeline systems and Special Publication Z662.1-07,Canadian Standards Association

API, 2005, In-line Inspection Systems Qualification Standard, API Standard 1163, First Edition, American Petroleum Institute, Washington, D.C.

Brown, M. Nessim, M and Greaves, H. 1995. Pipeline Defect Assessment: Deterministic and Probabilistic Considerations, Second International Conference on Pipeline Technology, Ostend, Belgium, September

Fu,B and batte, A.D, 1999, New Methods for Assessing the Remaining Strength of Corroded Pipelines. Proceedings EPRG/PRCI 12 th Biennial Joint Technical Meeting on Pipeline Research, May

Nessim, M.A and Zhou, W. 2006, Reliability Based Design and Assessment for Location Specific Failure Threats. Proceedings of IPC 2006, International Pipeline Conference, Paper No. IPC06-10097, Calgary, October.

Maher Nessim, Howard Yue, Joe Zhou, 2010, Application of Reiliabilty based Design And Assessment To Maintenance and Protection Decisions For Natural Gas Pipelines, Proceeding of the 8th International Pipeline Conference,

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50254-8

Strengthened SOCP Relaxations for ACOPF with McCormick Envelopes and Bounds Tightening

Michael Bynum[a], Anya Castillo[b], Jean-Paul Watson[b], Carl D. Laird[a,b*]

[a]*Purdue University, Davidson School of Chemical Engineering, 480 Stadium Mall Drive, West Lafayette, IN 47907, United States*
[b]*Center for Computing Research, Sandia National Laboratories, Albuquerque, NM 87185, United States*
cdlaird@sandia.gov

Abstract

The solution of the Optimal Power Flow (OPF) and Unit Commitment (UC) problems (i.e., determining generator schedules and set points that satisfy demands) is critical for efficient and reliable operation of the electricity grid. For computational efficiency, the alternating current OPF (ACOPF) problem is usually formulated with a linearized transmission model, often referred to as the DCOPF problem. However, these linear approximations do not guarantee global optimality or even feasibility for the true nonlinear alternating current (AC) system. Nonlinear AC power flow models can and should be used to improve model fidelity, but successful global solution of problems with these models requires the availability of strong relaxations of the AC optimal power flow constraints. In this paper, we use McCormick envelopes to strengthen the well-known second-order cone (SOC) relaxation of the ACOPF problem. With this improved relaxation, we can further include tight bounds on the voltages at the reference bus, and this paper demonstrates the effectiveness of this for improved bounds tightening. We present results on the optimality gap of both the base SOC relaxation and our Strengthened SOC (SSOC) relaxation for the National Information and Communications Technology Australia (NICTA) Energy System Test Case Archive (NESTA). For the cases where the SOC relaxation yields an optimality gap more than 0.1 %, the SSOC relaxation with bounds tightening further reduces the optimality gap by an average of 67 % and ultimately reduces the optimality gap to less than 0.1 % for 58 % of all the NESTA cases considered. Stronger relaxations enable more efficient global solution of the ACOPF problem and can improve computational efficiency of MINLP problems with AC power flow constraints, e.g., unit commitment.

Keywords: ACOPF, Second-order cone programming, McCormick, Bounds tightening.

1. Introduction

Reliable and economic operation of the electricity grid is vital for the economy, public safety, and health. Key practical problems relating to the operation of the grid include the Optimal Power Flow (OPF) and the Unit-Commitment (UC) problems. The OPF, also referred to as the ACOPF problem, seeks to find generator operating conditions for a single time period that minimize the total operating cost while satisfying physical power flow laws and other operational constraints, such as generator and transmission limits. The UC problem solves for a generator on/off schedule across multiple time periods, introducing discrete variables for these scheduling decisions. Both problems require a model for the AC power flow (ACPF), giving rise to a large number of

nonlinear, non-convex constraints. To remove these nonlinearities, the ACOPF problem is typically simplified to a linear approximation known as the DCOPF problem. This approximation is more computationally efficient; however, it does not guarantee global optimality or even feasibility for the system under consideration. For this reason, there has recently been extensive research in convex relaxations for the ACOPF model to support global optimization of ACOPF and UC with ACOPF constraints.

Lavaei and Low (2012) present a semidefinite programming (SDP) formulation for the dual of the ACOPF problem and provide a rank-one sufficient condition under which the duality gap is guaranteed to be zero. Coffrin et al. (2017) propose a quadratic relaxation on the polar form of the ACOPF problem. Jabr (2006) introduced a convex second-order Cone (SOC) programming relaxation of the ACOPF problem. Kocuk et al. (2016) proposed extensions to the SOC relaxation including (1) relaxations on the arctangent constraints relating the SOC variables to voltage angles differences and (2) linear cutting planes to incorporate information from the SDP relaxation. Liu et al. (2017) propose a piecewise outer approximation based global optimization algorithm for solving the ACOPF problem.

In this paper, we propose a strengthened SOC (SSOC) relaxation that relates the auxiliary variables in the SOC relaxation to the real and imaginary components of the voltages in the rectangular form through McCormick envelopes. The ACOPF problem has an infinite number of solutions unless the voltage angle is specified at an arbitrary bus, typically referred to as the reference bus. By relating the SOC auxiliary variables to the rectangular voltages, this reference bus specification may now be included in the proposed relaxation, providing tighter bounds on the voltage variables at the reference bus. Results for several widely studied ACOPF test cases demonstrate that this approach improves the effectiveness of optimality-based bounds tightening (OBBT) and strengthens the SOC relaxation significantly. In the next section, we present a review of the SOC relaxation; our SSOC relaxation is described in Section 3. Computational results comparing the optimality gap of the SOC and SSOC relaxations are presented along with results showing the impact of the reference bus on bounds tightening.

2. ACOPF Second-Order Cone Relaxation

A power system is modeled with a set of buses connected by transmission lines and/or transformers. Each bus may have one or more generators and loads. The rectangular form of the ACOPF includes real and imaginary components of the voltage, v^r and v^j respectively, at each bus. With the following definitions,

$$c_{b,b} \equiv (v_b^r)^2 + (v_b^j)^2 = v_b^2, \;\; c_{b,k} \equiv v_b^r v_k^r + v_b^j v_k^j, \;\; s_{b,k} \equiv v_b^r v_k^j - v_k^r v_b^j \tag{1}$$

the SOC relaxation of the ACOPF problem can be written as follows:

$$\min \sum_{g \in \mathcal{G}} [A_g^2 (p_g^G)^2 + A_g^1 p_g^G + A_g^0]$$

s.t.

$$\sum_{(b,k) \in \mathcal{L}_b^{in}} p_{b,k}^t + \sum_{(b,k) \in \mathcal{L}_b^{out}} p_{b,k}^f + G_b^{sh} c_{b,b} + P_b^D - \sum_{g \in \mathcal{G}_b} p_g^G = 0 \qquad \forall\, b \in \mathcal{B}$$

$$
\begin{aligned}
&\sum_{(b,k)\in\mathcal{L}_b^{in}} q_{b,k}^t + \sum_{(b,k)\in\mathcal{L}_b^{out}} q_{b,k}^f - B_b^{sh} c_{b,b} + Q_b^D - \sum_{g\in\mathcal{G}_b} q_g^G = 0 && \forall\, b \in \mathcal{B} \\
&p_{b,k}^f = G_{b,k}^{ff} c_{b,b} + G_{b,k}^{ft} c_{b,k} - B_{b,k}^{ft} s_{b,k} && \forall\, (b,k) \in \mathcal{L} \\
&q_{b,k}^f = -B_{b,k}^{ff} c_{b,b} - B_{b,k}^{ft} c_{b,k} - G_{b,k}^{ft} s_{b,k} && \forall\, (b,k) \in \mathcal{L} \\
&p_{b,k}^t = G_{b,k}^{tt} c_{k,k} + G_{b,k}^{tf} c_{k,b} - B_{b,k}^{tf} s_{k,b} && \forall\, (b,k) \in \mathcal{L} \\
&q_{b,k}^t = -B_{b,k}^{tt} c_{k,k} - B_{b,k}^{tf} c_{k,b} - G_{b,k}^{tf} s_{k,b} && \forall\, (b,k) \in \mathcal{L} \\
&(p_{b,k}^f)^2 + (q_{b,k}^f)^2 \le (S_{b,k}^{max})^2,\ (p_{b,k}^t)^2 + (q_{b,k}^t)^2 \le (S_{b,k}^{max})^2 && \forall\, (b,k) \in \mathcal{L} \\
&(V_b^{min})^2 \le c_{b,b} \le (V_b^{max})^2 && \forall\, b \in \mathcal{B} \\
&P_g^{G,min} \le p_g^G \le P_g^{G,max},\ Q_g^{G,min} \le q_g^G \le Q_g^{G,max} && \forall\, g \in \mathcal{G} \\
&c_{b,k} = c_{k,b},\ s_{b,k} = -s_{k,b} && \forall\, (b,k) \in \mathcal{L} \\
&c_{b,k}^2 + s_{b,k}^2 \le c_{b,b} c_{k,k} && \forall\, (b,k) \in \mathcal{L}
\end{aligned}
\tag{2}
$$

Here, capitalized symbols represent parameters and lower-case symbols represent variables. Real power is denoted by $\boldsymbol{p}$, reactive power by $\boldsymbol{q}$, and the relaxation variables by $\boldsymbol{s}$ and $\boldsymbol{c}$. The superscripts *G* and *D* denote generation and demand, and *f* and *t* denote the "from" and "to" ends of transmission lines. Subscripts denote indices of the sets of buses $\boldsymbol{\mathcal{B}}$, generators $\boldsymbol{\mathcal{G}}$, and transmission lines $\boldsymbol{\mathcal{L}}$. The optimization objective is to minimize the aggregate operating cost of all the generators. The constraints enforce real power balances at all the buses and reactive power balances at all of the buses, model real and reactive power flow on transmission lines at the "from" and "to" ends of the transmission lines, and enforce thermal limits on the transmission lines, voltage limits at the buses, and generator limits on real and reactive power. The last two constraints are derived from the definitions of *s* and *c*, while the final constraint is relaxed to be an inequality, forming the SOC relaxation. Eq.(2) is a convex optimization problem that may be solved efficiently with off-the-shelf solvers, e.g., Ipopt.

Note that bounds tightening does not improve the relaxation as written since none of the relaxations depend on variable bounds, as is the case when using McCormick envelopes. Further, the voltage variables have been removed from the model, so there is no direct way to incorporate a specification of the voltage angle at the reference bus. In the following section, we present our SSOC relaxation that incorporates McCormick envelopes for Eq.(1) into the SOC relaxation, allowing specification of the reference bus angle and providing a mechanism to effectively tighten the relaxation with OBBT.

3. McCormick Envelopes for Strengthening the SOC Relaxation

First, let $\widehat{xy}$ be an auxiliary variable for the bilinear term xy. We represent the set of McCormick envelopes for the bilinear term as $\widehat{xy} \in MCC(x,y)$. With this notation, Eq.(1) may be relaxed with McCormick Envelopes, as follows:

$$
\begin{aligned}
&c_{b,b} = \widehat{v_b^r} + \widehat{v_b^j}, \;\; c_{b,k} = \widehat{v_b^r v_k^r} + \widehat{v_b^j v_k^j}, \;\; s_{b,k} = \widehat{v_b^r v_k^j} - \widehat{v_b^j v_k^r} \\
&(v_b^r)^2 \leq \widehat{v_b^r} \leq (V_b^{r,\max} + V_b^{r,\min}) v_b^r - V_b^{r,\max} V_b^{r,\min} \\
&(v_b^j)^2 \leq \widehat{v_b^j} \leq (V_b^{j,\max} + V_b^{j,\min}) v_b^j - V_b^{j,\max} V_b^{j,\min} \\
&\widehat{v_b^r v_k^r} \in MCC(v_b^r, v_k^r), \;\; \widehat{v_b^j v_k^j} \in MCC(v_b^j, v_k^j) \\
&\widehat{v_b^r v_k^j} \in MCC(v_b^r, v_k^j), \;\; \widehat{v_b^j v_k^r} \in MCC(v_b^j, v_k^r)
\end{aligned}
\tag{3}
$$

The initial bounds on the real and imaginary voltage components are determined by the voltage magnitude limits provided in the problem specification. However, at the reference bus, the imaginary component is fixed to zero (i.e., the voltage angle is fixed to zero), and the bounds on the real component of the voltage at the reference bus can be tightened as with Eq.(4). The SSOC relaxation is then given by Eqs.(2-4).

$$
\begin{aligned}
&-V_b^{\max} \leq v_b^r, v_b^j \leq V_b^{\max} \\
&V_{ref}^{\min} \leq v_{ref}^r \leq V_{ref}^{\max}, \;\; v_{ref}^j = 0
\end{aligned}
\tag{4}
$$

To Further tighten the SSOC relaxation, OBBT is performed on the voltage variables at each bus. The relaxations in Eq.(3) are then updated, the SSOC relaxation is solved, and the process is repeated until either the optimality gap is less than 0.1 % or 20 iterations are performed with a cumulative reduction in the optimality gap of less than 0.1 %.

4. Results

4.1. Performance of the SSOC Relaxation

Table 1 compares the optimality gap for the SOC and SSOC relaxations for several of the NESTA test cases (Coffrin et al., 2014). The relaxations were modelled in Python 3.5.2 with Pyomo (Hart, 2011) and solved with Ipopt 3.12.5 (Wächter, 2006). The results show that the SSOC relaxation tightens the SOC relaxation significantly. Of the cases for which the optimality gap with the SOC relaxation is greater than 0.1 %, the SSOC relaxation with bounds tightening further reduces the optimality gap by an average of 67 % and overall reduces the optimality gap to less than 0.1 % for 58 % of the NESTA cases.

4.2. The Impact of the Reference Bus on Bounds Tightening

Figure 1 shows the importance of the reference bus and the corresponding bounds in Eq.(4). The figure shows the variable bound ranges for the real and imaginary components of the voltages averaged over all buses of the same shortest path in terms of the number of transmission lines from the reference bus. The iteration count for OBBT is shown on the x-axis. The figure shows the first four iterations of bounds tightening for nesta_case24_ieee_rts__api. As shown in the figure, before any bounds tightening is performed, the reference bus is the only bus with good bounds on the voltage. After one iteration of bounds tightening, the bounds on the voltages at buses connected directly to the reference bus are tightened dramatically. On the second iteration of bounds tightening, the bounds on the voltages at buses a distance of 2 from the reference bus

Table 1: Comparison of optimality gap of SOC and SSOC with OBBT.

Nesta Case	Optimality Gap (%)					
	Standard		Active Power Increase		Small Angle Difference	
	SOC	SSOC	SOC	SSOC	SOC	SSOC
case3_lmbd	1.32	0.40	3.30	0.07	4.28	0.05
case4_gs	0.00	0.00	0.65	0.07	4.90	0.01
case5_pjm	14.54	4.94	0.28	0.06	3.61	0.03
case6_ww	0.63	0.09	-	-	0.80	0.05
case6_c	0.30	0.10	0.35	0.05	1.36	0.01
case9_wscc	0.00	0.00	0.00	0.00	1.50	0.01
case14_ieee	0.11	0.06	1.34	0.17	0.06	0.06
case24_ieee_rts	0.01	0.01	20.75	1.98	11.42	0.06
case29_edin	0.14	0.11	0.44	0.43	34.47	2.02
case30_as	0.06	0.06	4.76	0.21	9.16	0.18
case30_fsr	0.39	0.29	45.97	41.53	0.62	0.14
case30_ieee	15.88	0.08	1.01	0.08	5.84	0.08
case39_epri	0.05	0.05	2.99	0.21	0.11	0.05
case57_ieee	0.06	0.06	0.21	0.09	0.11	0.09
case73_ieee_rts	0.03	0.03	14.39	14.36	8.37	2.51
case89_pegase	0.17	0.10	20.43	20.11	0.28	0.17
case118_ieee	2.07	1.75	44.15	29.89	12.89	5.16

are tightened more than the bounds on variables at other buses. The trend continues through all four iterations of bounds tightening shown. In this figure, it is clear that the effectiveness of bounds tightening at a given iteration is directly related to the distance of the corresponding bus from the reference bus. Additionally, although not shown in Table 1, the SSOC relaxation was solved with the bounds tightening procedure described in Section 3 but without the tight bounds on the reference bus voltages. The optimality gap did not improve at all beyond the gap obtained with the SOC relaxation for any of the test cases. This further indicates the necessity of the tight voltage bounds at the reference bus for effective bounds tightening.

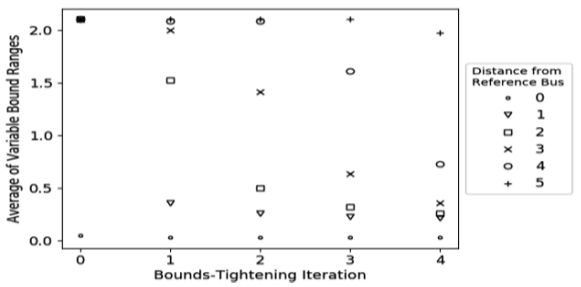


Figure 1:Impact of the reference bus on bounds tightening for nesta_case24_ieee_rts__api. The abscissa shows the iteration count, and the ordinate shows the average range between the upper and lower variable.

5. Conclusions

We describe the use of McCormick envelopes and OBBT to strengthen the well-known second-order cone (SOC) relaxation of the ACOPF problem. With the proposed relaxation, tight bounds on the voltages at the reference bus can be specified, and we demonstrate the necessity of this specification for bounds tightening to be effective. In Section 4.1, we showed that, of the NESTA test cases for which the optimality gap with the SOC relaxation is greater than 0.1 %, the SSOC relaxation with bounds tightening further reduces the optimality gap by an average of 67 % and ultimately reduces the optimality gap to less than 0.1 % for 58 % of all the cases considered. In Section 4.2, we demonstrated that the effectiveness of OBBT at a given iteration is directly related to the distance of the corresponding bus from the reference bus. In future work, this relationship may be exploited to develop more efficient strategies for bounds tightening.

6. Acknowledgements

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DENA0003525.

The work of Michael Bynum, Carl Laird, and Jean-Paul Watson was supported in part by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program under contract number KJ0401000 through the project "Multifaceted Mathematics for Complex Energy Systems". The work of Michael Bynum, Anya Castillo, Carl Laird, and Jean-Paul Watson was also supported in part by Sandia National Laboratories' Laboratory Directed Research and Development (LDRD) program.

References

C. Coffrin, D. Gordon, and P. Scott, 2014, NESTA, the NICTA Energy System Test Case Archive, arXiv preprint arXiv:1411.0359.

W. Hart, J.P. Watson, and D. Woodruff, 2011, Pyomo: Modeling and Solving Mathematical Programs in Python, Math Program Comput, 3, 3, 219-260.

H. Hijazi, C. Coffrin, and P. Hentenryck, 2017, Convex Quadratic Relaxations for Mixed-Integer Nonlinear Programs in Power Systems, Math Program Comput, 9, 3, 321-367.

R. Jabr, 2006, Radial Distribution Load Flow using Conic Programming, IEEE Trans Power Syst, 21, 3, 1458-1459.

B. Kocuk, S. Dey, and X. Sun, 2016, Strong SOCP Relaxations for the Optimal Power Flow Problem, Oper Res, 64, 6, 1177-1196.

J. Lavaei and S. Low, 2012, Zero Duality Gap in Optimal Power Flow Problem, IEEE Trans Power Syst, 27, 1, 92-107.

J. Liu, M. Bynum, A. Castillo, J.P. Watson, and C. Laird, A Multitree Approach for Global Solution of ACOPF Problems Using Piecewise Outer Approximations, Comput Chem Eng (2017), https://doi.org/10.1016/j.compchemeng.2017.10.018.

A. Wächter, and L. Biegler, 2006, On the Implementation of an Interior-Point Filter Line-Search Algorithm for Large-Scale Nonlinear Programming, Math Program, 106, 1, 25-57.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50255-X

Dual adaptation strategy for model-based operation

Cheng Yang, Kexin Wang, Zhijiang Shao*

College of Control Science and Engineering, Zhejiang University, 38 Zheda Road, Hangzhou 310027, China
szj@zju.edu.cn

Abstract

Model-based operations are frequently challenged by unsatisfactory model accuracy, which hampers finding the true plant optimum. Integrated parameter estimation and optimization methods were proposed to handle the problem. However, they did not focus on improving model accuracy or confining optimization to a valid operating range. Dual adaptation strategy is thus presented in this paper to integrate local parameter model and real-time optimization through trust-region framework. On one hand, application range of the local model is decided adaptively so as to avoid invalid operations. On the other hand, the local model is updated adaptively for purpose of reducing plant-model mismatch in moderate cost. The above two aspects contribute to dual adaptation concept. It is guaranteed that the optimal operations derived from this dual adaptation strategy converge to the plant optimum. Case study of significant load change for HTR-PM demonstrates effectiveness of the proposed method.

Keywords: plant-model mismatch, dual adaptation strategy, trust-region method, iterative method.

1. Introduction

Model-based applications are required to calculate valid operations for complex systems. The main difficulty lies in that plant-model mismatch is unavoidable, as it can arise from simplifications in modelling, characteristics variation during the operation course, and limited as well as contaminated data for deriving the model. Due to the mismatch, model-based operation is prevented from converging towards the plant optimum. Therefore, the desirable model applicable to the full operating range and strategy which guarantees convergence to the plant optimum in the presence of plant-model mismatch are indispensable in model-based applications.

Plant-model mismatch should be handled properly. This problem was considered in the framework of classical iterative two-step (ITS) method (Marlin and Hrymak, 1997), where parameter estimation (PE) and real-time optimization (RTO) are performed iteratively. However, the consistence of the model-based solution and the plant optimum cannot be guaranteed. Methods which modified the ITS framework were therefore developed to force convergence of model-based operations towards the true plant optimum (Brdys and Tatjewski, 2005; Marchetti, Chachuat, and Bonvin, 2009), where the formulation of the optimization problem was improved by taking advantage of KKT systems of the plant. Although convergence to the plant optimum was proved, these methods have nonanalytic input correction in each iteration.

Following the ITS concept, we propose a trust-region based strategy with dual adaptation. A sequence of local models is applied in the sought of optimal operations, at the same time, proper amount of input correction in each iteration is determined by

precise optimization. Moreover, the local model is updated only if necessary to reduce computational effort.

2. ITS method for model-based operation

ITS method intends to decrease the plant-model mismatch and provide model-based operations through performing PE and RTO iteratively, which can be expressed as follows.

Step 1: Apply the current operations to the plant and solve a PE problem to determine the model parameters using the resulting measurements.

Step 2: Calculate the operations for optimizing some performance index by solving a RTO problem based on the model with the updated parameters.

Step 3: If no further improvement is implemented, stop. Otherwise, go to Step 1.

However, model-based operations by ITS method rarely converge to the true plant optimum due to structural mismatch (Yip and Marlin, 2004). Moreover, sufficient data are usually unavailable and the local model derived from limited data sets fails to characterize the process over the full operating range.

3. Dual adaptation strategy to address plant-model mismatch

Dual adaptation strategy is designed in the framework of trust-region method that confines each optimization problem to a valid local model. As the local model derived from limited measurements may not be applicable to a wide range, we meant to extend its application range by constructing a more predictive model. At the kth iteration, parameters p can be approximated around the current conditions u_k by

$$m_k^p(u) = \hat{p}_k + \left(\frac{\partial p}{\partial u_k}\right)^T \cdot (u - u_k) \tag{1}$$

where $\hat{p}_k$ are the estimates from measurements at current conditions, the linearization term $\partial p/\partial u_k$ represent the sensitivities of parameters p with respect to the inputs u at the kth iteration and are obtained from perturbations to the system inputs.

The model parameters in Eq.(1) should be restricted to a proper region, which gives rise to an extended model, compared to the model at a single operation point, with restricted operating range. Based on such a model, the RTO problem for dual adaptation strategy is formulated in

$$\begin{aligned} \min_{u} \quad & \varphi(u) \\ s.t \quad & f(u, m_k^p(u)) = 0 \\ & c(u) \le 0 \\ & \|u - u_k\|_\infty \le \Delta_k \end{aligned} \tag{2}$$

where $\varphi(u)$ denotes the RTO objective function, f represents the system model, and c are constraints on the inputs. As the states and outputs are determined by inputs, they are omitted for simplicity. Δ_k is iteration related and limits operation u to a region around the current conditions u_k. This added constraint defines the application range of

the approximated model and allows us to integrate PE and RTO through trust-region framework.

In trust-region methods, both the model and trust-region radius are updated in each iteration based on model evaluation (Biegler, Lang and Lin, 2014). As it is expensive in this paper to calculate the plant derivatives to approximate the parameters, model update should be called only when necessary. This gives rise to dual adaptation strategy, where both the approximated model and its application range are updated adaptively based on model evaluation. The algorithm is stated below, where the merit function ϕ is exploited to balance between reducing the objective function and improving feasibility.

Algorithm: Dual adaptation strategy based on trust-region framework

Step 0 (Initialization): Give the current operation conditions u_0 and the initial trust-region radius Δ_0. Choose constants $0 \le \eta_0 \le \eta_1 < 1$ (with $\eta_1 \neq 0$), $\eta_0' = 2\eta_0$, $\eta_1' = 2\eta_1$, $0 < \gamma_0 \le \gamma_1 < 1$, and termination tolerance ε_t. Set $k \leftarrow 0$, and $\mathcal{E}x = \phi$.

Step 1: Estimate the parameters to obtain $\hat{p}_k$.

Step 2: If $k-1 \in \mathcal{E}x$, then go to Step 3. Calculate sensitivities $\partial p/\partial u_k$. Set $l = k$ and denote $V_l = \partial p/\partial u_k$. Build the local property model with $m_l^p(u) = \hat{p}_k + V_l^T \cdot (u - u_k)$.

Step 3: If $\|\nabla\phi(u_k, \hat{p}_k\| \le \varepsilon_t$, stop. Otherwise, solve problem (2), substituting m_l^p for the property model. Obtain solution u_{k+1}.

Step 4: Apply u_{k+1} to the plant and compute

$$\rho_k = \frac{\phi\big(u_k, p(u_k)\big) - \phi(u_{k+1}, p(u_{k+1}))}{\phi(u_k, \hat{p}_k) - \phi(u_{k+1}, m_l^p(u_{k+1}))} \tag{3}$$

Step 5: Consider the following conditions.

- If $\rho_k < \eta_0$ and $k = l$, then $u_{k+1} \leftarrow u_k$, $\Delta_{k+1} \leftarrow \gamma_0 \Delta_k$.
- If $\rho_k < \eta_0'$ and $k \neq l$, then $u_{k+1} \leftarrow u_k$.
- If $\eta_0 \le \rho_k < \eta_1$ and $k = l$, then accept u_{k+1} as the new iterate, $\Delta_{k+1} \leftarrow \gamma_1 \Delta_k$.
- If $\eta_0' \le \rho_k < \eta_1'$, $k \neq l$ and Eq.(4) holds, then accept u_{k+1} as the new iterate.
- If $\rho_k \ge \eta_1$ and $k = l$, or $\rho_k \ge \eta_1'$, $k \neq l$ and Eq.(4) holds, then accept u_{k+1} as the new iterate, add k to $\mathcal{E}x$, $\Delta_{k+1} \leftarrow \Delta_k$.

Step 6: $k \leftarrow k+1$. If u_{k+1} is accepted, go to Step 1; else if the sensitivities of the current iterate has been calculated, go to Step 3; otherwise go to Step 2.

For convergence of this algorithm, we focus on the iterates where derivatives are not updated. Consider the sequence $\{u_k, u_{k+1}, u_{k+2}\}$, where derivatives are calculated at u_k, which lead to the new iterate u_{k+1}; from this point, u_{k+2} is generated and accepted without updating the derivatives. We have the following results for iterations corresponding to this sequence, which can be easily extended to sequences where more iterates are involved before derivative update.

Remark 1 $\langle \nabla\phi(u_k, p(u_k)), u_{k+2} - u_{k+1} \rangle \le 0$. This is required to ascertain $d' = u_{k+2} - u_k$ is a descant direction.

Remark 2 Based on the above result, the gradient relatedness of d' can be established, i.e. $|\langle \nabla\phi(u_k, p(u_k)), u_{k+2} - u_k\rangle|/\|u_{k+2} - u_k\| = 0$ only if $\|\nabla\phi(u_k, p(u_k))\| = 0$. This is essential for convergence as we do not always calculate and search along the plant derivatives.

Remark 3 The descent along d' is sufficient. Specifically, suppose $\rho_{k+1} \geq \eta_0'$. If for some constant $b \geq 0$,

$$\frac{\phi\left(u_{k+1}, m_k^p(u_{k+2})\right) - \phi\left(u_k, m_k^p(u_{k+2})\right)}{\phi(u_k, p(u_k)) - \phi(u_{k+2}, p(u_{k+2}))} \leq b \tag{4}$$

then

$$\frac{\phi(u_k, p(u_k)) - \phi(u_{k+2}, p(u_{k+2}))}{\phi(u_k, \hat{p}_k) - \phi\left(u_k, m_k^p(u_{k+2})\right)} \geq \eta_0 \tag{5}$$

Setting $b = 1/\eta_0'$ reveals $\eta_0' = 2\eta_0$. Moreover, the same result holds for η_1' and η_1. According to this result, if a new iterate is generated without updating the derivatives and then is accepted, it will be accepted by standard trust-region methods, as well.

With these results, the first order convergence theory of standard trust-region methods is applicable to the dual adaptation algorithm.

4. Numerical results

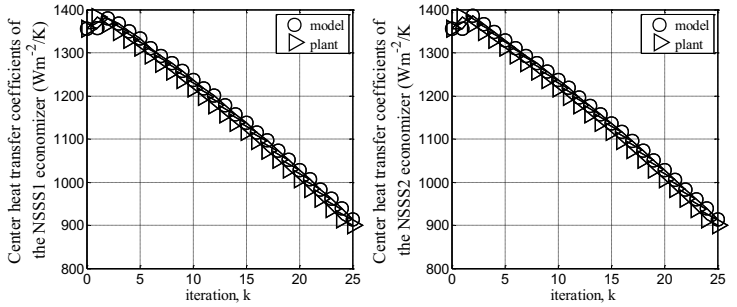
Dual adaptation strategy is applied to implement significant load change of high-temperature gas-cooled reactor pebble-bed module (HTR-PM) (Li, Huang and Zhang, 2008), where both reactors in two nuclear steam supply systems (NSSS1 and NSSS2) are expected to change from 100%RFP (reactor full power) to 50%RFP synchronously. Heat transfer coefficients in HTR-PM model vary with load so that the local model with estimated parameters may not be applicable to the full operating range. Meanwhile, key outputs of HTR-PM are required to stay close to the design values. Consequently, the task should be completed step by step due to safety and operation limits.

The RTO objective function for load change of HTR-PM is defined as the deviations between the model predicted and design values of the key outputs at the target load level. The key outputs include relative nuclear power, outlet helium temperature of the core, outlet steam temperature and pressure of the steam generator, steam temperature and pressure just before the turbine. The freedom for operation leaves to external reactivity, helium inlet flow rate, and feedwater flow rate, and opening of the valve prior to the turbine. At the *k*th iteration, heat transfer coefficients are approximated by Eq.(1) where their estimates $\hat{p}_k$ are calculated from measurements at u_k. The approximated parameters contain process information and affect RTO as expressed in Eq.(2). For comparison, ITS method is also used to realize the same task. Different from dual adaptation strategy, ITS method obtains the model through parameter estimation, based on which it defines the application range of the local model by directly restricting the maximal load change in each RTO step instead of limiting the operation region. Larger step of 8%RFP and smaller of 2%RFP for ITS method are implemented, respectively.

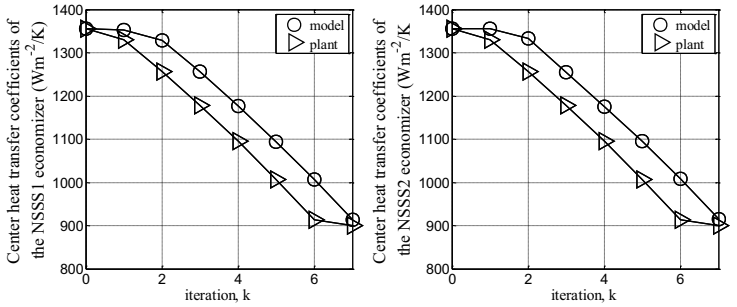
The resulting parameters are presented in Figure 1. As can be seen, ITS method with stepsize of 8%RFP and 2%RFP suffer the largest and smallest parameter mismatch,

respectively. Parameter mismatch from dual adaptation strategy is moderate. Model update is called twice in dual adaptation strategy, which is marked by '*'. One is at the beginning of load change, and the other is after the 3rd iteration when parameter mismatch becomes large.

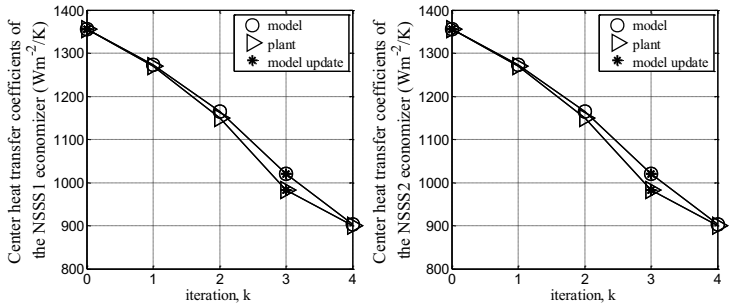
Numerical results by the three strategies are listed in Table 1. In particular, #iter denotes the number of iterations to realize the load change task, and stepsize denotes the sequence of load change in each iteration. Only the maximal parameter mismatch each strategy suffers is listed. The ITS method with stepsize of 8%RFP violates the safety requirements on steam temperature just before the turbine, under the maximal mismatch 10.2%. This indicates the local model in that step is not applicable to the maximal load change of 8%RFP in the next step. Although it is fast with least CPU time and moderate number of iterations, this strategy is not safe for load change of HTR-PM. On the other hand, the other two strategies satisfy the safety requirements during the whole process, where the ITS method with stepsize of 2%RFP suffers the smallest mismatch (i.e. 2.2%), while takes the most CPU time and iterations, which is conservative and inefficient. Dual adaptation strategy takes 4 iterations, which is the least among the three strategies, to realize the significant load change. This can be attributed to its ability in control and maintaining of model accuracy. Note that it costs more CPU time than the ITS methods per iteration, duo to calculation of plant derivatives and possible solution of the RTO problem with more restricted operation regions.



(1a) ITS method with stepsize of 2%RFP



(1b) ITS method with stepsize of 8%RFP



(1c) Dual adaptation strategy

Figure 1. Comparison of model accuracy - heat transfer coefficients

Table 1. Numerical results comparison

Strategy	#iter	Stepsize (%RFP)	Maximal mismatch	Output violation	CPU time (sec)
ITS-2%RFP	25	2	2.2%	-	1019
ITS-8%RFP	7	8 (first 6 steps) →2	10.2%	steam temperature	324
Dual adaptation	4	11→14→17→8	3.8%	-	620

5. Conclusions

We propose dual adaptation strategy to enforce correct convergence of model-based operations when plant-model mismatch is present. Different from ITS method, dual adaptation strategy is engaged in improving model accuracy and determining proper application range of the local model, so as to offer valid operations. In particular, the RTO step is adjusted adaptively to confine optimization to an appropriate operating range, and the local model is updated adaptively as well to avoid big mismatch at reasonable cost. The numerical results of implementing significant load change of HTR-PM shows that dual adaptation strategy is efficient in handling plant-model mismatch, and thus deriving operations with better performance in practice.

References

L.T. Biegler, Y.D. Lang, W. Lin, 2014, Multi-scale optimization for process systems engineering, Computers and Chemical Engineering, 60, 2, 17-30

M.A. Brdys, P. Tatjewski, 2005, Iterative Algorithms for Multilayer Optimizing Control, Imperial College Press, London, UK.

H. Li, X. Huang, L. Zhang, 2008, Operation and control simulation of a modular high temperature gas cooled reactor nuclear power plant, IEEE Transactions on Nuclear Science, 55, 4, 2357-2365

A. Marchetti, B. Chachuat, D. Bonvin, 2009, Modifier-adaptation methodology for real-time optimization, Industrial and Engineering Chemistry Research, 48, 13, 6022-6033

T.E Marlin, A.N Hrymak, 1997, Real-time operations optimization of continuous processes, AIChE Symposium Series - CPC V, 93, 316, 156-164

W.S Yip, T.E Marlin, 2004, The effect of model fidelity on real-time optimization performance, Computers and Chemical Engineering, 28, 1-2, 267-280

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50256-1

Integrated Design Strategy for Optimization of Utility Supply and Carbon Utilization: Multi-period Deterministic Model

Yuchan Ahn[a*], Jaewon Byun[a], Juyeon Kim[a], Jeehoon Han[a, b]

[a]*School of Semiconductor and Chemical Engineering, Chonbuk National University, Jeonju 54896, Korea*
[b]*School of Chemical Engineering, Chonbuk National University, Jeonju 54896, Korea*
ahnyuchan@jbnu.ac.kr

Abstract

Integrated design strategy for optimization of utility supply and CO_2 mitigation (USCM) is an important task for companies dispersed spatially in an industrial complex. The integrated strategy considers utility supply sub-strategy such as fossil fuel-based Rankine cycle, and CO_2 mitigation sub-strategy such as CO_2 capture, storage, and utilization systems. This work presents a multi-period deterministic model of the integrated design strategy for optimization of USCM. The proposed model determines where and how much (1) utilities are supplied one company to the others, (2) amount of CO_2 to be captured and transported, and (3) amount of CO_2 to be stored or utilized about the excess value over emissions permits allocated to each company over a multi-period planning horizon. The real scenario of Yeosu Industrial Complex in the Republic of Korea is applied to validate the integrated model. The optimization results will help to determine investment strategies for USCM.

Keywords: deterministic model; multi-period; utility supply; optimization; CO_2 storage; CO_2 utilization.

1. Introduction

In an industrial complex (IC), significant amounts of carbon dioxide (CO_2) have been emitted by the production of many chemicals or fuels, simultaneously consuming large amounts of utility such as steam, electricity, and water (Jung et al., 2012). Utility consumption has to be minimized to reduce the CO_2 emission. Many studies have been carried out to reduce utility consumption by designing the optimized supply chain network for exchanging utilities among companies in the IC (Han and Lee, 2014; Kim et al., 2010). To reduce CO_2 emission directly with indirect reduction of CO_2 caused by utility optimization of supply chain network, carbon capture and storage (CCS) system (Li et al., 2013) have to be considered to meet the CO_2 mitigation target of the IC. Many studies for the carbon mitigation were presented considering the mitigation demand and economics for the systems (Lima et al., 2016; Karimi and Sibudjing, 2016). Among CO_2 migration methods, CCS system, especially CO_2 capture using monoethanolamine (MEA) absorption, consumes large amounts of utility (Sethi, 2017). Moreover, the CCS poses a significant cost burden with the geological limitation on storage of CO_2. As an alternative, carbon capture and utilization (CCU) system is aimed to use CO_2 as a feedstock for production of commodity chemicals or fuels (Gollodi et al., 2017). Microalgae-based biodiesel production via hydrocracking has considered as the primary

one of CCU system, and it also produces by-products, such as naphtha and electricity (R Davis, 2009; Ryan Davis et al., 2011). Biodiesel is a conventional fuel to use in the IC, and by-products from biodiesel production can improve the economics of the IC by sending them as raw material and utility to companies in the IC. Therefore, an integrated network between utility supply (US) and CO_2 mitigation (CM) is needed to reduce CO_2 emission in the IC, and CO_2 utilization as the CM needs to consider for the optimization of the integrated network.

This study presents the development of a deterministic model for the optimal design of an integrated network between US sub-strategy such as fossil fuel-based Rankine cycle and CM sub-strategy such as CO_2 capture (CC), storage (CS), and utilization (CU) systems. The objective of the proposed model is to minimize the expected total cost of the integrated network for meeting USCM demands considering US and CM sub-strategies.

2. Problem definition and formulation

2.1. Problem statement

In this study, a decision-making framework defined as the optimal way is found between the different decision levels, such as USCM strategies over the multi-period planning horizon. A source company with sub-nodes such as chemical plant, US system, and CM strategies, and a sink company with the same sub-nodes excluding the US system are considered in this study. CM strategies can be classified into three types: CC system captures CO_2 directly from the off-gas in each company (MEA-based absorption); CS system transports and stores CO_2 captured from CC system (storage in depleted oil and gas reservoirs); CU system transports and utilizes CO_2 captured from CC system to produce CO_2-based products (biodiesel).

2.2. Model formulation

2.2.1. Objective function

$$\text{Minimize } \text{Cost}^{\text{total}} = \text{Cost}^{\text{utility supply}} + \text{Cost}^{\text{CO}_2\text{ mitigation}}. \quad (1)$$

The objective function consisted of utility supply and CO_2 mitigation network cost and all of the model notations are shown in Table 1.

2.2.1.1 Utility supply network

$$\text{Cost}^{\text{utility supply}} = \text{Cost}^{\text{utility onsite}} + \text{Cost}^{\text{utility offsite}}. \quad (2)$$

$$\text{Cost}^{\text{utility onsite}} = \text{FICU} + \text{FOCU} + \text{VICU} + \text{VOCU} + \text{RMCU}. \quad (3)$$

$$\text{Cost}^{\text{utility offsite}} = \text{ESC}. \quad (4)$$

Utility supply network cost consisted of onsite and offsite utility supply network cost including fixed capital and operating, variable capital and operating, raw material, and external steam cost.

2.2.1.2 CO_2 mitigation network

$$\text{Cost}^{\text{CO}_2\text{ mitigation}} = \text{Cost}^{\text{CC}} + \text{Cost}^{\text{CS}} + \text{Cost}^{\text{CU}}. \quad (5)$$

$$\text{Cost}^{\text{CC}} = \text{ICC} + \text{OCC} + \text{RMCU}. \quad (6)$$

$$\text{Cost}^{\text{CS}} = \text{TCC} + \text{SCC}. \quad (7)$$

Table 1. Model notation

b	utilization site	i	company	r	green emission
d	boiler	l	stage	t	time
h	company	p	steam	u	carbon capture

gamma	Emission allocated rate (%)
$N_{t,i,a,m}^{initial_{inventory}}$	Initial inventory of fuel (t)
$x_{t,i,r}^{ir}$	Flue gas amount at company (t/y)
$Cost^{total}$	Total cost of proposed model ($/y)
$Cost^{utility\ supply}$	Total cost of utility supply network ($/y)
$Cost^{CO_2\ mitigation}$	Total cost of CO_2 mitigation ($/y)
$Cost^{utility\ onsite}$	Total cost of utility onsite supply network ($/y)
$Cost^{utility\ offsite}$	Total cost of utility offsite supply network ($/y)
FICU, FICCU	Fixed investment cost of US and CU system ($/y)
FOCU, FOCCU	Fixed operating cost of US and CU system ($/y)
VICU, VOCU	Variable investment, operating cost of US system ($/y)
RMCU, RMCC, RMCCU	Raw material cost of US, CC, and CU system ($/y)
$Cost^{CC}$, $Cost^{CS}$, $Cost^{CU}$	Total cost of CC, CS, and CU system ($/y)
ICC, OCC	Investment and operating cost of CC system ($/y)
TCC, TCCU	Transportation cost of CC and CU system ($/y)
SCC	Storage cost of CO_2 in CS system ($/y)
BCU, ESC	Benefit and external steam cost of CU system ($/y)
$RI_{t,i,a,m}$	Required amount of fuel (t/y)
$x_{m_{t,i,a,m,"OUT"}}^{a}$	Requirement of fuel at the fuel tank (t/y)
$x_{t,i,d,r,l}^{dr}$	GHG amount at boiler (t/y)
$x_{p_{t,i,e,p,l}}^{e}$, $x_{n_{t,i,e,wt,l}}^{e}$	Steam and water amount at turbine (t/y)
$z_{t,i,u,r,l}^{ur\ cc}$	Captured gas amount at CC system (t/y)
$z_{t,i,h,r,l}^{ur\ cs}$, $z_{t,i,b,r,l}^{ur\ cu}$	Captured gas amount entering CS and CU system (t/y)

$$Cost^{CU} = FICCU + FOCCU + RMCCU + TCCU - BCU. \quad (8)$$

CO_2 mitigation network cost consisted of CC, CS and CU system cost including invest, operating, raw material, transportation, storage cost and benefit.

2.2.2. Constraint

2.2.2.1 Boiler and turbine

$$RI_{t,i,a,m} = x_{m_{t,i,a,m,"OUT"}}^{a} - N_{t,i,a,m}^{initial_{inventory}}. \quad (9)$$

$$x_{p_{t,i,e,"HS","IN"}}^{e} = x_{p_{t,i,e,"MS","OUT"}}^{e} + x_{p_{t,i,e,"LS,"OUT"}}^{e} + x_{n_{t,i,e,"wt","OUT"}}^{e}. \quad (10)$$

The required amount of fuel is calculated between the requirement of fuel and initial inventory at each company. And the amount of high-pressure steam entering turbine is equal to the summation of medium and low-pressure steam and water leaving turbine.

2.2.2.2 CO_2 emission

$$\left(1 - \frac{gamma}{100}\right)\left(\sum_d x_{t,i,d,r,l}^{dr} + x_{t,i,r}^{ir}\right) = \sum_u x_{t,i,u,r,l}^{ur\ cc}. \quad (11)$$

The amount of CO_2 emission from production process of target product and utility system to the atmosphere can be emitted by emission permits allocated to each company.

2.2.2.3 CO_2 mitigation

$$\sum_u z^{ur\ cc}_{t,i,u,r,"OUT"} = \sum_{pt} z^{ur\ cs}_{t,i,pt,"CO2","IN"} + \sum_{brr} z^{ur\ cu}_{t,i,b,"CO2","IN"}. \quad (12)$$

The amount of CO_2 from CC system can be mitigated by CS and CU systems.

3. Results and discussion

3.1. Case study

To demonstrate the validity of the deterministic model considering USCM strategies, the case study of the Yeosu Industrial Complex (YIC), which is adopted as a benchmark to apply the real scenario in the Republic of Korea from the previous studies (Han and Lee, 2014; Kim et al., 2010), conducted to satisfying the USCM demands over multi-period in a year. This study considers (1) multi-site both demands, (2) company types with different utility systems, (3) CO_2 mitigation methods, and (4) single site for storage and multi-site for utilization. The companies located in YIC are composed five types (A, B, C, D, and E) with different specifications of the utility system. From A to D types of companies have various kinds of boilers and turbines, and they can produce utilities and meet their US demand by themselves. However, E type of the company does not have boiler and turbine and must meet its US demand by an external source. CM strategies jointly use the CC system and decide whether to use CS or CU system to mitigate CO_2. The simulation results were evaluated by applying the model using GAMS (CPLEX 12.6) on a computer with Intel® Core ™ i7-7700 CPU@3.60 GHz and 16 GB RAM. The optimization results for the deterministic model of CSCU sub-strategy were compared with the CS-only sub-strategy regarding (1) optimal cost (Table 2), (2) amounts of CO_2 mitigation (Table 3), and network flows for steam and CO_2 (Fig. 1).

3.2. Optimal cost

Table 2 showed USCM costs within total cost for CS-only and CSCU sub-strategies. Total cost for CSCU sub-strategy (US$ 181.19 x 10^6/y) was 8.5 % lower than the CS-only sub-strategy (US$ 197.96 x 10^6/y).

Table 2. Comparison of optimal costs between only-CS and CSCU strategies.

		Strategy (million US$/y)	
System	Cost	Only-CS	CSCU
Utility supply	Raw material	127.40	126.77
	Investment	0.52	0.57
	Operating	7.17	7.23
CC	Raw material	41.87	42.78
	Investment	19.24	21.23
	Operating	0.09	0.10
CS	Operating	1.67	1.55
CU	Raw material	-	31.10
	Investment	-	45.01
	Operating	-	19.33
	Benefit	-	-114.47
Total		197.96	181.19

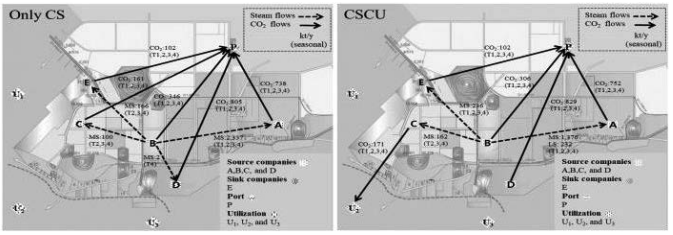


Figure 1. Comparison of network flows between five companies for only-CS and CSCU strategies.

The cost of US for CSCU sub-strategy (US$ 134.58 x 10^6/y) was 0.37 % lower than the CS-only sub-strategy (US$ 135.08 x 10^6/y). Notably, there was a distinct cost difference for the cost of CC system between CS-only and CSCU sub-strategies. The cost of CC system for CSCU sub-strategy (US$ 64.11 x 10^6/y) is 4.74 % (US$ 2.9 x 10^6/y) higher than the CS-only sub-strategy (US$ 61.21 x 10^6/y).This results showed that the amount of utilities used by the particular company (C type) is determined to provide the amounts of CO_2 demand in CU system and the larger capacity of CC system deployed in the other company (D type) than the previous capacity in order to satisfy the USCM demands of all companies through the network.

3.3. Network structure

In Table 3, the results showed that the CM amounts for CSCU strategy decreased compared to the result for only-CS strategy. The amount of CO_2 used in the CU system is 0.17 million t/y for only one of biodiesel production systems. Currently, the proposed model in the study was developed to select one system between CS and CU systems. If CU system was selected in a certain company, a utility system included in the company is implemented to emit the amount of CO_2 needed in one number of biodiesel production system. The details in steam and CO_2 flows among five companies, one port, and three utilization sites for only-CS and CSCU strategies were shown in Fig. 1. It shows that the USCM strategy has changed for satisfying the USCM demands in the IC. When comparing the only-CS and CSCU strategies, the specific portion is that the number of steam and CO_2 flows in CSCU strategy was different from only-CS strategy.

The number of steam flows decreased (only-CS: 4; CSCU: 3), and CO_2 flow to storage site decreased (only-CS: 5; CSCU: 4) and increased for utilization site (only-CS: 0, CSCU: 1) (Fig. 1). To provide the CO_2 required by the CU system, the CO_2 captured by the company C was provided to U_2 utilization site, and changes for the amount of utility exchanged and CO_2 captured can be confirmed to meet all USCM demands in the IC.

3.4. Sensitivity analysis for variation of biodiesel price

The most important parameter in this study is biodiesel price because it has the high influence to the benefit from the selling of biodiesel produced from the CU system. The CU system could be constructed in three sites, where the number of systems that could be constructed was limited to one. According to the limitation of construction for CU system, this system can produce the limited amount of the biodiesel.

Table 3. Comparison of the amounts of CO_2 mitigation between only-CS and CSCU strategies.

	Strategy (million t/y)	
	Only CS	CSCU
Storage	2.15	1.99
Utilization	-	0.17

Table 4. Sensitivity analysis of biodiesel prices for CSCU strategy.

Price ($/gallon)	Total cost (M$/ton)	CU system
11.73	181.19	O
11	187.93	O
10	197.19	O
9.9	197.96	X

Therefore, the construction of the CU system to produce biodiesel is determined by the price of biodiesel (Table 4). When the price of biodiesel is less than US$ 10/gallon, CS system is more economical than CU system in order to mitigate CO_2 because the cost of CU system is higher than that of CS system.

4. Conclusion

To find an optimal solution for the integrated network for USCM strategies, the deterministic model was developed. The goal of the proposed model is to minimize the expected total cost of the integrated network to meet the USCM demands. Comparison of the results between CS-only and CSCU sub-strategies was presented. To utilize captured CO_2, changes in the amount of utility transport and CO_2 capture occurred in the IC, and the biodiesel price has a significant influence on the use of CSCU strategy for CO_2 mitigation. As a further study, it will be possible to expand the model considering the alternatives in US system (renewable energy-based fuels) and CM system (biomass-based chemicals).

References

B. Li, Y. Duan, D. Luebke,B. Morreale, 2013, Advances in CO 2 capture technology: a patent review, *Applied Energy*, 102, 1439-1447.

F. Lima, P, Daoutidis, and M. Tsapatsis, 2016, "Modeling, optimization, and cost analysis of an IGCC plant with a membrane reactor for carbon capture." *AIChE Journal* 62.5, 1568-1580.

H. Han, I. Lee, 2014, A systematic process integration framework for the optimal design and techno-economic performance analysis of energy supply and CO 2 mitigation strategies, *Applied Energy*, 125, 136-146.

I. Karimi, and S. Kawi, 2016, "Technoeconomic perspectives on sustainable CO 2 capture and utilization.", In: Springer, 22223-22225.

G. Collodi, G. Azzaro, N. Ferrari, and S. Santos, 2017, "Demonstrating Large Scale Industrial CCS through CCU–A Case Study for Methanol Production." Energy Procedia 114, 122-138.

R. Davis, 2009, Techno-Economic Analysis of Microalgae-Derived Biofuel Production. Golden, CO: US Department of Energy, National Renewable Energy Laboratory.

R. Davis, A. Aden, P. T. Pienkos, 2011, Techno-economic analysis of autotrophic microalgae for fuel production, *Applied energy*, 88, 10, 3524-3531.

S. Jung, K. An, G. Dodbiba, T. Fujita, 2012, Regional energy-related carbon emission characteristics and potential mitigation in eco-industrial parks in South Korea: Logarithmic mean Divisia index analysis based on the Kaya identity, *Energy*, 46, 1, 231-241.

S. Kim, S. Yoon, S. Chae, S. Park, 2010, Economic and environmental optimization of a multi-site utility network for an industrial complex, *Journal of environmental management*, 91, 3, 690-705.

V. Sethi, 2017, "Low Carbon Technologies (LCT) and Carbon Capture & Sequestration (CCS)—Key to Green Power Mission for Energy Security and Environmental Sustainability." Carbon Utilization. Springer, 45-57.

Y. Ahn, J. Han, 2015, Strategic planning design of microalgae biomass-to-biodiesel supply chan in network: multi-period deterministic model, *Applied energy*, 154, 528-542.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50257-3

An optimization-based design and analysis of a biomass derived hydrogen energy system

Seolhee Cho[a†*], Wangyun Won[b†], Seulki Han[a], Sunghoon Kim[a], Chanhee You[a], Jiyong Kim[a*]

[a]*Department of Energy and Chemical Engineering, Incheon National University, Incheon, 22012, South Korea*
[b]*School of Civil, Environmental and Chemical Engineering, Changwon National University, Gyeongsangnam-do, 51140, South Korea*
[†] *equal contribution to this work.*
jykim77@inu.ac.kr

Abstract

In this study, we aim to design and analyze a comprehensive biomass-derived hydrogen energy system considering all processes from biomass collection to hydrogen storage or transport, and different biomass types. In addition, we compare the economics of this system with that of other non-renewable/renewable hydrogen energy systems to evaluate the feasibility of the proposed energy system. To achieve this goal, we first develop an optimization model using a mixed-integer linear programming (MILP) technique, then apply this model to the future hydrogen economy of Korea. We confirmed that biomass availability for hydrogen production is one of the critical factors in determining the configuration of the proposed energy system, contrary to a typical hydrogen energy system controlled mainly by hydrogen demand. Hydrogen supply cost is $23.1/GJ, and the main cost driver is the annualized capital cost of technology (35 %). While the hydrogen supply cost of the proposed energy system is higher than that of conventional routes ($10.3–14.7/GJ), it is more efficient than other renewable options (wind power: $25.5/GJ, Photovoltaic (PV): $34.0/GJ).

Keywords: Biomass, Hydrogen, Optimization model, Economic analysis, Korea

1. Introduction

Hydrogen is one of the most promising energy carriers in that it can provide both chemical and environmental advances: high specific energy (~142 MJ/kg) and zero carbon dioxide (CO_2) emission during combustion (Kim and Kim, 2016). Harnessing renewable energy sources (RES) for hydrogen production can maximize the strengths of a hydrogen energy system by integrating the merits of hydrogen and RES. Among different RES, biomass can be a main resource since it has the capability of being sustainably mass produced due to its regional abundance (Cho and Kim, 2015). Because of such an advantage, there are several studies on biomass-based hydrogen energy systems. Some researchers have analyzed the efficiency of technologies that convert biomass to hydrogen (Duman et al., 2014) and others have conducted techno-economic analysis of conversion technologies (Abuadala and Dincer, 2012). The previous studies focused mainly on improving the conversion efficiency or assessing the economics of hydrogen production technologies. However, to design the biomass-based hydrogen energy system as an alternative energy

system, researchers should perform an all-inclusive study taking other processes into consideration (e.g. cultivation of biomass crops and collection or hydrogen storage and transport) as well as the conversion process. Furthermore, they should include several biomass types such as residue or energy crop to ensure a stable biomass supply and to assess the feasibility of the biomass-based hydrogen energy system. Hence, in this study, we design a comprehensive biomass-derived hydrogen energy system using an optimization model and estimate the competitiveness by comparing the economics of the proposed energy system with that of other systems.

2. A biomass derived hydrogen energy system

A biomass-derived hydrogen energy system (Fig. 1) comprises different materials (i.e. biomass and hydrogen) and technologies (i.e. biomass storage, hydrogen production, hydrogen storage).

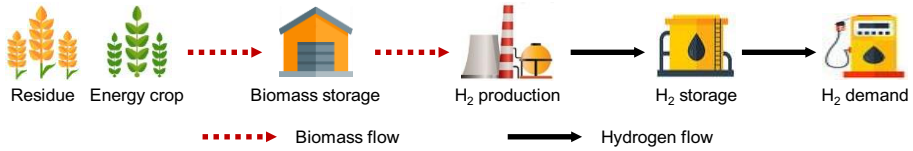


Fig. 1 The structure of the biomass derived hydrogen energy system

In this study, we consider two types of biomass: residue and energy crop. Residues are wastes obtained from other activities: e.g., a stump from a lumber field or woodchip from the paper industry. Although such resources are easily obtained, the amounts available for hydrogen production are limited by the fact that residues are widely used for different purposes (e.g. as a feed or a fertilizer). Energy crops mean additional resources which are not currently utilized but could be used for the sole purpose of hydrogen production. Even though, unlike residues, these types of biomass are freely utilized without limitation, several activities such as cultivation or crop harvesting are needed to make them fit for use as resources. Note that we do not consider food crops which can create a conflict between food and energy. There are three types of technologies. The properties of biomass (e.g., humidity or size) are maintained in biomass storage, and the handled biomass is transported to production technology. In this study, we choose a gasification process including gas cleaning, reforming, water gas shift reaction, and pressure swing adsorption as the conversion technology due to its easy operation and high economics. The hydrogen that is produced is stored in a hydrogen storage and used to satisfy the final hydrogen demand. As a hydrogen storage technology, we choose a liquid hydrogen storage because it has both high energy density and cheap transportation cost (Almaraz et al., 2014).

3. Optimization model

In this study, we develop an optimization model based on a typical hydrogen supply chain model. The objective function is to minimize the total annual cost required for establishing the proposed energy system.

$$TAC = BSC + ACC + TOC + TRC \tag{1}$$

where TAC is the total annual cost, BSC is the biomass supply cost required for collecting residue or utilizing energy crops, ACC is the annualized capital cost of technology, TOC

is the total operating cost of a technology, and *TRC* is the transportation cost. In this study, we develop a snapshot model that does not consider the time changes. There are four main constraints: 1) mass balance, 2) demand satisfaction, 3) resource availability, and 4) technology capacity.

3.1. Mass balance & demand satisfaction

$$B_{il} = \sum_{j \in J^B} \sum_{l' \in L} Q_{iljl'} \quad \forall i \in I, l \in L \tag{2}$$

where B_{il} is the amount of biomass I from all regions L, $Q_{iljl'}$ is the transported amount.

$$P_{jl'} = \sum_{i \in \left\{I, J^B, J^H\right\}} \sum_{l \in L} \alpha_j Q_{iljl'} \quad \forall j \in \left\{J^B, J^H, J^S\right\}, l' \in L \tag{3}$$

where the amount ($P_{jl'}$) produced in each technology (biomass storage J^B, hydrogen production J^H, and hydrogen storage J^S) is calculated by the transported amount from the previous process; for example, the amount of hydrogen produced in J^H is calculated by the amount transported from biomass storage J^B to hydrogen production. α_j is the process efficiency.

$$\sum_{j \in J^S} P_{jl'} \geq \rho_{l'} \quad \forall l' \in L \tag{4}$$

where the amount of hydrogen stored at J^S in all regions must be greater than or equal to the hydrogen demand ρ in all regions.

3.2. Resource availability

$$B_{il} = R_{il} + E_{il} \leq \upsilon_{il} \quad \forall i \in I, l \in L \tag{5}$$

where R_{il} is the amount of residues, E_{il} is the amount of energy crops, and υ_{il} is the available amount of both residue and energy crops in all regions.

3.3. Technology capacity

$$C_j^{min} N_{jl'} \leq P_{jl'} \leq C_j^{max} N_{jl'} \quad \forall j \in J, l' \in L \tag{6}$$

where the produced amount of biomass or hydrogen is subject to the minimum (C_j^{min}) and maximum capacities (C_j^{max}) of the corresponding technologies, and $N_{jl'}$ is the installed number of technologies in all regions.

4. Case study: Korean hydrogen economy from biomass

4.1. Hydrogen demand and biomass availability

In this study, we selected 15 Korean provinces as a case study and modified the demand used in our previous work (Kim and Moon, 2008)**.** The value is presented in Table 1. Here, residues can be divided into two types: woody residue and non-woody residue. While woody residue consists of lignocellulosic biomass like stumps or wood chips, non-woody residue, as an herbaceous biomass, includes straws, husks and reeds. The unit supply cost of residues is assumed to be $30/dry ton, and the total amounts available for hydrogen production can be found in Table 1. We also consider five types of energy crops:

corn stover, sugarcane bagasse, switchgrass, *Miscanthus*, and poplar. The information on energy crops can be found in Table 2.

Table 1. Hydrogen demand and biomass availability in different regions

Region	L1	L2	L3	L4	L5	L6	L7	L8
Hydrogen (ton/yr.)	130	42	30	41	20	23	16	200
Biomass (kton/yr.)	0.4	0.3	0.3	0.2	0.2	0.1	0.2	2.2
Region	L9	L10	L11	L12	L13	L14	L15	
Hydrogen (ton/yr.)	18	19	29	19	19	32	44	
Biomass (kton/yr.)	5.6	1.4	1.5	1.7	2.2	4.5	2.2	

Table 2. Technical and economic data of energy crops (Cheng and Timilsina, 2011)

	E1	E2	E3	E4	E5
Unit supply cost ($/ton)	74.5	192.7	58.3	41.3	130.2
Cultivation yield (ton/km^2)	667	934	1,357	1,334	1,112
By-product credit ($/ton)*	0.2	0.3	-	-	-

E1: corn stover, E2: sugarcane bagasse, E3: switchgrass, E4: Miscanthus, E5: popular
*: by-product credit means an additional benefit gained by selling the food crops.

4.2. Technology data

The economics of hydrogen production technology is largely contingent dependent on the type, capacity, and location. Hence, in this study, we adapted the technical values by taking into account the scale of the case study. The relevant data on the technologies is shown below in Table 3.

Table 3. Technical and economic data of all technologies (Won et al., 2017)

	B1	B2	H1	H2	S1	S2
Capacity (1,000 ton/yr.)	52	189	175	350	7	14
Capital cost (million $)	5	10	69	105	8	12
Operating cost ($/ton)	15	15	140	140	15	15
Lifetime (yr.)	20	20	15	15	20	20

B: biomass storage, H: hydrogen production, S: hydrogen storage

5. Results

We designed the biomass derived hydrogen energy system using a mathematic model. Table 4 shows the cost contributions of the proposed energy system. The total annual cost is 2.2 billion $/year, and the unit hydrogen supply cost is $23.1/GJ. Among all cost drivers, annualized capital cost of technology (ACC) accounts for 35 % of total annual cost. From this result, we identified that the economics of the proposed energy system largely depend on the capital cost of technologies.

Table 4. Cost distribution of the proposed energy system (million $/year)

	Biomass supply cost (BSC)	Annualized capital cost (ACC)	Technology operating cost (TOC)	Transport cost (TRC)	Total annual cost (TAC)
Value	449	778	304	372	2,233

As shown in Fig. 2, all technologies were installed in 14 regions, but not in L1. In particular, hydrogen production technology (H1 or H2) is primarily installed in L9 and

L14, where biomass availability is high. In addition, biomass or hydrogen storage facilities are installed close to production technology. From these results, we confirmed that installing technology in each region is more economical than transporting all materials between regions.

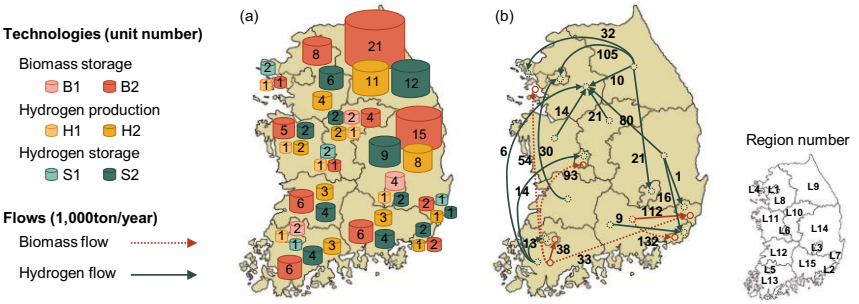


Fig. 2 (a) System configuration and (b) biomass or hydrogen flows

In this study, we suggest a strategy regarding energy crop supply detailing which type of energy crops should be cultivated in order to secure a stable supply. We increased the hydrogen demand up to 24 % from the baseline to identify the change in selected types. As shown in Fig. 3, residues are used as much as possible regardless of hydrogen demand, and the unmet demand is fulfilled by energy crops. To elaborate, when the demand of energy crop is relatively small, energy crops with by-product credit (E2) and low supply cost (E4) are selected. On the other hand, when the demand is high, the energy crop with a high cultivation yield (E3) is selected to produce as much hydrogen as possible.

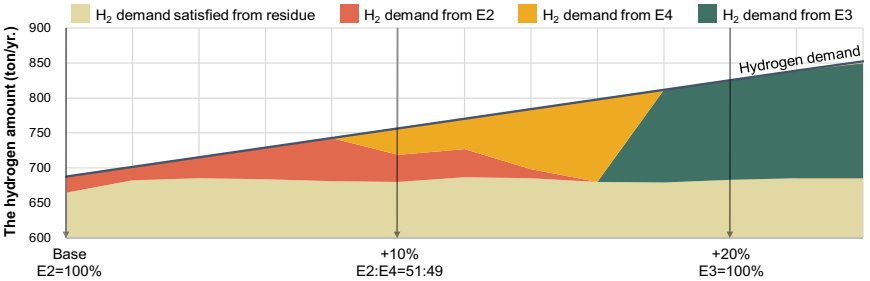
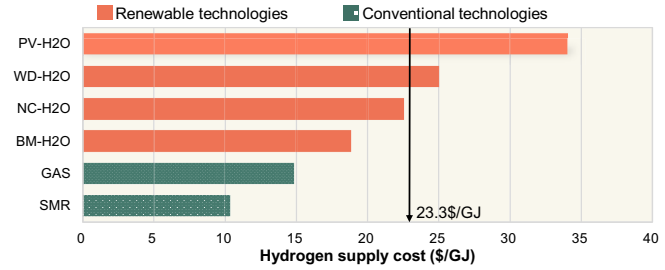


Fig. 3 Supply strategies of energy crops with demand change

Even though basic design and analysis are critical in establishing the biomass derived hydrogen energy system, it is also crucial to evaluate the feasibility of the proposed energy system by comparing the economics with those other systems.



PV-H2O: H_2 via PV electrolysis, WD-H2O: H_2 via wind electrolysis, NC-H2O: H_2 via nuclear electrolysis, BM-H2O: H_2 via biopower electrolysis, GAS: coal gasification, SMR: steam methane reforming

Fig. 4 Comparison of hydrogen supply cost in different technologies

According to Fig. 4, while using steam methane reforming (SMR) process is the cheapest pathway, using Photovoltaic-power (PV) for water electrolysis is required but has the highest supply cost. The hydrogen supply cost in this study is $23.1/GJ, and the value is cheaper than wind and PV based routes, but is more expensive than others. From this, we concluded that the biomass derived hydrogen energy system can be a viable alternatives based on the assumption that the conversion efficiency will be improved.

6. Conclusions

In this study, we designed and analyzed the comprehensive biomass-derived hydrogen energy system, considering all processes from biomass collection to hydrogen storage or transport and different biomass types. The major findings and contributions of this study are as follows:

- Most technologies are located in the regions where biomass availability is high rather than in the regions with high hydrogen demand, and the main cost driver of the proposed energy system is the technology capital cost.
- When the required amount of energy crop is small, the crops with subsidiary benefits and low supply cost are preferred. However, in cases of high hydrogen demand, the crops with high cultivation yield are selected to obtain as much hydrogen as possible.
- Compared to other renewable hydrogen options, the biomass-derived hydrogen energy system may be a promising technology.

This study provides a baseline model for the biomass-based hydrogen energy system. Although such fundamental design and analysis are useful, further rigorous study should be conducted to evaluate practical obstacles to the proposed energy system. In future, we will revise the optimization model in terms of several critical variables like demand uncertainty, change of biomass productivity, and improvement in conversion efficiency.

References

A. Abuadala and I. Dincer, (2012) A review on biomass- based hydrogen production and potential applications, Int. J. Energy Res, 36, pp. 415-455

S.D. Almaraz, C. Azzaro-Pantel, L. Montastruc, and S. Domenech, (2014) Hydrogen supply chain optimization for deployment scenarios in the Midi-Pyrénées region, France, Int J Hydrogen Energy, 39, pp. 11831-11845

J.J. Cheng and G.R. Timilsina (2011) Status and barriers of advanced biofuel technologies: A review, Renew. Energy, 36, pp. 3541-3549

S. Cho and J. Kim, (2015) Feasibility and impact analysis of a renewable energy source (RES)-based energy system in Korea, Energy, 85, pp. 317-328

G. Duman, M.A. Uddin, and J. Yanik, (2014) Hydrogen production from algal biomass via steam gasification, Bioresource Technology, 166, pp. 24-30

J. Kim and I. Moon, (2008) The role of hydrogen in the road transportation sector for a sustainable energy system: A case study of Korea, Int. J. Hydrogen Energy, 22, pp. 7326-7337

M. Kim and J. Kim, (2016) Optimization model for the design and analysis of an integrated renewable hydrogen supply (IRHS) system: Application to Korea's hydrogen economy, Int. J. Hydrogen Energy, 41, pp. 16613-16626

W. Won, H. Kwon, J. Han, and J. Kim, (2017) Design and operation of renewable energy sources based hydrogen supply system: Technology integration and optimization, Renew. Energy, 103, pp. 226-38

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50258-5

Resource Utilization & Supply Chain Optimization for Liquefied Gaseous Products

Shamik Misra[a], Mangesh Kapadi[b], Ravindra D. Gudi[a*]

[a]*Chemical Engineering Department,Indian Institute of Technology, Bombay,Mumbai-400076, India*
[b]*Praxair India Private Limited, Bangalore-560103, India*
ravigudi@iitb.ac.in

Abstract

In this paper, a resource utilization and supply chain optimization framework is presented. The reported MILP framework incorporates novel formulation techniques to address several complexities, such as, container utilization, vehicle capacity utilization, transit time, warehouse capacity & license limitation etc., which are part of any industrial supply chain. The results demonstrates the efficacy of the proposed framework in addressing several limitations & complexities that can arise in liquefied gaseous product supply chain. However, the presented approach requires more computationally efficient formulation techniques and that paves the way for future research in this domain.

Keywords: Stock Keeping Unit, Inventory Management, Distribution Scheduling.

1. Introduction

The liquefied gaseous products are transported in different type of cylinders or equivalent containers (such as, tonner, drums etc.).These containers vary in size, capacity, storing mechanisms etc. These containers are used for a specific products and have their own unique tracking IDs (contains the information regarding the product, manufacturer, etc.) attached to improve the efficiency and transparency of the inventory system. A specific container attached with uniquely identifiable bar code are called as, stock keeping units (SKU) (Sawaya and Giauque 1986). These SKUs are bottled in a plant, then transported to various warehouses using trucks or ships. The filled SKUs are sent to customers and empty SKUs are picked up from the customer and returned to the warehouse. Warehouse stores both filled & empty SKUs. As the SKUs are reusable, these empty SKUs are sent back to the plant for refilling purpose. The entire upstream and downstream network can become very large and complex, and often it has been noted that total transportation and distribution costs become quite higher compared to the actual production cost of the product (Samuelsson 2016). So, to attain an overall minimum transportation and distribution cost while maintaining a sustainable supply chain that ensures an efficient utilization of the assets, an optimum logistics & distribution schedule is necessary.

In recent past a few attempts has been made, to optimize the supply chain network for packaged gas. Singh et al. (2015) proposed a decomposition based framework to calculate optimum plant locations, production capacities, inventories & number of packages to be routed in distribution network. However, their algorithm does not focus on optimum logistics or distribution schedule. In a similar tone, Samuelsson (2016) presents a framework to optimally design the logistics network of a leading air separation industry, using an MILP framework. However, the approach does not provide any delivery

schedules. It also directly considers the product & does not provide any insight regarding SKU capacity or vehicle capacity utilization.

There are various bottlenecks in the above-mentioned supply chain which can be improved considerably using optimization techniques. The main characteristics and limitations of this supply chain are: 1) the delivery period can range from days to months, 2) limited storage space at warehouse as well as at customer location, 3) inventory for both filled and empty SKUs at warehouse needs to be maintained as the SKUs are reusable, 4) filled SKUs are sent to customer locations according to their demand & consumption profiles and demands need to fulfilled at any cost, 5) filled SKUs are sent to the customers using trucks or other vehicles in supply trip and empty SKUs needs to be carried back on the return trip, 6) SKU transport capacity for every vehicles according to the types and configurations needs to be respected, 7) Some liquefied gaseous products can be inflammable or hazardous and need special licenses to stock at warehouse.

The main objective of the optimization of this supply chain is to reduce the logistics (plant to warehouse transportation by ship or trucks) and distribution (warehouse to customers by truck or smaller vehicles) costs by proper utilization of available resources (SKUs) in the supply chain while respecting above mentioned constraints and limitations.

A mathematical formulation, to represent the above mentioned characteristics, is presented in the following. Though the formulation below is posed as an MILP, the framework is integer variable dominant, as the variables, to represent the number of trips to be executed or number of SKUs needed to be transported, have to be integer variables.

2. Formulation

A set of SKUs $k \in K$ carrying a set of products/items $i \in I$ are stocked in a set of warehouses $m \in M$. The SKUs are product specific, i.e. an SKU k can only be used for an item i. So, in the following formulation k will denote both the SKUs and the associated products. The empty SKUs in the warehouses are sent to the packaging centres $s \in S$ using ships. In the plant location the empty SKUs are filled and sent back to the warehouses by the same ships. So, the total travel time includes the onward & return journey time between the warehouses and the plants and the time required for filling the SKUs. The filled SKUs from the warehouses are then sent to the customer locations $c \in C$ using the delivery trucks. Both trucks & ships are part of a set of delivery vehicles denoted as $v \in V$. The trucks deliver the filled SKUs at customer locations and bring the used and empty SKUs back on its return journey. The waiting time required for delivering filled and collecting empty SKUs from customer locations are assumed negligible.

2.1. Stock Balance - Warehouse

Both filled & empty SKUs stocks are stored at warehouse. To maintain a steady supply chain, the more SKUs the warehouse can stock is the better. However, the trade-off is with the storage area expansion cost. Moreover, as the products are generally hazardous, special licenses, specifying maximum SKUs that can be stored, needs to be acquired.

$$FMQ_{m,k,d} = FMQ_{m,k,d-1} + \sum_{\substack{s=1 \\ (:MS_{m,s}=1)}}^{S} \sum_{\substack{v=1 \\ (:VT_v=0)}}^{V} SMSup_{s,m,v,k,(d-\sum_{a=1}^{A} TT_{m,s,v,a})} + \sum_{s=1(:MS_{m,s}=1)}^{S} SMPastSup_{m,s,k,d} - \sum_{r=1(:MR_{m,r}=1)} \sum_{v=1(:VT_v=1)}^{V} FMRDis_{m,r,v,k,d} + Add_{m,k,d} \quad \forall m,k(:MK_{m,k}=1), d(:d=1..ND) \tag{1}$$

Equation 1 describes the filled SKU inventory balance at warehouse. The variable, $FMQ_{m,k,d}$ indicates the inventory of SKU type k at a given day d at warehouse m. Similarly, the inventory balance for empty SKUs are expressed in Equation 2,

$$EMQ_{m,k,d} = EMQ_{m,k,d-1} - \sum_{\substack{s=1 \\ (:MS_{m,s}=1)}}^{S} \sum_{\substack{v=1 \\ (:VT_v=0)}}^{V} SMSup_{s,m,v,k,d} + \sum_{c=1(:MC_{m,c}=1)}^{C} RMPastDis_{m,r,k,d} + \sum_{r=1(:MR_{m,r}=1)} \sum_{v=1(:VT_v=1)}^{V} EMRDis_{m,r,v,k,d} \quad \forall m,k(:MK_{m,k}=1), d(:d=1..ND) \tag{2}$$

The empty SKUs collected from the customers are added in the empty SKU storage. These SKUs are then sent to the suppliers for refilling. The framework presented here can also include an option to facilitate the sale of extra empty SKUs in the supply chain. The maximum number of SKUs (both filled & empty) that can be stored in a warehouse at any day is limited by two factors: 1) storage capacity, 2) license for maximum storage. The former one is generally a hard constraint, while the second one is a soft constraint, i.e. the maximum stock limitation dictated by the license that can be increased by incurring suitable costs. The following equation calculates the maximum license violation in the entire horizon, which will then be minimized as a part of the objective function with suitable costs.

$$XML_{m,l} \geq \sum_{\substack{k=1 \\ (:KL_{k,l}=1)}} (FMQ_{m,k,d} + EMQ_{m,k,d}) - MLMax_{m,l} \qquad \forall m,l,d(d=1..ND) \tag{3}$$

The variable $XML_{m,l}$ denotes the extra licenses that need to be bought for a specific type of SKU $l \in L$. In this algorithm SKUs are considered as product specific, so similar type of SKUs, used for different products, are considered as different entities. The warehouses also required to maintain a minimum stock of filled SKUs to avoid disruption in the supply chain. The safety stock limitation constraints are represented using Equation 4.

$$FMQ_{m,k,d} \geq MinMK_{m,k} \qquad \forall m,k(:MK_{m,k}=1), d(:d=1..ND) \tag{4}$$

2.2. Vehicle Capacity Utilization

The entire distribution area associated with certain warehouse are divided into multiple regions based on the customer locations. This segregation of the distribution area reduces the rigorous vehicle routing effort which is needed to account single and multi drop scenarios. The assumption here is, once the vehicle reaches a particular region, all the customer in that region will be catered within a day. In this framework regions act as a virtual secondary distribution centres from where the customers can be catered without any delay. So, instead of a direct customer to warehouse suitability, connections between warehouse & region and region & customers are established. Each vehicle can carry a certain amount of SKUs in a trip. The minimum & maximum number of SKUs that can carried by a vehicle for a certain trip is dictated by the configurations. We introduce an

integer variable $y_{m,r,v,x,d}$ which indicates the number of trips that has been made from plant m to region r in vehicle type v with configuration x day d .The SKU quantity that can be carried on that trip is limited by the following equations,

$$\sum_{k=1(:KG_{k,g}=1)}^{K} FMRDis_{m,r,k,v,d} \geq \sum_{x=1(:VCap_{v,x,g}^{min}>0)} y_{m,r,v,x,d} * VCap_{v,x,g}^{min} \quad (5)$$
$$\forall m, r(:MR_{m,r}=1), v(:VT_v=1 \text{and} MRV_{m,r,v}=1), g(:VG_{v,g}=1), d(:d=1..ND)$$

$$\sum_{k=1(:KG_{k,g}=1)}^{K} FMRDis_{m,r,k,v,d} \leq \sum_{x=1(:VCap_{v,x,g}^{max}>0)} y_{m,r,v,x,d} * VCap_{v,x,g}^{max} \quad (6)$$
$$\forall m, r(:MR_{m,r}=1), v(:VT_v=1 \text{and} MRV_{m,r,v}=1), g(:VG_{v,g}=1), d(:d=1..ND)$$

Similar equations can be written for empty SKUs. As, the same vehicle picks up the empty SKUs from the customer locations, vehicle configurations at the return trips (customer to warehouse) are same as the onward trips (warehouse to customer). However, the number of SKUs carried in each vehicle can be different in return trips from the onward trips. The limitations in the logistics part can also be described using equation similar to Equation 5 & 6. The SKUs supplied from the warehouse reach the customer regions after a stipulated time indicated by the parameter $RTT_{m,r,v,a}$.

$$\sum_{m=1(:MR_{m,r}=1)}^{M} \sum_{v=1(:VT_v=1)}^{V} FMRDis_{m,r,v,k,(d-\sum_{a=1(:ATT_{ta}=1)}^{A} RTT_{m,r,v,a})} = \sum_{c=1(:CR_{c,r}=1)}^{C} FRCDis_{c,k,d} \quad (7)$$
$$\forall r, k, d(:d=1..ND)$$

Similarly, the empty SKUs from all the customers in that region are collected and sent to the warehouses associated with that region.

$$\sum_{m=1(:MR_{m,r}=1)}^{M} \sum_{v=1(:VT_v=1)}^{V} EMRDis_{m,r,v,k,(d+\sum_{a=1(:ATT_{ta}=0)}^{A} RTT_{m,r,v,a})} = \sum_{c=1(:CR_{c,r}=1)}^{C} ERCDis_{c,k,d} \quad (8)$$
$$\forall r, k, d(:d=1..ND)$$

2.3. Stock Balance – Customer Location

The inventory balance for filled and empty SKUs are also formulated for each customer locations. The continuous variables $FCQ_{c,k,d}$ and $ECQ_{c,k,d}$ indicate the filled & empty SKU inventory, at each customer location at every day, respectively in Equation 9 & 10.

$$FCQ_{c,k,d} = FCQ_{c,k,d-1} + FRCDis_{c,k,d} - DD_{c,k,d} + \sum_{m=1}^{M} MCPastDis_{m,c,k,d} \quad (9)$$
$$\forall c, k(:CK_{c,k}=1), d(:d=1..ND)$$

$$ECQ_{c,k,d} = ECQ_{c,k,d-1} - ERCDis_{c,k,d} + DD_{c,k,d} \quad (10)$$
$$\forall c, k(:CK_{c,k}=1), d(:d=1..ND)$$

The parameter $DD_{c,k,d}$ indicates the daily consumption of the product at every customer location in terms of SKU. The following constraint addresses the safety stock limitations at customer locations.

$$FCQ_{c,k,d} \geq UP_{c,k} - 1 \quad (11)$$

The parameter $UP_{c,k}$ indicates the number of use points available to connect a specific SKU type k at customer location c .

2.4. Objective Function

The main objective of the abovementioned framework is to attain 1) minimum logistics cost, 2) minimum distribution cost while ensuring optimum usage of the transportation vehicle capacity. The objective function for the abovementioned framework can be written as,

$$\min f = \sum_{m=1}^{M}\sum_{s=1}^{S}\sum_{v=1(:LC_{m,s,v}>0)}^{V} LC_{m,s,v} * \sum_{d=1}^{ND}\sum_{x=1(:YX_{v,x}=1)}^{X} y_{m,s,v,x,d} + \sum_{m=1}^{M}\sum_{r=1}^{R}\sum_{v=1(:SC_{m,s,v}>0)}^{V} DC_{m,r,v} * \sum_{d=1}^{ND}\sum_{x=1(:YX_{v,x}=1)}^{X} y_{m,r,v,x,d} +$$
$$\sum_{m=1}^{M}\sum_{l=1(:LC_{m,l}>0)}^{L} LC_{m,l} * XML_{m,l} + PC * \sum_{d=1}^{D}\sum_{v=1}^{V}\sum_{g=1}^{G}\left(\sum_{x=1(:VCap_{v,x,g}^{min}\neq VCap_{v,x,g}^{max})} VCap_{v,x,g}^{max} - \sum_{r=1}^{R}\sum_{m=1}^{M}\sum_{k=1(:KG_{k,g}=1)}^{K} FMRDis_{m,r,v,k,d} \right)$$

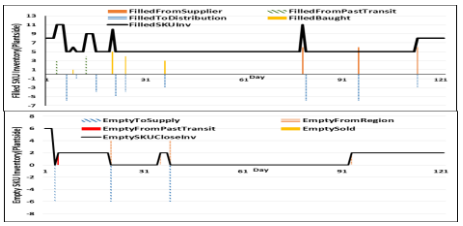
A detailed model incorporating more operational limitations, such as, 1) increasing or decreasing demands, 2) new and departing customers etc. are excluded from this paper due to brevity and will be a subject matter of a future publication (Misra *et al.* (2017)).

3. Result & Discussion

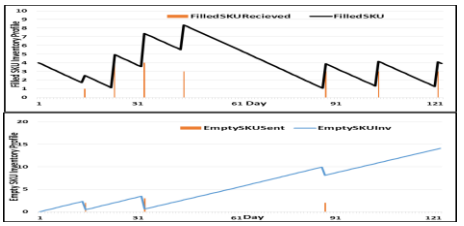
The abovementioned MILP formulation is tested upon a simple case study. Two liquefied gaseous products say A and B are stored in a warehouse in `tonner' type SKUs. Two suppliers are connected with the warehouse. Empty SKUs are sent to suppliers using ship for refilling purpose. 4 customer (say C1 to C4) are associated with the warehouse in which two customers has quoted demand for product A & the rest wants product B. The customer quotes demand for each month & the demands are converted to an average daily consumption pattern. Two regions are created based on customer locations & vehicle availability. C1 & C2 are located in region R1 whereas, C3 & C4 are in region R2. Two types of trucks are available & according to the vehicle customer SKU suitability, those are used for distribution purpose. The horizon is taken as 4 months starting from April to July. The entire horizon is discretized on a day basis.

The MILP formulation is solved for two different scenarios using Fico© Xpress Optimization suite using 'mmxprs' module version 2.8.1 on a Intel® Core™ i7(3.6 GHz) machine with 16GB RAM. Each of the scenarios contains 7885 constraints, 7323 variables in which 2440 are integer type, and is optimized for 15 minutes, and the optimality gap reached are 11.34% and 15.36 % respectively. After the abovementioned simulation time, there are no significant improvement in the solutions. As, the results presented in the following suggests, in both the scenarios, the framework aptly addresses the constraints regarding storage at warehouse or customer locations while meeting demands. The delivery time constraints are also respected at both logistics and distribution part as shown in Figure 1 & 2. Figure 1a and 2a also indicates that the framework exercises buying filled SKUs from other sources option to sustain the supply chain at the starting of the horizon. So, according to the result the framework aptly handles the limitations and provide optimum delivery schedules. However, the two scenarios presented here are also designed to focus on another important aspect of this supply chain: the circulation of SKUs. In the first scenario, there are no maximum limits on the customer storage. So, as the horizon progresses the empty SKUs are getting stored in the customer location as there are no incentives to bring those back to the warehouse. Moreover, there is a license limitation for the storage of SKUs at warehouse location. These two factors influence the decision behind stacking empty SKUs at customer location at the end of the horizon. The resultant warehouse and customer inventory profiles for both filled & empty SKUs are shown in Figure 1. The plots shown here only consider one customer and one SKU for brevity. The case in scenario 1 is not an ideal

one as, the SKUs need to be recirculated in the supply chain. This idea can be achieved by adding a storage capacity limitation at the customer locations. In scenario 2, the vehicles carry the empty SKUs back to the warehouse on their return trip opposed to scenario 1, where most of the return trips are empty i.e. vehicles are not carrying any empty SKUs specially near the end of the horizon. The results also indicates that the transit times among warehouse & customers and also among supplier and warehouses are respected, which can be seen from Figure 1 & 2.

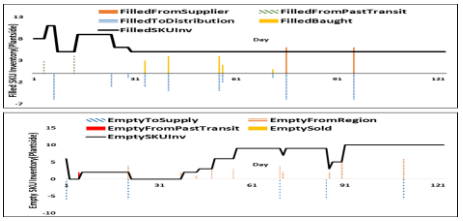


(a) Filled & Empty SKU Inventory at Warehouse

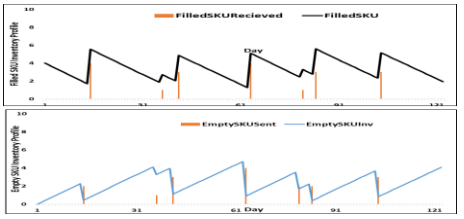


(b) Filled & Empty SKU Inventory at Customer Location

Figure 1: Filled & empty SKU inventory at customer location for scenario 1



(a) Filled & Empty SKU Inventory at Warehouse



(b) Filled & Empty SKU Inventory at Customer Location

Figure 2: SKU inventory profile at warehouse & customer location for scenario 2

4. Conclusions

In this paper, a novel MILP formulation is introduced to represent a resource utilization and supply chain optimization framework for liquefied gaseous products. The results show that the framework is formulation is rigorous and using the same a near optimal delivery and logistics schedule can be reached while respecting the limitations regarding warehouse and customer inventory, transit time and cost of the vehicles. The main bottle neck however is that even if the framework is optimized for longer time, the optimality gaps remains quite high, and requires more computationally efficient approach. As, the formulation is integer variable dominant, constraint programming approach might prove useful. A more systematic approach to improve the efficacy of the proposed framework is currently under investigation.

References

Samuelsson, Björn. 2016. “Estimating Distribution Costs in a Supply Chain Network Optimisation Tool, a Case Study.” *Operational Research* 16 (3): 469–99.

Sawaya, W J, and W C Giauque. 1986. *Production and Operations Management*. Harcourt Brace Jovanovich.

Singh, T. P., N. Neagu, M. Quattrone, and P. Briet. 2015. “Network Design for Cylinder Gas Distribution.” *Journal of Industrial Engineering and Management; 8 (1)*

Misra, S, M. Kapadi and R.D.Gudi. 2017 . "Distribution Scheduling and Resource Optimization of Packeged Gas Supply Chain ", *Manuscript under preparation*.

Guéret, C, C. Prins, M. Sevaux, and S. Heipcke, 2002. *Applications of Optimization with Xpress-MP*. Dash Optimization.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50259-7

Optimal design under uncertainty of carbon capture, utilization, and sequestration network considering benefit, environmental impact, and preference on risk

Suh-Young Lee[a*], In-Beum Lee[a], Jeehoon Han[b]

[a]*Department of Chemical Engineering, Pohang University of Science and Technology, Pohang, 37673, Korea*
[b]*School of Chemical Engineering, Chonbuk National University, Jeon-ju, 54896, Korea*
suhyoung@postech.ac.kr

Abstract

This study proposes a decision-making tool for stochastic planning of a carbon capture utilization and sequestration (CCUS) network by optimizing conflicting objectives under uncertainties. A scenario-based two-stage stochastic model is formulated to handle various sources of uncertainty, such as fluctuation of CO_2 emission level, variability of construction and operation cost, and unpredictable events. In the model, a concept of downside risk is introduced to evaluate the risk caused by uncertainties. This model holds its strength in that it provides an intuitive decision-making algorithm for financial impact, environmental impact, and risk on a CCUS network within a single framework. To evaluate environmental impact, Life Cycle Assessment method, specifically Eco-indicator 99, considering all contributions caused by installation and operations over a whole chain of CCUS is applied. By formulating as a stochastic multi-objective optimization model, it can evaluate the trade-off between total annual benefit with financial risk considering an environmental impact. This problem is formulated as a mixed integer linear program model with linear constraints. A case study for the industrial complex on the east coast of Korea in 2030 is presented to illustrate the applicability of the proposed model.

Keywords: CCUS, multi-objective optimization, Life Cycle Assessment, stochastic model, downside risk

1. Introduction

Carbon capture and sequestration (CCS) is seen world-wide and considered as the most promising technology to mitigate CO_2 released from industrial point source or power plant (Karim and Kawi, 2016 and Lima et al., 2016). However, the past years have shown that significant potential techno-economic bottlenecks are associated with CCS options mainly because of its expensive capture cost and geological sequestration limitation. For this reason, this paper gives an insight of alternative CO_2 mitigation solution: carbon capture utilization and sequestration (CCUS).

One of the key success factors for CCS projects is economics of scale which operated as a large-scale (1 Megaton CO_2 per year) based on Han et al. (2012). While potential capture plants are located near big cities where industrial complex is developed, geological sequestration site can be located in a limited area due to expensive onshore

storage cost and problems with public acceptance for possibility of leakage by uncertainties. To find ways to adapt to these problems, a nation-wide CCUS network should be introduced.

To tackle environmental uncertainties such as leakage from storage site and erosion of pipes, the model is formulated as a multi-objective problem based on a stochastic method (Ozturk and Turkay, 2016). To minimize the total cost and the associated environmental impact, there are recent studies conducted based on multi-criterion optimization (Guillen-Gosalbez et al., 2014). The model aims to maximize CCUS profit and minimize environmental impact by using Life Cycle Assessment (LCA), specifically Eco-Indicator 99 method (EI 99).

Moreover, to take into account the preference of risk caused by uncertainties, the model enables decision makers to choose their own tolerance of risk of a CCUS network by applying 2-phase-2-stage stochastic multi-objective problem (2P2SSMOOP) algorithm developed by Lee et al. (2017). Here is the main difference with previous studies that can derive a fixed risk-neutral optimal solution (M. Hasan et al., 2014).

This study addresses the following novel challenge, which differs from previous research: (a) introducing the concept of risk to a CCUS network, (b) incorporating financial risk problems with environmental concerns using LCA, and (c) enabling decision-makers to choose their own tolerance for risk by uncertainties.

2. Problem Statement

The goal is to find the optimum plan of CCUS infrastructure by developing a supply chain network model considering the preference of a decision maker. This paper presents a mathematical model for optimal design of a CCUS network to maximize profit while minimizing environmental impact and uncertainty risk as a solution to mitigate CO_2. The network superstructure for CCUS infrastructure is based on Han et al. (2011).

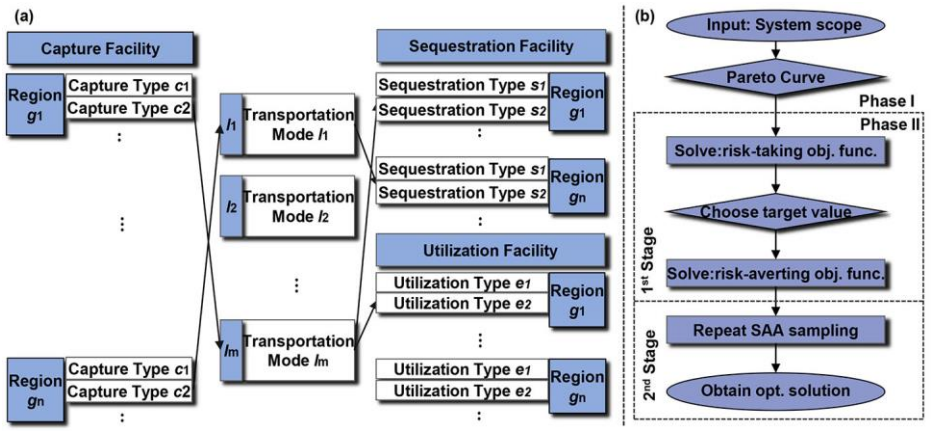


Figure 1: (a) CCUS network superstructure. (Components are described in the text)

(b) Solving algorithm for a 2P2SSMOOP

Given are the following data: A) the target amount of CO_2 reduction at each scenario, B) cost data for each type of facility construction and operation, C) capital charge rate and learning rate of facilities, D) capacity data of each facility, E) transportation availability

between regions. The model determines: (1) location and type of capture facilities, (2) location, type, and number of sequestration facilities, (3) type, capacity, and number of pipe lines, (4) location, type, and number of utilization facilities, and (5) flows of CO_2 between capture, utilization, and sequestration regions.

3. Mathematical Model

The mathematical formulation is presented as a mixed inter linear programming (MILP). The first stage variables of a two stage stochastic problem of Phase II (see Fig. 1) are implemented as either binary variables or integer variables; other variables are continuous variables for the second stage of Phase II. The mathematical model and notation of variables and parameters are based on the CCS network model developed by Lee et al. (2017) which optimizes the cost and environmental impact using a risk-considering approach. Several constraints used in the previous model are also applied in this model; these include constraints on facility capacity, transportation capacity, reduction target, cost, IE 99, and environmental impact management. The following additional equations are included in the CCUS network optimization model developed here.

Eq.(1) states overall mass balance of a certain region *g,* in a physical from CO_2 *i,* and for a scenario *sc*. $U_{sc,i,p,g}$ is a continuous variable for consumed amount of CO_2 by utilization in a physical form CO_2 *i,* by a utilization facility type *p*, in region *g*, for a scenario *sc*.

$$\sum_{c}\sum_{si}\sum_{sp} C_{sc,i,c,si,sp,g} = \sum_{l}\sum_{g'} (Q_{sc,i,l,g,g'} - Q_{sc,i,l,g',g}) + \sum_{s} S_{sc,i,s,g} + \sum_{p} U_{sc,i,p,g} \quad \forall sc, i, g \qquad (1)$$

Eq.(2) is a capacity constraint for CO_2 utilization facility. $BP_{e,p,g}$ is a binary variable which holds a value 1 if investment of utilization facility in a product form *e*, by a product facility type *p*, in region *g* is installed. $PCap_{e,p,g}$ is minimum and maximum availability of utilization facility form *e* by product facility type *p* in region *g*.

$$PCap^{\min}_{e,p,g} BP_{e,p,g} \le P_{ep,p,g,sc} \le PCap^{\max}_{e,p,g} BP_{e,p,g} \qquad \forall\, sc,\ e,\ p,\ g \qquad (2)$$

Eq.(3) and Eq.(4) are reduction target constraints. $CUF_{e,p}$ is a vale of CO_2 use factor in product form *e* by product facility type *p*. Eq.(3) states that the sum of CO_2 consumed amount by sequestration and utilization should be larger than the target reduction amount.

$$\sum_{i}\sum_{g}\sum_{s} S_{sc,i,s,g} + \sum_{i}\sum_{g}\sum_{p} U_{sc,i,p,g} \ge T_{sc} \qquad \forall sc \qquad (3)$$

$$U_{sc,i,p,g} = \sum_{e} CUF_{e,p} P_{sc,e,p,g} \qquad \forall sc, i, p, g \qquad (4)$$

Eq.(5) and Eq.(6) state cost of a CCUS network. $PCC_{e,p,g}$ is a capital cost parameter of establishing utilization facility of product form *e* by product facility type *p* in region *g*. $UPC_{sc,e,,p,g}$ is a unit operating cost parameter of utilization facility of product form *e* by product facility type *p* in region *g* for a scenario *sc.*

$$FCC = \sum_{g}\left[\frac{CCRf}{LR}\left\{\sum_{i}\left(\sum_{c}\sum_{si}\sum_{sp} CCC_{i,c,si,sp,g} BC_{i,c,si,sp,g} + \sum_{s} SCC_{i,s} NS_{i,s,g}\right) + \sum_{e}\sum_{p} PCC_{e,p,g} BP_{e,p,g}\right\}\right] \qquad (5)$$

$$FOC_{sc} = \sum_{g}\sum_{i}\sum_{c}\sum_{si}\sum_{sp} UCC_{sc,i,c,si} C_{sc,i,c,si,sp,g} + \sum_{g}\sum_{i}\sum_{s} USC_{sc,i,s} S_{sc,i,s,g} + \sum_{e}\sum_{p} UPC_{sc,e,p,g} P_{sc,e,p,g} \quad \forall sc \tag{6}$$

Eq.(7) states total annual selling benefits (TAB_{sc}) of CCUS of a scenario *sc* and Eq.(8) states total annual profit (TAP_{sc}) of a scenario *sc* .

$$TAB_{sc} = \sum_{e}\sum_{p}\sum_{g} USB_{sc,e,p} P_{sc,e,p,g} \qquad \forall sc \tag{7}$$

$$TAP_{sc} = TAB_{sc} - TAC_{sc} \qquad \forall sc \tag{8}$$

Eq.(9) and Eq.(10) state operation and installation environmental impact for each.

$$\begin{aligned} IO_{sc,n,x,g} = & \sum_{i}\sum_{c}\sum_{si}\sum_{sp}\sum_{b1} wo^{Ca}_{sc,b_1,c} v_{sc,n,x,b_1} C_{sc,i,c,si,sp,g} + \sum_{e}\sum_{p}\sum_{b_1} wo^{Ut}_{sc,b_1,s} v_{sc,n,x,b_1} P_{sc,e,p,g} \\ & + \sum_{i}\sum_{s}\sum_{b_1} wo^{Sq}_{sc,b_1,s} v_{sc,n,x,b_1} S_{sc,i,s,g} + \sum_{i}\sum_{l}\sum_{b_1} wo^{Tr}_{sc,b_1,l} v_{sc,n,x,b_1} \sum_{g'} Q_{sc,i,l,g,g'} \quad \forall sc,n,x \end{aligned} \tag{9}$$

$$\begin{aligned} II_{n,x,g} = & \sum_{i}\sum_{c}\sum_{si}\sum_{sp}\sum_{b_2} wo^{Ca}_{b_2,c} v_{n,x,b_2} BC_{i,c,si,sp,g} + \sum_{i}\sum_{s}\sum_{b_2} wo^{Sq}_{b_2,s} v_{n,x,b_2} NS_{i,s,g} \\ & + \sum_{e}\sum_{p} wo^{Ut}_{b_2,c} v_{n,x,b_2} BP_{e,p,g} + \sum_{i}\sum_{l}\sum_{b_2} wo^{Tr}_{b_2,l} v_{n,x,b_2} \sum_{g'} NT_{i,l,g,g'} \quad \forall n,x,g \end{aligned} \tag{10}$$

The environmental objective function (EI 99) is based on Eq.(11).

$$EI99_{sc} = \sum_{g}\sum_{r}\sum_{n} \vartheta_{r,n}\eta_n \left(\sum_{x} IO_{sc,n,x,g} + \sum_{x} II_{n,x,g}\right) \qquad \forall sc \tag{11}$$

To derive a CCUS infrastructure network considering the risk tolerance, the objective function for risk-taking and risk-averting can be formulated as follows: Eq.(12) states objective function for risk-taking decision makers and Eq.(13) and Eq.(14) state risk-averting objective function introducing the concept downside risk.

$$\min E[TAP_{sc}] = \sum_{sc} prob_{sc} TAP_{sc} \tag{12}$$

$$FDRisk(x, \Omega^{Fin}) = \sum_{sc} prob_{sc} \delta^{Fin}_{sc}(x, \Omega^{Fin}) \tag{13}$$

$$\delta^{Fin}_{sc}(x, \Omega) = \begin{cases} \Omega^{Fin} - TAP_{sc} & if \ \ TAP_{sc} < \Omega^{Fin}, \\ 0 & otherwise. \end{cases} \qquad \forall sc \tag{14}$$

4. Case Study

The presented mathematical model is used to assess a CCUS deployment scenario for the industrial complex on the east coast of Korea in 2030 the same as the previous case study reported by Han et al. (2011). As the reduction target, we set 38% of the estimated CO_2 emission from industrial plants in Korea. The cost data and capacity parameters used in this case study are calculated based on the Chemical Engineering Plant Cost Index using six-tenths factor rules. Environmental impact by constructing a CCUS network is evaluated by LCA analysis suggest by Goodkoop et al. (2001).

Table 1: Summary of the capture, sequestration, transportation, and utilization technologies

Activity	Type
Industry type	Power generation / Steel / Oil refinery / Petrochemical
Capture technology	Absorption using Monoethanolamine (MEA)
Sequestration technology	Depleted Gas Reservoir (DGR) / Saline Aquifer Storage (SAS)
Transportation mode	Compressed gaseous CO2 via pipeline
Utilization	Green-polymer / Bio-butanol

5. Results

A case study (see Fig. 2) was solved using the proposed mathematical model to plan the CCUS infrastructure on the east coast of Korea. The proposed model was implemented on GAMS 24.5 and solved using the CPLEX 12.0 MILP solver on an Intel 4.20-GHz computer. All solutions were obtained with zero optimality gaps. The case study was conducted to compare the finance risk-taking choice and the financial risk-averting choice considering environmental impact by using ε-constraint method widely used for a multi-objective programming (M. Ehrgott 2006).

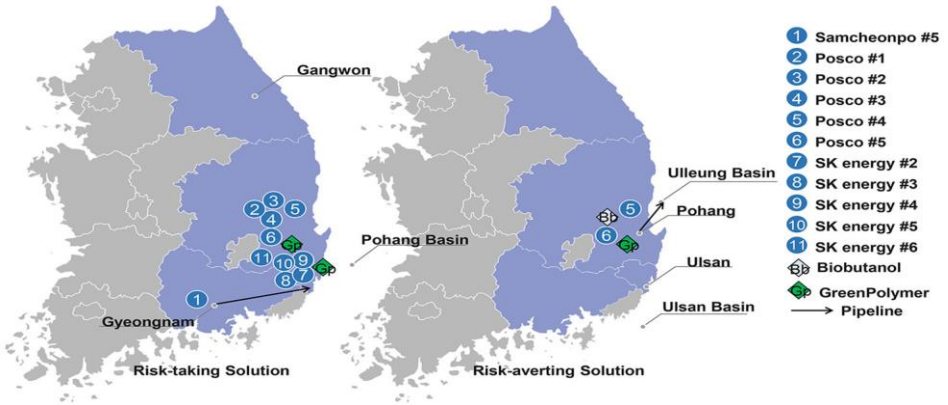


Figure 2: Case study results: financial risk minimization at target profit M$ 1,000 with ε: 0.1

Major difference between risk-taking decision and risk-averting decision is choice between utilization and sequestration. Since risk-averting decision makers avoid the worst case scenario which earns profit under a given target value, they tend to capture less amount of CO_2 and less utilize it. Moreover, to minimize transportation cost, they choose not utilization but sequestration. Since utilization is cost effective compared to sequestration, risk-taking decision makers try to capture and utilize CO_2 as much as they can in the boundary of the environmental burden they can afford of. However, this strategy is not effective for risk-averting decision makers because risk-taking decision leads to a wide range of profit results for each scenario. Thus, the downside risk of it is larger than that of risk-averting ones.

In both cases, green polymer is preferred as utilization technology and utilization facilities are established in Pohang city. This is due to the fact that the land price of Pohang is less expensive compare to Ulsan. Land price is the major key factor of capital cost for

utilization facility, thus capturing plant are determined in the same city where utilization facilities are located to reduce transportation cost.

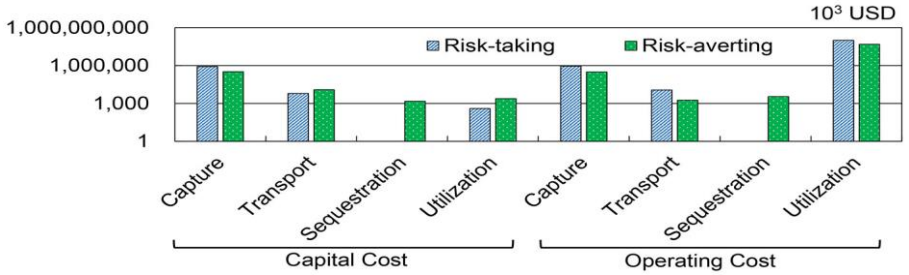


Figure 3: Case study results of capital, operating cost (10^3 USD) of CO_2 infrastructure planning

6. Conclusion

A mathematical modelling has been developed as a tool for planning a CCUS network by optimizing conflicting objectives under uncertainty. The goal is developing an intuitive decision-making algorithm for a stochastic programming problem with multiple objectives. The suggested model holds its novelty in that it provides the optimum CCUS network plan taking into account financial impact, environmental impact, and risk caused by uncertainty within a single framework. Conducted case study shows risk-taking decision makers invest more and earn more profit, while risk-averting decision makers capture less amount of CO_2 and prefer sequestration method lather than invest another utilization plant. Also, the location of capture plants is determined by the location of utilization plant which is highly dependent on land price. The overall model is addressed as MILP, which can be solved by a commercial linear optimization solver. The model is applied to the scenario of the east coast of Korea of 2030.

References

A. Ozturk, M. Turkay, 2016, AIChE J., 62, No. 9, 3473-3483

F. V. Lima, P. Daoutidis, and M. Tsapatsis, 2016, AIChE J., 62, No.5, 1568-1580.

I. A. Karimi, S. Kawi, 2016, Environmental Science and Pollution Res., 23, No. 22, 22223–22225

J. Han, I. Lee, 2011, Ind. Eng. Chem. Res., 50, No.10, 6297-6315.

J. Han, Y. Ahn, J. Lee, I. Lee, 2012, Korea J. Chem. Eng. 29, No. 8, 975-984.

M. Ehrgott, 2006, Springer Science & Business Media.

M. Goodkoop, R. Spriensma, 2001, PreConsultants BV.

M. Hasan, E. First, F. Boukouvalva, C. Floudas, 2014, Computer Aided Chemical Engineering, 34, 98-107

P. Vaskan, G. Guillen-Gosalbez, M. Turkay and L. Jimenez, 2016, Ind. Eng. Chem. Res., 53, No. 50, 19559-19572

S. Lee, J. Lee, I. Lee, J. Han, 2017, Applied Energy, 189, No. 1, 725-738.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50260-3

A supply chain analysis and design method based on the value of information

Nobuaki Ishii[a*], Masaaki Ohba[b]

[a] *Kanagawa University, 3-27-1 Rokkakubashi, Kanagawa-ku, Yokohama, 221-8686, Japan*
[b] *Nihon University, 1-3-2 Misaki-cho, Chiyoda-ku, Tokyo, 101-8360, Japan*
n-ishii@kanagawa-u.ac.jp

Abstract

A supply chain is a system consisting of different entities, which add value from the original resource sources to the final products while exchanging information on market, production, and so on. Although each entity is indispensable to obtain the final product, the degree of importance of each entity varies depending on the value of information each entity produces. In this paper, we apply the page ranking idea used in ranking Web pages to evaluate the value of information each entity produces in the supply chain. The developed method searches an information network of the supply chain according to the index, developed based on the value of information, for evaluating the entire supply chain. The accuracy of the information is also evaluated using a value map, then the method modifies the information network of the supply chain.

Keywords: eigenvector, matrix analysis, page rank, supply chain management.

1. Introduction

Because of the increasing severe competition in the global market, today's companies, including the process industry, have keen interests in making their supply chain (SC) more competitive than those of competitors. SC is a complex system consisting of entities, such as suppliers, manufacturers, warehouses, distributers, retailers, and so on. Thus, to design and operate SC effectively, Supply Chain Management (SCM) that focuses on the combination of strategies and tools to integrate all the entities of the SC (Bonfill et al., 2008) is expanded. Although a variety of solutions have been developed for SCM (Simchi-Levi, et al., 2007), almost all of them take advantage of information associated with SC in many ways.

In this paper, we develop a method of analysis and design of SC by focusing on the value and accuracy of information each entity produces. In the method, the value of information is calculated based on the same idea used for ranking Web pages. Then the method searches the information network of SC according to the ranking of each entity in order to control SC effectively based on the assumption that the higher ranked entity has a higher value of information, thus has stronger inference to the SC operation. In addition, the method modifies the information network of SC in consideration of the accuracy of the information each entity produces.

2. Evaluation of supply chain based on the value of information

Both researchers and practitioners, including the field of process systems engineering, have recognized that the information associated with SC, such as inventory, demand,

production plan, and so on, is key for effective management of the entire SC. In fact, literature on SCM has focused on this kind of information in many ways. For example, Lee et al. (2000) showed the value of information sharing in a two-level supply chain. Vegetti et al. (2005) proposed the design of an ontology that provides the foundations for the specification of information logistics processes in supply chains. Zamarripa, et al. (2014) showed a model that integrates third parties to optimize the planning decision in supply chains. Rasello and Manca (2014) focused on market information of oil prices, and presented a methodology to determine dynamic models of prices of different oils and the derived distillates in SC.

However, these types of literature have focused on tools and applications that use information associated with SC, and not on the analysis and design using the value and accuracy of information each entity produces in SC. In this paper, the method of analysis and design of SC in consideration of the value and accuracy of information is developed based on the principle shown below.

2.1. Principle of page rank calculation

The page rank is used for ranking Web pages, and it is considered that the page of a higher rank has a higher value of information. In the page rank calculation, $\boldsymbol{R}$ that holds the relationship of Eq.(1) is a vector representing the page rank of each web page (Page et al., 1999). That is, $\boldsymbol{R}$ is the eigenvector of the matrix $\boldsymbol{A}$ representing the structure of page links, and the element r_i (i =1 to n) of $\boldsymbol{R}$ is the page rank of page i. Where n is the total number of Web pages.

$$\boldsymbol{R} = c\,\boldsymbol{A}\,\boldsymbol{R} \tag{1}$$

In Eq.(1), $\boldsymbol{A}=(a_{v,u})$ is a square matrix in which row v and column u correspond to the links between Web pages, and its elements are determined as follows: where N_u is the total number of links from web page u to other web pages. From page u to page v, if there is a link: $a_{v,u} = 1 / N_u$, otherwise: $a_{v,u} = 0$.

2.2. Value of information of each entity in the supply chain

Since information is key to operate SC effectively, it is necessary to share and exchange information on production, inventory, transportation, etc. among entities constituting SC. Therefore various links of information among the entities are formed, and those links affect information each entity produces differently. Based on the idea of page ranking, it can be assumed that the entity ranked higher produces information that has a stronger affect than that of the lower ranked entity, thus we can say the entity has higher value.

In this paper, to calculate the value of information each entity produces by applying Eq.(1), SC is expressed in an $n \times n$ square matrix $\boldsymbol{A}$ from its entity number n, and the value of information each entity produces is calculated as its eigenvector $\boldsymbol{R}$. Where for element r_i of $\boldsymbol{R}$, $r_i \geqq 0.0$, $\Sigma_{i=1,n}\, r_i = 1.0$. The elements of $\boldsymbol{A}$ are obtained in Eq.(2) in consideration of the importance of information links from each entity.

$$a_{v,u} = W_{v,u}/N_u \tag{2}$$

where N_u is the total number of links from the entity u to other entities v. $W_{v,u}$ is a real number from 0 to N_u indicating relative importance of the links from entity u to v, and $\Sigma_{v=1,n} W_{v,u} = N_u$ (u=1 to n). If there is no link, $W_{v,u}$=0.0.

It is also clear that the different amount of values are added to the products incrementally at each entity via SC. Since we can say that the entity adding the higher value should have a higher value of information to manage the entire SC effectively, the difference between the value of information and the added value in each entity can be an index that measures the soundness of the entire SC from the view point of controllability. In other words, we can say that SC should be designed for the effective operation based on the index as Eq.(3) which determines the evaluation index of the entire SC, i.e., *SCV*, as the distance between $\boldsymbol{R}$ and $\boldsymbol{VA}$. Where $\boldsymbol{VA}$ is the vector of the relative value added by each entity i, and for element va_i of $\boldsymbol{VA}$, $va_i \geqq 0.0$, $\Sigma_{i=1,n}\, va_i = 1.0$.

$$SCV = |\boldsymbol{R} - \boldsymbol{VA}| \tag{3}$$

In addition, when the management level of the high ranked entity, where many pieces of critical information gather, is low, inaccurate production plans or inventory data etc. propagate to the entire SC. This will cause poor decision making in each entity. In other words, it can be said that the management level of the entity having high value of information among entities should be kept high.

However, all the entities within SC are not always within the framework that can manage SC effectively. It is difficult to manage adequately the entities outside the controllable framework, such as an external organization, and so on. In such a case, increasing the value of information of the manageable entity is required to reduce the influence to the entire SC by the inaccurate information of the uncontrollable entities.

3. Method of supply chain analysis and design based on the value of information

Based on the observations up to the previous section, we develop the analysis and design method of SC consisting of two stages. The method searches the information network of SC according to the *SCV* index in the first stage, then the information network is modified in consideration of the accuracy of information in each entity of SC.

3.1. First stage: Design of information network based on SCV index

Based on *SCV*, the developed method searches the information network of SC so that the *SCV* index is as small as possible based on the following steps:

Step 1.1: Calculate the *SCV* by Eq. (3) from the eigenvector $\boldsymbol{R}$ calculated by Eq.(1) and the vector $\boldsymbol{VA}$ showing the relative value added in each entity i in SC.

Step 1.2: Attempt to minimize the *SCV* by the following evolution rules R1 and/or R2.

(a) R1: For the element k with the largest difference between $\boldsymbol{R}$ and $\boldsymbol{VA}$, review the value of $a_{k,u}$ (u=1 to n) to reduce the difference between the value of information and relative value added to k.

(b) R2: For the element k with the largest difference between $\boldsymbol{R}$ and $\boldsymbol{VA}$, where $r_k < va_k$, makes new links from elements not currently connected to k.

3.2. Second stage: Modify the information network based on accuracy of information

The information network of SC designed in the first stage is modified by the accuracy of information according to the following steps:

Step 2.1: From the relative value of information in each entity and its relative accuracy, as shown in the value map of Figure 1, the position of each entity, i.e., Area I, Ⅱ, Ⅲ, and Ⅳ, is determined. The relative value of the information is obtained by Eq.(4). Where r_i is the rank of entity i, and n is the total number of entities in SC.

$$r_i \cdot n - 1 \tag{4}$$

Step 2.2: In the value map, the entity whose accuracy is relatively low, despite the high value of information, is located in Area Ⅳ, thus the following actions are taken based on the controllability of the entity to improve its accuracy.

(a) Information accuracy is controllable in the relevant entity: Improve management level of the relevant entity in order to move the entity in the direction of Area Ⅱ.
(b) Information accuracy is NOT controllable in the relevant entity: Establish a new entity that has equivalent function of the relevant entity, or construct a new information link in SC in order to move the entity in the direction of Area Ⅲ.

Step 2.3: If there is an entity in Area IV, repeat Steps 2.1 and 2.2. Otherwise, terminate all the Steps.

4. Numerical examples and discussion

To illustrate the developed method of analysis and design of SC, we show numerical examples using a simple SC consisting of five entities, i.e., A, B, C, D and E. Figures 2 and 3 show the flow of materials and the initial information links among entities in the example SC, respectively. In Figure 3, the real-valued numbers indicate the ranks of each entity, i.e., the value of each element of $\boldsymbol{R}$. The values of the fraction indicate the importance of information links from each entity, i.e., the value of each element of $\boldsymbol{A}$.

First, in Step 1.1, the importance of information links from each entity is expressed as each element of a 5×5 matrix $\boldsymbol{A}$ as Eq.(5), then $\boldsymbol{R}$ is calculated. We assume that the matrix $\boldsymbol{A}$ can be set by evaluating the information exchanged among entities in SC. In this example, as shown in Figure 3, $\boldsymbol{R}$ is calculated as (0.308, 0.189, 0.396, 0.0651, 0.0437). In addition, if the relative value added $\boldsymbol{VA}$ by each entity i is set to (0.25, 0.35, 0.2, 0.1, 0.1), the SCV is calculated as 0.264 by Eq.(3).

$$R = c \begin{pmatrix} 0 & 1/4 & 2/3 & 0 & 0 \\ 1/4 & 0 & 1/6 & 0 & 1 \\ 3/4 & 1/2 & 0 & 1 & 0 \\ 0 & 0 & 1/6 & 0 & 0 \\ 0 & 1/4 & 0 & 0 & 0 \end{pmatrix} R \tag{5}$$

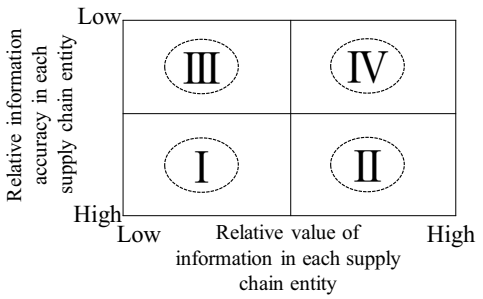


Figure 1. Value map of supply chain entity.

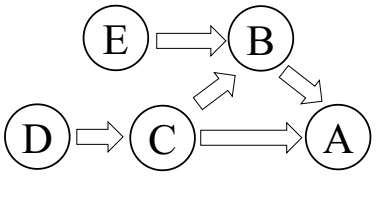


Figure 2. A materials flow of a supply chain.

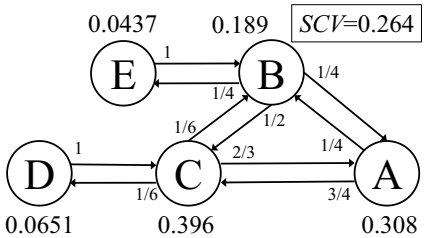


Figure 3. Information links and ranks (Case 0).

Next, in Step 1.2, *SCV* is reduced by rules R1 and/or R2. In Case 1, the importance of the information link from entity A to B is increased from 1/4 to 1/2 by rule R1. As a result, as shown in Figure 4, ***R*** and *SCV* are calculated as (0.286, 0.264, 0.330, 0.0549, 0.0659) and 0.169, respectively. Namely *SCV* is reduced by about 36% compared to Case 0. In addition, a new link of information from D to B is applied in Case 2 by rule R2. As shown in Figure 5, it reduced *SCV* to 0.140. In Step 2.2, the information network is modified according to the information accuracy of each entity so that the entity categorized in Area IV moves in the direction of Area II or III. For example, in Case 2 shown in Figure 5, the relative values of information of IV in the value map shown in Figure 1. Furthermore, if it is difficult to manage the information on entity C by SC, the method considers setting up entity F as an entity having the same function as entity C. Figure 6 shows the new information network of SC, its new ranking, and *SCV* when entity F is added. From Figure 6, it can be seen that the addition of a new entity F lowered the rank of entity C from 0.309 to 0.193. The relative value of information decreased from 0.546 to 0.156. However, *SCV* increases from 0.140 to 0.203. The results suggest the necessity of the process that repeats the first and the second stages of the developed method, or integrates the stages to design the information network that minimizes *SCV*.

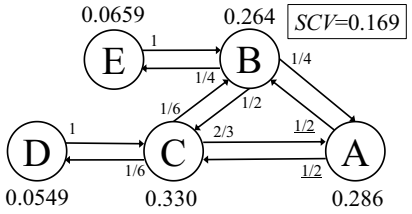


Figure 4. Information links and ranks (Case 1).

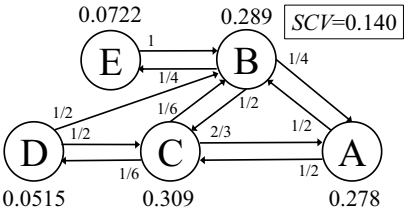


Figure 5. Information links and ranks (Case 2).

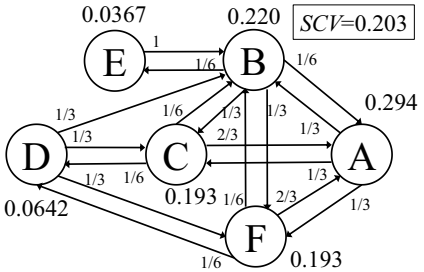


Figure 6. Information links and ranks (Case 3).

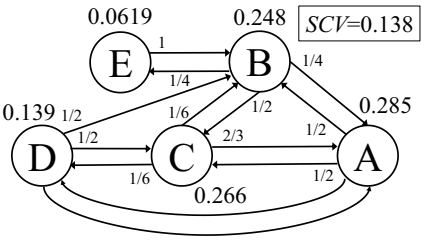


Figure 7. Information links and ranks (Case 4).

In addition, the rankings of Case 4 that makes new information links between entities D and A are shown in Figure 7. In this case, the ranking of entity C has decreased from 0.309 in Case 2 to 0.266, and the relative value of information has decreased from 0.546 to 0.327, thus the category of entity C moves from Area Ⅳ to the Area Ⅲ direction on the value map. The *SCV* also changed from 0.140 to 0.138.

5. Conclusions

In this paper, we have addressed a method of analysis and design of SC based on the value and accuracy of information each entity produces. In the method, we apply the page rank idea used in ranking Web pages to evaluate the value of information each entity produces. The effectiveness of the proposed method is demonstrated via numerical experiments using a simple SC model.

There are several issues that require further investigation. For example, the required data and process to determine the importance of each link within the information network of SC should be investigated. In addition to the importance of information, diverse factors must be considered, such as cost, risk, technical feasibility, and so on, to make a well-designed SC. The method that integrates these factors should be developed.

References

A. Bonfill, A. Espuña, L. Puigjaner, 2008, Decision support framework for coordinated production and transport scheduling in SCM, Computers and Chemical Engineering, 32, 6, 1206-1224.

H.L. Lee, K.C. So, S.S. Tang, 2000, The value of information sharing in a two-level supply chain, Management Science, 46, 5, 626-643.

L. Page, S. Brin, R. Motwani, T. Winograd, 1999, The PageRank citations ranking: bringing order to the Web, Technical Report, Stanford InfoLab.

R. Rasello, D. Manca, 2014, Stochastic price/cost models for supply chain management of refineries, Computer Aided Chemical Engineering, 33, 433-438.

D. Simchi-Levi, P. Kaminsky, E. Simchi-Levi, 2007, Designing and Managing the Supply Chain: Concepts, Strategies, and Case Studies, McGraw-Hill; Thrid Edition.

M. Vegetti, S. Gonnet, G. Henning, H. Leone, 2005, Information logistics for supply chain management within process industry environments, Computer Aided Chemical Engineering, 20, 1231-1236.

M. Zamarripa, K. Hjaila, J. Silvente, A. Espuña, 2014, Tactical management for coordinated supply chains, Computers and Chemical Engineering, 66, 110-123.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50261-5

Strategic planning for managing municipal solid wastes with consideration of multiple stakeholders

Alicia Danae Diaz-Barriga-Fernandez[a], Jose Ezequiel Santibañez-Aguilar[b], J. Betzabe González-Campos[a], Fabricio Nápoles-Rivera[a], Jose M. Ponce-Ortega[a*], Mahmoud M. El-Halwagi[c]

[a]*Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Mich., 58060, México*
[b]*Tecnológico de Monterrey, Monterrey, Nuevo Leon, 64849, Mexico*
[c]*Texas A&M University, College Station, Texas, 77843, USA.*
jmponce@umich.mx

Abstract

Municipal solid waste management (MSWM) is associated with the control of generation, storage, collection, transportation, processing and disposal of solid waste in accordance with the principles of public health, economics, engineering, conservation, aesthetics and other environmental considerations. However, the management of municipal solid waste (MSW) involves multiple stakeholders such as government agencies, suppliers, consumers, providers of treatment/recycle services and transporters. Furthermore, the product prices constantly change due to offer and demand, and the composition of the generated MSW depends on many factors including the place where they are produced, the season and the economic development. An optimal management strategy should take into account the uncertainties in the supply chain of MSW as well as the multiple participants involved. This way, an optimum and more realistic waste management system that creates synergistic opportunities to benefit the multiple stakeholders can be obtained. This paper presents a multi-objective optimization approach for the strategic planning of a municipal solid waste management system under uncertainty. The formulation considers tasks such as recycle, reuse, transportation, separation and distribution. The proposed approach also takes into account the multiple stakeholders, with the objective of maximizing the benefit of all participants. A case study from Mexico is analyzed to probe the applicability of the proposed model. Furthermore, this model considers the financial risk by analyzing the optimistic, mean and pessimistic cases. This way, the obtained results provide additional information about the system to improve accuracy in decision-making.

Keywords: Multi-objective optimization; Multi-stakeholders, Municipal solid waste; Uncertainty; Waste management.

1. Introduction

Municipal solid waste management is an important environmental concern around the word. Solid wastes are constantly produced in mega-cities, small towns and large villages, and if residues do not receive the right management, they can cause several affections on the environment and human health. These affections include greenhouse gases emissions, soil contamination, bad odours, underwater contamination, and spread of diseases, among others. Commonly, the treatment of MSW is not carried out because this requires a substantial proportion of government budget without any economic remuneration. Due to

most of the projects for MSWM depend on the stakeholders interested in founding the project, and all stakeholders are interested in satisfying their expectation, it is needed to propose an optimization approach for the MSWM that takes into account the multi-stakeholders involved. Furthermore, as the production and the composition of the municipal solid waste (MSW) depend on many factors including the population growth, consumption patterns, season and climatic conditions, the uncertainties associated to the MSWM must be considered. In this way, several contributions have been published, Habibi et al. (2017) developed a multi-objective optimization model for the supply chain of MSW in Tehran, Entezaminia et al. (2016) presented a model for the production planning in a green supply chain considering the traditional production system with recycling hubs. Other publications have focused on predicting the generation of the MSW through linear and nonlinear models, considering certain variables that affect the amount of MSW generation (Sun and Chungpaibulpatana, 2017), or analysing the economic viability of a power generation project from MSW (Pin et al., 2017). The dependence on time in the supply chain of MSW is an important issue, which has been recently considered (Santibañez-Aguilar et al., 2017; Nguyen-Tron et al., 2017; Zhang and Huang, 2013; Dai et al., 2012). To determine the incentives for the government and recycling industries to invest in the MSWM, Zheng et al. (2014) analysed the application of the landfill gas-fired and MSW incineration to produce power, and presented a series of preferential policies and regulations to encourage the expansion of MSW to energy. Wang et al. (2018) introduced a multi-attribute decision analysis method for prioritizing the MSW treatment alternatives based on the interval-valued fuzzy set theory. Nevertheless, none of the papers mentioned above has taken into account the uncertainty in the availability of the MSW, the demand of the products obtained from them, or the multi-stakeholders involved at three different level of financial risk (optimistic, mean and pessimistic). Therefore, this paper proposes a model formulation to provide to the involved stakeholders a tool for decision making from a wider perspective.

2. Model Formulation

To develop the mathematical model, a superstructure of the analysed system was proposed (Figure 1). Figure 1 shows a set of neighbouring cities, where the MSW are collected to be distributed to different processing facilities. Each industry has a set of technologies to treat a certain type of residue. It is assumed that only one type of product is produced from each type of technology. Then, the products are sent to different hubs located in specific places to be sent to markets or an industry as feedstock. The proposed mathematical model is describe as follows.

To analyse the involved financial risk, and determine the probability to obtain the best results, the next relationships are used. The optimistic objective function must be the greatest of all the objective functions o in the different scenarios e.

$$OptFunction_{o} \geq Function_{o,e} \qquad \forall o \tag{1}$$

It is necessary to limit the value of the optimistic function through the next set of relationships.

Equation (2) stays that only one binary variable can be one.

$$\sum_{e} y_{o,e}^{op} = 1 \qquad \forall o \tag{2}$$

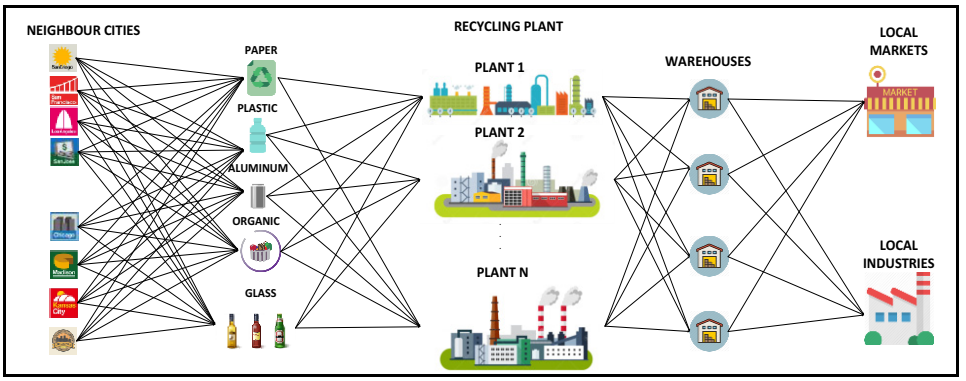


Figure 1. Superstructure for the municipal solid waste management system.

Furthermore, Equation (3) states that if the value of the binary variable is 0, then the optimistic function will be always lower than de right side of the equation. Thus, the parameter $OptF^{max}$ is an upper limit. Otherwise, if the binary variable takes a value of 1, the relationship in the disjunction is activated.

$$OptFunction_{o,e} \leq Function_{o,e} + OptF^{max}\left(1 - y_{o,e}^{op}\right) \qquad \forall o,e \tag{3}$$

The mean case is the average of the each objective function, in other words, it is equal to the sum of the objective functions for each scenario divided by the number of scenarios:

$$MeanFunction_o = {\sum_e Function_{o,e}} \Big/ {card(e)}, \quad \forall o \tag{4}$$

The pessimistic scenario corresponds to the lowest value for the objective function o between all the objective functions of a certain type o for all the scenarios.

$$PessimisticFunction_o \geq Function_{o,e} \qquad \forall o \tag{5}$$

Once the optimistic, mean and pessimistic cases are analysed, it is important to normalize the objective function for these three cases to formulate an equation that involves the multi-stakeholders. Equation (6) describes the normalized optimistic function, which is obtained by the difference between the greatest value obtained for the objective function when it was maximized individually (upper bound) and the optimistic function, divided by the difference between the upper bound for the optimistic function minus the lower bound for the optimistic function:

$$NOF_o = \frac{OF_o^{UB} - OptFunction_o}{OF_o^{UB} - OF_o^{LB}} \qquad \forall o \tag{6}$$

Similarly, the normalized mean function is equal to the relation between the upper bound for the mean function minus the mean function and the upper bound for the mean function minus the lower bound for the mean function:

$$MOF_o = \frac{MF_o^{UB} - MeanFunction_o}{MF_o^{UB} - MF_o^{LB}} \qquad \forall o \tag{7}$$

In the same way, for equation (8) the normalized pessimistic function is equal to the difference between the upper bound for the pessimistic function and the pessimistic function, divided by the difference between the upper bound for the pessimistic function minus the lower bound for the pessimistic function:

$$POF_o = \frac{PF_o^{UB} - PessimisticFunction_o}{PF_o^{UB} - PF_o^{LB}} \quad \forall o \tag{8}$$

The overall compromise function is the sum of all the normalized objective functions for the three cases (optimistic, mean and pessimistic).

$$OCS = \sum_o \left(NOF_o + NMF_o + NWF_o \right) \tag{9}$$

3. Results and discussions

Three different levels of financial risk for each objective function were analysed, then the upper and lower values for each objective function were calculated (see Table 1). These values were used to implement the multi-stakeholder approach. The results for the overall compromise solution (OCS) are shown in Figure 2.

Table1. Upper and lower bounds for the objective functions according to the financial risk.

OBJECTIVE FUNCTION	OPTIMISTIC		MEAN		PESSIMISTIC	
	UPPER BOUND	LOWER BOUND	UPPER BOUND	LOWER BOUND	UPPER BOUND	LOWER BOUND
%CONSUMED	99.96	0.46	99.96	4.10	99.95	4.039
ONP	-10.61	-14.75	-11.24	-14.70	-11.32	-14.72
NPP	96.67	-525.28	96.67	-517.55	67.85	-524.35
NPM	0.00	-11.94	0.00	-11.94	0.00	-11.95
NPG	2.15	-2.50	2.15	-2.50	2.15	-2.50
NPRP	7.94	-4.87	7.94	-4.87	7.24	-4.86
NPO	0.00	-14.26	0.00	-14.26	0.00	-16.20
GOVEXP	99.95	-483.75	99.95	-483.80	80.01	-483.75

The objective functions with the highest values are the percentage of consumed MSW (%CONSUMED), the overall net profit (ONP) and the net profit from organics (NPO): 97.66%, 73.69$MM/y and 81.22$MM/y, respectively. The curve for the %CONSUMED is constant at first, just two of the ten considered scenarios show a different value with consumptions of 65% and 45%. While the curves for ONP and NPO are similar, they tend to decrease almost constantly, except for one scenario that shows the lowest value for this function with zero for the NPO and a negative value for the ONP. The curves with the lowest values, but still with profits, are the net profit for paper (NPP) and the net profit for metals (NPM). These curves keep almost a constant linear trend, the highest values from them are 7.3 and 1.7$MM/y. For most of the scenarios, these curves take values very similar at the end. The curve for the net profit from recycled plastic (NPRP) takes a value of zero, since recycling plastic is not feasible for this problem; this also can be seen

in Table 1, where the upper bound for NPRP is zero. Now, analysing the results for the government expenditures (GOVEXP), the curves show a linear behaviour with a smooth slope. The range of values taken for this curve are between -10.9 and -13.48$MM/y.

Figure 3 shows the dissatisfaction percentage for each objective function when the OCS was minimized. The functions with less percentage of dissatisfaction, in most of the scenarios, were the NPRP, the NPM and the government expenditures; the best values for them were 0 %, 3.2 % and 3.7 %, respectively, and the worst 2.7 %, 17 and 17.1 %, respectively. For the % CONSUMED, the percentage is low in most of the scenarios, just in two of ten the percentage increase drastically from 2.4 % to 56.8 %. Similarly, the NPM and NPP have the lowest value of 7.8 % and 4.81 %, respectively; then, they tend to increase to 46.3 % and 61.9 %. The objective function with the greatest percentage was the NPG with a value between 13.6 -16.7 % for most of the scenarios.

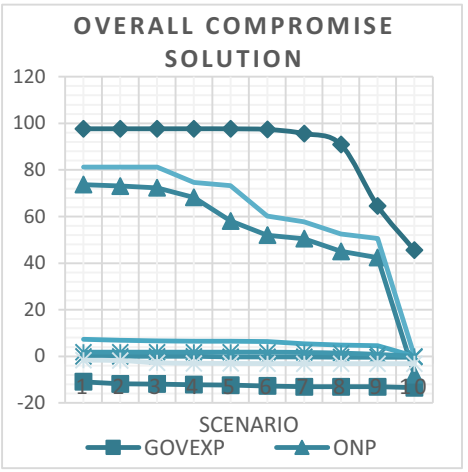


Figure 2. Results for each objective function.

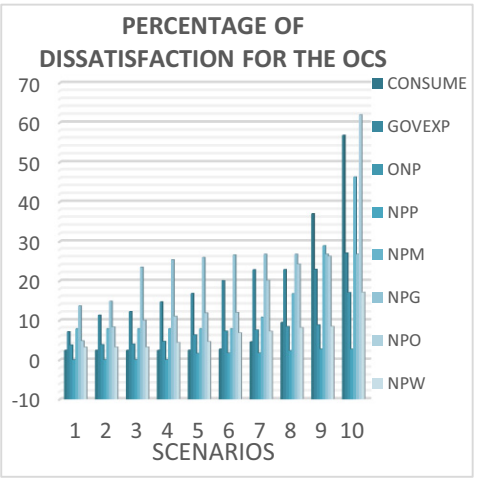


Figure 3. Percentage of dissatisfaction.

4. Conclusions

This paper proposed an optimization approach for the strategic planning of the municipal solid waste management system. The optimization accounts for the involved uncertainty, where three different financial risks (optimistic, mean and pessimistic) through a multi-stakeholder formulation were considered. This way, an optimum and more realistic waste management system that creates synergistic opportunities to benefit the multiple stakeholders can be obtained. It should be noticed that in this work the three financial risks were optimized simultaneously. The number of scenarios taken into account was used to reduce the computation time; additionally, the results for 100 scenarios showed a similar trend. The scenarios were generated from historical data (standard deviation) using the GAMS software through a deterministic algorithm (called a generator).

A case study from Mexico is presented to show the applicability of the proposed approach. The results show that it is possible to obtain attractive solutions for all the involved stakeholders with a low level of financial risk. In addition, the proposed approach allows determining the level of dissatisfaction of the involved stakeholders, and this way the proposed approach can be a useful framework for decision makers.

Finally, implementing the multi-stakeholder optimization approach to the proposed model allows obtaining solutions that look for satisfying all the involved parts. Therefore, the proposed approach promotes the economic and environmental incentives for the MSWM. The proposed approach can be applied to other problems and it represents a good alternative to consider simultaneously several objective functions.

References

C. Dai, Y.P. Li, and G.H. Huang, 2012. An interval parameter chance-constrained dynamic programming approach for capacity planning under uncertainty. Resources, Conservation and Recycling. 62,37-50.

A. Entezaminia, M. Heydari, and D. Rahmani, 2016. A multi-objective model for multi-product multi-site aggregate production planning in a green supply chain: Considering collection and recycling centers. Journal of Manufacturing Systems. 40, 63-75.

F. Habibi, E. Asadi, S.J. Sadjadi, and F. Banrzinpour, 2017. A multi-objective robust optimization model for site-selection and capacity allocation of municipal solid waste facilities: A case study in Tehran. Journal and Cleaner Production. 166, 816-834.

K. Nguyen-Trong, A. Nguyen-Thi-Ngoc, D. Nguyen-Ngoc, and V. Dinh-Thi-Hai, 2017. Optimization of municipal solid waste transportation by integrating GIS analysis, equation-based, and agent-based model. Waste Management. 59, 14-22.

B.N.R. Pin, R.M. Barros, E.E.S Lora, and I.F.S. Dos Santos, 2017. Waste management studies in a Brazilian microregion: GHG emissions balance and LFG energy project economic feasibility analysis. Energy Strategy Reviews. https://doi.org/10.1016/j.esr.2017.11.002, (in press).

J.E. Santibañez-Aguilar, A. Flores-Tlacuahuac, M. Rivera-Toledo, and J.M. Ponce-Ortega, 2017. Dynamic optimization for the planning of a waste management system involving multiple cities. Journal of Cleaner Production. 165, 190-203.

N. Sun, and S. Chungpaibulpatana, 2017. Development of an appropriate model for forecasting municipal solid waste generation in Bangkok. Energy Procedia. 138, 907-912.

Z. Wang, J. Ren, M.E. Goodsite, and G. Xu, 2017. Waste-to-energy, municipal solid waste treatment, and best available technology: Comprehensive evaluation by an interval-valued fuzzy multi-criteria decision making method. Journal of Cleaner Production. 172, 887-899.

X. Zhang, and G. Huang, 2013. Optimization of environmental management strategies through a dynamic stocastic possibilistic multiobjective program.Journal of Hazardous Materials. 246-247, 257-266.

L. Zheng, J. Song, C. Li, Y. Gao, P. Geng, and B. Qu, 2014. Preferential policies promote municipal solid waste (MSW) to energy in China: Current status and prospect. Renewable and Sustainable Energy Reviews. 36, 135-148.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50262-7

A Game Theory Approach to Design and Optimization of Decentralized Supply Chains under Uncertainty

Jiyao Gao, Fengqi You*

Cornell University, Ithaca, New York, 14853, USA

Abstract

Supply chains are normally managed in a decentralized way by multiple stakeholders pursuing distinct objectives. However, most existing supply chain studies rely on centralized models and neglect the uncertain behaviors of stakeholders in the decision-making process. In this work, we propose a novel game theory based stochastic model that integrates two-stage stochastic programming with a single-leader-multiple-follower Stackelberg game scheme. The aim is to address the optimization problem of decentralized supply chains considering multiple stakeholders under uncertainty. The resulting model is formulated as a stochastic mixed-integer bilevel nonlinear program, which can be further reformulated into a tractable single-level stochastic mixed-integer linear program by applying KKT conditions and Glover's linearization method. To illustrate the applicability of proposed modeling framework, a case study of a large-scale shale gas supply chain is presented, which demonstrates the advantages of the proposed modeling framework and efficiency of the solution algorithm.

Keywords: Game theory, supply chain, uncertainty, stochastic, bilevel program.

1. Introduction

The management of supply chains normally involves multiple stakeholders. Each stakeholder may pursue different objectives, thus leading to compromised solutions (Cachon and Netessine, 2004). Nevertheless, most existing studies on the optimal design and operations of supply chains still rely on centralized models (Garcia and You, 2015). Consequently, the optimal solutions of centralized models can be suboptimal or even infeasible in a decentralized supply chain (Gao and You, 2017a). To address this research challenge, multiple game theoretic models are developed to explicitly account for the interest of each stakeholder. For instance, there are models for the optimization of cooperative multi-enterprise supply chains based on the generalized Nash bargaining solution approach (Gjerdrum et al., 2001; Yue and You, 2014a). On the other hand, optimization models integrating game theories of Stackelberg game and Nash-equilibrium are proposed for the design and operations of noncooperative supply chains (Yue and You, 2014b; Yue and You, 2017; Gao and You, 2017b; Gao and You, 2017c). However, these models assume that all the information among different players is deterministic, so that players can always make decisions that benefit them the most based on perfect knowledge. In practice, there are always time delays between the design and operational decisions for each player, and various types of uncertainties, such as price fluctuation, change of productivity, etc., can occur during the operation phase. These uncertainties in turn may significantly affect the rational behaviors of other players (Hjaila et al. 2016). Therefore, it is necessary to develop a holistic game

theoretic model that can systematically optimize the design and operational decisions of different stakeholders with uncertainty behaviors.

In this work, we propose a novel modeling framework that integrates the leader-follower Stackelberg game with two-stage stochastic programming approach into a holistic game theoretic model. Specifically, this modeling framework allows consideration of one leader and multiple followers. Following the sequence of time, decision variables for both the leader and the followers are classified into design decisions that must be made "here-and-now" and operational decisions that are postponed to a "wait-and-see" mode after the realization of uncertainties from both sides of players (Birge and Louveaux, 2011). As a result, both players will interact with each other to determine their optimal design strategies at the first stage. Uncertainties from both the leader and the followers are then realized, and both players go to the second stage to determine their best operational strategy based on their previous design decisions and given uncertainty information. Following the stochastic programming approach, uncertainties are depicted with discrete scenarios with given probabilities. The objectives of the leader and the followers are to maximize their own expected net present values. The resulting problem is formulated as a stochastic mixed-integer bilevel programming (MIBP) problem, where the upper-level problem corresponding to the leader's optimization problem is formulated as a mixed-integer nonlinear programming (MINLP) problem, and the lower-level problems corresponding to the optimization problems of the followers are formulated as linear programs. We further reformulate this MIBP problem into a single-level mixed-integer nonlinear programming (MINLP) problem by replacing the lower-level problems with their equivalent Karush-Kuhn-Tucker (KKT) conditions. To make this MINLP problem more tractable, we apply the Glover's linearization approach to reformulate it into an equivalent MILP problem (Glover, 1975). The applicability of the proposed modeling framework is illustrated by a large-scale case study of shale gas supply chains based on Marcellus Shale. According to the optimization results, both players tend to choose more conservative strategies when considering the uncertainties in the noncooperative supply chain optimization problem.

2. Problem Statement

As mentioned in the previous section, the proposed modeling framework integrates the leader-follower Stackelberg game with two-stage stochastic programming approach. We aim to simultaneously optimize the design and operations decisions of different stakeholders under uncertainty. The Stackelberg game is selected to depict the sequential decision-making process among different players. Specifically, the shale gas producer in the shale gas supply chain is identified as the leader that enjoys decision-making priority. The leader's decisions comprise "here-and-now" design decisions, including drilling schedule at each candidate shale site, instalment of gathering pipelines, and processing contract selection, and operational decisions that are made "wait-and-see", namely production profile of shale gas, amount of raw shale gas transported to each existing processing plant, and water management strategy. Meanwhile, the midstream shale gas processing companies are identified as followers in the supply chain. After the observation of leader's decisions, the follower will react rationally and make corresponding decisions, including design decisions regarding the unit process fee associated with the processing contracts, and operational decisions on the planning of processing and distribution activities.

Following the stochastic programming approach, uncertainties are depicted with discrete scenarios with given probabilities. Thus, both the leader and the followers strive to maximize their own expected net present value (NPV).

3. Model Formulation and Solution Algorithm

The general model formulation is presented as follows.

$$\max\ TEP^{leader} = \mathbb{E}_{\xi}\left[TP^{leader}\left(x^u, y^u, x^l, \xi\right)\right] \tag{1}$$

$$\text{s.t. } A_R(\xi)x^u + A_Z(\xi)y^u + B_R(\xi)x^l < r(\xi) \tag{2}$$

$$x^u \in \mathbb{R}_+^{m_R},\ y^u \in \mathbb{Z}_+^{m_Z}$$

where x^l solves:

$$\max\ TEP_p^{follower} = \mathbb{E}_{\xi}\left[TP_p^{follower}\left(x^l, \xi\right)\right], \quad \forall p \in P \tag{3}$$

$$\text{s.t. } Q_R(\xi)x^u + Q_Z(\xi)y^u + P_R(\xi)x^l < s(\xi) \tag{4}$$

$$x^l \in \mathbb{R}_+^{n_R}$$

where TEP^{leader} denotes the leader's total expected NPV. TP^{leader} denotes the leader's total NPV associated with realization of uncertainty ξ. $TEP_p^{follower}$ denotes the follower p's total expected NPV. $TP_p^{follower}$ denotes the follower p's total NPV associated with realization of uncertainty ξ. x^u and y^u indicate the continuous variables and discrete variables corresponding to the leader's decisions, respectively. x^u involves decisions including the shale gas production rate at each shale site in each time period, the amount of shale gas transported from shale sites to processing plants, capacity of gathering pipelines, amount of freshwater required for each shale site, and amount of wastewater treated by different technology options. y^u involves discrete decisions on the number of wells drilled at certain shale site in each time period, and whether a gathering pipeline should be installed or not. x^l indicates the continuous variables corresponding to follower's decisions, including the unit processing fee of processing contracts, processing schedule of shale gas, and amount of natural gas and NGLs extracted from raw shale gas. As can be observed, the leader's and the follower's decisions variables appear in both the upper level problem and the lower level problem. This indicates that the leader's optimization problem is partially dependent on the follower's decisions, and the follower's optimal strategy will be determined based on the leader's decisions. The final optimum of such a bilevel program follows the solution concept of Nash equilibrium.

As can be seen, the resulting problem is a stochastic MIBP that cannot be tackled directly by general-purpose MINLP solvers. Therefore, we first replace the lower-level optimization problems with their equivalent KKT conditions, so the original bilevel problem is reformulated into an equivalent single level MINLP problem with bilinear terms. To improve the computational efficiency, we further adopt the Glover's linearization approach to reformulate this single level MINLP problem into an equivalent MILP problem (Glover, 1975), which can be solved directly by MILP solver CPLEX 12.6.3.

4. Application to a Decentralized Shale Gas Supply Chain

To illustrate the applicability of the proposed modeling framework and solution approach, we consider a case study of a decentralized shale gas supply chain based on Marcellus Shale. A shale gas producer is considered as the leader, who oversees five potential shale sites. Each shale site allows for drilling of up to four to eight shale wells. The wastewater can be handled by three different water management options, including underground injection into Class-II disposal wells, centralized wastewater treatment (CWT), and onsite treatment and recycling with multi-stage flash, multi-effect distillation, or reverse osmosis technologies. The produced raw shale gas can be transported to three shale gas processing plants. Alternatively, the raw shale gas can be sold directly to the market. Each processing plant represents a follower, who will charge the producer for their processing service based on the processing contracts. Specifically, three types of fee-based processing contracts are considered, corresponding to 50%, 75%, and 100% of total processing capacity in a shale gas processing plant. A 10-year planning horizon is considered, which is divided into 10 time periods. Detailed input data are taken from Gao and You (2015). We consider two types of uncertainties in this shale gas supply chain, namely the uncertain productivity of a shale well at a certain time period and the uncertain processing capacity that can be provided by the processing plants. We assume normal distributions for both uncertainties following existing studies (You et al. 2009). By applying sampling approximation approach, the number of scenarios is finalized as 200 scenarios with 98% confidence interval (Shapiro and Homem-de-Mello, 1998). After the reformulation steps mentioned above, the resulting single level MILP problem has 341 integer variables, 115,016 continuous variables, and 65,557 constraints. All the models and solution procedures are coded in GAMS 24.7.3 on a PC with an Intel® Core™ i7-6700 CPU and 32GB RAM.

In Figure 1, we summarize the leader's total NPV and the followers' NPV in different scenarios to demonstrate the impacts of uncertainties. Each point in this figure indicates a specific solution point associated with a scenario. The straight line indicates the total expected NPV throughout the 200 scenarios considered in this case study. Meanwhile, by comparing with the traditional deterministic models, we note that the optimal solution of the proposed model is relatively conservative, but it can effectively hedge against extreme cases with worse economic performances.

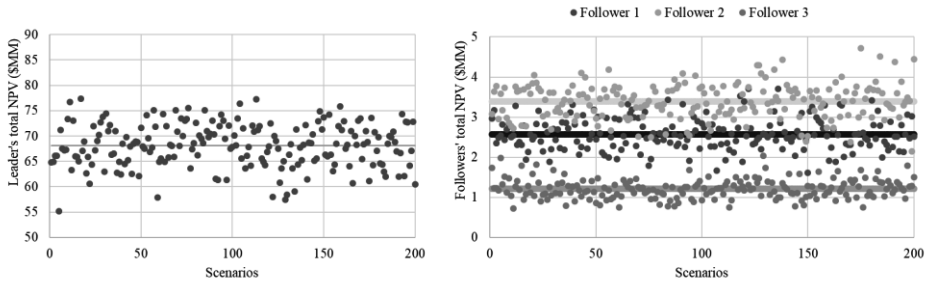


Figure 1. Stakeholders' total NPV in 200 scenarios.

In Figure 2, we present the optimal drilling schedule determined by the leader. The corresponding average shale gas production profile based on the 200 scenarios is included as well. As can be seen, a total of 27 shale wells are drilled within the 10-year planning horizon. Specifically, 10 shale wells are drilled in the beginning at shale sites

1, 2 and 3 to satisfy the initial natural gas demand. Later, based on the forecasted price of natural gas, there will be a price increase in the fourth year. As a result, the leader determines to drill 12 extra shale wells at shale sites 1, 3, and 4 to make more profit. Due to the decreasing feature of productivity of shale wells, 5 more shale wells are drilled in the sixth year to compensate for the production decrease. Despite the production uncertainty at each shale well, this optimal drilling schedule sets the basic tune of shale gas production profile as shown in the right figure.

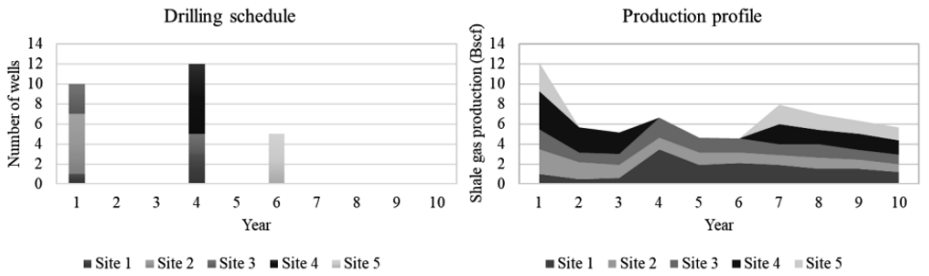


Figure 2. Optimal drilling schedule and shale gas production profiles at shale sites.

The optimal shale gas processing strategies of the leader and the followers are summarized in Figure 3, including the leader's selection of processing contracts and the followers' price setting decisions. As can be observed, in most years, the leader chooses to sign a processing contract with only one processing plant, and 50% of the total processing capacity is considered enough for the raw shale gas output. Since different processing plants have distinct processing capabilities, the leader may switch to other processing plants when production fluctuation is encountered. In the fourth year, with 12 extra shale wells being developed, additional processing capacity is required. Thus, the leader signs contracts with both processing plants 2 and 3 to purchase 75% and 100% of their processing capacities, respectively. Notably, the leader's strategy is also affected by the follower's price setting decisions. We can easily identify that: (a) processing plants tend to offer a lower unit processing price for smaller processing capacity contracts; (b) processing plants have higher chance to be selected by the producer when their unit processing fee is lower than other competitors.

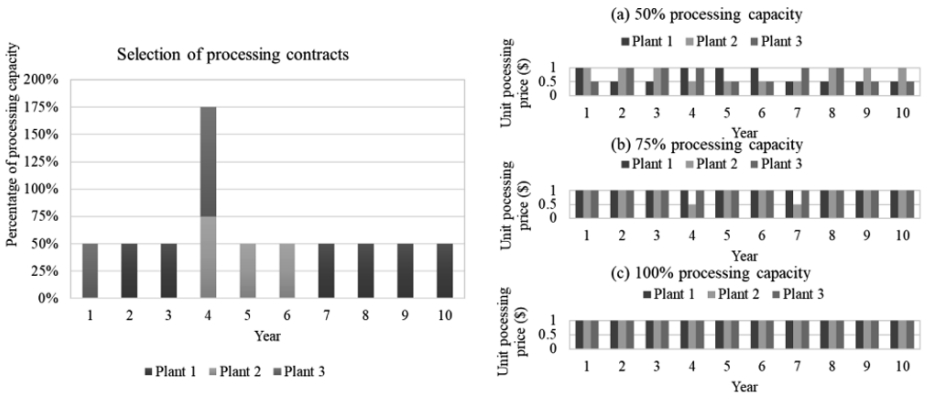


Figure 3. Optimal shale gas processing strategies of the leader and the follower.

5. Conclusions

An optimization framework is proposed that integrates the Stackelberg game with two-stage stochastic programming. This framework enables us to investigate the optimal design and operations of decentralized supply chains involving multiple stakeholders under uncertainty. To illustrate the application, we consider a case study of Marcellus shale gas supply chain. The optimization results show that stakeholders tend to choose more conservative strategies when considering uncertainties in decentralized supply chains.

References

J. Birge, F. Louveaux, 2011, Introduction to stochastic programming, Springer.
G. Cachon, S. Netessine, 2004, Game theory in supply chain analysis, in: D. Simchi-Levi, S.D. Wu, Z.-J. Shen (Eds.), Handbook of quantitative supply chain analysis, Springer US, 13-65.
J. Gao, F. You, 2015a, Shale Gas Supply Chain Design and Operations toward Better Economic and Life Cycle Environmental Performance: MINLP Model and Global Optimization Algorithm. ACS Sustainable Chemistry & Engineering, 3, 1282-1291.
J. Gao, F. You, 2015b, Optimal Design and Operations of Supply Chain Networks for Water Management in Shale Gas Production: MILFP Model and Algorithms for the Water-Energy Nexus. AIChE Journal, 61, 1184-1208.
J. Gao, F. You, 2017a, Design and optimization of shale gas energy systems: Overview, research challenges, and future directions, Computers & Chemical Engineering, 106, 699-718.
J. Gao, F. You, 2017b, Game theory approach to optimal design of shale gas supply chains with consideration of economics and life cycle greenhouse gas emissions, AIChE Journal, 63, 2671-2693.
J. Gao, F. You, 2017c, Economic and environmental life cycle optimization of noncooperative supply chains and product systems: Modeling framework, mixed-integer bilevel fractional programming algorithm, and shale gas application, ACS Sustainable Chemistry & Engineering, 5, 3362-3381.
D. Garcia, F. You, 2015, Supply chain design and optimization: Challenges and opportunities, Computers & Chemical Engineering, 81, 153-170.
J. Gjerdrum, N. Shah, L. Papageorgiou, 2001, Transfer prices for multienterprise supply chain optimization, Industrial & Engineering Chemistry Research, 40, 1650-1660.
F. Glover, 1975, Improved linear integer programming formulations of nonlinear integer problems, Management Science, 22, 455-460.
K. Hjaila, J. Laínez-Aguirre, L. Puigjaner, A. Espuña, 2016, Scenario-based dynamic negotiation for the coordination of multi-enterprise supply chains under uncertainty, Computers & Chemical Engineering, 91, 445-470.
A. Shapiro, T. Homem-de-Mello, 1998, A simulation-based approach to two-stage stochastic programming with recourse, Mathematical Programming, 81, 301-325.
F. You, J. Wassick, I. Grossmann, 2009, Risk management for a global supply chain planning under uncertainty: Models and algorithms, AIChE Journal, 55, 931-946.
D. Yue, F. You, 2014a, Fair profit allocation in supply chain optimization with transfer price and revenue sharing: Minlp model and algorithm for cellulosic biofuel supply chains, AIChE Journal, 60, 3211-3229.
D. Yue, F. You, 2014b, Game-theoretic modeling and optimization of multi-echelon supply chain design and operation under stackelberg game and market equilibrium, Computers & Chemical Engineering, 71, 347-361.
D. Yue, F. You, 2017, Stackelberg-game-based modeling and optimization for supply chain design and operations: A mixed integer bilevel programming framework, Computers & Chemical Engineering, 102, 81-95.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50263-9

Optimal planning of a waste management supply chain

Maryam Mohammadi[a], Iiro Harjunkoski[a,b*], Susanna Mikkola[a], Sirkka-Liisa Jämsä-Jounela[a]

[a]*Aalto University, School of Chemical Technology, Process Control and Automation Research Group, P.O. Box 16100 FI-00076 Espoo, Finland*
[b]*ABB AG, Corporate Research, Wallstadter Str. 59, 68526 Ladenburg, Germany*
iiro.harjunkoski@aalto.fi

Abstract

In this paper, a mixed integer linear programming (MILP) model for the optimization of a municipal solid waste (MSW) system in a supply chain (SC) network is presented with the aim of maximizing the total net profit over a specified time horizon. The proposed model covers waste collection at sites, separation at dumps, and processing for waste recycling and waste to energy (WtE) production at plants, as well as selling the end products. A case study is illustrated to demonstrate the feature and validity of the model. MSWs including biomass materials such as paper and food waste, non-biomass combustible materials such as plastic, and non-combustible materials such as glass and metal are managed by a combination of disposal in landfill sites, recycling to produce useful products, and incineration to produce power.

Keywords: municipal solid waste, supply chain, mixed integer linear programming.

1. Introduction

The generation of MSWs is increasing worldwide due to the rapid population and economic growth, and change in life style behaviors and consumption patterns, which leads to the need for proper waste management strategies. Inappropriate management of MSW systems results in serious environmental risks such as human health damage, harm to ecosystems and resource availability, and water, soil and air pollution, as well as negative economic and social effects. Waste management generally contains six hierarchical steps: waste prevention, reduction, reuse, recycle, energy recovery and disposal. Among all the available options, WtE is still one of the most viable option for many countries as an effective waste management solution. WtE technologies can solve the dilemma of energy demand and provide the valuable energy in forms of electricity and heat, reduce the waste volume and increase the landfill space, and diminish greenhouse gas emissions.

1.1. Municipal solid waste supply chain

The MSW management can be considered as a strategic SC problem as it generally consists of waste collection, separation, disposal, processing and transportation. Although much consideration has been devoted to developing the mathematical models for optimization of MSW systems, specifically in production and distribution sectors, most of these models have considered partial integration of the functions in an SC. Only a few studies focused on the optimization of the MSW systems in an integrated SC network. An integrated network involves all enterprises of the entire process of SC

including MSW collection stations, waste distribution centers, waste disposal/treatment facilities, and end product distribution centers, which will result in smooth and cost-effective allocation of waste flows, extracting the maximum value from the produced waste, efficient logistics, as well as providing a cost breakdown across different levels of the SC. More detailed information related to the optimization of the entire SC for MSW management can be found in Nwe et al. (2010), Santibañez-Aguilar et al. (2013) and Zhang et al. (2014). Based on the existing literature, little attention has been paid to the optimization of an MSW system in a multi-level SC considering multiple time periods and products. Thus, there is a need for developing an integrated mathematical model and solution approach to address these shortcomings. The target is to develop a generic MSW model that is granular enough to bring the SC planning closer to operations.

2. Mathematical model formulation

In this paper, an optimal planning of an MSW management system in a multi-period multi-product multi-level SC is presented. The resulting optimization problem is modeled as an MILP. The MSW model covers waste collection at sites, separation at dumps, processing for recycling and incineration at plants, as well as selling the end products. The modeled mass flows include waste transportation from sites to dumps, waste delivery from dumps to processing plants and product delivery from plants to the markets at sites. The proposed model can determine the quantity of waste types distributed from sites to dumps and from dumps to plants, amount of wastes in plants transferred to different technologies, and amount and types of generated products distributed from plants to sites.

2.1. Notation

Indices:

w	Waste type, $w \in \{1, 2, \dots, W\}$
s	Waste collection site, $s \in \{1, 2, \dots, S\}$
d	Dump, $d \in \{1, 2, \dots, D\}$
p	Waste processing plant, $p \in \{1, 2, \dots, P\}$
j	Waste treatment technology, $j \in \{1, 2, \dots, J\}$
i	Product, $i \in \{1, 2, \dots, I\}$
t	Time period, $t \in \{1, 2, \dots, T\}$

Parameters:

A_{wst}^{prod}	Amount of waste w produced in site s in period t
DL_{is}^{low}	Lower limit for demand of product i in site s
DL_{is}^{up}	Upper limit for demand of product i in site s
SL_{wd}^{d}	Storage limit for waste w in dump d
SL_{wp}^{p}	Storage limit for waste w in plant p
TL_{pj}^{low}	Lower capacity limit for technology j to be used in plant p
TL_{pj}^{up}	Upper capacity limit for technology j to be used in plant p
E_i	Price of product i
CD_d	Disposal cost in dump d
CS_w	Separation cost of waste w
CO_{wpj}	Operation cost of waste w sent to technology j to produce a product in plant p
CT_w^{sd}	Transportation cost of waste w from sites to dumps
CT_w^{dp}	Transportation cost of waste w from dumps to plants

CT_i^{ps} Transportation cost of product *i* from plants to sites
CF_{wpj} Fixed cost to purchase technology *j* for processing waste *w* in plant *p*
CV_{wpj} Variable cost associated with the capital of technology *j* to process waste *w* in plant *p*
dis_{sd}^1 Distance from site *s* to dump *d*
dis_{dp}^2 Distance from dump *d* to plant *p*
dis_{ps}^3 Distance from plant *p* to site *s*
α_{wd} Separation factor for waste *w* in dump *d*
β_{wij} Conversion factor of waste *w* to produce product *i* by technology *j*
ε Annualization factor for the fixed cost

Decision variables:
q_{wsdt}^{dist} Quantity of waste *w* distributed from site *s* to dump *d* in period *t*
q_{wdt}^{in} Quantity of waste *w* inlet to dump *d* in period *t*
q_{wdt}^{sep} Quantity of waste *w* separated in dump *d* in period *t*
q_{wdt}^{stor} Quantity of waste *w* stored in dump *d* in period *t*
q_{wdt}^{out} Quantity of waste *w* outlet from dump *d* in period *t*
q_{wdt}^{land} Quantity of waste *w* landfilled in dump *d* in period *t*
a_{wdpt}^{dist} Amount of waste *w* distributed from dump *d* to plant *p* in period *t*
a_{wpt}^{in} Amount of waste *w* inlet to plant *p* in period *t*
a_{wpt}^{stor} Amount of waste *w* stored in plant *p* in period *t*
a_{wpjt}^{tech} Amount of waste *w* in plant *p* transferred to technology *j* in period *t*
n_{ipt}^{p} Number of product *i* produced in plant *p* in period *t*
n_{ipst}^{dist} Number of product *i* distributed from plant *p* to site *s* in period *t*
n_{ist}^{s} Total number of product *i* transferred to site *s* in period *t*
y_{pj} Binary variable = 1 if plant *p* uses technology *j* (else = 0)

2.2. Objective function

The objective function of the proposed model is shown in Eq. (1), where the revenue from sales should be maximized from which the total cost of the entire SC including disposal, separation, operation, transportation from site to dumps, from dumps to plants and from plants to sites, and capital costs is deducted.

$$\begin{aligned} Max\ f = & \sum_i \sum_s \sum_t n_{ist}^s E_i - \left(\sum_w \sum_d \sum_t q_{wdt}^{land} CD_d + \sum_w \sum_d \sum_t q_{wdt}^{in} CS_w \right. \\ & + \sum_w \sum_p \sum_j \sum_t a_{wpjt}^{tech} CO_{wpj} + \sum_w \sum_s \sum_d \sum_t q_{wsdt}^{dist} CT_w^{sd} dis_{sd}^1 \\ & + \sum_w \sum_d \sum_p \sum_t a_{wdpt}^{dist} CT_w^{dp} dis_{dp}^2 + \sum_i \sum_p \sum_s \sum_t n_{ipst}^{dist} CT_i^{ps} dis_{ps}^3 \\ & \left. + \varepsilon \sum_w \sum_p \sum_j y_{pj} CF_{wpj} + \sum_w \sum_p \sum_j \sum_t a_{wpjt}^{tech} CV_{wpj} \right) \end{aligned} \tag{1}$$

2.3. Constraints

i. Waste collection

Equation (2) expresses that each site generates a given amount of waste, which is distributed to different dumps during each period. The amount of waste inlet to each dump per period is shown in Eq. (3).

$$A_{wst}^{prod} = \sum_{d} q_{wsdt}^{dist} \quad \forall w, s, t \tag{2}$$

$$q_{wdt}^{in} = \sum_{s} q_{wsdt}^{dist} \quad \forall w, d, t \tag{3}$$

ii. Waste processing at dumps

The amount of waste being separated from the inlet in each dump during each period is calculated in Eq. (4). Equation (5) shows the amount of waste stored at each dump per period and Eq. (6) indicates the amount of waste landfilled in dumps in each period. The storage limit for wastes in dumps is defined by Eq. (7).

$$q_{wdt}^{sep} \leq \alpha_{wd} q_{wdt}^{in} \quad \forall w, d, t \tag{4}$$

$$q_{wdt}^{stor} = q_{wdt-1}^{stor} + q_{wdt}^{sep} - q_{wdt}^{out} \quad \forall w, d, t \tag{5}$$

$$q_{wdt}^{land} = q_{wdt}^{in} - q_{wdt}^{sep} \quad \forall w, d, t \tag{6}$$

$$q_{wdt}^{stor} \leq SL_{wd}^{d} \quad \forall w, d, t \tag{7}$$

iii. Waste processing at plants

Equation (8) defines the amount of wastes distributed from dumps to plants. The amount of wastes inlet to plants during each period is indicated by Eq. (9) and Eq. (10) computes the amount of waste stored in each plant per period, which is limited by Eq. (11).

$$q_{wdt}^{out} = \sum_{p} a_{wdpt}^{dist} \quad \forall w, d, t \tag{8}$$

$$a_{wpt}^{in} = \sum_{d} a_{wdpt}^{dist} \quad \forall w, p, t \tag{9}$$

$$a_{wpt}^{stor} = a_{wpt-1}^{stor} + a_{wpt}^{in} - \sum_{j} a_{wpjt}^{tech} \quad \forall w, p, t \tag{10}$$

$$a_{wpt}^{stor} \leq SL_{wp}^{p} \quad \forall w, p, t \tag{11}$$

The amounts of waste distributed to processing technologies in each plant are limited by Eq. (12) and in Eq. (13) the amounts of products produced by different technologies at each plant per period are computed.

$$TL_{pj}^{low} y_{pj} \leq \sum_{w} a_{wpjt}^{tech} \leq TL_{pj}^{up} y_{pj} \quad \forall p, j, t \tag{12}$$

$$n_{ipt}^{p} = \sum_{w} \sum_{j} \beta_{wij} a_{wpjt}^{tech} \quad \forall i, p, t \tag{13}$$

iv. Products distribution

Equation (14) presents the number of end products obtained within each plant distributed to different sites in each period. Equation (15) shows the total amount of per product

shipped to each site per period, which is limited by the demand of product as shown in Eq. (16).

$$n_{ipt}^{p} = \sum_{s} n_{ipst}^{dist} \quad \forall i, p, t \tag{14}$$

$$n_{ist}^{s} = \sum_{p} n_{ipst}^{dist} \quad \forall i, s, t \tag{15}$$

$$DL_{is}^{low} \leq n_{ist}^{s} \leq DL_{is}^{up} \quad \forall i, s, t \tag{16}$$

3. Results and discussions

The problem features considered in this paper are taken from a real case study conducted in Mexico (Santibañez-Aguilar et al., 2015). However, their proposed model is modified by considering the intermediate storages and multiple time periods, and data is revised to fit the model requirements presented in this paper. In order to validate our model, the input data from the above study was used. For the variables which were missing in their study, the value of zero was assigned. After running the proposed model, the same result for the objective function was obtained as acquired by the previous model.

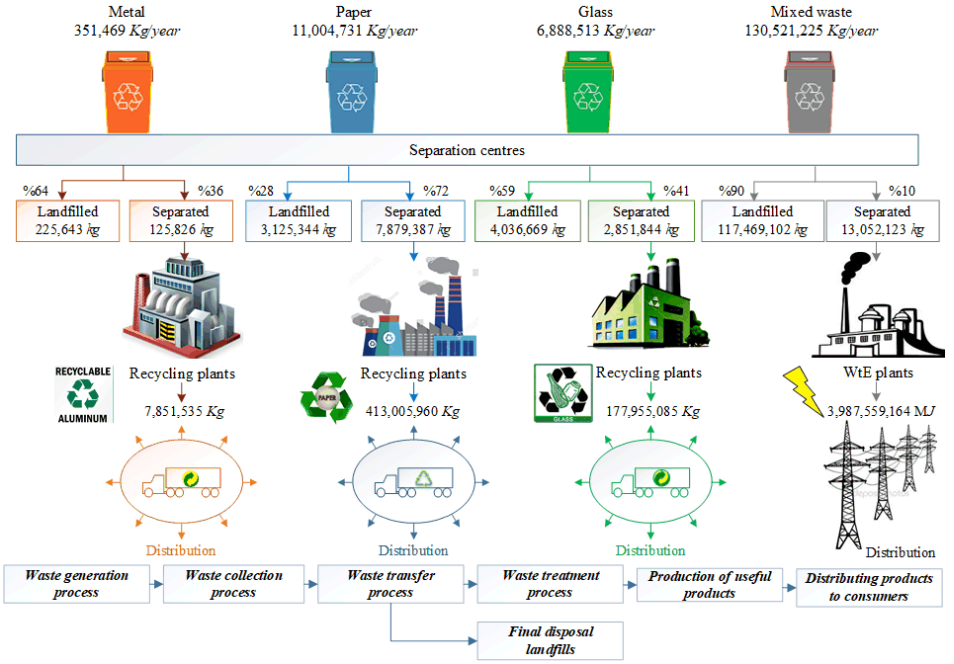


Figure 1. Schematic of considered municipal solid waste management supply chain network

Based on the collected data, it was observed that the MSW generation varies over time. During summer months, due to the change in consumption habits caused by the hot weather and higher consumption of drinks and fresh foods and vegetables, higher amount of glass wastes is produced, whereby less kitchen waste is generated. In winter, generally more solid wastes such as metal, paper and mixed waste are generated. The considered MSW system consists of three sites, two dumps and three plants. The problem is split into weekly planning periods with a total time horizon of one year. During one year, about

149 kt MSWs of metal (0.24 %), paper (7.40 %), glass (4.63 %) and mixed waste (87.73 %) are produced. The technologies used to process the waste include recycling technologies for metal, paper and glass, and pyrolysis arc gasification technology to generate electricity. The produced products are aluminum, paper, glass, and electricity.

The MILP model is formulated and solved using GAMS/CPLEX (v.24.9.1) on a laptop with an Intel Core i5/2.40 GHz and 8 GB of RAM. The model includes 9,784 constraints, 11,864 continuous variables and 12 binary variables. It was solved in 31.52 seconds. The results show that the model produces an annual net profit of -88.68 M€. The total cost is 402.81 M€, mainly consisting of capital cost (76.29 %), operation cost (14.63 %) and separation cost (8.68 %). Disposal and transportation costs construct only a small portion of the total cost. The total annual revenue obtained from selling the recyclable waste and produced energy is 314.13 M€. It is worth mentioning that consumption of the total available waste is not plausible since some wastes cannot be recycled and reused, and each waste has a certain percentage of separation. Figure 1 presents the separated and landfilled amounts of each type of waste and total amount of produced products during a year. It is indicated that mixed waste was the most valuable waste category in terms of production of useful product (electricity).

4. Conclusions

This paper presented an MILP model to analyze the profitability and SC network design of the MSW system. The structure of the proposed model is multi-level, multi-facility, multi-product and multi-period. A case study was selected to test the effectiveness and efficiency of the proposed model. The objective of the model is to find the optimal amount of waste distributed from waste production sites to dumps, from dumps to processing plants, as well as products transported from plants to consumers, so that the demand can be satisfied with the lowest cost over a given planning horizon. Furthermore, selling the recycled products and energy at sites was considered. The presented MILP model was solved efficiently in a reasonable computational time.

References

E. S. Nwe, A. Adhitya, I. Halim, & R. Srinivasan, 2010, Green Supply Chain Design and Operation by Integrating LCA and Dynamic Simulation. Computer Aided Chemical Engineering, 28, 109-114

J. E. Santibañez-Aguilar, J. M. Ponce-Ortega, J. B. González-Campos, M. Serna-González, & M. M. El-Halwagi, 2013, Optimal Planning for the Sustainable Utilization of Municipal Solid Waste, Waste Management, 33, 2607-2622

J. E. Santibañez-Aguilar, J. Martinez-Gomez, J. M. Ponce-Ortega, F. Nápoles-Rivera, M. Serna-González, J. B. González-Campos, & M. M. El-Halwagi, 2015, Optimal Planning for the Reuse of Municipal Solid Waste Considering Economic, Environmental, and Safety Objectives, AIChE Journal, 61, 6, 1881-1899

Y. Zhang, G. H. Huang, & L. He, 2014, A Multi-Echelon Supply Chain Model for Municipal Solid Waste Management System, Waste Management, 34, 553-561

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50264-0

An Improved Approach to Scheduling Gasoline Blending and Order Delivery Operations

Nur Hussain[a], Jie Li[a*], Li Sun[a], Xin Xiao[b], Cuiwen Cao[c]

[a]*School of Chemical Engineering and Analytical Science, The University of Manchester, Manchester, M139PL, UK*
[b]*Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, China*
[c]*Key Laboratory of Advanced Control and Optimization for Chemical Processes, East China University of Science and Technology, Shanghai, 200237, China*
jie.li-2@manchester.ac.uk

Abstract

Scheduling gasoline blending and order delivery operations is an important routine task in an oil refinery since gasoline can account for as much as 60-70% of total revenue. In this work we improve the mixed-integer nonlinear programming (MINLP) model of Li et al. (2016) to ensure optimality and incorporate nonlinear property correlations. To solve such nonconvex MINLP model to ε- global optimality, a global optimization method is proposed. It is shown that our improved model and proposed method can solve industrial examples to 1%- global optimality and generate the same or better solutions with less CPU time than those from Cerda et al. (2016). Using nonlinear property correlations could lead to more accurate prediction than linear correlations.

Keywords: Scheduling, gasoline, blending, global optimization, nonlinear prediction

1. Introduction

Scheduling gasoline blending and order delivery operations is an important routine task in an oil refinery since gasoline can account for as much as 60-70% of total revenue (Li and Karimi, 2011). Optimal scheduling can avoid ship demurrage, minimize quality give-aways, reduce transitions, etc. However, this scheduling problem is a nonconvex mixed-integer nonlinear programming (MINLP) problem. Li and Karimi (2011) and Castillo-Castillo and Mahalec (2016) developed MINLP models for this problem, incorporating many realistic features such as non-identical parallel blenders, minimum run length and amount, constant blending rates, multipurpose product tanks, piecewise constant profiles of blend component qualities and feed rates. Li et al., (2016) improved the MILP relaxation of Li and Karimi (2011) and proposed a global optimization method. However, their model may lead to suboptimality. Although Cerda et al. (2016) improved efficiency, their model results in infeasibility or suboptimality in some cases.

In this work we improve the nonconvex MINLP model of Li et al. (2016) to ensure optimality and incorporate nonlinear property correlations instead of property indices. The nonlinearities come from not only modelling constant blending rate in a run, but also the property correlations. To solve such model to ε- global optimality, a global optimization method is proposed, which first allows blending rates to vary and removes nonlinear property correlations, then uses adjustment procedure, solves a feasibility problem with nonlinear property correlations and an iterative MILP and NLP procedure, and finally employs deterministic global optimization solver ANTIGONE (Misener and

Floudas, 2014) to find ε- global solutions. It is shown that our improved model and proposed method solve all examples to 1%- global optimality and generate the same or better solutions with less efforts than those from Cerda et al. (2016). Using nonlinear property correlations could lead to more accurate prediction than linear correlations.

2. Problem statement

Figure 1 illustrates a general gasoline blending, storage and order delivery process in a typical oil refinery. It involves I blend components in exactly I dedicated component tanks (i = 1, 2, …, I) , B blenders (b = 1, 2, …, B), J multipurpose product tanks (j = 1, 2, …, J), and some lifting ports. P (p = 1, 2, …, P) grades of gasoline are produced and O orders (o = 1, 2, …, O) is to be fulfilled in the scheduling horizon H. Given the scheduling horizon and data on components, blenders, component and product tanks, orders and economics, the scheduling problem is to determine (i) blending schedule; (ii) charging schedule for blenders to product tanks; and (iii) order delivery schedule. The objective is to minimize total operating cost including component, transition, demurrage, and delivery costs. More details can be referred to Li and Karimi (2011).

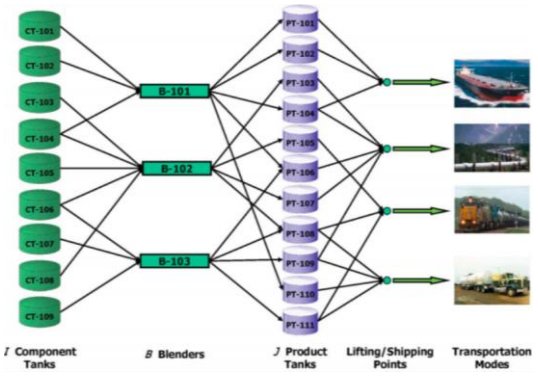


Figure 1 Schematic of a general gasoline blending, storage and order delivery in an oil refinery

3. Motivating Example

Consider an example from slight modifications of Example 1 in Li and Karimi (2016). The blending rate ranges are modified to 10-15 kbbl/h and the capacities of product tanks are reduced to 55 kbbl except for PT-101 whose capacity is reduced to 95 kbbl. The initial product in PT-104 changes to P1. It is found that no feasible solution is generated from the model of Cerda et al. (2016), whilst a feasible solution of 5470.422 K$ is obtained from the model of Li et al. (2016). In addition, a better solution of 5139.929 K$ for Example 1 is obtained from Cerda et al. (2016). All these motivate to improve the model of Li et al. (2016) to ensure optimality for any cases.

4. Improved MINLP formulation

In this section, only improvements for Li et al. (2016) is presented. We first remove eqs. 11a-b in Li et al. (2016) which do not allow a blender to process the same product in two consecutive blend runs. Then, 0-1 variables $xe(b,n)$ denoting the end of a blend run are imposed as binary. To account for transition cost correctly when empty event points exist in two consecutive blend runs, 0-1 auxiliary continuous variables $xx(b,p,n)$ and $xxe(b,n)$ are employed, which have the following relationship with $x(b,p,n)$.

$$xx(b,p,n) \geq x(b,p,n) \qquad \forall n, (b,p) \in \mathbf{BP} \qquad (1)$$

$$xx(b,p,n) \geq x(b,p,n-1) - \sum_{p':(b,p)\in\mathbf{BP}} x(b,p',n) \qquad \forall n, (b,p) \in \mathbf{BP} \qquad (2)$$

$$\sum_{p:(b,p)\in\mathbf{BP}} xx(b,p,n) = 1 \qquad \forall n, (b,p) \in \mathbf{BP} \qquad (3)$$

$$xxe(b,n-1) \geq xx(b,p,n-1) - xx(b,p,n) \qquad \forall (b,p) \in \mathbf{BP}, n \qquad (4)$$

$$xxe(b,n-1) \geq xx(b,p,n) - xx(b,p,n-1) \qquad \forall (b,p) \in \mathbf{BP}, n \qquad (5)$$

$$xxe(b,n-1) + xx(b,p,n-1) + xx(b,p,n) \leq 2 \qquad \forall (b,p) \in \mathbf{BP}, n \qquad (6)$$

$$\sum_{b=1}^{B}\sum_{n=1}^{N-1} xxe(b,n) \leq NP - NB \qquad (7)$$

The nonlinear prediction correlations for RON and RVP are given below,

$$\theta(b,p,s,n) = \overline{r}(b,p,n) + \overline{a}_1 \cdot [\overline{r}_s(b,p,n) - \overline{r}(b,p,n)\cdot\overline{s}(b,p,n)] + \overline{a}_2 \cdot [\overline{o}_{sq}(b,p,n) - \overline{o}(b,p,n)^2]$$
$$+\overline{a}_3[\overline{\mathrm{a}}_{\mathrm{sq}}(b,n,p)^2 - 2\overline{\mathrm{a}}_{\mathrm{sq}}(b,n,p)\cdot\overline{a}(b,p,n)^2 + \overline{a}(b,n,p)^4] \quad \forall (b,p) \in \mathbf{BP}, n, s = RON \qquad (8)$$

$$\theta(b,p,s,n)^{1.25} = \sum_i recipe(i,b,p,n)\cdot\theta(i,s)^{1.25} \qquad \forall (b,p) \in \mathbf{BP}, n, s = RVP \qquad (9)$$

$$\theta^L_{PS}(p,s)\cdot x(b,p,n) \leq \theta(b,p,s,n) \leq \theta^U_{PS}(p,s) \quad \forall (b,p) \in \mathbf{BP}, n, s \in \{RON, RVP\} \quad (10a,b)$$

More details about eqs. 8-9 can be referred to Parkash (2003). The variables *recipe*(*i*,*b*,*p*,*n*) in eqs. 8-9 denote the recipe of a product, which is calculated as follows,

$$q\mathrm{l}(i,b,p,n) = Q_P(b,p,n)\cdot recipe(i,b,p,n) \qquad \forall (b,p) \in \mathbf{BP}, n, i \qquad (11)$$

To model constant blending rate in a run, the following constraints are enforced.

$$Q(b,n) = BR(b,n)\cdot BL(b,n) \qquad \forall b, n \qquad (12)$$

$$BR(b,n) - F^U_B(b)\cdot xe(b,n) \leq BR(b,n+1) \leq BR(b,n) + F^U_B(b)\cdot xe(b,n) \quad \forall b, n < N \quad (13\text{-}14)$$

Note that eqs. 12-14 are the same as eqs. 22-24 in Li et al. (2016). All other constraints are also the same as those in Li et al. (2016). The objective function is given as follows,

$$TC = \sum_{i=1}^{I}\sum_{b=1}^{B}\sum_{n=1}^{N} c(i)\cdot q(i,b,n) + \sum_{b=1}^{B}\sum_{n=1}^{N-1} CB(b)\cdot xxe(b,n) + \sum_{b=1}^{B} CB(b)\cdot xxe(b,0) + \sum_{o=1}^{O} DM(o)\cdot d(o)$$
$$+\sum_{j=1}^{J}\sum_{n=1}^{N-1} CT(j)\cdot ue(j,n) + \sum_{j=1}^{J} CT(j)\cdot ue(j,0) + \sum_{j=1}^{J}\sum_{o:(j,o)\in\mathbf{JO}}^{O}\sum_{n=1}^{N} COT(o)\cdot z(j,o,n) \qquad (15)$$

5. Global optimization method

Figure 2 shows the proposed global optimization method including solving an MILP underestimator, adjustment procedure, a feasibility model, an iterative MIP and NLP procedure, and global optimisation solver ANTIGONE to ensure ε- global optimality.

MILP underestimator

Nonlinear constraints (i.e., eqs. 12-14 and 8-11) are dropped from the original MINLP

model to allow varying blending rates and eqs. 16a-b are added to ensure blending rates vary within limits, which provides a lower bound (LB) for the original MINLP model.

$$F_B^L(b)\cdot BL(b,n) \le Q(b,n) \le F_B^U(b)\cdot BL(b,n) \qquad \forall b,\, n \qquad (16a,b)$$

Schedule adjustment procedure

The constant blending rate in a run is computed using the optimal solution from the MILP underestimator. It is then fixed in the original MINLP model as parameters with other variables varying to build an MILP model without eqs. 8-11, ensuring constant blending rate in a run but without satisfying eqs. 8-10 for RON and RVP. The values of 0-1 discrete variables from this MILP model are extracted to fix 0-1 discrete variables in the original MINLP model and build a NLP model with eqs. 8-11. The optimal solution from this NLP provides an upper bound (UB) for the original MINLP model.

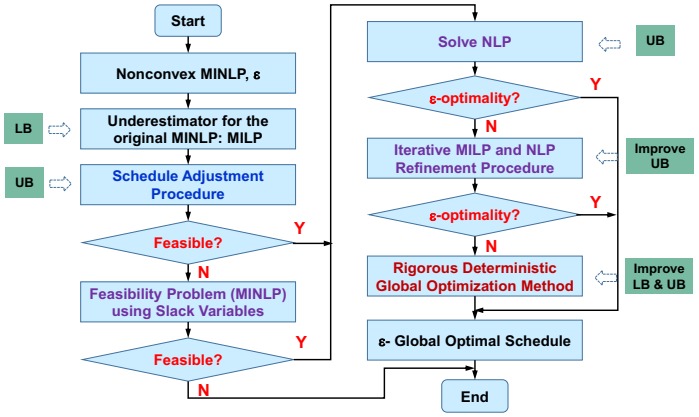


Figure 2 Schematic diagram of the proposed global optimization method

Feasibility problem

Two positive slack variables $s1(b,p,s,n)$ and $s2(b,p,s,n)$ are introduced to eqs. 8-9 in the original MINLP model.

$$\theta(b,p,s,n) = \overline{r}(b,p,n) + \overline{a}_1\cdot[\overline{r}_s(b,p,n) - \overline{r}(b,p,n)\cdot\overline{s}(b,p,n)] + \overline{a}_2\cdot[\overline{o}_{sq}(b,p,n) - \overline{o}(b,p,n)^2]$$
$$+\overline{a}_3[\overline{a}_{sq}(b,n,p)^2 - 2\overline{a}_{sq}(b,n,p)\cdot\overline{a}(b,p,n)^2 + \overline{a}(b,n,p)^4] + s1(b,p,s,n) - s2(b,p,s,n)$$
$$\forall (b,p) \in \mathbf{BP},\, n,\, s = RON \qquad (8')$$

$$\theta(b,p,s,n)^{1.25} = \sum_i recipe(i,b,p,n)\cdot\theta(i,s)^{1.25} + s1(b,p,s,n) - s2(b,p,s,n)$$
$$\forall (b,p) \in \mathbf{BP},\, n,\, s = RVP \qquad (9')$$

The objective is to minimize slack variables in eq. 15′. This feasibility model which is solved using ANTIGONE can ensure if the original MINLP model is feasible.

$$z = \sum_b \sum_{p:(b,p)\in\mathbf{BP}} \sum_{s\in\{RON,RVP\}} \sum_n s1(b,p,s,n) + s2(b,p,s,n) \qquad (15')$$

Note that UB is a feasible solution for the original MINLP model. An iterative MILP and NLP procedure is used to improve the quality of the UB, which is similar to that of Li et al (2016). The difference is that blending recipes as well as blending rates are fixed in the original MINLP model to construct an MILP model in the iterative procedure.

6. Computational studies

We revisit the motivating example with our improved model and global optimization method. The optimal solution of 5470.422 K$ is obtained, whilst it is infeasible from Cerda et al. (2016). We also solve 5 examples with single and multiple periods from Li and Karimi (2016) using CPLEX 12.6.3/GAMS 24.7.4 on a Desktop (Intel® XeonR CPU 3.2 GHz, 8 GB memory) running Windows 7. The results are given in Tables 1-2. From Table 1, we obtain the same optimal solution of 5139.929 K$ as Cerda et al. (2016) for Example 1. Although we obtain the same optimal solution as that from Cerda et al. (2016) for Example 10, the CPU time is reduced significantly. For others, we obtain better solutions. Similar conclusions can be made for multi-period from Table 2.

Table 1 Comparative results of ours with that of Cerda et al. (2016): single-period

Example	Model	Binary Variables	Constraints	Cost (K$)	CPU Time (s)	Gap (%)
1	Ours	175	2057	5139.929	0.748	0.0
	Cerda	285	1983	5139.929	0.297	0.0
5	Ours	306	3457	4556.673	0.452	0.0
	Cerda	628	4438	4558.173	0.484	0.0
10	Ours	714	8590	11286.102	84.725	0.0
	Cerda	1490	11092	11286.103	2521.694	0.0
11	Ours	1480	17319	13248.581	551.589	0.1
	Cerda	2382	17300	13249.788	3601.845	0.0
14	Ours	2141	26800	17930.669	1000.185	0.2
	Cerda	4016	28949	Infeasible	-	-

Table 2 Comparative results of ours with that of Cerda et al. (2016): multi-period

Example	Model	Binary Variables	Constraints	Cost (K$)	CPU Time (s)	Gap (%)
1	Ours	175	2057	5139.929	1.341	0.0
	Cerda	278	1895	5139.929	0.468	0.0
5	Ours	427	4705	4556.673	0.608	0.0
	Cerda	630	4446	4556.673	1.545	0.0
10	Ours	714	8590	11309.542	957.940	0.0
	Cerda	1500	11132	11309.542	300.583	0.0
11	Ours	1480	17074	13272.961	850.814	0.2
	Cerda	2396	17356	13284.396	3600.081	0.0
14	Ours	1485	18948	18102.555	9276.924	0.2
	Cerda	4043	29056	Infeasible	-	-

We solve Example 14 to 1%-global optimality using nonlinear correlations for RON and RVP instead of property indices. The optimal solution of 18164.576 K$ is generated within 6178 s. The quality values from nonlinear and linear models are given in Table 3. The RON and RVP values from the linear model is off-specification in some cases, whilst they are on-specification from the nonlinear correlations.

7. Conclusions

In this work we improved the model of Li et al. (2016) for scheduling gasoline blending and order delivery operations. Nonlinear property correlations were included. A global optimization method is proposed to solve the nonconvex MINLP model to ε- global

optimality. It is shown that our improved model and proposed global optimization method solved these examples to 1%- global optimality and generated the same or better solution with less efforts than those from Cerda et al. (2016). Using nonlinear property correlations could lead to more accurate prediction than linear correlations.

Table 3 The quality values obtained using both the linear and nonlinear methods for Example 14

Event point	Product	RON			RVP		
		Linear	Nonlinear	Min	Linear	Nonlinear	Min
1	P1	95.31	95.00	95.00	0.392	0.406	0.20
	P2	96.27	96.00	96.00	0.348	0.357	0.20
	P4	89.61	90.00	90.00	0.561	0.570	0.20
2	P3	92.66	94.00	94.00	0.304	0.347	0.20
	P4	90.04	90.00	90.00	0.520	0.532	0.20
	P5	97.78	98.00	98.00	0.350	0.357	0.20
3	P1	91.62	95.00	95.00	0.152	0.200	0.20
	P3	94.36	94.00	94.00	0.424	0.442	0.20
	P4	90.42	90.00	90.00	0.485	0.505	0.20
4	P1	95.31	95.00	95.00	0.391	0.405	0.20
	P4	90.37	90.00	90.00	0.655	0.702	0.20
5	P1	95.26	95.00	95.00	0.322	0.327	0.20
6	P1	94.75	95.00	95.00	0.160	0.200	0.20

Acknowledgements

Cuiwen gratefully acknowledge financial support from the National Nature Science Foundation of China (under Grant 61673175).

References

P. A. Castillo-Castillo, V. Mahalec, 2016, Improved Continuous-Time Model for Gasoline Blend Scheduling. Computers and Chemical Engineering, 84, 627-646.

J. Cerda, P. C. Pautasso, D. C. Cafaro, 2016, A Cost-Effective Model for the Gasoline Blend Optimization Problem, AIChE Journal, 62, 3002-3019.

J. Li, I. A. Karimi, 2011, Scheduling Gasoline Blending Operations from Recipe Determination to Shipping using Unit Slots, Ind. Eng. Chem. Res., 50, 9156-9174.

J. Li, X. Xiao, C. A. Floudas, 2016, Integrated Gasoline Blending and Order Delivery Operations: Part I. Short-Term Scheduling and Global Optimization for Single and Multi-Period Operations. AIChE Journal, 62, 2043-2070.

J. Li, I. A. Karimi, 2016, Corrections to “Scheduling Gasoline Blending Operations from Recipe Determination to Shipping Using Unit Slots”, , Ind. Eng. Chem. Res., 55, 7231-7237

R. Misener, C. A. Floudas, 2014, ANTIGONE: Algorithms for coNTinuous / Integer Global Optimization of Nonlinear Equations, Journal of Global Optimization, volume 59, 503-526

S. Parkash, 2003, Refining Processes Handbook, Gulf Professional Publishing.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50265-2

A Methodology for Development of a Pedagogical Simulation Tool used in Fermentation Applications

Simoneta Caño de las Heras[a,c], Seyed Soheil Mansouri[a], Stefano Cignitti[a], Hinrich Uellendahl[c], Charlotte Lærke Weitze[b], Krist V. Gernaey[a], Helle Rootzén[b], Ulrich Krühne[a]

[a] *PROSYS, Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads, Building 229, 2800 Kgs. Lyngby, Denmark*
[b] *LearnT, Technical University of Denmark, Richard Petersens Plads, Building 324, 2800 Kgs. Lyngby, Denmark*
[c] *Department of Chemistry and Bioscience, Section for Sustainable Biotechnology, Aalborg University, A.C. Meyers Vænge 15, 2450 Copenhagen, Denmark*

Abstract

In this work, a holistic methodology was developed and implemented in a computer-aided framework with the purpose of establishing a pedagogical simulation tool for bioprocess applications. The methodology integrates the development of a process simulator with the use of gamification elements to improve the pedagogical experience. Moreover, model development and reuse was incorporated into the framework through creating a generic model template. The application of the methodology was demonstrated by the simulation of a batch fermentation process of *Saccharomyces cerevisiae* cultivated under aerobic conditions with glucose as substrate, thereby, demonstrating mixed metabolism dynamics and considering the inhibition of ethanol consumption due to the presence of glucose with an integrated learning design and gamification elements. The learning experience, done with beta users with different previous levels of knowledge about fermentation processes, confirmed an interest for using gamification for training chemical engineering, biochemical engineering, and biotechnology students.

Keywords: Methodology, Simulator, Fermentation, Gamification, Learning experience.

1. Introduction

Institutions of higher education are facing the challenge of providing quality education to an increased number of students with a budget that has remained constant (Balamuralithara and Woods, 2009). A special area of concern is engineering education, in which students require extensive practical experience to acquire technical competences for their career. Acquiring this experience often comes with complications due to economic expenses and limitations of experimental teaching. One way to tackle these limitations is exploration through creating "what if" scenarios, or the possibility to simulate a variety of fictional environments, and consequently create "intuition" and provides a platform for learning by failure. Today, this is often partially achieved using process simulation tools. However, most of these tools do not target teaching and therefore, they come with educational limitations. For example, there is not a clear connection between visual/graphic features and underlying mathematical models, and this results in the majority of process simulators being experienced by the user as black-

box modules. Interestingly, research has shown that including game-elements into simulations of complex systems can provide an excellent framework for making the theoretical knowledge more approachable (Ebner and Holzinger, 2007).

In this work, the objective is to propose and implement a methodology for the development of pedagogical simulation tools. Furthermore, the application of the methodology is highlighted through a specific case of a fermentation process. The implemented software tool includes game-elements, and a user experience experiment is carried out amongst three beta groups with different educational backgrounds. This is done to measure the learning experience, to quantify the intrinsic motivation and to involve the future users in the design of the project.

2. Computer-aided Framework

The framework for the development of a pedagogical simulation tool including methods and tools and information flow is presented in Figure 1. The framework consists of five hierarchical steps. The framework is generic, and each step may be applied independently, provided the availability of input information, and it is described in the forthcoming sub-sections.

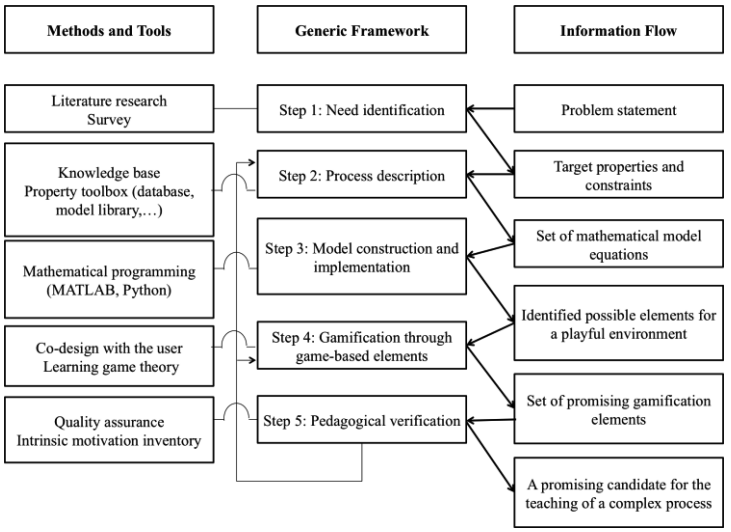


Figure 1. Framework for the development of a pedagogical simulation tool.

2.1. Step 1: Need identification

In this step, the problem statement is defined to identify the need that has to be fulfilled through a pedagogical approach. In a chemical engineering context, this need may be to provide sufficient training associated with any process or unit operation. Thereby, the problem statement is based on either user-need (often quantified through a survey) or a comprehensive literature survey. The objectives to be fulfilled as a result of satisfying the identified need may be qualitative in terms of learning experience or quantitative through a set of properties and constraints. The problem formulation and statement will then be used as an input to make structural decisions to describe a given process.

2.2. Step 2: Process description

In this step, the case study process is mathematically and pedagogically described based on the target properties and constraints that were included in the problem formulation in Step 1. In this step, there is a need for information about pure component properties, reaction rate constants, transfer coefficients and other relevant process parameters that are stored in a property toolbox, supported by information collected from literature or database. Meanwhile, the learning design is done using a didactic frame designed by Hiim and Hipee (1997), pursuing a learning based on experience, understanding, and action. Therefore, by completing this step all information that is needed to describe the process mathematically and pedagogically is obtained.

2.3. Step 3: Model construction and implementation

This step is intended to design, construct, solve and validate a process model. For this, a workflow is developed for the creation and use of a process template (Figure 2).

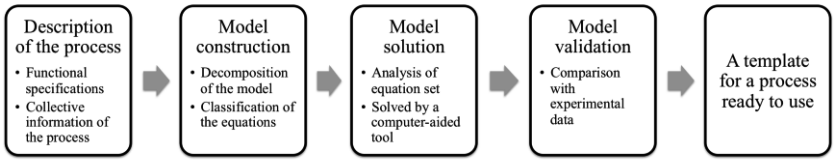


Figure 2. Workflow for the creation and use of a process template.

This workflow is based on the work of Fedorova (2015), for the module modelling due to template creation, and it is based on the idea of the combination of parts of previously developed models into a new set of equations for a specific case. Therefore, templates can be modified generating different versions of the base model, and all those options can be stored.

2.4. Step 4: Gamification through game-based elements

Game-based elements are introduced in this step to engage the users in the learning experience and in order to support their intrinsic motivation (Ebner, and Holzinger, 2007, Jayasinghe and Dharmaratne, 2013). Although gamification is usually developed in parallel with experimentation, it is initially based on the supportive framework of the Smiley Model (Weitze, 2016), which combines learning design and game design elements.

2.5. Step 5: Pedagogical verification

Finally, this module corresponds to model application and validation and is done by a learning experience, in which it is evaluated the learning design and the use of gamification elements.. To do so, a questionnaire for data collection can be distributed before and after the trial of the prototype amongst users, evaluating previous knowledge and the learning experience or by personal interview. This module should provide information that can help improve the simulation tool by modifying the process description (the theoretical content) or the gamification elements used.

3. Case Study: *Saccharomyces cerevisiae* cultivated under aerobic conditions with glucose as substrate in batch fermentation

For the application of the proposed methodology (see Figure 1), a prototype template of the cultivation of *Saccharomyces cerevisiae* under aerobic conditions and glucose as a carbon source in a batch system was used.

Step 1: Need identification: *S. cerevisiae* is one of the most common microorganisms in the biotechnological industry (Ostergaard *et al*., 2000) and therefore its understanding is a basic need in a bioprocess course.

Step 2: Process description: In its cultivation under aerobic conditions, and therefore considering liquid and gas phases actives, *S. cerevisiae* presents a mixed metabolism controlled by the availability of oxygen that involves the oxidation of glucose as substrate and formation of ethanol as a fermentation product (Sonnleitner and Käppeli, 1986). Ethanol consumption is inhibited by the presence of glucose, and by the reduction of glucose. On the other hand, the pedagogical design is based on a constructive learning, with an educational content for the complete description of a fermentation process.

Step 3: Model construction and implementation: – For the specific case of a fermentation, a sub-workflow for the construction of the model was developed and is shown in Figure 3.

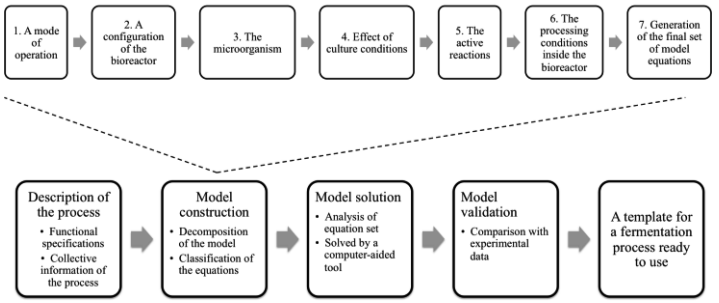


Figure 3. Workflow for the model construction of fermentation.

A set of model equations corresponding to the mathematical equations, shown in Table 1, are implemented as a computational problem and solved within Python while the user only interacts through a visual interface. The parameter values and the constants are based on the work by Fernandes *et al*. (2013).

Table 1. Set of mathematical equations used for the modelling of the fermentation process of *S. cerevisiae*

Definition	Equation
Mass balance for a batch bioreactor	$N_{A0}\frac{dx}{dt} = r_A V$
Volumetric flux of oxygen	$\dot{N_{O_2}} = k_L a\,(C^*_{O_2} - C_{O_2})$
Kinetic model	Sonnleitner and Käppeli (1986)

Step 4: Gamification through game-based elements: Although gamification is usually done in parallel with experimentation, the game elements used in the first step of the prototype were:

1. Goal and action space. The game is developed in the context of finding a job, and at the end of it, the user will be prepared to start working in a company.
2. Choice. Parameters and kinetic equations are easily modified, providing the opportunity of checking "what if" scenarios by the student.
3. Challenge. It is introduced in the way of questions that increase in level of difficulty with a correct answer, although there is also an answer option "No idea" that will teach the concept to the student.
4. Feedback.

Step 5: Pedagogical verification: This was done through evaluating the learning experience at three different pre-knowledge levels about fermentation processes in order to verify the pedagogical value of the prototype and involve the student in an what they want or what would be useful for them, using a paper prototype and personal interviews or questionnaires. The pre-knowledge levels were: novice, elemental knowledge of biotechnology at bachelor level and expert knowledge. In the learning experience with novices, done by personal interviews, an excessively high level of the unknown theory behind the game was perceived, and a higher number of gamification elements and hints were requested. However, in the case of previous elemental knowledge, the information obtained during the learning experience was non-uniform and, in some cases, conflictive as students had problems to answer some questions but considered the content as easy. Furthermore, students generally believed that the tool could be very helpful and considered gamification useful for teaching complex knowledge (see Figure 4). Meanwhile, the expert user-experience, to whom was provided the same questionnaire than the elemental knowledge users, showed initially some resistance to the use of gamification, but in the survey carried out after the learning experience, an interest was shown in its use. One key issue is the development of a more attractive interface and overcoming user reluctance in playing a gamified pedagogical simulation tool.

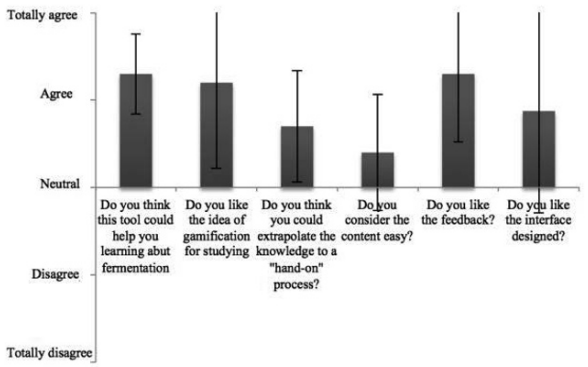


Figure 4. Survey about the usability and enjoyability of the prototype from a class of the second year of the bachelor in Sustainable biotechnology.

4. Conclusions

A computer-aided framework has been developed to support the creation of pedagogical simulation tools integrating the use of gamification elements to engage the learning process. The application of the proposed methodology has been highlighted through a case study of *Saccharomyces cerevisiae* cultivated under aerobic conditions with glucose as substrate in batch fermentation and is implemented in a computer-aided software tool. Furthermore, a learning experience survey, in which users were involved in the co-design, showed a general interest in the use of this tool for the study of fermentation, but also that the design should be adapted to the specific level of pre-knowledge of the user.

Acknowledgements
The project partly received financial support from Innovation Fund Denmark through the BIOPRO2 strategic research center (Grant number 4105-00020B).

References

Balamuralithara, B., & Woods, P. C. (2009). *Computer Applications in Engineering Education*, *17*(1), 108-118.

Ebner, M., & Holzinger, A. (2007). *Computers & education*, *49*(3), 873-890.

Fedorova, M., Gani, R., & Sin, G. (2015). Systematic Methods and Tools for Computer Aided Modelling. Kgs.Lyngby: Technical University of Denmark (DTU).

Fernandes RL, Bodla VK, Carlquist M, Heins AL, Eliasson Lantz A, Sin G, Gernaey KV (2013) Applying mechanistic models in bioprocess development. Adv Biochem Eng Biotechnol. 132:137-66.

Hiim, H., & Hippe, E. (1997). Learning through experience, understanding and action. *Gyldendal Uddannelse*.

Jayasinghe, U., & Dharmaratne, A. (2013). Game based learning vs. gamification from the higher education students' perspective. In *Teaching, Assessment and Learning for Engineering (TALE), 2013 IEEE International Conference on* (pp. 683-688). IEEE.

Ostergaard, S., Olsson, L., & Nielsen, J. (2000). Metabolic engineering of Saccharomyces cerevisiae. *Microbiology and molecular biology reviews*, *64*(1), 34-50.

Sonnleitner B, Käppeli O (1986) Growth of Saccharomyces cerevisiae is controlled by its limited respiratory capacity: formulation and verification of a hypothesis. Biotechnol Bioeng 28:927–937.

Weitze, C. L. (2016). Designing for Learning and Play-The Smiley Model as a Framework. *IxD&A*, *29*, 52-75.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50266-4

Transforming Instruction to Chemical Product Design

Ka M. Ng[a], Warren D. Seider[b*]

[a]*The Hong Kong University of Science and Technology, Department of Chemical and Biological Engineering, Hong Kong, Peoples Republic of China*
[b]*University of Pennsylvania, Department of Chemical and Biomolecular Engineering, Philadelphia, Pennsylvania, 19104-6393, United States*
seider@seas.upenn.edu

Abstract

This paper describes the progress of our efforts to lead the CACHE (Computer Aids for Chemical Engineering Education) Task Force in transforming from chemical process design toward chemical product design. Through CACHE, we are coordinating the development of a library of product-design case studies. Beginning with preliminary product designs created previously over several semesters, we are arranging for faculty experts, knowledgeable in the underlying technology platforms, to work with student groups to enrich the product designs. Over a 3-year period, a collection of approximately 25 case studies is being prepared. This article describes the research envisioned as innovative product designs are created, both regarding applications of new technologies, and product design evolution/evaluation; and in advancing strategies for teaching product design. The anticipated use of these case studies in departments worldwide for design courses taught by similar technology experts, just a few in each department, is discussed. These will align with the expertise of the faculty members, who will help undergraduate student design groups gain the technology skills to carry out specific design projects.

Keywords: product design, innovation, technology platforms, teaching assessment.

1. Introduction

1.1 Background

The creation of product and process designs are integral tasks in the commercialization of any chemical product. Traditional chemical engineering focuses on process design because commodity products tend to be well defined and product design is not required. For these mature products, the companies compete in the marketplace by reducing product cost. This is achieved by improving the production process or by switching to a cheaper feedstock. The latter often necessitates the development of a new process.

With the emergence of traditional chemical engineering in developing economies such as Saudi Arabia, China, South Korea, Vietnam, Thailand, and so on, companies in developed economies can no longer prosper by merely producing mature products. To attain a higher profit margin, the obvious way is to produce innovative consumer products that enhance human comfort, health, and convenience. For these consumer products, product design is crucial. The designer has to select the proper product ingredients and design the process to configure / process these ingredients in such a way that the final product provides the desired product attributes.

Since product design has not received much attention in the chemical engineering community until the late 90s (Stephanopoulos, 2003; Cussler and Wei, 2003; Gani, 2004; Wesselingh et al., 2007; Ng et al., 2007, Wei, 2007; Cussler and Moggridge, 2011; Seider and Widagdo, 2012; Bröckel et al., 2013), there are significantly less teaching materials and fewer case studies for teaching product design than for process design. At present, all chemical engineering curricula finish with a course (or two courses) on chemical process design – with a few curricula offering an elective on chemical product design. Process design (designing the manufacturing processes) builds upon process-oriented engineering-science courses (i.e., thermodynamics, fluid mechanics, heat and mass transfer, separations, and reactor design). Not surprisingly, most design projects, including the capstone design projects, in typical chemical engineering curricula are related to process design, simulation and optimization.

Over decades, process designers have built upon engineering-science fundamentals – with textbooks and process simulators widely available. Few handle both product and process design. Recently, Seider et al. (2017), in *Product and Process Design Principles* (4^{th} Ed., John Wiley), in addition to process design of traditional chemical plants, methods for designing and manufacturing new chemical products for consumers are also comprehensively covered. It offers a natural transition from process to product design and from commodity to consumer products.

Pedagogically, the product design strategies lead naturally to the process design strategies. However, product design often involves complex technology platforms that characterize pastes and creams, soft materials, tailor-made solvents, detergents, dehumidifiers, mosquito repellents, and so on. To carry out the product designs, faculty and students need examples of how to work with specific technology platforms. This leads to a key reason for carrying out this educational research – that is, it seems clear that a coupling is needed that links specific product-design problems with the expertise of specific faculty members. Our faculty collaborators are developing case studies closely aligned to their specific interests/expertise and to those of their departments in general. Also, we anticipate that just a few of the resulting case studies will be selected by faculty in each department worldwide, depending upon the interests/expertise of the faculty in each department.

This collaboration is timely because a gradual shift from traditional process design to new product and process design in academia has been taking place without fanfare for almost two decades. This can be seen in the recently hired chemical engineering faculty. Relatively few engage in traditional process engineering areas such as unit operations (distillation, crystallization, and so on), reaction engineering, or process design and control. The majority are bio- and materials-oriented projects. These teacher-researchers with their domain knowledge are well positioned to contribute to a product-process design curriculum suitable for the new generation of chemical engineers in advanced economies such as the U.S.

1.2 The Hong Kong Experience

Hong Kong is a global financial center engaging in the development and commercialization of non-commodity chemical products. Manufacturing is often carried out by OEM in Mainland China. For this reason, the curriculum at The Hong Kong University of Science and Technology (HKUST) has already significantly shifted from process design to product design. New product design courses have been developed for freshmen, seniors, and graduate students. We believe that, with the recent developments in fracking and the use of natural gas as feedstock for the chemical industry, the change from a primarily process design curriculum towards a more balanced product and process design curriculum will be delayed by several years in the U.S. But, it will happen.

Over the past four years, HKUST has tried different products in their product-design projects (Table 1). In each design project, a team of 3-5 students is given a project statement defining the scope of the project, relevant literature and consulting sources to help complete the product, guidelines for creating the written and oral reports, and an assessment rubric for evaluating student performance. Advice and comments are provided to the students throughout the project.

Table 1. Products Used for Design Projects in Hong Kong

• **Wine Aerator**	• Air Conditioner with Green Refrigerants
• **Faucet Water Filter**	• Building Air Conditioning System
• **Conductive Ink**	• Phase Transition Type Heating Pad
• **Vitamin C Tablets**	• Energy Recovery Ventilator
• **Hand Lotion**	• Conductive Paste for Printed Electronics
• **Laundry Detergent**	• Disposable Diaper
• **Shampoo**	• Desiccant Dehumidifier
• **Toothpaste**	• Refrigerant Dehumidifier
• **Mosquito Repellent Vaporizer**	• Respirator
	• Portable Seawater Desalination System
• **Mosquito Repellent Mat**	• Wearables
• **IR Blocking Smart Window**	• Powdered Milk
• **PM 2.5 Air Purifier**	• Controlled Release Granule
• **Wound Dressing**	• Membrane Bioreactor
• **Magnetocaloric Fridge**	• Biosensor
• **Refrigerator**	

At the end of the project, the team submits a written report and a video using a power point presentation to summarize all the main points. The items addressed in the final report are given in Table 2. In class, the video presentation is followed by a Q&A session of about 10 minutes. The student solutions to these design projects were the basis for our workshop on Process and Product Design Instruction at the ASEE Summer School for Chemical Engineering (Young) Faculty at North Carolina State University in the Summer 2017. Several solutions distributed to the attendees are available from the first author, Prof. Ng. Attendees of our workshop, and others, have been encouraged to help design groups develop improved product-design case studies, as discussed in Section 2.1.

Because of the open-ended nature of these design problems, they do not have unique and complete solutions. Indeed, different students often come up with very different products for the same problem statement. For example, the design of disposable diaper was one of the design projects. One group of students considered a product that focused on reusing part of the disposable diaper, while another group designed a product that comes with an RFID (Radio Frequency Identification) wetness sensor. Clearly, it is impossible for one single instructor to possess the necessary domain-specific knowledge to guide the students to explore deeply and broadly all the different products in Table 1. Without doubt, a team effort, involving many collaborators worldwide, is the best way to identify interesting and workable chemical products for student design projects.

Table 2. Items Addressed in the Student Written Report

- The product and its desired functions.
- The market size, major companies selling this and related products, competitive analysis, potential innovative products, and so on.
- The required product attributes to succeed in the marketplace.
- The product concept including product microstructure, macrostructure, types of ingredients, and so on.
- The product specifications
- The product manufacturing process; e.g., the flowsheet, equipment to be used, equipment operating conditions, material balances, and so on.
- A financial analysis accounting for the R&D cost, equipment cost, material cost, advertising cost, and so on.

2. Research

This section first describes the foci of the product-design case studies. Unfortunately, space is not available to present an example case study, showing examples of the research issues addressed. But, the research issues in teaching product design, which we are working to explore are included – emphasizing the research aspects of motivating and assessing the success of achieving our objectives with chemical engineering seniors and design instructors.

2.1 Product Design Case Studies

As described previously, up to 25 faculty members from different universities are being identified, many of whom have agreed to help design groups develop improved product-design case studies. Each technology expert will advise a student design team that expands on one of the case studies in Table 1, or creates a completely new product design project following the multidisciplinary, hierarchical approach described in Seider et al. (2017). Note that the products designed by student groups at HKUST were primarily formulated products, functional products, or devices, not including molecular products, which will be included going forward. For each of the 25 case studies, the faculty advisers will prepare the students and guide them toward designing more innovative products. After carrying out their product designs, several learning outcomes are anticipated, including the ability to:

- Perform a market study including competitive analysis using information in the public domain

- Generate new product ideas to enhance human comfort, health, and convenience in a sustainable manner and determine whether such a product is potentially profitable using methods such as business model canvas
- Identify product attributes and product specifications using methods such as house of quality
- select product ingredients using thermodynamic databases and computer-aided molecular design tools such as ICAS from Prof. Rafiqul Gani
- model product performance and adjust the design to meet product specifications
- design the manufacturing process
- perform prototyping
- set product price
- perform capital budgeting
- perform product sustainability analysis
- present a realistic product development plan both orally and in writing effectively

While this is not a requirement, ideally, the faculty adviser, responsible for teaching product/process design, will team up with another faculty member with expertise in the fundamental sciences and applications of the selected product.

2.2 Expected Advances in Teaching Product Design

Learning is achieved in many ways. As shown in Figure 1, the students will learn about the multidisciplinary, multiscale, hierarchical framework for product design from the product-design faculty adviser, perhaps in the classroom. Elements of the framework needed for performing product design – model-based methods, rule-based methods, tools for simulation and optimization, databases for material properties, equipment performance, and so on, and experiments – will be covered. Each design team is expected to learn about the underlying science and technology for their product. This will be achieved by self-study guided by the domain-knowledge faculty adviser.

This combination of experiential learning, self-learning, formal learning, and peer learning in a group setting is expected to improve the students' learning experience. We seek to understand how to use these different ways of learning synergistically and effectively for teaching chemical product design.

Overall, the success of this approach can be judged by the utilization of the product-design case studies worldwide. In addition, each individual product-design case study will be assessed as follows:

- Benefit to human comfort, health, and convenience
- Impact on sustainability and social responsibility
- Potential economic impact
- Innovation
- Technically correct and an appropriate amount of details
- Breadth and depth of the report Feedback from students will be documented. Items will include:
- Motivation to self-learning
- Stimulation to pursue entrepreneurship

- The learning outcome of the perspective, methods, and tools for product design Feedback will also be sought from faculty participating in this guided product-design project:
- The effectiveness in integrating process design faculty with faculty engaging in engineering sciences in a typical chemical engineering department
- The effectiveness of the project-based approach to teach product design

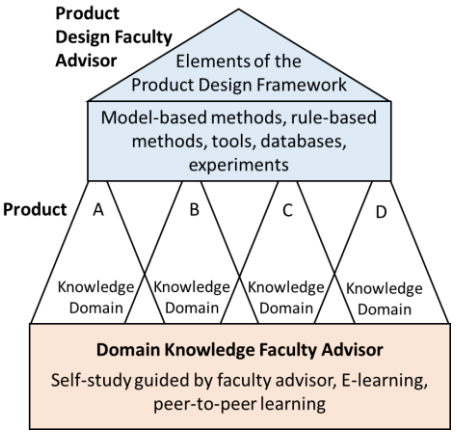


Figure 1. The students learn about how to perform product design and the domain knowledge underlying the product under consideration from the faculty adviser(s), E-learning, and so on. Each different product (of the four, A to D, shown here) is expected to involve a different knowledge domain although these domains are expected to overlap to some degree.

3. Conclusions

This manuscript documents our approach to generate a library of expert-driven, product-design case studies.

References

U. Bröckel, W. Meier, G. Wagner (editors), 2013. Product Design and Engineering: Formulation of Gels and Pastes, Wiley, Somerset, NJ.

E. L. Cussler, G. D. Moggridge, 2011. Chemical Product Design, 2nd ed. Cambridge University Press, Cambridge, U.K.

E. L. Cussler, J. Wei, 2003. Chemical product engineering. AIChE J., 49, 1072-1075.

R. Gani, 2004. Chemical product design: challenges and opportunities. Computers and Chemical Engineering, 28, 2441-2457.

K. M. Ng, R. Gani, K. Dam-Johansen (editors), 2007. Chemical Product Design: Toward a Perspective through Case Studies, Computer Aided Chemical Engineering.

W. D. Seider, S. Widagdo, 2012. Teaching chemical engineering product design. *COCHE*, 1, 472-475.

W. D. Seider, D. R. Lewin, J. D. Seader, S. Widagdo, R. Gani, K. M. Ng, 2017. Product and Process Design Principles: Synthesis, Analysis and Evaluation. 4th Ed., Wiley.

G. Stephanopoulos, 2003. Invention and Innovation in a Product-Centered Chemical Industry: General Trends and a Case Study, AIChE Institute Lecture.

J. Wei, 2007. Product Engineering: Molecular Structure and Properties, Oxford University Press, Oxford, New York, NY.

J. A. Wesselingh, S. Kiil, M. E. Vigil, 2007. Design & Development of Biological, Chemical, Food and Pharmaceutical Products, Wiley, Chichester, U.K.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50267-6

Reflections on embedding safety throughout the process engineering program

Michaela Pollock, Eva Sorensen*

Department of Chemical Engineering,UCL, Torrington Place, London, WC1E 7JE,United Kingdom
e.sorensen@ucl.ac.uk

Abstract

Safety is an important part of being a well-rounded, responsible process engineer. It not only covers fundamental scientific knowledge but also a way of thinking and culture in how engineers approach their work, and is continually developed throughout the working life of a process engineer. However, how this safety learning can start to be imparted to engineering students in an academic environment is a challenge for educators. In this work the systems approach that has been taken as part of UCL's Integrated Engineering Program (IEP) teaching framework is examined. Within this framework, safety is embedded into the curriculum from the start in Year 1 and is continually extended and advanced throughout the process engineering program. As the first cohort of students graduate we reflect on how this has been implemented and received.

Keywords: teaching, safety, embedded learning

1. Introduction

The significance of safety within process engineering is widely acknowledged in industry and education. However, the learning, and application of that learning, in an industrial and educational context is very different. In an industrial setting, where recent engineering graduates are applying their newly-gained fundamental knowledge to real-life problems within interdisciplinary teams of varying experience, it is easy to comprehend how the graduates can build on their fundamental knowledge to develop into well-rounded, safety-aware and responsible engineers. These real-life settings and awareness of the full context of safety, are challenging for educators to impart to students who are still developing their knowledge of core chemical engineering subjects and have limited if any industrial experience.

Ultimately, these challenges cannot be overlooked by educators aiming to prepare a young person for a career in engineering, as imparting an understanding and awareness of safety is a key learning outcome required by accreditation bodies globally. The IChemE's accreditation guidelines (IChemE, 2017) require graduating students to have gained an understanding of a variety of process safety learning outcomes, such as inherent safety, principles of risk assessment and methods of identifying process hazards. In addition to formally taught safety courses, the IChemE insists that students are instilled with appropriate attitudes to safety, health & the environment (SH&E). Similarly, the ABET criteria for chemical engineering programs (ABET, 2017) state that the curriculum must include the hazards associated with chemical, physical and/or biological processes. These open-ended and wide-ranging learning outcomes on safety are mirrored in numerous fundamental process safety textbooks, such as Crowl and Louvar (2011), who

propose that successful safety programs require several elements including fundamentals, experience and attitude.

The Department of Chemical Engineering at UCL offers both Bachelor and Masters programs in Chemical Engineering through its Integrated Engineering Programme (IEP) (Sorensen, 2016), with around 120-150 students per cohort. The IEP is a Faculty-wide, multidisciplinary program which started in 2014/15 and combines core disciplinary technical knowledge with interdisciplinary and/or research-based projects with strong emphasis on professional skills and academic learning connected with workplace learning. The program enables students to understand the fundamentals of their discipline, to practice the application of their core technical knowledge and to apply this to current complex global challenges such as energy, health *etc*. Students also have the opportunity to apply their fundamental learnings to industrially-relevant mini-design projects, known as scenarios. This focus on problem and/or project-based learning starts in Year 1 and continues throughout the program. As for most UK programs, most courses are compulsory in Years 1-3, and are taken by the entire cohort at the same time. It is therefore possible to properly plan the delivery of key concepts such as safety to ensure the material is introduced at the right time and that there is a clear progression from one course to the next or from one year to the next.

The IEP framework enables a systems approach to teaching safety to be undertaken. Firstly, from a depth perspective where safety is embedded into the curriculum immediately in Year 1 and the level of understanding is built upon in subsequent years. Secondly, from a breadth perspective, where different safety learnings are applied to an assortment of industrially-relevant problems throughout the program.

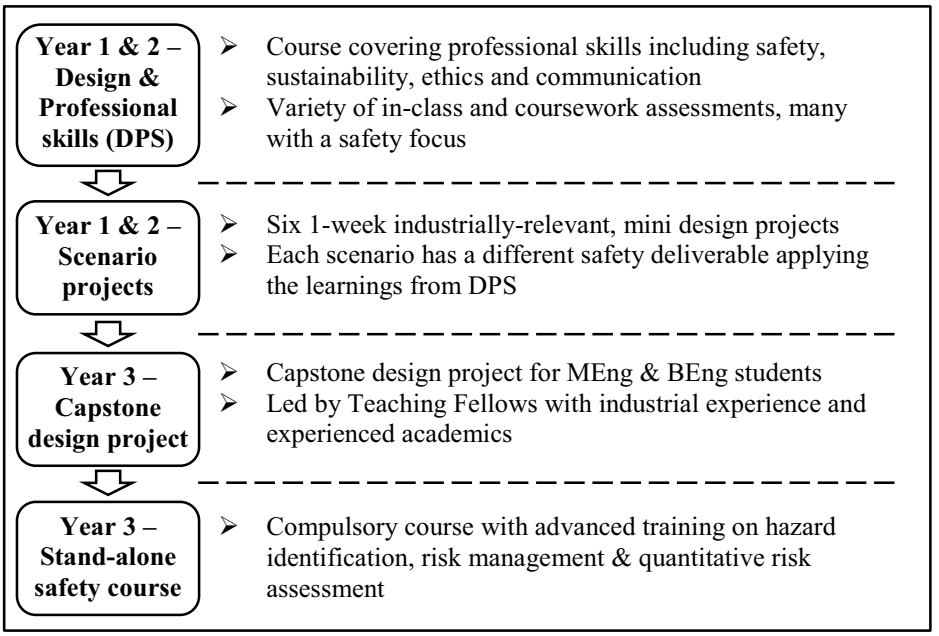


Figure 1: Illustration of how safety is embedded in the UCL Integrated Engineering Program (IEP) teaching framework.

Figure 1 illustrates how safety has been embedded throughout the IEP framework. In the remainder of this paper, the methods used to introduce and apply safety concepts in Years 1 and 2 are examined, and the approaches taken to extend this to advanced application and learning in Year 3 are then outlined. As the first cohort of students graduate from the IEP framework, reflections on embedding safety teaching and student feedback are examined, followed by conclusions on the approach.

2. Introducing safety concepts in Years 1 and 2

As soon as students arrive on campus they are introduced to the concepts associated with safety as one element of the 'Design and Professional Skills' course. Topics covered include risk, hazards, inherent safety and hazard evaluation techniques. Initially these concepts are introduced in Year 1 by considering everyday situations to which students can easily relate. For example, examining a photo of a staircase and considering safety strategies that can be adopted, or performing a HAZID (hazard identification) on everyday activities such as the commute of a cyclist in London. In Year 1, this is also extended to the context of the process industries through lecture examples which briefly introduce past accidents and utilise simple process examples such as an example of a HAZOP (hazard and operability study) on a fired heater. As in Shallcross (2013), assessment is introduced in terms of examining past accidents and considering issues that led to, and lessons learned from, the incident.

As students proceed through the second year of the IEP program, safety concepts and hazard evaluation techniques continue to be examined with a greater emphasis on process industries and their application to different stages in the process design timeline. As suggested by Shallcross (2014), brief 5 minute safety shares, which reflect industrial practice of focusing safety discussions at the start of meetings, are used to introduce safety topics covered in the lectures. Topics examined range from dangers in the home, major process incidents such as Bhopal and lessons learned from other industries such as the sinking of the Herald of Free Enterprise.

Supporting material for lecture content comes from traditional textbooks such as AIChE CCPS (2008) and Mannan and Lees (2012). This is supplemented by safety videos from the IChemE Safety Centre (ISC) (Kerin, 2016) and the Chemical Safety Board (CSB) (Horowitz and Gilmour, 2007) to highlight particular topics covered within lectures and to illustrate the context of process industries to students who potentially have not yet set foot on a process plant. One particular topic covered in the second year is hazards and control of exothermic reactions, fulfilling the recommendations from the CSB following the T2 Laboratories accident that awareness of reactive hazards is incorporated into the curriculum (Crowl and Louvar, 2011).

3. Applying safety learnings in Years 1 and 2

Ultimately, the subjects of engineering and safety are applied subjects. As such, students on the IEP program are given the opportunity to apply their safety learnings through a safety deliverable incorporated into each of the six industrially-relevant, week-long mini design projects, known as scenarios, which are completed in the first two years of the IEP degree program. As an example, in a scenario on air separation for the production of nitrogen for food packaging, students produced a safety newsletter examining a past accident involving cryogenic distillation and reported findings from a HAZID performed on their nitrogen production plant. The form of the deliverable, a safety newsletter,

replicated a common industrial safety document. In addition, this scenario increased students' awareness of technical documentation and standards through the freely available publications of the European Industrial Gases Association (EIGA) accessed via EIGA (2017).

The variety and format of the six scenarios completed during the first two years allows students to apply safety learnings on processes with which they are already familiar. This is representative of experiences of process engineers in industry where engineers perform hazard evaluations on processes well known to them as well as to unfamiliar processes. For example, in a Year 1 scenario, students performed a HAZOP on a reactor involved in styrene monomer production based on a process flow diagram (PFD) which they constructed from a process description provided in the student brief. In Year 2, the design of the styrene monomer process was further developed into a detailed PFD with some initial control strategies by teaching staff. The students were again asked to perform a HAZOP, this time on a distillation column and surrounding secondary units. By embedding safety into the curriculum, students are able to perform hazard evaluations on a process they have already encountered and on engineering documentation with varying levels of detail reflecting different stages in the process design timeline. Furthermore, students learn about fundamental chemical processes, such as distillation and reactors, and concurrently consider safety issues related to the design and operation of these units.

4. Advancing safety learning in Year 3

The safety learnings are further embedded in the third year in two parallel approaches, the capstone design project and a stand-alone Advanced Safety and Loss Prevention course. These courses, and their scheduling together, encourages in-depth learning of safety concepts through advanced application of hazard evaluation methods and deeper learning of the consequences of hazards, their quantification and safety management systems.

The capstone design project, which accounts for a quarter of the third year and where students work in teams of 6-7 students, is arranged in a manner so that students follow the development of a process from research and conceptual design, through preliminary, and then detailed design. As discussed in Towler and Sinnott (2012), the engineering information available, and the hazard evaluation methods that can be employed to assess the hazards related to the process, differ at every stage of the project. This is one of the key learnings for students where at every stage of the design project a different hazard evaluation, or safety element, is performed.

During the initial research and conceptual design stage, students investigate what is known to them in the form of a safety data sheet (SDS) review and are recommended to review literature on past accidents for similar processes. They then use this knowledge to perform a preliminary hazard analysis (PreHA) of the conceptual design in their design teams, focusing on potential hazards related to feedstock, major units, utility and support systems and other categories as relevant to the chosen project. This approach is flexible and allows students to focus on areas relevant to their process, for example, major units such as exothermic reactors in vinyl chloride monomer production or storage of feedstock in the conversion of biomass into bioethanol.

The design project then moves into preliminary design as students develop their process further with a detailed PFD and process simulation. At this stage, following formative and summative feedback on their PreHA study and their second year HAZOP workshop,

students perform a HAZOP on the process they are developing. Knowledge from the chemical review and any review of past accidents again aid in the hazard evaluation of the process and recommendations from the HAZOP often give an indication of potential control system requirements.

Students then move onto the individual component of the capstone design project where they are required to develop a detailed design of a major process unit, typically a separation unit, such as a distillation or absorption column, or a reactor unit. As they take their individual unit design through the process design stage and into detailed design, students develop a control system for their unit which they then represent on a piping and instrumentation diagram (P&ID). To complete the detailed design stage, students perform a safety integrity level (SIL) analysis where they quantitatively assess the reliability of their system. They identify the major hazards associated with their unit by looking back over findings from the PreHA, HAZOP and past accident reviews. For the selected major hazard they then identify a tolerable risk threshold for the hazard. To evaluate the reliability of their individual process unit design, students then develop a Fault Tree to identify the causes that can lead to the hazard and quantify the risk of the hazard occurring. Following comparison with the tolerable risk threshold, students then formulate recommendations of further actions, if any, required for the process design.

The safety teaching within the IEP framework concludes with a compulsory stand-alone Advanced Safety and Loss Prevention course extending the students' existing knowledge and awareness of safety. The course has a strong focus on risk management, management systems and quantitative risk assessment. The causes and physical effects of hazards are examined in detail and the course concludes by examining the concepts of reliability and availability. Assessment is split between 80% exam-based and 20% coursework-based.

5. Reflections on embedding safety within the IEP framework

At the end of 2016/17, as the first cohort of students were about to graduate with a Bachelor's in chemical engineering following the IEP framework, the students were asked for feedback on their perceptions of the IEP approach to embedding safety learning. Over 90 % of the respondents *Somewhat* or *Strongly* agreed with the statements that they understood the importance of safety in society in general, in chemical engineering and the concepts of inherent safety. When asked in which year they had been first introduced to the concepts of risk, hazard and accidents or had the opportunity to apply their safety knowledge, over 50 % of respondents indicated in their first year. While this does indicate that some further emphasis of safety teaching and its application is still needed in the first year, we can see that we are well on our way to achieving our goals of developing graduate engineers well prepared for a career in engineering with one student commenting "*courses are taught with a safety aspect in mind so that safety becomes part of our everyday critical thinking".*

6. Conclusions

In conclusion, embedding safety throughout the process engineering curriculum has been successfully achieved using a systems approach at UCL Engineering as part of the Integrated Engineering Program (IEP) teaching framework. This has been validated through the successful accreditation of the UCL IEP chemical engineering degree by the IChemE in 2016. Furthermore, student feedback has indicated significant understanding of the importance of safety in chemical engineering and in society in general. Finally, it

is important to emphasise that such an all-encompassing program can only be achieved through the efforts of staff, many of whom are chartered engineers, with a wide range of academic, industrial and lab experience which mirrors the wide-ranging background and experience of engineers that graduates will work with in future roles.

References

ABET, 2017, Criteria for accrediting engineering programs, http://www.abet.org/wp-content/uploads/2017/12/E001-18-19-EAC-Criteria-11-29-17-FINAL.pdf [accessed 2 December 2017].

AIChE Center for Chemical Process Safety (CCPS), 2008, Guidelines for hazard evaluation procedures, CCPS, New York, N. Y., Wiley-Interscience, Hoboken, N. J.

D. A. Crowl and J. F. Louvar, 2011, Chemical Process Safety Fundamentals with Applications, Prentice Hall, Upper Saddle Hall, N. J., 3rd Ed.

European Industrial Gases Association (EIGA), 2017, Catalogue of EIGA publications, https://www.eiga.eu/index.php?eID=dumpFile&t=f&f=149&token=3af8d1141e323b02944281b731348c5fa28071da [accessed 1 December 2017].

D. M. Horowitz and S. Gilmour, 2007, CSB Safety Videos: A new and effective communication tool for accident investigation findings, Process Safety Progress, 26, 3, 177-179.

IChemE, 2017, Accreditation of chemical engineering programmes A guide for higher education providers and assessors, http://icheme.org/accreditation/university-accreditation/guidance.aspx [accessed 2 December 2017].

T. Kerin, 2016, A new way of learning from incidents – ISC's new case-study based training eliminates hindsight bias, The Chemical Engineer, 900, 51.

S. Mannan and F. P. Lees, 2012, Lee's loss prevention in the process indusstries: hazard identification, assessment, and control, Butterworth-Heinemann, Boston, Elsevier, Amsterdam.

D. C. Shallcross, 2013, Safety education through case study presentations, Education for Chemical Engineers, 8, e12-e20.

D. C. Shallcross, 2014, Safety shares in the chemical engineering class room, Education for Chemical Engineers, 9, e94-e105.

E. Sorensen, 2016, Changing the World – Educating students differently with a more scenario and problem-based engineering curriculum, The Chemical Engineer, 904, 27-30.

G. Towler and R. K. Sinnott, 2012, Safety and Loss Prevention in Chemical Engineering Design: Principles, Practice and Economics of Plant and Process Design, Elsevier Science, Amsterdam, 2nd Ed.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50268-8

Advantages of Flipping Multiple PSE Courses (to the same students)

Daniel R. Lewin[a*] and Abigail Barzilai[b]

[a]*Department of Chemical Engineering, Technion I. I. T., Haifa 32000, Israel*
[b]*Center for the Promotion of Learning, Technion I. I. T., Haifa 32000, Israel*
dlewin@technion.ac.il

Abstract

In this paper, we present a case for injecting active learning methodology into PSE-related courses. In particular, we discuss the teaching of both the capstone process design course and the control course in flipped format, focusing specifically on the associated benefits and costs, both from the perspective of the teacher and from those of the students. While not all of the students have taken to the flipped approach, the vast majority agree that they learn more by flipping than by the conventional lecture-based approach. The bottom line is an increased average grade, a normal, unimodal grade distribution, accompanied by a reduced failure rate.

Keywords: Flipped Classroom, Capstone Design Course, Control Course, Learning Assessment, Active Learning.

1. Introduction

Most chemical engineering curricula feature at least two Process Systems Engineering components: a capstone process design course and a course on process dynamics and control. Often, these courses are taught in the same semester, which makes sense given their interrelated nature. In this paper, we present the benefits of taking an active-learning approach, employing flipping, to teaching these two courses, both from the perspective of the teacher and from those of the students.

Student engagement in the learning process is a prerequisite for a successful outcome. Deslauriers et al. (2011) and Bishop and Verleger (2013) present ample evidence that students who study in a learner-centered environment with an emphasis on active student engagement acquire a better conceptual understanding than their peers in a teacher-centered learning environment, where the emphasis is on transferring knowledge from teacher to learner. The lecture model is suitable for conveying simple information, but fails in response to the need to provide customized challenges and feedback required for a thorough understanding of complex problems. Student-centered environments are perceived by students as supportive to their learning and characterized by interaction between the lecturer and the students, active learning and the provision of effective feedback and time-management. These environments can accommodate different student skills and learning methods, enabling them to conduct discussions in the classroom and work in teams on projects, under conditions that ensure positive interdependence and personal responsibility, and were found to improve student achievements and learning experience, increasing class attendance and helping to reduce failure rates (Freeman, 2014).

As shown in Figure 1, our flipped classroom model (Lewin and Barzilai, 2017) features three main components: (a) On-line lessons, taking students up to three hours to cover at home in their own time, and in advance of the class-session. A lesson contains a

sequence of between 5 and 12 segments, with each segment comprised of a short video clip (up to 15 minutes per clip, with most being a lot shorter) and an associated quiz question; (b) A two-hour class-meeting in which portions of that week's on-line lessons which were found to have been difficult for most of the students are explained in more detail, but mostly allowing for in-class activities (i.e. clicker questions and collaborative open-ended problem solving); (c) A three-hour active tutorial session, in which groups of students solve classwork (previously called homework). This pedagogical model has been implemented in both the design and control courses.

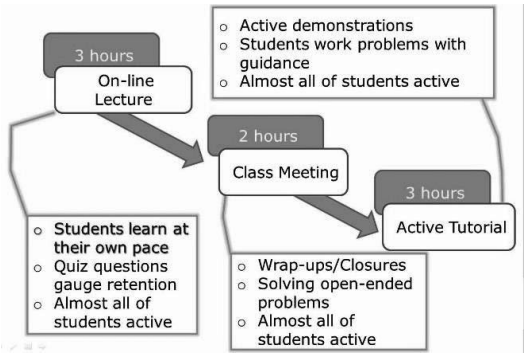


Figure 1: A weekly cycle of our flipped classroom model.

This paper summarizes the experiences after two years of "double-jeopardy," that is, after having two consecutive years of undergraduates take both the design and control courses in the same semester in flipped format, it having been suggested that a single flipped course is so exhausting for students that no more than one such course is recommended per semester. In particular, we shall analyze the students' opinions regarding their level of confidence in the materials taught, how they perceived the flipped approach impacted on their workload and well as the final results.

2. Curriculum and teaching objectives

The design course, closely following Seider et al. (2017), consists of four segments: (a) An overview of the design process itself, and including discussion of ethics, safety and environmental protection issues; (b) Plant economic analysis, focusing on capital cost estimation and profitability analysis; (c) Systematic design methods such as separation system synthesis methods and heat exchanger network synthesis; and (d) Plantwide control, HAZOP and HAZAN. An integral part of the design course is a design project, worth 30% of the total course grade, in which teams of five students compete with each other on the most profitable process for a given production rate of a target chemical, with 25% of the project grade proportional to the venture profit scored by each team. On completing the course, a successful student will be able to:

- Simulate a chemical process, including its optimization, using UNISIM.
- Perform an economic assessment of a chemical process.
- Synthesize a separation system.
- Synthesize a heat exchanger network (HEN) for a chemical process such that the maximum energy is recovered and/or the minimum number of exchangers is used.
- Prepare a Piping and Instrumentation Diagram (P&ID).
- Conceptually design a plantwide process control configurations using qualitative methods.
- Perform HAZOP/HAZAN analysis on a process based on its P&ID.

The control course provides the tools for developing dynamic process models (focusing only on lumped parameter systems), and the use of these models to develop process control systems. Taught in parallel with the design course and to the same cohort of students, the course largely follows Seborg et al. (2016), and consists of two parts, as illustrated in Figure 2. The first part of the course, taking six weeks of the semester, covers the modelling of process dynamics as well as mathematical analysis: Laplace transforms, open-loop transfer functions and linear dynamics. The second part of the course, lasting another seven weeks, covers control system design: frequency domain analysis, derivation of closed-loop transfer functions, stability analysis, root locus design, PID tuning, and cascade and feedforward control system design. On completing the course, the student will be able to:

- Formulate a lumped parameter model describing the process dynamics.
- Generate an approximate linear model of the process in its standard forms, either from fundamental models or empirical models fitted to plant data.
- Design a control system for the process that meets desired performance criteria.
- Use MATLAB/SIMULINK for the design and testing of control strategies.

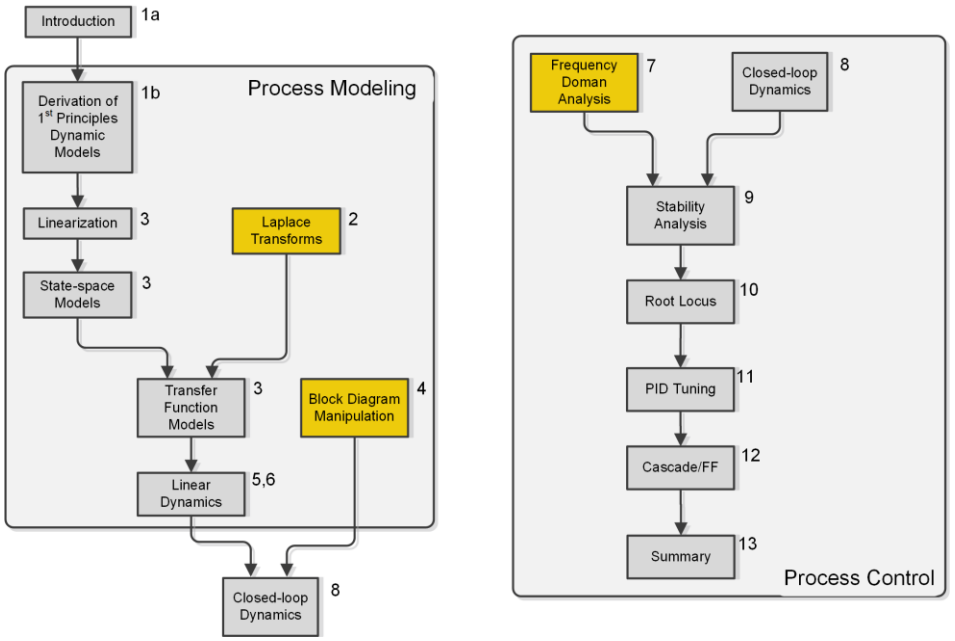


Figure 2. Flow diagram for control course (numbers indicate week of activity).

3. Differences between conventional and flipped teaching

For students to master a course, they need to apply themselves to independently working problem sets covering the course material. Traditionally at the Technion, this has been left to the student to do on his own at home. This leads to several problems: (a) Most of the contact time in the course involves students listening, first to the lecturer and then to the teaching assistant, during which most students are passive. Thus, the first time they are asked to be active is at home, where there is no help; (b) After submitting their solutions, typically if feedback is provided (often there is insufficient manpower available to do this effectively), it comes far too late to be of much use; (c) The feedback on student

performance is received by the lecturer too late for him to make the necessary corrections to his teaching.

A much better alternative is to convert the exercise sessions to "active tutorials," where students working in teams solve the homework in class time. This ensures that: (a) All students who participate in the sessions are actively involved in working problems; (b) Assistance can be provided by the course staff, and by the students themselves; (c) Both students, assistants and the lecturer all receive feedback in a timely fashion. In our experience, in every course in which the exercises have been switched to this mode, there has been a significant improvement in the performance of the students in the final exams. As an illustration, consider the depth of coverage possible in three weeks of the control course by keeping the students actively engaged, shown in Table 1.

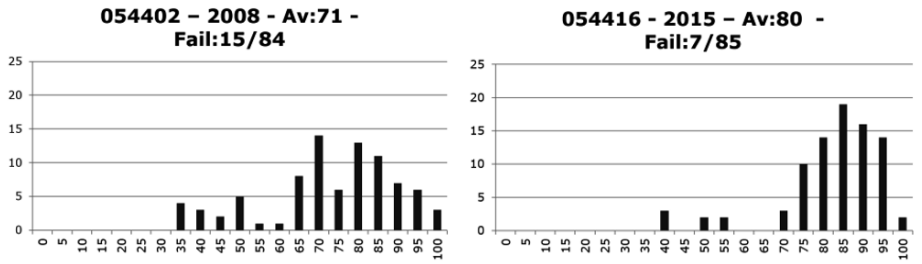
Table 1: Subjects and concepts taught and exercised in the three-week sequence between weeks 8 and 10 in the control course covering control system design.

Week	Subject	Concepts	Exercises
8	Closed-loop Dynamics	– Introduction to PID controller – Closed-loop transfer functions	– Computing actual closed loop transfer functions – Closed loop response using MATLAB/Simulink
9	Stability Analysis	– Stability tests (Routh Hurwitz, Direct, Bode) – Internal stability	– Practical examples – Controller tuning using MATLAB/Simulink
10	Root Locus	– Root locus method – Designing to meet specifications	– Designing for specifications – Open-loop unstable systems – Root locus design using MATLAB/Simulink

By design, flipped learning environments encourage continuous opportunities to assess student learning, both from the point of view of the student and of the lecturer. While watching the on-line lectures, the students participate actively, receiving immediate feedback on the on-line quiz questions, thus reinforcing their learning. They come to the class meeting where additional quiz questions are posed to them with clickers used to collect the responses of the entire class, and used to further drive discussion. Open-ended problem-solving further prepares them for the next step – the active tutorials, where they tackle classwork problems in cooperative effort. There are several features of this approach that can also provide the lecturer with on-going assessment of the degree to which course objectives are attained, providing opportunities to continuously improve and modify delivery. First of all, by observing the class performance on the on-line quizzes ahead of the class meeting, the lecturer can tailor the materials delivered to focus the discussion to items/issues that most of the class found the most difficult. Using clicker questions as CAT (class assessment techniques) in the class meeting, provides additional opportunities for both additional testing of student capabilities, as well as nurturing discussion of the results, providing for deeper understanding. Finally, the active tutorials are the ultimate test of the capabilities of the students to handle problems on their own, and the presence of the staff (assistants as well as the lecturer) enable a true assessment these capabilities to be ascertained well ahead of the final exam.

Figure 3 compares a typical grade distributions on the final exam before the design course was switched to active tutorials, with that obtained in the first flipped instance of the course, in which both flipping and active tutorials increased the amount of time students were actively engaged in their learning. It is clear that vast majority of the

students did very well in the course, as indicated by the increased average grade, and the reduced variance and number of failed students. Irrespective of how the lecturer performs, the very best students in the class will probably do well in any course, with the least capable students are likely to face unsurmountable difficulties. That leaves a huge number of students in the middle, whose performance can be significantly enhanced if the lecturer takes the time and makes the effort to reach out and motivate them. In our opinion, that is one of the yardsticks by which one should measure the degree of success, and the data in Figure 3 seems to confirm that the flipped approach achieves it.



(a) (b)

Figure 3: Typical final grade distributions for the design course: (a) before introduction of the active tutorials and flipping; (b) after introduction flipping and the active tutorials.

A word about the costs associated with the flipped approach is in order. While the total amount of time invested by the average student on a flipped course is the same as that expected in a conventional course (assuming the homework time is included), the same is not true for the load on the lecturer. The planning, and production of the on-line lecture videos as well as the on-line quizzes requires considerable time and attention from the lecturer, but this is mainly a one-time investment, and need not involve high costs. Indeed, as opinioned by Marlin (2017), "if you can prepare a PowerPoint, you can prepare an e-lesson for the internet." What is true for both students and lecturers, is that the time spent working together on the course is used far more effectively.

Table 2: Overall performance record for the design course over time.

Year	Mean (%)	STD (%)	Failure Rate	Remarks
2008	71	17	15/84 = 18%	
2009	75	16	6/68 = 9%	
2010	75	17	13/74 =18%	
2011	79	13	4/77 = 5%	First active tutorial
2012	82	11	7/80 = 9%	
2013	78	13	6/67 = 9%	
2015	80	13	7/85 = 8%	First flipped class
2016	76	20	9/71 = 13%	
2017	78	15	9/87 = 10%	

4. Teaching outcomes and student feedback

Table 2 summarizes the student performance in the design course since 2008, noting that the tutorial went "active" in 2011, and the course was switched to flipped format in 2015. It is noted that the move to more active involvement of the students resulted in increased average grade and a reduced failure rate. The reduction in the failure rate is often

accompanied by the transformation of a multimodal grade distribution (e.g., Figure 3a) into a normally distributed one (e.g., Figure 3b), as many of the poorer-performing students improve their mastery and thus become higher performers.

It is therefore fair to claim that flipping and active tutorial leads to improved class outcomes, but what do the students think? Student participants of our flipped courses are polled to ascertain their preference between the flipped/conventional approaches, and to gauge the level of their confidence in the course materials at each stage of the flipped approach. By the end of the courses, only a small minority of the students still think that the conventional approach is the most appropriate, in contrast to about two-thirds of the class that conclude that they prefer the flipped approach. Interestingly, usually about a quarter of the class are in strong disagreement that the flipped approach suits them, maintaining this opinion throughout the semester. It should be noted, though, that approximately half of these "die-hards" usually concede that they learn better using the flipped approach. The majority of the students consider the active tutorial component to be the most helpful in helping them attain mastery of course content.

5. Conclusions

Experience with the flipped-class approach indicates that overall, engagement with the materials and ongoing assessment throughout the semester improved the students' level of confidence in their mastery of the curriculum. This is principally because they come to class and to the tutorials better prepared than with the conventional version of the course. These observations could explain the improved performance in the final exams, as demonstrated. Table 2 shows the progressively increasing students' competence, as indicated by the sustained and improved average grade and reduced failure rates. This progression has been achieved in concert with the course's gradual transformation from conventional teacher-centered one (i.e., lecture-based) to student-centered ones (i.e., active tutorials and flipping). Having students take more than one flipped course in a semester does not overload them; in fact the teaching both the design and control courses together in this format only enhances the degree to which students engage as they also grasp the necessary synergy between the design and control of chemical processes.

References

Bishop, J. L. and M. A. Verleger, 2013, "The Flipped Classroom: A Survey of the Research," *Proceedings of the ASEE Annual Conference & Exposition. 2013*, 23.1200.1-23.1200.18.

Deslauriers, L., E., Schelew and C. Wieman, 2011, "Improved Learning in a Large-enrollment Physics Class," *Science*, **332**, 862-864.

Freeman, S., S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, J. Hannah and M. P. Wenderoth, 2014, "Active Learning Increases Student Performance in Science, Engineering, and Mathematics," *Proc. Natl. Acad. Sci.,* **111**(23), 8410-8415.

Lewin, D. R. and A. Barzilai, 2017, "Flipping the Capstone Process Design Course," *Proc. of the 27th European Symposium on Computer Aided Process Engineering*, 2923-2928.

Marlin, T. E., 2017, "Flipping the Chemical Engineering Process Control Class with e-Lessons," *Proceedings of the ASEE Annual Conference & Exposition. 2017*, 12385-12411.

Seborg, D. E., T. F. Edgar, D. A. Mellichamp and F. J. Doyle, 2016, "Process Dynamics and Control," 4th Edition, John Wiley and Sons, Inc.

Seider, W. D., D. R. Lewin, J. D. Seader, S. Widagdo, R. Gani and K. M. Ng, 2017, "Product and Process Design Principles," 4th Edition, John Wiley and Sons, Inc.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50269-X

Cyber Incident Exercise Admitting Inter-Organization for Critical Infrastructure Companies

Akihiro Tsuchiya*, Yuitaka Ota, Yuma Takayama, Tomomi Aoyama, Takashi Hamaguchi, Yoshihiro Hashimoto, and Ichiro Koshijima

Nagoya Institute of Technology, Nagoya 466-8555, JAPAN
a.tsuchiya.054@nitech.jp

Abstract

To improve cyber security awareness for Critical Infrastructure (CI) companies, Kaspersky Interactive Protection Simulation (KIPS) is provided by Kaspersky Lab. Through the KIPS exercise, participants simulate countermeasures while experiencing the effect of a cyberattack on a virtual CI company. However, an organizational structure that makes decisions relative to incident responses via communication between multiple departments is not explicitly specified in the KIPS exercise. Therefore, to increase awareness of the complexity of decision-making processes, a non-technical exercise method that divides a virtual company into headquarters and production plant administrator groups is proposed. The proposed exercise method was implemented using KIPS, and a trial was conducted with participants involved in CI companies.

Keywords: Industrial Control System, Cyber Security, Exercise, Risk Communication

1. Introduction

Cyberattacks on Critical Infrastructure (CI) have been recognized as a significant problem since the discovery of the Stuxnet worm in July 2010 [1]. The Stuxnet worm was primarily designed and developed to target an Industrial Control System (ICS). ICSs are used in gas pipelines, power plants, and chemical and petrochemical plants. The attackers designed Stuxnet to inflict damage by reprogramming programmable logic controllers (PLC) to control ICSs. After the discovery of Stuxnet, similar malware that targets ICSs has been discovered.

Cyber incidents that target ICSs are security, safety, and business problems. Such abnormal events affect physical devices, such as actuators and sensors. If a cyberattack results in manufacturing operations being shut down, a company will lose significant revenue. In addition, if a cyberattack targets systems that require safety operations, the operators will be endangered. For example, cyberattacks on an iron furnace have been reported [2]. In addition to safety risks, cyberattacks continue to pose serious financial risk for companies [3]. Therefore, cyberattacks should be prevented to ensure corporate resilience.

In addition to awareness of potential cyberattacks, the need for cybersecurity training has increased. The National Institute of Standards and Technology (NIST) specifies that incident response teams should be assigned and trained to develop incident response capabilities against cyberattacks. NIST SP 800-61 [4] describes the ability required for

an incident response as follows: “Managers should be technically adept and have excellent communication skills.” A capability for an incident response requires technical skill and non-technical skill. Several authorities have developed tabletop cybersecurity exercises to improve technical security awareness. Tomomi et al. described a technical exercise from the perspective of non-technical skills [5]. However, most security exercises focus on the technical aspects of incident responses. In other words, they are not designed to evaluate a team’s non-technical skills even though it is obvious that a team’s non-technical skills will affect overall performance.

In this paper, a tabletop exercise that involves participant communication skills and decision-making processes in a complex environment is proposed. The remainder of this paper is organized as follows. The basic exercise is explained in Section 2. The proposed exercise structure is described in Section 3. A prototype implementation and initial trial are discussed in Section 4, and conclusions are presented in Section 5.

2. Kaspersky Interactive Protection Simulation Exercise

2.1. Exercise Overview

The Kaspersky Interactive Protection Simulation (KIPS) was developed by Kaspersky Lab [6]. KIPS is a hybrid game with action cards and a game simulator that is intended to deepen the common understanding of the timeline of cyber incidents. Through KIPS exercises, players practically simulate an incident response while experiencing the effects of a cyberattack on a virtual CI company. Players acting as a security administrator for a virtual company determine countermeasures against cyberattacks within time and cost constraints. The goal of the game is to maximize revenue when responding to cyber incidents.

The KIPS exercise for multiple players comprises a game board, action cards, and a game console. The game board represents the plant and network configuration of the virtual company. Players use the game board to understand how the plant works and the devices related to the plant’s operations. The game board also includes space for enabled action cards. Once a player enables an action card, it is placed in an applicable space. Thus, players can observe which action cards have been used. An action card represents a set of cybersecurity countermeasures. There are thirty types of action cards, e.g., a network disconnection card. Each action card represents a countermeasure and shows the required time and costs. Some action cards are added in some cases. Player can combine action cards according to the situation, such as plant status, and the available budget and time. The game console is used to simulate the game, and it provides players with information about the virtual company. In addition, players send their selected action cards to a game moderator.

2.2. Scenario for CI Company

The KIPS provides two CI-related scenarios, i.e., a water purification plant and a combined cycle power plant. The water purification plant has two production lines, each of which comprises a precipitation tank, sand filter, disinfection tank, and drinking water tank. The power plant has two turbines, i.e., a gas turbine powered by burning fuel and a steam turbine powered by boiling water. The water is heated by exhaust gases. Then, the exhaust gas is emitted through a gas filter. In addition, the steam is changed to liquid water by cooling water.

Here, PLCs control both plant operations, and the PLCs are connected to a server in the control network. In the control network, there are various devices, such as a Human Machine Interface, a Data Historian, and an Engineering Workstation. Process data are sent to the headquarters over the Internet. The goal is to protect the devices using action cards.

2.3. Game Simulation

The game consists of a message phase, an action phase, a revenue phase, and a report phase. These four phases are cycled five times to complete the game. Prior to starting the four phases, the moderator explains the rules of KIPS and shows the participants threats of the same industry as news. The moderator operates a dedicated game console to advance each phase. In the message phase, players receive various information, such as news from the same industry and the status of the plant. Next, in the action phase, players evaluate the current situation and use action cards as countermeasures using the game console. The action phase is finished after the moderator has received action cards from all teams. The administrator console calculates each team's revenue according to their actions. The results of a team's actions and their revenue are sent to the applicable team in the report phase. Then, a card assistant distributes additional action cards to some groups that chose an action card which leads new event. In the report phase, all players review their team's result. Figure 1 shows game console of KIPS by four phases.

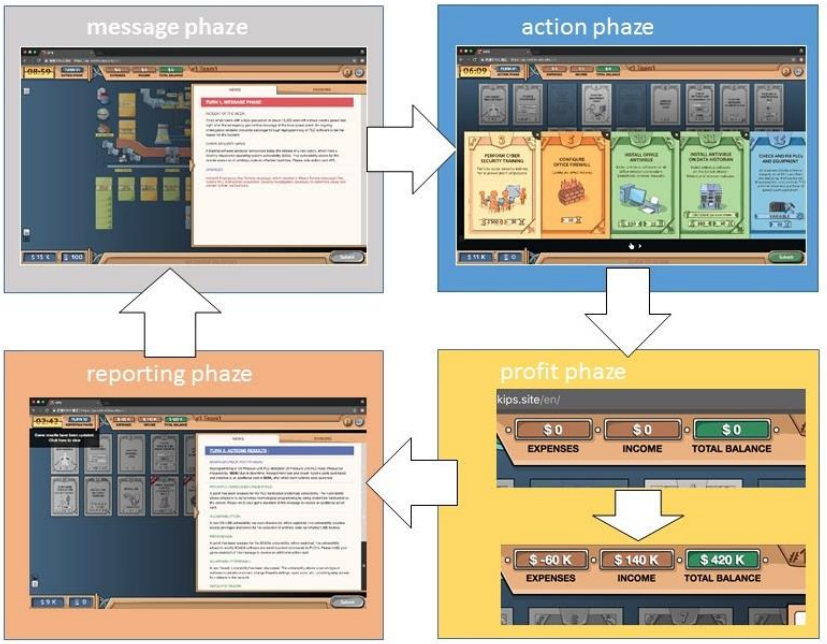


Figure 1 Game console of KIPS

At the end of the game, the moderator shows the total revenue and budget left after the five game cycles. In addition, bonuses are added to the revenue depending on the actions taken. The total revenue and remaining budget can be used to evaluate how security countermeasures contribute to the company's performance. Figure 2 shows the relationships among the KIPS stakeholders.

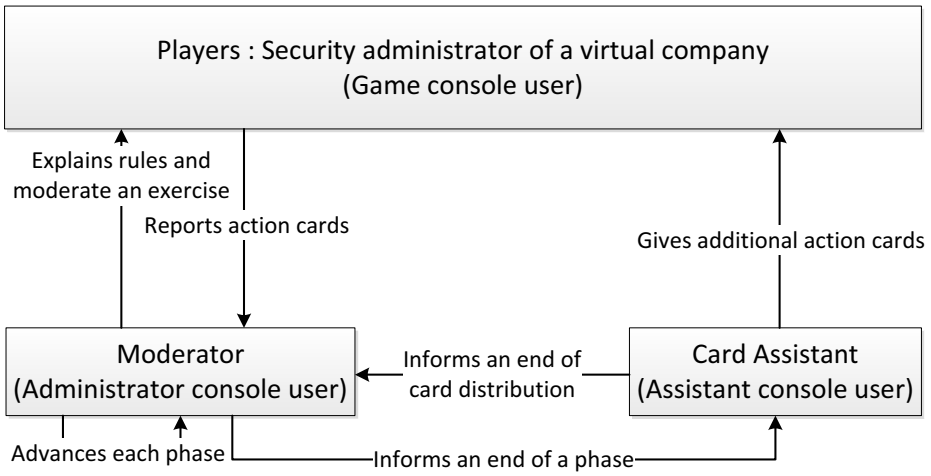


Figure 2 Relationships among KIPS stakeholders

3. Structure of Proposed Cyber Incident Exercise

3.1. Inter-organization cooperation

KIPS was designed to aware importance of inter-organizational incident response through game simulation. KIPS participants play the role of a security administrator. However, compared to real CI companies, an incident response is performed by several departments because both business and safety objectives should be considered simultaneously relative to a cyber incident. However, these objectives sometimes have a low affinity of a response due to differences among the policies of different departments. Therefore, we incorporate a cooperative inter-organization perspective into KIPS.

Therefore, we consider following mechanisms to design KIPS from the perspective of inter-organization cooperation.

- Separate one team into several groups

  An information gap is created by dividing a team into several groups. This information gap results in more complex decision-making scenarios. A group may communicate with other groups to acquire a group's unique information. Then, players should consider the nature of the current situation and what information is required for the given situation.

- Observe player decision making

  A mechanism to evaluate non-technical skills is required, and the decision-making process should be observable.

3.2. Proposed Exercise

To create the information gap within a team, players form two groups, i.e., a plant administrator group and a headquarters administrator group. The former is responsible for maintaining the safety and security of the plants. The objective of the plant administrator group is to maintain stable plant operations through five turns regardless of the nature of the cyber incident. On the other hand, the headquarters administrator

group is responsible for the overall network security, the company's budget, and its profit. The objective of the headquarters administrator group is to maximize revenue. Here action cards are distributed to the groups based on their role. One group does not initially know the information about the other group's action cards. Then, both groups discuss their actions through a chat system. The chat system enables us to observe the decision-making process because it records the communication.

In the proposed exercise, the chat system is used by both the players and facilitator. The facilitator provides information about the message phase with the plant and headquarter administrator groups at the start of the action phase. Each group receives only the information related to their responsibility; however, the players can obtain information from each other using the chat system. When players determine the action cards they will play, the headquarters administrator group notifies the facilitator of the cards' IDs. After the facilitator enters the selected action cards into the game console, the moderator uses the administrator console to proceed to the revenue phase. The administrator console shows the temporary revenue and budget available after the revenue phase. Then, the facilitator checks the result of each team on the game console and sends the results to each group. The card assistant then gives an additional action card to an applicable group that chose an action card which leads new event in the report phase. The moderator then oversees the next message phase and cycles the above process five times. Figure 3 shows the relationships among the stakeholders in the proposed exercise.

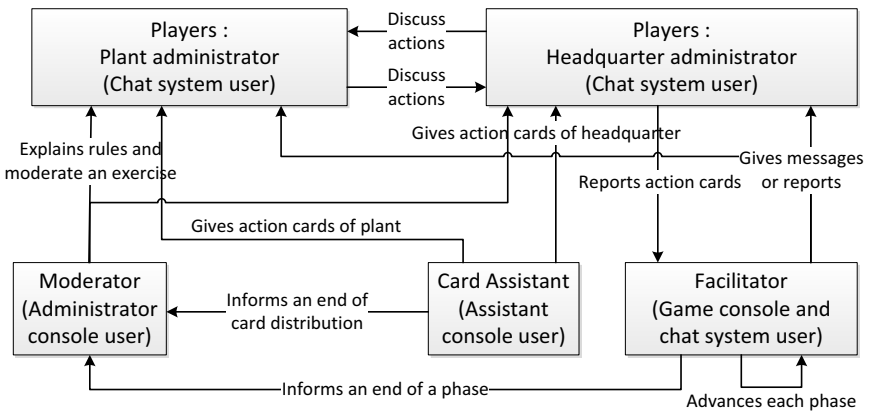


Figure 3 Relationships among stakeholders in proposed exercise

4. Implementation and Trial

4.1. Prototype Implementation

The water plant scenario was used for a prototype implementation. Here, 12 action cards were assigned as headquarters administrator cards and 18 action cards were assigned to the plant administrators.

Slack [7] was used as the chat system for the proposed exercise. Slack records user messages and can create channels for individual communication. Here, channel 1 was

between the headquarters administrators and the facilitator, channel 2 was between the plant administrators and the facilitator, and channel 3 was between the headquarters administrators and the plant administrators. Note that the facilitator could observe channel 3 to understand the teams' situations.

4.2. Trial

In September 2017, the proposed exercise was performed with 42 participants involved in CI companies. Seven teams were organized in this trial. Facilitator provide information with head quarter group and site group. Participants try a prototype game and evaluate if the game is good to make aware an importance of communication skill of incident response by a survey. As a result, 97% of the participants desired implementation of the proposed exercise at their company, which shows that this exercise is an effective evaluation tool.

5. Concluding Remarks

In this paper, we have proposed a new cyber incident exercise method that considers the complexity of decision making and communication skills. KIPS is an effective training tool for security awareness at CI companies; thus, we redesigned KIPS as an inter-organization exercise by implementing a chat system. An initial trial of the proposed exercise satisfied many participants who work at CI companies. In future, we plan to analyze the chat log of players that earned high revenue to determine effective communication processes.

Acknowledgement

Masato Matsuoka of Kaspersky Lab made significant contributions to our research. This research would not have been possible without his technical support and advice. We thank him for providing permission to use KIPS for the proposed exercise.

References

[1] K. Zetter, 2011, An Unprecedented Look at Stuxnet, the World's First Digital Weapon, https://www.wired.com/2014/11/countdown-to-zero-day-stuxnet/

[2] R. Lee, M. Assante, and T. Conway, 2014, Dec 30. German Steel Mill Cyber Attack, SANS ICS

[3] S. Smith, 2015 May 12, Cybercrime will cost business over 2 Trillion,

https://www.juniperresearch.com/press/press-releases/cybercrime-cost-businesses-over-2trillion

[4] P. Cichonski, T. Millar, T. Grance and K. Scarfone, 2012, Computer Security Incident Handling Guide, NIST, p16

[5] T. Aoyama, H. Naruoka, I. Koshijima and K. Watanabe, 2015, How management goes wrong?–The human factor lessons learned from a cyber incident handling exercise, 6th International Conference on Applied Human Factors and Ergonomics and Affiliated Conferences

[6] Kaspersky Lab. 2017, Kaspersky Security Awareness Building a safe corporate cyber-environment with gamified training, https://www.kaspersky.com/enterprise-security/security-awareness

[7] Slack, 2017, https://slack.com/

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50270-6

Supporting the use of PSE computational tools across a chemical engineering program

Mazaher Molaei Chalchooghi and Eva Sorensen*

Department of Chemical Engineering, UCL, Torrington Place, London, WC1E 7JE, UK
**e.sorensen@ucl.ac.uk*

Abstract

The use of Process Systems Engineering (PSE) computational tools across a chemical engineering curriculum is now an expectation rather than an exception, but may be challenging to deliver in a consistent and meaningful manner across an entire program. The Department of Chemical Engineering at UCL offers a variety of computational tools across our undergraduate programs, including MATLAB, GAMS, gPROMS ModelBuilder and ASPEN Plus. All modules across the program are expected to provide at least parts of the assignments in the form of PSE problems requiring the use of computation tools, often formulated in such a way that the students have a choice of tool to use. Our extensive use of computational tools is made possible by a dedicated faculty member whose sole responsibility is the development and delivery of regular lecture and tutorial material, support with assignments, as well as the development and maintenance of e-learning resources. In addition, we also have dedicated teaching assistants for each module who have domain expertise in the relevant tools, and who assist with tutorials and marking of assignments. Student feedback has shown that students highly value the extensive training in these commonly used tools, and that their leaning experience is made relevant by the application of challenging PSE problems.

Keywords:
Engineering education, process systems engineering, computational tools, e-learning

1. Introduction

Computational tools are now used extensively in most, if not all, chemical engineering programs. Most core modules or courses require some use of computational tools, from solving simple algebraic equations or complex differential equations systems, to plant design, control or optimization. Depending on the type of problem, and the licenses held by the department, the student can choose the most suitable tool from a variety of available commercial packages such as Excel, MATLAB, C++, Python, gPROMS, ChemCAD, ASPEN Plus *etc*.

The Department of Chemical Engineering at UCL offers both Bachelor and Masters programs in Chemical Engineering through its Integrated Engineering Programme (IEP) (Sorensen, 2016), with around 120-150 students per cohort. The IEP is a Faculty-wide, multidisciplinary program which combine core disciplinary technical knowledge with interdisciplinary and/or research-based projects with strong emphasis on professional skills and academic learning connected with workplace learning. The program enables students to understand the fundamentals of their discipline, to practice the application of their core technical knowledge and to apply this to current complex global challenges such as energy, health *etc*. This focus on problem and/or project-based learning starts in Year 1 and continues throughout the program. For each project, the students will be

working in teams of 6-8 students in solving an open-ended and complex problem. In Year 1, students will consider two five-week Engineering Challenges, which are considered alongside their other modules. Later in the year, and in Year 2, we suspend all other teaching for a week and work fulltime on a specific problem that is linked to the modules the students are currently studying. In total, six such scenarios are considered. In addition to providing practical examples of real-life problems within mass & heat transfer, thermodynamics, separation processes, reaction engineering *etc*., the scenarios also help develop the students' transferable skills such as team working, presentation, technical writing and time management as well as their professional skills in terms of understanding of, for instance, ethics, safety and sustainability.

One of the main aspects of the IEP is a strong emphasis on modelling and design, with two compulsory module or classes dedicated to mathematical modelling & analysis for all Faculty of Engineering Sciences programs in Years 1 and 2. The department continues this with another module on computational modelling & analysis in addition to a number of other modules through Years 2-4 (see Table 1). With this approach, the use of computational tools is embedded within the curriculum and within the individual modules. The students learn to solve problems based on traditional approaches, such as McCabe-Thiele diagrams, but also learn how to set up and run Aspen Plus simulations of standard unit operations, and more importantly, how to critically evaluate the results based on the assumptions made in the definition of the problem.

As for most UK programs, most modules are compulsory in Years 1-3, and are taken by the entire cohort at the same time. It is therefore possible to properly plan the use of a tool to ensure the material is introduced at the right time and that there is a clear progression from one module to the next or from one year to the next.

2. Integrating computational tools in undergraduate programs

The main computational tools used within our program are: MATLAB - for general purpose computation, GAMS - for solving algebraic equation systems and optimization, gPROMS ModelBuilder - for custom modelling of steady-state or dynamic mixed algebraic and differential equations, and Aspen Plus - for process flowsheeting and plant design. Other computational tools are mainly used for research in Year 4 such as ANSYS Fluent and STAR-CCM+ for computational fluid dynamics, and DynoChem scale up for process development.

Each tool is introduced in an introductory lecture that outlines the main capabilities of the tool, its usage in the chemical industries and beyond, as well as its basic features. The lecture is followed by one or more tutorials in a large cluster room where the cohort is split into smaller groups from 30 to 70 students. The tutorials are essential to ensure the students are able to first set up a basic problem and then gradually work though a number of problems of increasing complexity. The lecture is given by a dedicated faculty member who has extensive knowledge of each tool. The same person runs the tutorials but now supported by a number of teaching assistants, usually PhD students who are using the tool in their own research.

Table 1. Chemical engineering program at UCL (modules in grey are based on modeling and make explicit use of computational tools).

	Term 1	Term 2
Year 1 (Freshman)	Engineering Challenges	Transport Phenomena I
	Design & Professional Skills I	Thermodynamics
	Mathematical Modelling & Analysis I	Physical Chemistry
	Introduction to Chemical Engineering	**Computational Modelling & Analysis**
Year 2 (Sophomore)	Design & Professional Skills II	Design & Professional Skills II
	Mathematical Modelling & Analysis II	**Mathematical Modelling & Analysis II**
	Engineering Experimentation	**Particulate Systems & Separation Processes II**
	Process Heat Transfer	Chemical Reaction Engineering I
	Separation Processes I	Minor I
Year 3 (Junior)	**Process Design Project**	**Process Design Project**
	Process Dynamics & Control	Transport Phenomena II
	Chemical Reaction Engineering II	Advanced Safety & Loss Prevention
	Minor II	Minor III
Year 4 (Senior)	**Research Project**	**Research Project**
	Process Systems Engineering & Design	Elective
	Elective	Elective
	Elective	Elective

Extensive additional resources are available as e-learning resources and cover all the aspects of modelling that the students may encounter in their homework assignments and examinations, for instance, heat exchanger networks, column scheduling, pump sizing *etc*. The resources are mainly in the form of pdf versions of Powerpoint presentations and are available on the student intranet via Moodle to all students, researchers and staff within the department. Each module also has a dedicated Moodle site, with a dedicated Question & Answer Forum where students can ask questions related to the computational tool or their assignments, and where answers are provided by the teaching faculty to the entire cohort. If additional support is needed, which is very often the case, this is arranged on an as-and-when basis in the form of helpdesk sessions in a smaller computer cluster, again by the same faculty staff. The students can sign up for these via the relevant Moodle site.

2.1. MATLAB

MATLAB (Mathworks, 2017) is introduced to all Faculty of Engineering Sciences students in the first compulsory Mathematical Modelling & Analysis module in the first term. The module is taught by a number of academics from across the Faculty, with a weekly 2 hr Faculty lecture followed by a weekly 2 hr tutorial in the home department. The focus is very much on engineering problems, which by their nature can be described by a set of algebraic and/or differential equations. The students learn how to formulate the problems as well as the mathematical concepts required in solving them.

Whilst the lecture presents examples from across all engineering disciplines, the tutorials are based on problems from within the discipline.
MATLAB is used for a number of the assignments hence support needs to be available for nearly 800 students. This is provided at Faculty level involving teaching staff, but also a large group of specially trained teaching assistants. The use of MATLAB is continued in the second year in Mathematical Modelling & Analysis II, which runs over the entire academic year with a 2 hr lecture at Faculty level one week and a 2 hr tutorial in the home department the following week. In addition to the compulsory use of MATLAB in these two modules, students may use MATLAB in other assignments throughout their program, for instance in the design of reactors in Chemical Reaction Engineering or for detailed unit design in their capstone Design Project in Year 3. Support is provided by the department in the form of helpdesk sessions for such work but only at the request from students.

2.2. GAMS

The first Mathematical Modelling & Analysis module is followed by Computation Modelling & Analysis which is taught in the department and which follows on from the first module by demonstrating how larger systems of equations can be solved by computation tools rather than by hand or using MATLAB. The module is taught in two weekly two hour lectures with tutorials in addition some weeks. GAMS (GAMS, 2017) is introduced first and is used to solve steady-state problems consisting of a number of algebraic equations. The students are given assignments which must be completed as a combination of hand-calculations, Excel spreadsheets and using GAMS. Examples are to solve the mass balances on the form $\mathbf{A}\ \mathbf{x} = \mathbf{b}$ for a three column systems used to separate benzene, styrene, toluene and xylene, or to maximize the profit for a gas phase reaction system consisting of a mixer, a reactor, two separation units and a recycle stream. As for MATLAB, students may also use GAMS whenever they prefer for other assignments throughout their program.

2.3. gPROMS ModelBuilder

The department has used gPROMS (Process Systems Enterprise, 2017) in undergraduate teaching for about 15 years. Until four years ago, gPROMS was only taught in Year 4 as part of our advanced design project. With the introduction of the Integrated Engineering Program (IEP), it was decided to use gPROMS from Year 1 and to cover increasingly complex aspects of the tool term by term and year by year. gPROMS is now introduced in the second part of Computational Modelling & Analysis and is used to solve sets of differential and algebraic systems. Again, the students are given assignments which must be completed as a combination of hand-calculations and using gPROMS. An example is to solve a set of differential equations which describes an aquatic ecosystem and to analyze how pollutants have an impact on water, plants, soil and fish.

The use of gPROMS is continued in Year 2, and is incorporated into the two modules considering separation processes. In the first module, students use the tool to investigate the dynamic performance of, for instance, a distillation column stage subject to different disturbances. The students need to simulate and explain the dynamic responses based on their fundamental knowledge of volatility and mass balances for distillation. In the second separation module, students solve a problem related to Acetylsalicylic Acid (ACA) crystallization from Ethanol solution with a continuous MSMPR process. They

are free to choose which modelling tool they prefer and many use gPROMS for this instead of MATLAB or GAMS.

In Year 3, students consider Process Dynamics and Control, and again use gPROMS to investigate the dynamic behavior of a common unit operation such as a distillation column or a reactor. In Chemical Reaction Engineering, students solve a number of reactor design problems and many use gPROMS for this.

The fourth year advanced design project, Process Systems Modelling & Design, builds on the capstone project in Year 3 and considers more advanced design issues such as dynamics, optimization and design under uncertainty. gPROMS is the main tool for this and students consider different design options for a given process, for instance styrene production, and determine the optimal design for both a reaction section and a separation section. They investigate the impact of the main design parameters, such as reactor temperature, scheduling of the distillation columns *etc.*, and they consider the open-loop performance of the plant from both a safety and a production point of view and also propose a suitable control scheme for the overall plant. Some of students will also use gPROMS in their final year research project. gPROMS is the computational tool of choice for the department due to its robust capabilities in simulating large and complex equation systems, but more importantly because it is an easy tool to introduce students to, the learning curve is not too steep and it is easy to introduce the tool in bite-size portions throughout the program.

2.4. ASPEN Plus

ASPEN Plus (Aspentech, 2017) is the other main tool used in our program. The students are introduced to the tool in Year 2 and will solve small design problems in each of the two separation modules, e.g. designing a distillation column and consider the impact of tray efficiency on the design. The tool is used extensively in the capstone design project in Year 3, and in the advanced design project in Year 4, for flowsheeting and individual unit design. Some students will also use ASPEN Plus in their research projects.

2.5. Use of computational tools in Scenarios and research-based work

In addition to the regular modules and the assignments required for these, students also make extensive use of computational tools in the six week-long scenarios which run throughout Years 1 and 2. The open-ended problem descriptions are generally written in such a way that the students have a choice of which tool to use, which better mirrors the experience they will have as practicing engineers. All Masters' students will undertake a research project in their final year, and the majority of them will use a computation tool for this. In addition to the main tools mentioned above, they may also make use of tools developed in-house or more advanced tools such as STAR_CCM+ for CFD simulations.

3. e-learning Resources

Although all our computational tools are introduced and taught in formal lectures as well as supported by tutorials, extensive e-learning resources are also provided to further support the students as all the nuances and details of a computational tool cannot be conveyed in the form of lectures/tutorials. The e-learning material has been developed mainly in-house and aims to lead the students through different examples and to teach them different aspects of the tools that have not been fully covered during the face-to-face sessions. As an example, for ASPEN Plus we do not cover the detailed use

of all the available unit operations during the lectures/tutorials, but rather we give the students a brief overview in the face-to-face sessions and leave the details to the e-leaning material. We update the e-Learning material every year based on what we experience as being difficult for the students, as well as based on direct student feedback. The development of the e-learning material has primarily been done through student summer internships which we have run for three summers with a number of undergraduate students every year. This approach has worked well, as the student interns know better than teaching staff what they themselves found difficult and where more help and support is required.

4. Discussion

The use of Process Systems Engineering (PSE) tools across a chemical engineering curriculum may be challenging to deliver in a consistent and meaningful manner across an entire program. First and foremost, faculty that are experienced in the use of the tools is essential, as is teaching assistants with a good working knowledge. Secondly, there is a balance between teaching the tool via lectures and tutorials, based on student self-study using e-learning resources and books, or as a combination. There is generally limited time available on the time table to provide in-depth training, and there is often an expectation that tools such as MATLAB or ASPEN Plus are easy to pick up and require little support, however, from a student perspective, these tools are not intuitive and are often a challenge to learn, particularly for weaker students. As a result, the use of PSE computational tools are often not considered as widely as they could, and should, be. Our decision to have a faculty member dedicated to the support of computational tools has not only provided a much better service to our students but has also reduced the workload on colleagues who now receive proper help to formulate their PSE related assignments. One final year student comments: "*Across the years I have been at UCL I have been able to see the changes that has been made, and the course has definitely improved. It is particularly noticeable when looking at … and the quality and frequency of programming tutorials*."

5. Conclusion

The use of computational tools in a chemical engineering undergraduate program at UCL has been presented. All the main tools are directly embedded into the curriculum and are supported by lectures, tutorials and e-learning resources. This is made possible by a dedicated faculty member who delivers all the training across the program, who works with colleagues in integrating the use of PSE problems and tools, and who supports the students directly in their work.

6. References

Aspentech, 2017, http://home.aspentech.com/products/engineering/aspen-plus (viewed Dec 1, 2017)

E. Sorensen, 2016, Changing the World – Educating students differently with a more scenario and problem-based engineering curriculum, The Chemical Engineer, 904, 27-31.

GAMS, 2017, https://www.gams.com/ (viewed Dec 1, 2017).

Process Systems Enterprise, 2017, https://www.psenterprise.com/products/gproms (viewed Dec 1, 2017).

Mathworks, 2017, https://www.mathworks.com/products/matlab.html (viewed Dec 1, 2017).

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50271-8

Sustainable Manufacturing Education Modules for Senior Undergraduate or Graduate Engineering Curriculum

Debalina Sengupta[a], Yinlun Huang[b], Cliff I Davidson[c], Thomas F. Edgar[d], Mario R. Eden[e], Mahmoud M. El-Halwagi[a,f*]

[a] *Gas and Fuels Research Center, Texas A&M Engineering Experiment Station (TEES), College Station, TX, USA*
[b] *Department of Chemical Engineering and Materials Science, Wayne State University, Detroit, MI, USA*
[c] *Department of Civil and Environmental Engineering, Syracuse University, Syracuse, NY, USA*
[d] *McKetta Department of Chemical Engineering, Tthe University of Texas at Austin, Austin, TX, USA*
[e] *Department of Chemical Engineering, Auburn University, Auburn, AL, USA*
[f] *Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX, USA*
**el-halwagi@tamu.edu*

Abstract

Sustainable manufacturing may be defined as the creation of manufactured products using non-polluting, energy and natural resources conserving, and economically sound and safe processes. Recently, there has been renewed interest in United States manufacturing and several industrial and governmental endeavours have been launched for sustainable manufacturing initiatives. To support such initiatives and to prepare the next generation of scientists and engineers, academic institutions have a responsibility to introduce educational programs and tools in the area of sustainable manufacturing. This work provides the engineering community with structured modules for introducing the topic of sustainable manufacturing in the curriculum. Working through an NSF funded SEES RCN program, the SMART CN project delivered 21 educational modules covering various aspects of engineering with the common theme of computer aided process design. This paper summarizes the concepts for module based learning, and a brief review of these educational modules for wider use and dissemination to the academic community.

Keywords: Sustainable engineering, Education for sustainable development, Sustainable manufacturing, Curriculum innovation, Module-based learning

1. Introduction

The Manufacturing Sector has been a key factor to influence a nation's growth. Over the past decade, United States has faced competitive trends due to global manufacturing competition. A shift towards fast implementation, frequent product transitions, just-in-time model-based manufacturing, and technical personnel shifting has been observed. Uncertain energy feedstock prices, potential environmental impacts and their associated burdens are a major source of concern for industries. The country has an ongoing effort

in restructuring their focus to revitalize the sector by investing in advanced manufacturing through several institutes operating under the umbrella of Manufacturing USA®. In academia, advanced manufacturing theory, alternative energy, nanotechnology, infrastructure resiliency, composite materials, etc. have become very active, but the connections to sustainability are not well defined. Research coordination between academic community industries are insufficient and lack depth. Thus, sustainable development of the manufacturing sector becomes an important topic for academia, industry and the government in the area of coordinated research. More importantly, it has become imperative to develop instructional material suitable for educating the next generation of engineers and scientists, while retraining the current workforce.

The term "sustainable development" was defined as "Development that meets the needs of the present without compromising the ability of future generations to meet their own needs" by the World Commission on Environment and Development's report Our Common Future[1]. One of the first efforts for education in sustainable development (ESD) was led by UNESCO[2]. New ideas in method and content were combined with existing ideas for the purpose of rethinking educational programs and systems and equip them to support unsustainable societies. ESD efforts have focused on topics like biodiversity, climate change, poverty reduction, disaster risk reduction, cultural diversity, sustainable lifestyles, gender equality, peace and human security, water, health promotion, and urbanization. Filho, et al. [3] discuss the key issues in introducing sustainable development in higher education in which were identified in Rio and the UN Decade of Education for Sustainable Development.

Perhaps the biggest challenge in implementing sustainability in education, particularly engineering education, is the ability to quantify sustainability concepts into mathematical expressions. Engineering Design courses typically act as the culmination point for various theoretical topics, and our approach to educating in sustainability focuses on quantification in this area Problems that consider the complexity, boundary, and various scales of manufacturing processes' and products' sustainability are considered for education modules in this paper.

A systematic modular structure is used to develop instruction ready material for sustainable manufacturing. This paper discusses the overview of development objectives, structure, content, and applicability of these modules. A detailed explanation of the theory and basis of the module development is given by Sengupta, et al. [4]. The modules are available from the Computer Aids for Chemical Engineering website for free download, along with a detailed paper on the needs of module based learning and its impact on higher education for teaching sustainability in manufacturing..

2. Background and Objectives

In recent times, the concept of learning and associated learning methods have gone beyond the definition of the university in terms of physical boundaries and flexible lecture class times accommodates a wider variety of students[5]. Addressing these changes, it becomes imperative to decide on the selection and construction of knowledge that needs to be represented in universities (epistemologically and from the perspective of learners). This also raises the need for the organization of this knowledge (institutionally and from the perspective of teachers) through effective ways of teaching and learning. Bridges [5] identifies five competing but eminent epistemological changes in higher education and focuses on the importance of deconstruction of the subject

where it is packaged in new ways to enable the modularization and offer flexible pattern of study. The modularization concept offers a three way impact: the deconstruction of traditional subjects as the epistemological units of study; the deconstruction of a time bound undergraduate course as the chronological unit of study; and the deconstruction of the single university as the topographical location of the unit of study.

The modularization of the units of learning in higher education enables the expression of two different educational principles. Firstly, a student-centred curriculum can be offered where a student can assemble a degree program according to their personal interests and aspirations. Secondly, it can satisfy the social expectation that university programs might serve more directly the needs of employers, where an employer guides the creation of modules best suited for their needs. Thus, the modularization concept has gained considerable impetus owing to the commitment to flexibility and student choice. A key consideration in involving the web helps to supplement the classroom experience, and provide unique learning opportunities only available through these online learning methods. These new trends in education have capability to reach a very wide audience by eliminating boundaries of classroom, geographical location, and even time. This brings the concept of flexibility in teaching and learning methods[6].

3. Module Structure and Use

Module based learning requires a focus on the quality of learning that can be achieved by students using them, and not on the specific content of the module and how they are intended to be taught. In the process of devising a module, the key is to forge educationally sound and logical links between learner needs, aims, learning outcomes, resources, learning and teaching strategies, assessment criteria and evaluation[7]. Svanström, et al. [8] discussed the key learning outcomes in education of sustainable development, which also apply to sustainable manufacturing.

The SMCS modules are designed for introducing basic and core concepts related to sustainable manufacturing, combine the concepts from the core modules to produce a methodology for sustainability analysis of manufacturing processes, and dedicated tools that incorporate some of these methods. The modules have flexibility to be combined with other courses and certainly carries the uniqueness of the module developers, who mostly were faculty consultants.

The modules were developed to have an overview of the module contents, a slide set for intended delivery or self- learning of the material, and case studies that provide the computer aided design component for the topics. A text document is provided which is intended to be a guide for both the instructor and the learner. It is assumed that the instructor will be well equipped with the traditional engineering concepts, but not conversant or expert in the sustainability aspects of a particular module. This document is aimed at connecting traditional engineering knowledge to the sustainability concepts provided by the module, providing the instructor with a starting point for engaging in a new topic related to sustainable manufacturing. The presentation slides for in-class delivery are intended to provide the instructor with a perspective of presenting the background information and the case studies with minimal modification. These slides can also be used as-is for self-learning. Finally, the case study files enable the instructor to deliver questionnaires or homework problems and the supporting design files or tools help to delve deeper into the topic for complete understanding of the module.

There are several possible ways to use these modules. Three such uses for education related to sustainable manufacturing have been envisioned. The first use is where these modules complement existing core courses in undergraduate or graduate studies. The second is where a dedicated course is proposed for introducing sustainable manufacturing, and can be complementary to engineering design. The third type is to cater to the needs of sustainability inclusion in a specific manufacturing sector of the industry.

4. Module Content and Description

Table 1 summarizes the key information for the modules available in the website till date. The modules reflect the outcomes of the SMART Roadmap Development Workshop in Cincinnati. These modules are broadly divided into Learning Tools, Case Studies and Methods, and Process Examples. A brief description of the modules is provided in the Table 2. There are 11 modules in this table, but 21 have been developed and will be made available soon.

Table 1 Summary of the SMART Education Modules

Module Name	Module Developer	Module Content
Learning Tools		
Early Stage Sustainability Analysis Tool - EarlySim	Akshay Patel, SustAnalyze and Utrecht University	Provides an early stage chemical process assessment tool. The tool can be used for sustainability assessment in the areas of economic constraints, environmental impact of raw materials, process costs and environmental impact, EHS index, and Risk aspects.
Case Studies and Methods		
Assessment of the Presidential Green Chemistry Award Winners using Green Chemistry Metrics	Christopher L. Kitchens, Clemson University	Evaluates the work that has received the Presidential Green Chemistry Challenge Award using green chemistry metrics, principles, and design strategies.
Life Cycle Assessment for Sustainable Manufacturing	Debalina Sengupta, Texas A&M University	Provides overview of life cycle assessment methodology as outlined in the ISO standards, Emphasize the utility for the LCA methods for manufacturing sustainability
Sustainability Metrics and Sustainability Footprint Method	Debalina Sengupta, Texas A&M University	Provides overview of methods to compute sustainability metrics. It also gives a method compute overall sustainability by aggregating metrics.
Green Chemistry to Manufacture Specialty Chemicals from Renewable Resources	Jeffrey R. Seay, University of Kentucky	Introduces the concept of green chemistry for green design of processes, gives three methods for assessing "greener" processes: The WAR Algorithm for potential environmental impact (PEI) of a process, Life Cycle Assessment, and inherently safe process design.
Sustainability Root Cause Analysis (SRCA)	Helen H. Lou, Lamar University	Demonstrates Sustainability Root Cause Analysis (SRCA) as a tool to determine the bottlenecks for a system's progress towards sustainability. The framework is built on the combination of Pareto chart and the Fishbone diagram, in conjunction with a set of

		sustainability metrics (economics, environmental and safety).
Optimization and Uncertainty for Green Design and Industrial Symbiosis	Urmila Diwekar, VRI and Yogendra Shastri, IIT Bombay, India	Demonstrates the use of optimization methods for sustainable manufacturing. Incorporates systems theory as a valuable tool to enable the integration of multi-scale, multi-disciplinary components using an informational and computational platform.
Atomic Layer Deposition Nano-Manufacturing Technology	Chris Yuan, University of Wisconsin, Milwaukee	Module on atomic layer deposition (ALD) focuses on the study of energy usage and exergy efficiency, simulate reactions inside ALD system and analyze ALD deposition and emissions.
Process Examples		
Optimal Design and Operation of Reverse Osmosis Desalination	Mingheng Li, California State Polytechnic	Specific energy consumption (SEC) in reverse osmosis (RO) desalination is considered for sustainability of the water treatment process. The module focuses on case studies that help in the optimal design for RO with the sustainability concerns in energy consumption addressed.
Sustainable Additive Manufacturing	Karl Haapala/Oregon State University	Provides a module that covers additive manufacturing as a means for sustainable manufacturing. This module explains the basics of additive manufacturing, and explores energy analysis as a metric to establish the benefits of AM.
Sustainable Mitigation of Carbon Dioxide to Chemicals	Debalina Sengupta, Texas A&M University and Sherif Khalifa, Drexel University	This module explores CO_2 mitigation strategies through the utilization of CO_2 into high value chemicals. A superstructure optimization model is formulated and solved for different scenarios.

5. Module Dissemination and Assessment Plan

These modules have been made available through the CACHE Corporation[9], and all future modules will also be disseminated through this website and the website of the Center for Sustainable Engineering[10]. These modules have undergone rigorous testing and review from the SMART team members, and are continuously updated based on reviews and responses from users. CACHE is a not-for-profit organization consisting of 21 universities and 7 industry members. Its purpose is to promote cooperation among universities, industry and government in the development and distribution of computer-related and/or technology-based educational aids for the chemical engineering profession. CACHE also operates short courses and research conferences on subjects of interest to the Process Systems Engineering community.

The Accreditation Board for Engineering and Technology requires well defined Program Educational Objectives, Student Outcomes, Assessment, and Evaluation planning of engineering programs[11]. The Sustainable Manufacturing Modules were described according to the terminology suggested by ABET. For the assessment of the modules, an assessment rubric was created to evaluate each module. Through a self-study exercise for graduate students from various institutions, this rubric was used to evaluate the modules on how effective were the modules in achieving the objectives and student outcomes; whether the modules were suitable as self-study material; and what basic information needed to be added to the modules for making them self-sufficient.

The modules were presented at AIChE Conferences, used in a two day tutorial for a larger group of students through lecture type delivery, and in the ASEE Summer School for Faculty to introduce them to the modules. A positive feedback was received for the modules from all these entities. Based on these assessment methods, the modules were considered effective to be made public.

6. Conclusions

The need for integrating a complex topic like sustainability and the need for introduction and assimilation in current engineering courses led to the Module based learning methods explained in this paper. Sustainable manufacturing case study modules having a computer aided design component were developed to highlight the various aspects for sustainable manufacturing as a part of the Sustainable Manufacturing Advances in Research and Technology Coordination Network. The specific module structures have been discussed and the suggested application of these modules has been proposed. Finally, the module content and dissemination plan have been discussed in this paper. The modules are self-contained sources of information that can be used by instructors and learners alike. The modules are freely available for download and use from the CACHE corporation website [9]. Users are requested to provide a feedback to the SMART executive team (authors of this paper) to improve these modules.

7. Acknowledgment

This work was supported in part by U.S. National Science Foundation (Grant No. 1140000).

References

1. Brundtland, G. H., *Report of the World Commission on environment and development: "our common future."*. Oxford University Press, London: United Nations, 1987.
2. Wals, A. E., *Shaping the education of tomorrow: 2012 full-length report on the UN decade of education for sustainable development*. UNESCO Paris: 2012.
3. Filho, W. L.; Manolas, E.; Pace, P., The future we want. *International Journal of Sustainability in Higher Education* **2015,** *16* (1), 112-129.
4. Sengupta, D.; Huang, Y.; Davidson, C. I.; Edgar, T. F.; Eden, M. R.; El-Halwagi, M. M., Using module-based learning methods to introduce sustainable manufacturing in engineering curriculum. *International Journal of Sustainability in Higher Education* **2017,** *18* (3), 307-328.
5. Bridges, D., Back to the Future: The higher education curriculum in the 21st century. *Cambridge Journal of Education* **2000,** *30* (1), 37-55.
6. Palmer, S., Engineering flexible teaching and learning in engineering education. *European Journal of Engineering Education* **2001,** *26* (1), 1-13.
7. Donnelly, R.; Fitzmaurice, M., Designing Modules for Learning. In *Emerging issues in the practice of University Learning and Teaching*, G. O'Neill, S. M., B. McMullin, Ed. All Ireland Society for Higher Education (AISHE): Dublin, 2005.
8. Svanström, M.; Lozano-García, F. J.; Rowe, D., Learning outcomes for sustainable development in higher education. *International Journal of Sustainability in Higher Education* **2008,** *9* (3), 339-351.
9. CACHE Computer aids in Chemical Engineering. http://cache.org/super-store (accessed 11/27/2017).
10. CSE Center for Sustainable Engineering. http://www.csengin.org/csengine/csevl/csevl.html (accessed 12/04/2017).
11. ABET Criteria for Accrediting Engineering Programs. http://www.abet.org/wp-content/uploads/2017/12/E001-18-19-EAC-Criteria-11-29-17-FINAL.pdf (accessed 12/03/2017).

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50272-X

Teaching data-analytics through process design

Fani Boukouvala*, Jianyuan Zhai, Sun Hye Kim, Farida Jariwala

Chemical & Biomolecular Engineering, Georgia Institute of Technology, 311Ferst Dr., Atlanta, GA, 30332, USA
**fani.boukouvala@chbe.gatech.edu*

Abstract

Data-analytics is becoming a very influential tool for decision-making today both in industry and academia. The incorporation of data-driven concepts in the core chemical engineering curriculum would be very beneficial to our graduates, making them competitive in today's market. The main objective of this work is to develop tools and case studies that: (a) increase students' understanding and appreciation of data for decision making, (b) introduce students to basic data analytics and machine learning methods, (c) introduce students to basic data visualization tools, and (d) exhibit that the interactions between rigorous simulations and data could lead to improved solutions. This work shows that process design is an appropriate course for the introduction of this material, since students are at a senior level and have obtained the necessary mathematical background, and the development of process flowsheets using commercial simulators creates opportunities for inexpensive data generation to create many case studies.

Keywords: data-analytics, data-driven modelling, process design, simulation, optimization, machine-learning.

1. Introduction

Undoubtedly, we have entered an era, during which data are everywhere and the power and capabilities of data-driven decision-making is being increasingly adopted in many fields of engineering (Beck et al., 2016). The terms "data science", "data-analytics" and "Big-Data" have started to emerge in an increasing number of research papers and conference presentations from both academics and industry in chemical engineering (Venkatasubramanian, 2009, Qin, 2014). Data-scientists have the ability to combine statistics, mathematics and computer programming to capture data trends and find emerging patterns, via data pre-processing, visualization and analysis. Data-analytics is the science of applying algorithmic methods to raw data, in order to draw useful conclusions, which oftentimes may verify or disprove an existing insight. In this work, we aim to show how to incorporate data-analytics into the course of process design, since our main aim is to teach our students how to use information in the form of input-output data to validate existing insight, or gain new insight that would be hard to recognize without data-analytics. Due to the fact that data-sets are often extremely large, with high dimensionality and numerous measurements from advanced sensor technology, the term "Big-Data" has also been predominantly used to capture the volume, variety, velocity and veracity of data and data generation (Qin, 2014). The concepts that are described in this work do not require terabytes of data (which is less likely to be the case in most chemical engineering applications), but rather aim to introduce students to systematic data-driven decision-making tools, which can be applied to a wide range of data-set sizes.

This work is motivated by the vision that chemical engineers with a data-analytics background can make a significant impact, by knowing how to use data in tandem with their first-principle knowledge, to design better systems. A recent sample survey of data-related education showed that statistics and data-analytics do not currently have a big role in most chemical engineering programs (Braatz, 2015). This work will show examples of how process design can be one of the potential ideal courses in every undergraduate curriculum, through which one can introduce data-driven methods and tools, through simple case studies that can be formulated and solved using any simulation software, or process design project.

Three modules for teaching different aspects of data-analysis have been developed and are introduced in this paper. The first module aims to introduce students to multivariate analysis techniques for analysing noisy, unstructured data, and for identifying true dimensionality of data sets. The second module introduces students to the development of reduced-order regression models, using data from their simulations, in order to represent correlations of multiple inputs to outputs of interest. The final module allows students to experiment with data-driven optimization concepts, ranging from simple grid-search, to more sophisticated trust-region optimization. The modules are developed in Matlab and they are not reliant on a specific simulation package, as long as the simulation can communicate with Matlab via generation of text files. These modules are currently being taught and evaluated at the School of Chemical & Biomolecular Engineering at Georgia Institute of Technology, in the course of Process Design and Economics I.

2. Integration of Data-Analytics through the course of Process Design

Despite the acknowledged need to increase our students' understanding of statistics, data-handling, and data-analytics, it is often hard to create new courses that can be incorporated within the core curriculum, due to total credit-hour limitations. As a result, a very effective and efficient approach is to incorporate data-analytics within existing courses throughout the undergraduate curriculum. There are many opportunities for this seamless integration, since the concepts of empirical models, analytical methods and parameter estimation are all embedded within our core undergraduate courses already.

One very timely opportunity for integration of instruction of data-analytics is through the course of process design and economics. During this course, students are introduced to the complexity of decision-making when open-ended design problems need to be solved, using their accumulated knowledge in chemical engineering fundamentals and computer simulations. This course has the objective of teaching students how to develop rigorous simulations of their conceptual designs, generate alternatives, and finally, follow a trial-and-error approach to optimize their design with objectives that range from energy minimization to maximization of economic profitability measures, such as net present value (Seider et al., 2015). The most commonly used flowsheet simulators in process design course are ASPEN, gPROMS, ProMax, HYSYS, COCO, Unisim, and more (Adams, 2018).

The three modules that have been implemented within this course, are appropriately designed to follow the structure of the course, and supplement the learning objectives of the existing syllabus (Figure 1). In the first component, a data-analytics approach is introduced that helps students identify the true dimensionality of a data-set and to visualize the correlations in the data (Module 1). In the second case study, students are introduced to best practices and state-of-the-art methods in machine learning for fitting

multivariate regression models (Module 2) to represent a complex output as a function of a set of inputs. Finally, students are introduced to basic optimization formulations and solvers, and are asked to build their own adaptive optimization solver to identify optimal solutions of their rigorous design simulation (Module 3). Each module is described in more detail in the next subsections.

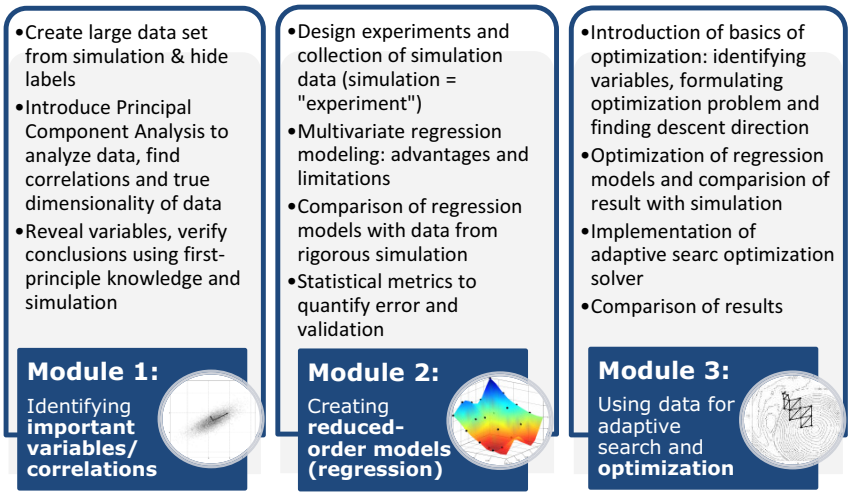


Figure 1. Summary of three modules for data-analytics in process design

2.1. Multivariate analysis for identification of correlations and dimensionality reduction

One of the first challenges that students face when they have completed a rigorous simulation of a fully integrated process, containing multiple units (i.e., reactors, separators, heat exchangers, pumps, compressors, etc.) is identifying which variables to tune, in order to achieve an improved operation. Teaching process design is challenging due to the fact that students may fall in the trap of blindly relying on the simulator, without attempting to use their first-principle insight for making informed perturbations to the system. The former is our first goal, and for this reason students are first expected to explore their design via trial-and-error and sensitivity analysis tools, in order to gain better understanding of the performance of the design, as well as to learn how to apply chemical engineering fundamentals to explain perturbation effects to the system.

However, even when students reach the point of understanding the effect of a single (or a few) key variables within a single unit to the outputs of that unit, it becomes more challenging to understand and quantify how changes in several variables across units, may propagate and affect the operation of other units within a fully integrated design with recycle streams. In addition to the above, it is true that a large fraction of our students in the future, may be faced with the challenge of trying to understand, analyse and predict the performance of a system, by using large historical data-sets.

Through the steps of Module 1 (Figure 2), we take advantage of the existence of large integrated simulations, for the given project statement in order to create a large data set, which aims to mimic a historical data set of a process design. More specifically, we select a large number (M_1) of manipulated design variables, feed conditions, and operating conditions from every unit in the design and identify a reasonable range of operation of these variables. In addition to the set of manipulated variables, we also select a set of

monitored variables (M_2) that may or may not be highly sensitive to the manipulated variables. We then randomly perturb the M_1 manipulated variables, in order to create a set of N simulated case studies of the design, and generate a data matrix $X = [N \times (M_1 + M_2)]$. Various sampling techniques can be used to generate the samples, such as Monte Carlo sampling or space-filling designs (i.e., Latin Hypercube sampling, orthogonal arrays) (Wang et al., 2006). Some of the columns (dimensions) of this data set are correlated (by design), but this is not revealed to the students. The module also has an option of adding noise to the data set X, in order to better mimic noisy historical data sets. After students are introduced to Principal Component Analysis (PCA) in a lecture (Jolliffe, 2002), they are provided with the data set and are asked to use PCA to identify correlated columns, as well as the true dimensionality of the data which they feel truly describes the variability in the data. Students are asked to complete exercises in which they generate score plots and are asked to identify correlated dimensions. Once these exercises are completed, the real labels of the data set are revealed, and students find that their conclusions based solely on data, often agree with their intuition. Through this exercise, it is truly fascinating to observe that his reinforcement of intuition helps students trust the data-driven approach. Other times, the data analysis points students towards a correlation that they did not identify through their trial-and-error approach, which they are always asked to explain using their engineering insight. Despite the fact that our aim is to introduce data-analytics, the most powerful learning outcome of this module is the realization that both rigorous simulation based on chemical engineering fundamentals, coupled with data-analytics can collectively lead to better analysis and understanding of complex systems.

Step 1: Indentify set of manipulated variables (M_1) (feed conditions, operating conditions, design parameters) and their ranges.
Step 2: Identify set of monitored variables (M_2) (outlet conditions, conversions, energy requirements, etc.) across units.
Step 3: Randomply perturb manipulated variables within their ranges, and perform $(N > M_1 + M_2)$ simulations, generating data set $X = [N \times (M_1 + M_2)]$.
Step 4 (optional): Add random noise to data set $X = X + \varepsilon$.
Step 5: Provide students with X, omitting labels/descriptions of data. Assignement includes PCA analysis of data, to identify correlations and reduce dimensionality.
Step 6: Collect results, reveal data labels and compare data-driven conclusions to engineering insight. [Questions posed: Did you verify your insight? Did you learn something new from data?]

Figure 2. Outline of steps of Module 1

2.2. Regression and interpolation using simulation data

Despite the fact that chemical engineering students oftentimes use empirical correlations, their understanding of regression does not often go beyond linear or polynomial regression in one dimension. Throughout the progression of a process design course, students can often benefit from using multidimensional regression models for analysis, visualization and optimization. One of the most commonly encountered example of the need to use regression functions (or else parametric models, surrogate models or metamodels) to represent phenomena is to allow for the incorporation of a rigorous model from a flowsheet simulator within a synthesis optimization formulation (Caballero et al., 2008). Through this module, students are taught about computer experimental design (Sacks et al., 1989), and subsequently they learn how to use the collected input-output data to minimize the least-squares error of prediction between the data and a variety of different regression functions: linear, quadratic, polynomial (Bhosekar and Ierapetritou, 2018) and more advanced state-of-the-art machine-learning inspired functions, such as

support vector regression (Smola et al., 1996). The importance of over-fitting and validation is conveyed to the students through a series of examples and comparisons of fitted models. The final assignment for this module is the collection of data from the actual simulation using a designed experiment; the fitting of a variety of regression models (Figure 3); the selection of the best model using statistical measures; plotting of the selected function in surface plots for visualization; and finally the validation of these fitted models for a set of data points that was not used to train them.

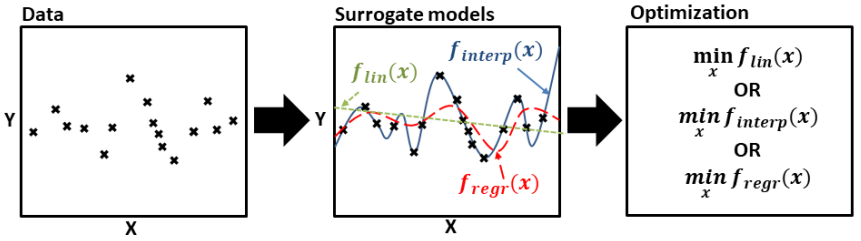


Figure 3. Concept of fitting parametric models (regression, interpolation) to data and finally using these functions for optimization.

2.3. Data-driven Optimization for flowsheet optimization

During the final lectures of the process design course, students learn about optimization fundamentals, such as identifying the degrees of freedom, formulating the objective function, formulating the constraints, linear optimization vs. nonlinear optimization, use and complexity of discrete variables to represent logical constraints, and formulation of synthesis problems. Students quickly identify the gap between being able to solve rigorous optimization problems, and the use of rigorous simulations. Several simulation packages contain built-in optimization solvers, through which a flowsheet can be optimized directly (i.e., ASPEN optimization tool). However, the goal of this work is to develop modules that are independent of a simulation platform, and also to teach the students the fundamentals of optimization by allowing them to develop their own simplified solver. During the first stage of this module, students are asked to use their results from Module 1 and Module 2 for the purpose of optimization. From Module 1, they have identified the most important and influential variables, while in Module 2 they have developed regression models. First, students use their fitted regression model to perform local optimization in Matlab (fmincon) (Figure 3). Subsequently, students have to simulate the identified optimum within their simulation software, in order to compare the accuracy of the result. Oftentimes, the predicted optimum is not identical to the optimal value obtained by the rigorous simulation. Thus, students quickly comprehend the concept of adaptively sampling their rigorous simulator in 'promising' locations, in an effort to identify better solutions. This teaching material starts with the simple concept of grid-search, and ends with the teaching of a more advanced trust-region-search algorithm towards descent directions (Boukouvala et al., 2016, Kolda et al., 2003). The students are asked to implement this algorithm, which iteratively exchanges data with their simulation in order to search for optimal locations. The advantage of implementing a customized solver, is the ability to account for simulation instances that fail to converge due to numerical issues, which may hinder the applicability of certain solvers. For simplicity, this analysis is performed for a relatively low number of optimization variables (~2-3), which have been previously identified to be significant. Finally, the students compare their results with in-built optimization solvers, if those are available.

Regardless of the result, through this exercise students truly understand fundamental optimization concepts, as well as the difference between local and global optimization

3. Conclusions

Three case studies have been presented, through which various data-driven analysis, modelling and optimization concepts are introduced to chemical engineering seniors, through the course of process design. The goal of these modules is to prepare graduates to enter the workforce, better equipped and competitive in today's Big-Data era, by introducing them to data-analytics tools that have gained significant applicability in industry and academia. The goal of the modules is to teach students how to use multivariate input-output data to improve their decision-making process, while never forgetting their chemical engineering fundamentals. These modules will continue to be tested and enhanced through additional case studies in the future, and will soon be made available to the community.

References

T. Adams II, 2018, Learn ASPEN Plus in 24 hours, McGrawHill.

D.A.C. Beck, J. M. Carothers, V. R. Subramanian, J. Pfaendtner, 2016, Data Science: Accelerating Innovation and Discovery in Chemical Engineering, AIChE J, 62(5), p. 1402-1416.

A. Bhosekar and M. Ierapetritou, 2018, Advances in surrogate based modeling, feasibility analysis, and optimization: A review. Computers & Chemical Engineering, 108, p. 250-267.

F. Boukouvala, R. Misener, and C.A. Floudas, 2016, Global optimization advances in Mixed-Integer Nonlinear Programming, MINLP, and Constrained Derivative-Free Optimization, CDFO. European Journal of Operational Research, 252(3), p. 701-727.

R. Braatz, 2015, Sampling of Data Education in ChE Curricula. Available from: cache.org/files/Sampling-of-Data-Education-in-ChE-V3.0.pptx

J. A. Caballero, J.A. and I.E. Grossmann, 2008, An algorithm for the use of surrogate models in modular flowsheet optimization. Aiche Journal, 54(10), p. 2633-2650.

I.T. Jolliffe, 2002, Principal Component Analysis, Springer Series in Statistics, 2nd Edition, Springer, NY.

T. G. Kolda, R.M. Lewis, and V. Torczon, 2003, Optimization by direct search: New perspectives on some classical and modern methods. Siam Review, 45(3), p. 385-482.

S. J. Qin, 2014, Process data analytics in the era of big data. AIChE Journal, 60(9), p. 3092-3100.

J. Sacks, W.J. Welch, T.J. Mitchell, H.P. Wynn, 1989, Design and analysis of computer experiments, Statistical science, 4(4), p. 409-423

W. Seider, D.R. Lewin, S. Widagdo, R. Gani and K. M. Ng, 2016, Product and Process Design Principles: Synthesis, Analysis and Evaluation, 4th edition, Wiley.

A.J. Smola and B. Schölkopf, 2004, A tutorial on support vector regression. Statistics and Computing, 14(3), p. 199-222.

V. Venkatasubramanian, 2009, DROWNING IN DATA: Informatics and modeling challenges in a data-rich networked world. AIChE Journal, 55(1), p. 2-8.

G. Wang, S.S. Shan, 2006, Review of Metamodeling Techniques in Support of Engineering Design Optimization. ASME. J Mech. Des, 129(4), p. 370-380.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50273-1

Development of the Cyber Exercise for Critical Infrastructures Focusing on Inter-Organization Communication

Hidekazu Hirai[a*], Yuma Takayama[b], Tomomi Aoyama[b], Yoshihiro Hashimoto[b], Ichiro Koshijima[b]

[a,b]*Nagoya Institute of Technology, Department of Industrial Management Engineering, Showa-Ku, Gokiso-Cho, Nagoya, 4668555 Aichi, Japan*
cjr17067@nitech.jp

Abstract

In recent years, the risk that cyber incidents occur is increasing in Industrial Control System (ICS) which is the core function of Critical Infrastructures (CIs). As a result, companies holding CIs must improve the organization's incident response ability to cyber incidents. Because cyber incidents in ICS cause not only problems to cybersecurity but also problems to plant's safety and company's business, the company-wide response including not only the IT department but also the ICS department and management layer is essential. To smoothly promote this company-wide response, Incident Commanders who adjust between departments are essential. However, many companies holding CIs do not adequately educate Incident Commanders. Therefore, in this research, the authors developed the cyber exercise aimed at educating Incident Commanders.

Keywords: Cyber Security, Critical Infrastructure, Cyber Exercise, Incident Commander.

1. Research Background

In recent years, Cyberattacks to Industrial Control System (ICS), the core functions of Critical Infrastructures (CIs) such as power plants and chemical plants has been increasing year by year. In 2010, the uranium enrichment centrifuge fell into dysfunction at Iran's nuclear facility (Nobutaka Oguma, 2011) and the blast furnace was damaged in the German steelworks plant in 2014 (Motohiro Tsuchiya, 2015). Also, two large power outages occurred in Ukraine in 2015 and 2016 (Toshio Miyachi, 2016).

In Japan, the Tokyo Olympic and Paralympic Games are scheduled to be held in 2020. It is said that the 2012 London Olympic Games could have resulted in the loss of lighting at the opening ceremony site due to cyberattacks (BBC NEWS, 2013). For this reason, it is urgent for companies holding CIs in Japan to secure cybersecurity in ICS for the Tokyo Olympic and Paralympic Games (Nihon Keizai Shimbun, 2017).

Cyberattacks against ICS affect not only cyberspace but also the physical world. As a result, there are concerns about destruction of the equipment itself, leakage of dangerous substances, and explosion accidents. Also, in recent years as the sophistication of cyberattacks methods increases, preparing cybersecurity measures on the premise of receiving cyberattacks is regarded as important. One of the cybersecurity measures is a rapid and appropriate response by the organization to mitigate damage and bring the situation under control after cyberattacks occur. Thus, it is essential to formulate a response plan in advance for events caused by cyberattack (from now on referred to as

"cyber incidents"). For this reason, the authors have developed the cyber exercise to assist companies holding CIs in formulating their cyber incident response plans (Hidekazu Hirai, 2017).

2. Problem Setting

Even if companies holding CIs prepare the cyber incident response plan, it is not easy for companies holding CIs to implement cyber incident response plan that does not have experience. Cyber incidents in ICS cause not only problems with cybersecurity but also problems with plant's safety and company's business. Therefore, against cyber incidents in ICS, multiple departments such as IT department and ICS department must respond in parallel. However, when multiple departments respond in parallel, there may be errors in decision-making due to lack of information sharing among departments and conflicts due to differences in response objectives between departments. Therefore, under cyber incidents in ICS, Incident Commanders (from now on referred to as the "commanders") to coordinate and make decisions between departments are especially important. Many companies have Computer Security Incident Response Team (CSIRT) as commanders against cyber incidents in IT. However, few companies have specialized response teams against cyber incidents in ICS. Therefore, if cyber incidents occur in ICS, the leaders of divisions or departments in the existing organization may have to become commanders. However, it is challenging for anyone who has never experienced cyber incidents in ICS to properly perform the commander role. For this reason, companies holding CIs need to educate position who have to become commander under cyber incidents in ICS. Therefore, in this research, the authors will develop training for Incident Commander's educational support for companies holding CIs. However, in cyber incidents in ICS, abnormal situations in each company is unique to each company occurs. The role that commanders should be in cyber incidents varies widely from company to company, and therefore, the content of their training needs to be customized for each company. However, it is inefficient for us to develop customized training for each company. For this reason, the training is developed in this research must include the mechanism that each company can customize on its own. Thus, cyber exercise is developed in this research is training for commanders, including the mechanism that each company can customize in-house.

3. Cyber Exercise For Training Incident Commanders

3.1. Form of Training

To customize the training for each company, the training components need to be divided into parts that are common to all critical infrastructure companies and parts that must depend on the individual characteristics of each company. Customizing the training is to change the part of the training that depends on the individual nature of each company, which is a component of the training, into one that is unique to each company. To do this, training in the form of computer games is useful. Computer games consist of programs and data embedded in programs. For this reason, in computer games, the programs are a common part that can be used by all critical infrastructure companies, and the data can be viewed as a part that depends on the individuality of each company. Therefore, incorporating company-specific information into the data creates a customized computer game for each company. Therefore, the fact that the training is in the form of computer games is effective from the viewpoint that the training can be customized for each

company, and therefore, the training be developed in this research is in the form of computer games.

3.2. Requirements for the Computer Game

The purpose of the computer game is to enable participants to learn about the role of commanders and to perform that role properly. For this purpose, simulated training methods that allow participants to experience the role of commanders are useful. For this reason, it is essential to have mechanisms that allow participants to simulate the role of commanders in the computer game. To clarify what commanders should do cyber incidents to clarify the mechanism required for the computer game. Cyber incidents in ICS cause not only cybersecurity issues but also safety and business issues simultaneously. For this reason, not only IT departments, but also ICS departments, management, and others need to take action at the same time. To facilitate this company-wide response, it is necessary for the leaders of each department to act as commanders, to coordinate among departments, and to direct the company-wide response. To do this, each commander must perform the following:

- Collection of information on cyber incidents (information on damage status of ICS, network logs, operation status, etc.).
- Sharing information with other commanders and external organizations (customers and affiliates).
- Decision making based on information obtained and task instructions to workers and other commanders based on contents of decision making.

Thus, for the participants to simulate the role of commanders, the following mechanisms need to be established in the computer game.

- Mechanism by which more than one participant can participate in the computer game as another commander (ICS Director, IT Director, Management, etc.)
- Mechanism by which participants can collect information, share information, and issue task instructions as commanders.

To realize these mechanisms, the computer game with the configuration shown in Figure 1 is proposed.

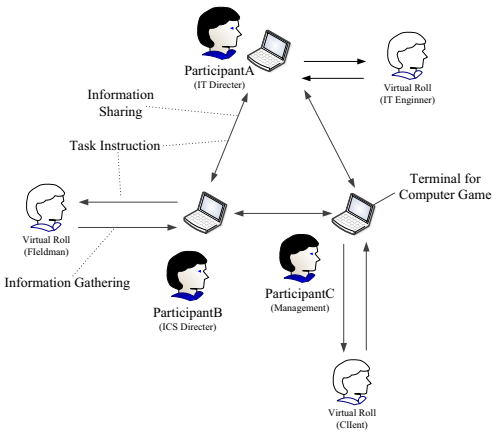


Figure 1 Composition of the Computer Game

Each participant has the terminal for the computer game and participates in the computer game as the separate commander. Each participant uses the terminal to collect information, share information, and issue task instructions among the participants. However, the commander needs to collaborate not only with other commanders but also with workers who are subordinates of the commander and external organizations (customers, affiliates, etc.). For this reason, it is necessary to prepare workers and external organizations as virtual roles on the computer game, which have the function of returning responses to collecting information, sharing information, and task instructions from participants. This allows each participant to interact not only with other participants but also with his subordinates and external organizations within the computer game. Also, to play the computer game, scenarios for proceeding the computer game is required. The game scenarios consist of cyberattacks scenario and an ICS state transition scenario affected by cyberattacks. Based on the game scenarios, virtual roles must respond to the participant. Participants understand the situation by using information from the virtual roles. The computer game also requires the mechanism whereby the instruction is reflected in the game scenarios when participants issue the task instructions to the virtual roles. For example, if the ICS Director instructs the fieldman to operate the ICS, then the status of the ICS in the gaming scenarios must change to the state in which the instruction is reflected. This mechanism is essential in the computer game, and with this mechanism, participants can aim for convergence of cyber incidents in ICS, which is the goal of the computer game. Thus, the requirements to be met by the computer game are as follows.

- Some participants can participate in the computer game as separate commanders.
- Each participant can collect information, share information, and perform task instructions to other participants and virtual roles.
- Based on the game scenarios, the virtual roles interact with the participants.
- Task instructions to the participants' workers are reflected in the game scenario.

Additionally, what is indispensable in training is feedback after training. In feedback, participants should assess whether they were able to adequately perform their commander roles, identify problems, and discuss remedial measures. However, to provide feedback, a workflow is required to record what the participants did in the computer game. Therefore, the computer game also requires the ability to output workflows.

3.3. Development of Prototype of the Computer Game

We developed the prototype computer game that satisfies the requirements described in the previous section. The programs of the computer game were described in Java.

Figure 2 is User Interface (UI) of terminals used by participants in the computer game. Participants can use this UI to collect information, share information, and issue task instructions. In UI, the communication sent and received is arranged in chronological order. Therefore, it is displayed on this UI when it communicates with other participants or virtual roles, or when itself receives communication from other participants or virtual roles. Pressing the buttons above the tray in which the communication in Figure 2 is arranged opens the box shown in Figure 3. When communication contents, address, etc. are set in this box and transmitted, the information is sent to the specified address.

Figure 4 shows an example of the result of a computer game.Artifact is output in workflow format, and records who, when, and what do. Participant training is not completed only by computer game execution. Training must include extracting the

current problem (slow decision on plant stoppage etc) based on deliverables and deciding how to improve the function of commanders. Also commanders training is not complete at once. Commanders cannot maintain its function unless it receives regular training. Commanders cannot maintain its function unless it receives regular training. And, further training provides new learning to participants every time. Therefore, it is important that Commander repeats training, always maintaining and improving functions.

Figure 2 User Interface

Figure 3 Box for Creating Communication

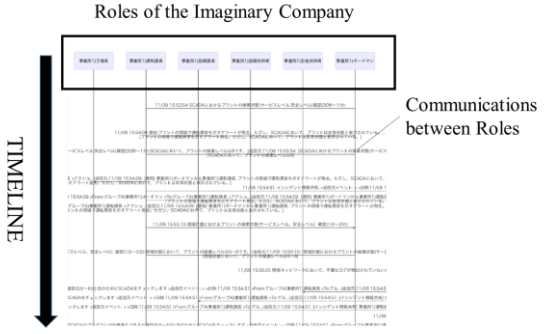


Figure 4 Typical Deliverable of the Participants

3.4. Implementation of Prototype of the Computer Game

In September 2017, the authors held the trial event of the prototype for 38 people from companies holding CIs. The data of prototype was developed based on the virtual

company used in the current cyber exercise. As commanders of the virtual company, participants experienced the prototype with the goal of convergence of cyber incidents.

In the post-implementation questionnaire, 17 out of 38 respondents answered that they would be trained on the question, "Can this game be used to train the Incident Commanders?" Therefore, it can be said that this game was evaluated as effective to a certain extent.

4. Conclusion

In this research, the authors developed the cyber exercise in the form of a computer game for educating incident commanders in CI owner companies. Through workshops with OT and cyber security experts, the developed prototype was successfully evaluated by using the participants' questionnaire. To apply this prototype to each company, it is necessary to customize the data to reflect the situation (such as the organization structure, the jurisdiction scope, the corporate culture and so on) of each company.

At present, however, the complexity of the data integrity makes it difficult to prepare unified template to simplify the customization. Accordingly, we will continue to develop the utility tools to promote our exercise.

5. Acknowledgement

This research is partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (A), No.16H01837 (2016) and Council for Science, Technology and Innovation (CSTI), Cross-ministerial Strategic Innovation Promotion Program (SIP), “Cyber- Security for Critical Infrastructure” (Funding Agency: NEDO), however, all remaining errors are attributable to the authors.

References

BBC NEWS (2013), “The 'cyber-attack' threat to London's Olympic ceremony”, available at: http://www.bbc.com/news/uk-23195283 (accessed 20 October 2017)

Hidekazu Hirai (2017), Framework for Cyber Incident Response Training, WIT Transactions on The Built Environment, Volume 174, pp. 273-283.

Nihon Keizai Shimbun (2017), “Cyber attack, Tokyo is several times as much as Rio?”, available at: https://style.nikkei.com/article/DGXMZO21989500W7A001C1SHB000?channel=DF220420167266&n_cid=LMNST011 (accessed 10 October 2017)

Motohiro Tsuchiya (2015), “With the cyber attack, German steel mills were suffering enormous damage”, available at: https://www.newsweekjapan.jp/tsuchiya/2015/09/post-2_1.php (accessed 10 October 2017)

Nobutaka Oguma (2011), “Stuxnet - first malware aiming at the control system-”, available at: https://www.jpcert.or.jp/ics/2011/20110210-oguma.pdf (accessed 10 October 2017)

Toshio Miyachi (2016), “Present and Future of Control System Security”, available at: https://www.jpcert.or.jp/present/2016/20160217_CSC-JPCERT01.pdf (accessed 10 October 2017)

13TH INTERNATIONAL SYMPOSIUM ON PROCESS SYSTEMS ENGINEERING (PSE 2018)

PART C

Edited by
MARIO R. EDEN
MARIANTHI G. IERAPETRITOU
GAVIN P. TOWLER

COMPUTER-AIDED CHEMICAL ENGINEERING, 44

13TH INTERNATIONAL SYMPOSIUM ON PROCESS SYSTEMS ENGINEERING (PSE 2018)

PART C

This page intentionally left blank

13TH INTERNATIONAL SYMPOSIUM ON PROCESS SYSTEMS ENGINEERING (PSE 2018)

PART C

Edited by

Mario R. Eden
Department of Chemical Engineering
Auburn University
Auburn, Alabama, USA

Marianthi G. Ierapetritou
Department of Chemical and Biochemical Engineering
Rutgers University
Piscataway, New Jersey, USA

Gavin P. Towler
Honeywell Performance Materials and Technologies
Des Plaines, Illinois, USA

ELSEVIER

Amsterdam – Boston – Heidelberg – London – New York – Oxford
Paris – San Diego – San Francisco – Singapore – Sydney – Tokyo

Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, USA

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the Library of Congress

ISBN (Part C): 978-0-444-64245-5
ISBN (Set): 978-0-444-64241-7
ISSN: 1570-7946

Publisher: Joe Hayton
Acquisition Editor: Kostas Marinakis
Editorial Project Manager: Emily Thomson
Production Project Manager: Paul Prasad Chandramohan
Designer: Greg Harris

Typeset by SPi Global, India

Contents

Contributed Papers: Sustainable and Renewable Systems Engineering

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50274-3

Mathematical Modeling and Optimization for Central Refrigeration Systems Design Using Low-Grade Energy in Industrial Parks

Qiao Q. Tang, Kai Liu, Qing L. Chen, Ming Pan, Chang He, Bing J. Zhang*

School of Chemical Engineering and Technology, Guangdong Engineering Center for Petrochemical Energy Conservation, Sun Yat-Sen University, No. 135, Xingang West Road, Guangzhou, 510275, China
zhbingj@mail.sysu.edu.cn

Abstract

Industrial parks are competitive because of the concentrated utilization of resource and energy. In industrial parks, power plants are energy hubs to supply power and various levels of steam to other enterprises. However, it is still challenging to centralized produce and supply cold energy in industrial parks. We present a process superstructure for the retrofit of power plants to centralized produce and supply cold energy for industrial parks. The energy sources for refrigeration include cold energy from liquefied natural gas, pressure energy of natural gas and low-pressure steam in power plants. A mixed integer nonlinear programming model is formulated to optimize the process superstructure to minimize the total annual cost. The model is solved using the global solver BARON in GAMS environment. The solution results, total annual cost and different price of electricity are analyzed and compared. The best alternative is a hybrid central refrigeration system that includes different energy sources for chilling, and the power consumption of cold energy is reduced by 36%.

Keywords: Optimization, Mathematical modeling, Industrial park, Refrigeration.

1. Introduction

Energy conservation has been a routine behavior for process industries to reduce operating costs and emissions under the requirement of sustainable development (Zhang *et al.*, 2016). Industrial parks require considerable energy to transform crude feeds into final products. In industrial parks, power plants are energy hubs to supply power and various levels of steam to other enterprises. At the same time, most of the low-grade heat (LGH) is emitted directly. Using the unrecovered LGH to produce cold energy is an effective way to improve energy efficiency and reduce emissions for industrial parks.

Nowadays, due to the rising energy prices and stringent environmental protection requirements, the utilization of LGH is significantly investigated, such as absorption refrigeration (Ponce-Ortega *et al.*, 2011), desalination (Ammar *et al.*, 2012), and power generation (Lin *et al.*, 2014). Because of the high energy recovery efficiency of absorption chiller, absorption refrigeration is widely used to converting LGH for cooling. Wu *et al.* (2002) introduced a new method for synthesizing integrated refrigeration systems. Hammond and Norman (2014) used absorption chillers to convert heat energy for cold energy to fulfill chilling demand. Zhang *et al.* (2016) proposed a mixed integer nonlinear programming (MINLP) model to finish the optimization and economic analysis for the LGH utilization in petrochemical industrial parks and a good economic performance is

obtained. Salmi *et al.* (2016) presented a model to optimize an absorption refrigeration system using waste heat (WH) of ship as energy source. Wang *et al.* (2017) presented a model for the match between heat sources and absorption refrigeration to improve energy recovery efficiency of waste heat. Nevertheless, it is still challenging to centralized produce and supply cold energy in industrial parks.

In this study, the design optimization of central refrigeration systems (CRSs) using LGH and other energy sources in industrial parks is proposed. A MINLP model is formulated to optimize the process superstructure to minimize the total annual cost (TAC). An industrial case is then investigated and analyzed.

2. Problem Description and CRS superstructure

Natural gas power plants in industrial parks supply heat and electricity to other plants. However, cold energy is often separately produced and used in individual plants by using electricity. Thus, a CRS based on power plants is attractive for industrial parks to improve energy performance and profitability.

The typical flowsheet of natural gas power plants in industrial parks is shown in the upper part of Figure 1. The pressure regulating station, regasification station and low-pressure steam form the steam turbine can supply energy for refrigeration. The cold energy is then sent to cold sinks in industrial parks. In order to minimize the TAC, we need to determine: (1) which energy sources in the power plant is used for refrigeration? (2) which level of steam from the power plant is used for refrigeration? (3) how many refrigeration stations are built to produce cold energy? and (4) how much cold energy can be supplied from a refrigeration station to a cold sink?

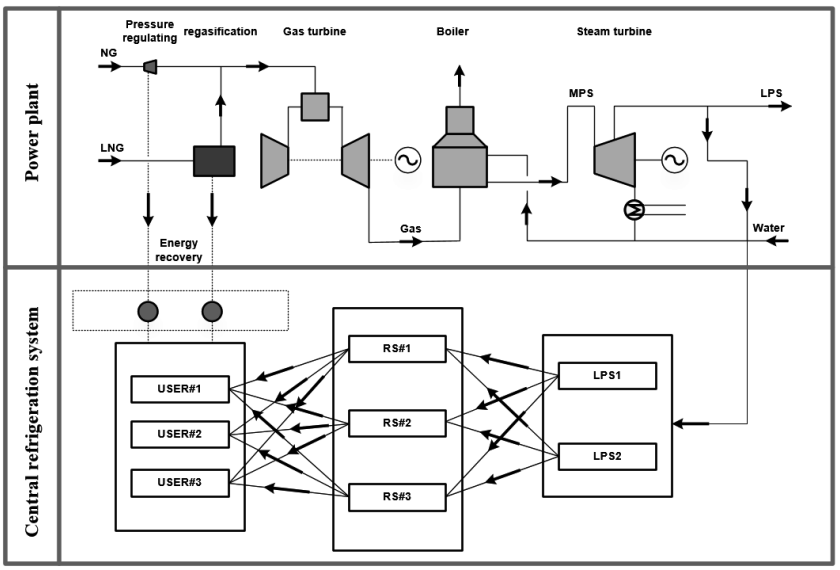


Figure 1. Flowsheet of natural gas power plants and superstructure of CRS

In order to address the above problems, a CRS superstructure is extracted and shown in the lower part of Figure 1. The possible flowsheets of CRS are represented as three levels. The first level is the configuration of energy sources from the power plant. We can use anyone of the three energy sources. The second level is refrigeration stations. A

refrigeration can collect energy from any sources, and convert it into cold energy, and then sent the cold energy to any cold sinks. The last level is cold sinks. A cold sink in industrial parks can receive cold energy from any refrigeration stations. Note that when different energy sources or flowsheets is employed, the operation of power plant varies.

3. Mathematic Modeling

A MINLP mathematical model is introduced in this section to optimize the superstructure. According the three levels of CRS superstructure, the mathematic model includes power plants, waste energy or low-grade heat recovery, refrigeration and pipeline network.

3.1 *Modeling for power plants*

Natural gas (NG) or liquid natural gas (LNG) after treatment enter into gas turbines to generate power and exhaust high temperature and pressure flue gas. Then, the flue gas goes into boilers to generate middle pressure steam. The middle pressure steam is introduced into steam turbines to generate power and obtain waste steam. Hence, the mass and energy balances of power plants are represented in Eqs. (1)- (5). Eq. (1) is the energy balance of fuel combustion. Q_{com} is the combustion heat of CH_4, F_{NG} and F_{LNG} are the mass flowrate of NG and LNG, respectively. $C_{CH_4,NG}$ and $C_{CH_4,LNG}$ are the mass fraction of CH_4 in NG and LNG, respectively. Q_{CH_4} is the low calorific value of CH_4. Eqs. (2) and (3) are the mass and energy balances for gas turbines, respectively. F_{air} is the mass flowrate of air enter gas turbine and F_{gas} is the mass flowrate of flue gas. Q_{gas} is the heat of gas from gas turbine, Q_{ele} is the power generated by gas turbine and Q_{loss} is the heat of low temperature flue gas. Eq. (4) is the model for boilers, and Eq. (5) is the model for steam turbines (Shang and Kokossis, 2004). In Eqs. (4) and (5), F is the mass flowrate of steam, ΔH_{steam} is the enthalpy of steam, TE_t is the power generated by steam turbine, D_t is the isentropic enthalpy drop of steam turbine, and α, β, δ and ε are correlation coefficients.

$$Q_{com} = (C_{CH_4,NG}F_{NG} + C_{CH_4,LNG}F_{LNG}) \times Q_{CH_4} \tag{1}$$

$$F_{gas} = F_{NG} + F_{LNG} + F_{air} \tag{2}$$

$$Q_{com} = Q_{gas} + Q_{loss} + Q_{ele} \tag{3}$$

$$Q_{gas} = [\alpha_b FBL_b^{max} + (1 + \beta_b)FBL_b] \times \Delta H_{steam} \tag{4}$$

$$TE_t = \frac{6}{5}\frac{1}{\delta_t}\left(D_t - \frac{\varepsilon_t}{F_t^{max}}\right)\left(F_t - \frac{1}{6}F_t^{max}\right) \tag{5}$$

3.2 *Modeling for recovery of waste energy*

The gas streams entering power plants include NG and LNG. NG is at high pressure and need to be decompressed at pressure regulating stations. The energy released from decompression processes can be recovered. The regasification of LNG also emits cold energy. The energy balance for the two processes are shown in Eqs. (6)-(7). Q_{NG} and Q_{LNG} are the capacity of refrigeration by NG pressure regulating and LNG regasification, respectively; CP_{NG} and CP_{LNG} are the specific heat capacity of NG and LNG, respectively;

T_{NG}^{in} and T_{NG}^{out}, T_{LNG}^{out} and T_{LNG}^{in} are the inlet and outlet temperatures of NG and LNG, respectively; η_{expand} is the efficiency of expander, and X_{NG} and Y_{LNG} are 0-1 variables to identify whether waste energy is recovered or not. Because of the throttling expansion for NG, the outlet temperature of NG can be calculated by Eq. (8). P_{NG}^{in} and P_{NG}^{out} are the inlet and outlet pressure of *NG*, respectively; k is the insulation index.

$$Q_{NG} = F_{NG} \cdot X_{NG} \cdot CP_{NG} \cdot \left(T_{NG}^{in} - T_{NG}^{out}\right) \cdot \eta_{expand} \tag{6}$$

$$Q_{LNG} = F_{LNG} \cdot Y_{LNG} \cdot CP_{LNG} \cdot (T_{LNG}^{out} - T_{LNG}^{in}) \tag{7}$$

$$\frac{T_{NG}^{in}}{T_{NG}^{out}} = \left(\frac{P_{NG}^{in}}{P_{NG}^{out}}\right)^{\frac{k-1}{k}} \tag{8}$$

3.3 *Modeling for refrigeration using steam*

Absorption chiller is widely used in refrigeration using low-grade heat. An absorption chiller is consisted of four main parts: generators, condensers, evaporators, and absorbers. We can model the absorption chiller by considering the energy balance in the four parts. The models of the four parts can be found in Literature 7. The total energy balance of the absorption chiller can be expressed in Eq. (9). Q_s^g is the energy of generator, Q_s^a is the energy of absorber, Q_s^c is the energy of condenser, and Q_s^e is the energy of evaporator. The coefficient of performance COP_s of the absorption chiller is represented in Eq. (10).

$$Q_s^g + Q_s^e = Q_s^a + Q_s^c \tag{9}$$

$$COP_s = \frac{Q_s^e}{Q_s^g} \tag{10}$$

3.4 *Objective function*

The TAC is used as the objective for the design of a CRS that covers all the cooling demands in industrial parks. The objective is represented in Eq. (11). TAC is the total annual cost, C^R is the investment of refrigeration stations, C^P is the investment of pipelines, C^E is the investment for valves and exchangers, C^D is the investment for expanders, θ is the equipment depreciation period, C^W is the cost for cooling water, P^C and P^E are the income for the cold energy and electricity, respectively. The correlation for the cost of equipment can be found in Literature 7.

$$\boldsymbol{min}\ \ TAC = \frac{C^R + C^P + C^E + C^D}{\theta} + C^W - (P^C + P^E) \tag{11}$$

4. Case Study

The MINLP model is applied to a case extracted from an industrial park at South of China. In the case, the total operation time per year is 8,000 hours, and there are two devices in the power plant to supply the energy and power to meet the industrial park's demand. The minimum annual demand of power in the entire park is 1.6 million MW. There are three

energy sources (the pressure regulating station, regasification station and steam turbine) in the power plant can be used for refrigeration. The specification of gas stream, the demand of chilled water from enterprises and the distances between cold sinks and refrigeration stations are all listed in Table 1.

Table 1. Basic data of the case

The original status for gas stream				The demand of cold energy sinks		
	Mass Flowrate	Temperature	Pressure		Quantity	Temperature
NG	20 t/h	273.15 K	4.0 MPa	User#1	10 MW	278.15 K
LNG	5 t/h	163.15 K	2.5 MPa	User#2	4 MW	278.15 K
				User#3	6 MW	278.15 K

The distances between steam turbines, sinks and refrigeration stations (m)			
	RS #1	RS #2	RS #3
Steam turbine #1	300	230	240
Steam turbine #2	300	230	240
User #1	2200	1200	3000
User #2	1300	470	1300
User #3	480	2200	1400

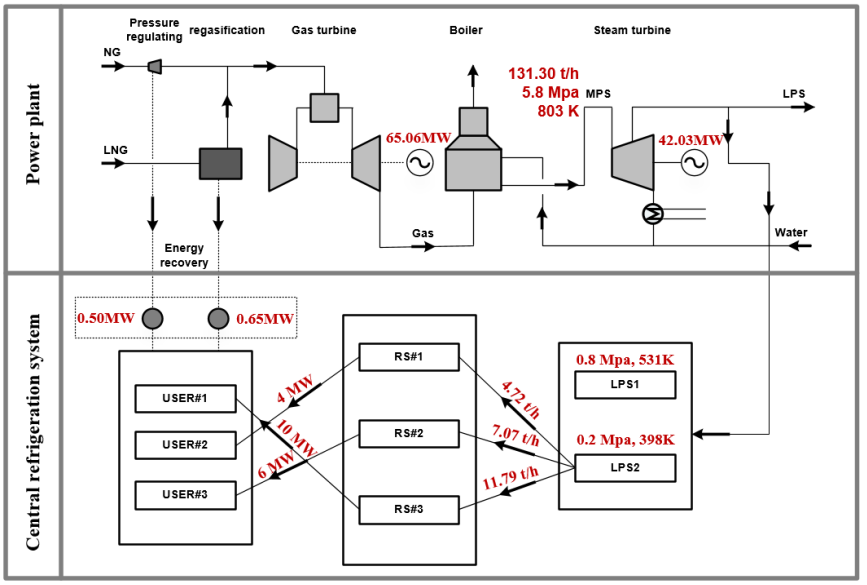


Figure 2. The optimal flowsheet and heat load of CRS

The MINLP model is coded in GAMS environments and solved using the global solver BARON, there are 217 variables and 22 binary variables, the computing time is 8s. The cold energy from pressure regulating stations and regasification stations is directly used by the power plant to produce chilled water. Low-pressure steam from steam turbine #2 is introduced into three refrigeration stations to produce chilled water. Cold energy from refrigeration station #1 is transferred into user #2, that from station #2 into user #3, and that from station #3 into user #1. The TAC of this case is $ 10.7 million dollars, the total power generated in the power plant is 1.7 million MW per year, the power can be generated by steam which is used to refrigeration is 2.55 MW and other solution results are shown in Figure 2.

5. Conclusions

A process superstructure was introduced for the design of CRSs based on natural gas power plants in industrial parks. A MINLP model was formulated to optimize the superstructure to obtain the minimum TAC. The model was applied to an industrial park. For 20 MW of cold energy, the total electricity consumption is 4.0 MW when using separate refrigeration processes, while the electricity generation in the power plant is only reduced by 2.55 MW when using the CRS. The CRS reduces the power consumption by 36%. The application demonstrates that the process superstructure and MINLP model can optimize the design of CRSs and improve energy efficiency.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (No. 21776323) and the project of Guangdong Provincial Natural Science Foundation of China (No. 2015A030313112).

References

Y. Ammar, H. Li, C. Walsh, P. Thornley, V. Sharifi, A. P .Roskilly, 2012, Desalination using low grade heat in the process industry: Challenges and perspectives, Appl. Therm. Eng., 48, 446−457.

G. P. Hammond, J. B. Norman, 2014, Heat recovery opportunities in UK industry, Applied Energy, 116, 387-397.

S. Lin, N. Y. Yip, T. Y. Cath, C. O. Osuji, M. Elimelech, 2014, Hybrid Pressure Retarded Osmosis-Membrane Distillation System for Power Generation from Low-Grade Heat: Thermodynamic Analysis and Energy Efficiency, Environ. Sci. Technol., 48, 9, 5306−5313.

J. M. Ponce-Ortega, E. A. Tora, J. B. González-Campos, M. M. El-Halwagi, 2011, Integration of Renewable Energy with Industrial Absorption Refrigeration Systems: Systematic Design and Operation with Technical, Economic, and Environmental Objectives, Ind. Eng. Chem. Res., 50, 16, 9667−9684.

W. Salmi, J. Vanttola, M. Elg, M. Kuosa, R. Lahdelma, 2016, Using waste heat of ship as energy source for an absorption refrigeration system. Applied Thermal Engineering. 115: 501-516.

Z. G. Shang, A. C. Kokossis, 2004, A transhipment model for the optimization of steam levels of total site utility system for multiperiod operation, Comput. Chem. Eng., 28, 1673–1688.

Y. Wang, C. Wang, X. Feng, 2017, Optimal match between heat source and absorption refrigeration, Computers & Chemical Engineering, 102, 268-277.

G. Wu, X. X. Zhu, 2002, Design of Integrated Refrigeration Systems, Ind. Eng. Chem. Res., 41, 3, 553−571.

B. J. Zhang, Z. L. Zhang, K. Liu, Q. L. Chen, 2016, Network modeling and design for low grade heat recovery refrigeration and utilization in industrial parks, Ind. Eng. Chem. Res., 55, 36, 9725-9737.

B. J. Zhang, J. Li, Z. L. Zhang, K. Wang, Q. L. Chen, 2016, Simultaneous design of heat exchanger network for heat integration using hot direct discharges/feeds between process plants, Energy, 109, 400-411.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50275-5

Influence of Humic Substances on the Solubility and Crystallization Behavior of Calcium Phosphates during Crystallization from Digester Supernatant

Liubov Vasenko*, Haiyan Qu

Department of Chemical Engineering, Biotechnology and Environmental Technology, University of Southern Denmark, Campusvej 55, Odense M 5230, Denmark
liva@kbm.sdu.dk

Abstract

Conditional solubility of dicalcium phosphate dihydrate (DCPD) and hydroxyapatite (HAp) was determined in distilled water and in digester supernatant experimentally and compared with the modelling results with PHREEQC. Conditional solubility of both phases in digester supernatant increases at the pH values > 5 – 6, and for HAp this increase is more significant. It was confirmed that the increase in conditional solubility is caused by the presence of natural organic matter (especially humic substances) by determination of conditional solubility of DCPD and HAp in the solution of humic acid (10 mg/L). The obtained solid-liquid equilibrium data served as the background for designing the separation process for recovery of phosphorus from digester supernatant. A two-step process with fast precipitation and recrystallization was developed, from which DCPD, HAp or their mixture were obtained. However, it was observed that the presence of humic compounds significantly inhibited the crystallization of DCPD and HAp and affected the purity of the obtained products. Thereafter, a novel hybrid separation process was designed and developed. The new process consists of an oxidation and a crystallization steps. It could successfully overcome the negative effects of humic compounds. Either DCPD or HAp with the high purity could be produced with a P-recovery rate of about 90 %.

Keywords: Phosphorus recovery, Calcium phosphates, Reactive crystallization, Conditional solubility.

1. Introduction

Recovery of phosphorus from digester supernatant in wastewater treatment plants is one of the most promising ways toward phosphorus sustainability (Egle et al., 2015). Currently, the major product in such P-recovery processes is struvite, which is sold as a low-prise slow releasing fertilizer. Production of high-purity calcium phosphate (CaP) can ensure a broad application of the product (industry, fertilizers. feed additives etc.), and thus can improve the economy of the process. However, crystallization of calcium phosphates from digester supernatant is a challenging task due to the complicated equilibrium between different phases and the influence of operation parameters and presence of different impurities. It was reported that humic substances (HS) inhibit the crystallization of DCPD and HAp (Amjad, 1989; Freche et al., 1992). Song et al. reported that HS are removed from solution simultaneously with CaP crystallization by adsorption on crystals' surface (Song et al., 2006). HS can as well create complexes

with Ca^{2+}, reducing its availability for the formation of CaP. Previously we reported that the appearance of DCPD or HAp during crystallization depends on nucleation pH, which is strongly affected by Ca/P and NH_4-N/P ratios (Vasenko and Qu, 2017a). Moreover, presence of ammonia also affected the kinetics of the crystallization process.

A conceptual approach was suggested in the present work for the design and development of a hybrid separation process for recovery of high purity CaP products from digester supernatant. Firstly, the key impurities in the digester supernatant that influence the conditional solubility of CaP and inhibit their crystallization were identified. Then a two-step process of fast precipitation and recrystallization was developed and tested to minimize the effects of the impurities. Finally, a new hybrid process with oxidation and crystallization was designed. This process could produce high purity CaP with a high P-recovery rate at about 90 %, and thus has a great potential in application at wastewater treatment plants as a new P-recovery technology (Vasenko and Qu, 2017 b).

2. Materials and methods

2.1. Materials

Reagent grade DCPD and HAp were used for conditional solubility determination. pH adjustment was done using 37 % HCl and 23 % NaOH solutions. All mentioned reagents were purchased from Sigma Aldrich. Humic acid solution with the concentration of 10 mg/L was prepared from crystalline humic acid with assay > 95 % (purchased from Alfa Aesar). Solid and liquid separation after experiments were done with vacuum filtration using an Advantec Glass Fiber filter with a porosity < 0.45 μm.

2.2. Experimental determination of conditional solubility

The conditional solubility of two CaP phases was determined in distilled water, digester supernatant and humic acid solution with concentration of 10 mg/L at the room temperature. pH of 50 ml liquid aliquotes was adjusted to different values in the range of 1 – 13. An excess amount of DCPD or HAp was suspended in these aliquots for 24 hours to attain the solid-liquid equilibrium. Thereafter, solid phase was separated using vacuum filtration, final pH and the total concentration of phosphate ($[P^*_{tot}]$) and calcium ($[Ca^*_{tot}]$) in the liquid phase were determined. The conditional solubility was calculated using Eq. (1).

$$logP_s = log([Ca^*_{tot}] \cdot [P^*_{tot}]) \quad (1)$$

The values of logPs were plotted versus the final pH of solutions to graphically represent the dependence of conditional solubility on pH. All pH measurements were done with a PHM210 Standard pH Meter with reference Meter Toledo Routine sensor. The electrode was calibrated with standard IUPAC buffer solutions with pH 4.00 and 7.00 at room temperature before each experiment and was stored in a 3M KCl solution between measurements.

2.3. Determination of conditional solubility in distilled water using PHREEQC

PHREEQC software was used for the thermodynamical modeling of speciation during the dissolution of DCPD or HAp in distilled water at different pH values edited by the introduction of HCl or NaOH. Speciation equations were taken from the phreeqc.dat database. The equation for DCPD speciation was added manually to the program as it is

not in the phreeqc.dat database and was used with the activity based K_{sp} value (Patel, et al., 1974).

2.4. Analysis of liquid phase

Total phosphate and calcium concentrations in liquid phase were determined using ion chromatography equipped with a 761 SD Compact IC with an 813 Compact Autosampler by Metrohm. Anions analysis was performed using a Metrosep A Supp 5 - 150 (6.1006.520) column and the eluent is 3.2 mM Na2CO3/1.0 mM NaHCO3 + 10 % acetone. A column for cations analysis is Metrosep Cation 4 (6.1010.000) and the eluent is 1.7 nM Nitric acid/0.70 mM dipicolinic acid.

2.5. Analysis of solid phase

All solid samples were analysed with FT-Raman and PXRD in order to identify the CaP phase. A Bruker MultiRAM FT-Raman Spectrometer equipped with 1064 nm laser was used with the laser power of 100 – 400 mW. A Rigaku Miniflex 600 X-ray diffractometer with a Cu source of X-rays and a Ni filter was used for the PXRD analysis with the voltage of 40 kV, current – 15 mA, speed – 10°/min, step size – 0.02, 2Θ range: 2 – 70°. A demo version of Match! Ver.3. was used with a COD-Inorg database to analyse the PXRD patterns.

3. Results and discussion

3.1. Effect of humic substances on conditional solubility of DCPD and HAp

Conditional solubility of DCPD and HAp in water and in digester supernatant are shown in Fig. 1 together with the modelled results of PHREEQC. Analysis of the solid recovered after the solubility measurement confirmed no transformations between DCPD and HAp phases during the measurement. Conditional solubility of DCPD in digester supernatant only slightly increases at pH > 5 – 6 compared to its solubility in distilled water. For HAp the increase of conditional solubility is more significant at pH > 6, while for pH < 5 the decrease of its conditional solubility can be observed. This change of conditional solubility can be explained by the presence of different impurities in digester supernatant and their possible complexation with either calcium or phosphate ions as well as by the increase of the ionic strength of the solution.

In order to determine whether the presence of humic compounds can have an effect on conditional solubility of these two phases, the conditional solubility was measured in the solution of the humic acid with the concentration of 10 mg/L (Fig. 1).

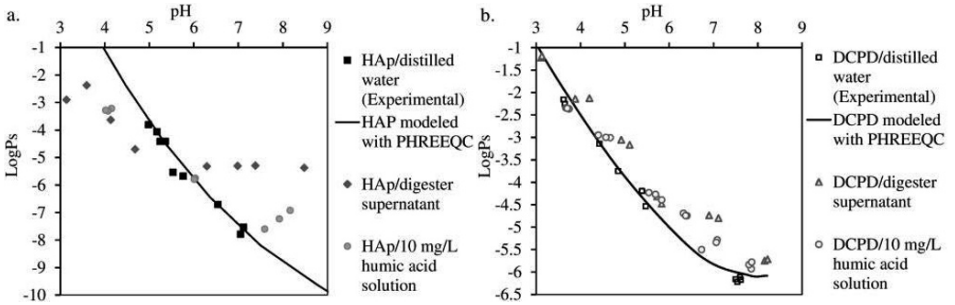


Figure 1. Conditional solubility of calcium phosphates in different media: a. HAp; b. DCPD.

Humic substances in digester supernatant have a major impact on the conditional solubility of HAp and DCPD, as the change of conditional solubility in digester supernatant follows the same trend as its change in the humic acid solution. It may also be true, that in digester supernatant the impact of humic substances is enhanced by the presence of other impurities such as carbonate etc. The change of conditional solubility described above can have a strong negative effect on the crystallization of DCPD or HAp as the yield of the crystallization process will be reduced. At the same time, the decrease in conditional solubility of HAp at pH below 5 may result in co-precipitation of mixture HAP and DCPD, due to the very similar conditional solubility of these two phases. Therefore, for the development of the robust technology, it is important to find the way to reduce the influence of humic substances.

4. Design of the P-recovery technology

4.1. Two-step fast precipitation/recrystallization technology

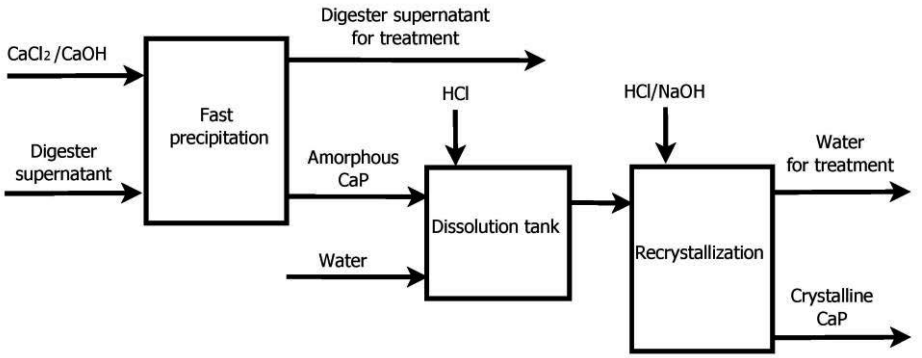


Figure 2. Two-step fast precipitation/recrystallization technology.

Based on the obtained solubility data a two-step process with fast precipitation and recrystallization was proposed, as it is shown in Fig 2. In this process, calcium is dosed to the digester supernatant at rather a high pH, so that amorphous CaP precipitates from the solution. The precipitation was performed within a short time so that the adsorption of humic substances onto the surface of CaP product can be minimized. Amorphous product later-on is dissolved in the water at low pH and can be re-crystallized into DCPD or HAp by controlling the operational pH and adding seeding.

This technology was tested in the laboratory scale (Vasenko and Qu, 2018). Aaby WTTP (Aarhus, Denmark) provided digester supernatant containing 0.008 mol/L of total-phosphate. Fast precipitation was performed at the initial pH of digester supernatant equal to 7.8 – 8.0 at Ca/P ratios 2 and 5 and total-phosphates concentrations 0.008 mol/L and 0.015 mol/L (increased by the addition of $NH_4H_2PO_4$). Batch recrystallization resulted in crystallization of HAp, DCPD or their mixture. Addition of DCPD seeding promoted crystallization of DCPD, but decreases the P-recovery rate, due to the higher solubility of DCPD than HAp. The highest obtained P-recovery rate was 69.5 % when HAp was crystallized from the solution. The drawback of this technology is that doubled amount of liquid waste is produced. Moreover, even though the products quality increases comparing to the direct crystallization of CaP from digester supernatant, it is a powder with light brown color that cannot be characterized as a high-purity product.

4.2. Novel hybrid technology

Based on the results obtained from the two-step process described above, a novel hybrid technology was developed. The novelty of this technology is that it takes the advantage of oxidation to degrade the NOM in digester supernatant and then crystallization is performed in an environment with minimal effects from the impurities (Vasenko and Qu, 2017b). This process is illustrated in Fig. 3.

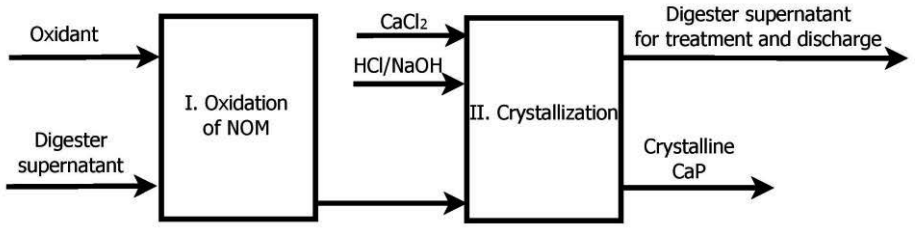


Figure 3. Hybrid oxidation/crystallization technology for high-purity CaP recovery.

During the oxidation step, digester supernatant is mixed with an oxidant for the time sufficient for degradation of NOM. Thereafter, crystallization of CaP is performed by dozing calcium to the oxidized supernatant and controlling the operational pH. This hybrid technology has been tested in laboratory scale with the application of H_2O_2 (20 ml of 30 w/w % per 1 L of digester supernatant) as an oxidant in combination with short ultrasound (US) treatment to enhance degradation of NOM. Sonication was on for 10 minutes at the beginning of oxidation step, which lasted for 24 hours. Thereafter, crystallization was performed with Ca/P=2 by dosing $CaCl_2$, adjusting pH until nucleation and keeping it stable throughout the process.

Optimization in terms of pH at oxidation step was performed with 4 pH values: 2.0, 3.9, 7.9 and 10.0. The color of digester supernatant was changing from dark brown to light yellow during the oxidation and the higher was the pH of oxidation the more significant was the improvement of the solution color. However, the performance of crystallization step was the highest, when the pH of oxidation was 7.9. At this condition, the highest P-recovery rate was achieved (89 %), and the purest product was obtained.

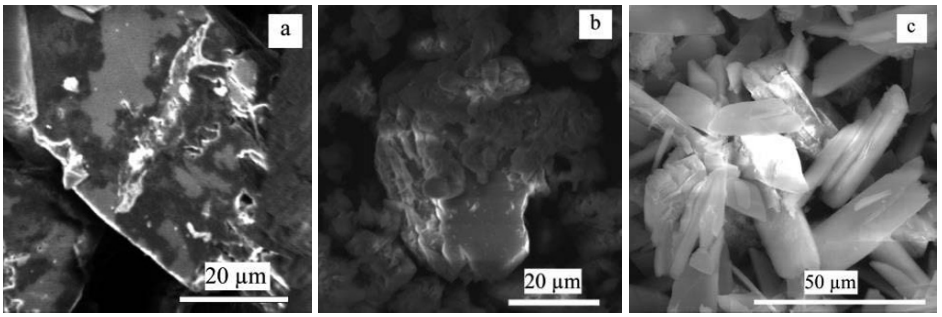


Figure 4. SEM images of selected samples: a. Direct crystallization of CaP from digester supernatant (Vasenko and Qu, 2017 c); b. Fast precipitation/recrystallization technology; c. Novel hybrid technology.

Improvement of the product quality especially the adsorption of HS on the CaP products can be observed from Fig. 4. On the surface of the DCPD product, crystallized directly

from digester supernatant, dark spots can be noticed, which are believed to be adsorbed HS (Fig.4.a). With the application of fast precipitation/recrystallization technology, no dark spots could be observed on the surface of the DCPD product, but the color of the crystals remains light brown (Fig. 4.b). When the novel hybrid oxidation/crystallization technology was used with the oxidation pH 7.9 white DCPD crystals with some minor amount of HAp were obtained (Fig. 4.c). ICP-MS analysis of the heavy metals contents in these products obtained with novel hybrid process indicated that the heavy metal concentration fulfils the EU requirement for food additives. These results proved the significant improvement in the purity of the products when the hybrid process is applied. Moreover, the amount of liquid wastes after the hybrid process is reduced compared to the fast precipitation/recrystallization technology.

5. Conclusions

The negative effect of humic substances on CaP crystallization from digester supernatant was proved by determination of conditional solubility of two target CaP phases in humic acid solution, digester supernatant, and distilled water. Based on the observation regarding the influence of humic substances on CaP crystallization, two processes were designed and tested in laboratory scale. The proposed two-step process with fast precipitation and recrystallization could slightly improve the quality of the produced CaP products. A more significant improvement of both the product quality and the P-recovery rate was achieved with the new hybrid process with oxidation and crystallization. There is a great potential of applying this new hybrid technology in phosphorus recovery in wastewater treatment plants.

References

Z. Amjad, 1989, Effect of precipitation inhibitors on calcium phosphate scale formation, *Can J Chem*, Vol 67 (5), 850-856

L. Egle, H. Rechberger, M. Zessner, 2015, Overview and description of technologies for recovering phosphorus from municipal wastewater, *Resour Conserv Recycl*, Vol. 105, 325-346

M. Freche, J.L. Lacout, 1992, Effect of humic compounds and some organic acids added during dicalcium phosphate dihydrate crystal growth process, *J Alloys Compd*, Vol. 188, 65 – 68

P.R. Patel, T.M. Gregory, W.E. Brown, 1974, Solubility of CaHPO4·2H2O in the Quaternary System Ca(OH)2 – H2Po4 – NaCl – H2O at 25 oC, Journal of Research of the National Bureau of Standards – A. Physics and Chemistry, Vol. 78A (6). 675 - 681

Y. Song, H.H. Hahn, E. Hoffmann, P. Weidler, 2006, Effect of humic substances in the precipitation of calcium phosphate, *J Environ Sci*, Vol. 18 (5), 852-857

L. Vasenko, H. Qu, 2017 a, Effect of NH4-N/P and Ca/P molar ratios on the reactive crystallization of calcium phosphates for phosphorus recovery from wastewater, *J Cryst Growth*, Vol. 459, 61-66

L. Vasenko, H. Qu, 2017 b, European Patent Application No. 17192930.0, “Recovery of high purity calcium phosphate”

L. Vasenko, H. Qu, 2017 c, Calcium phosphates crystallization from digester supernatant. *Proceedings of 24th International Workshop on Industrial Crystallization BIWIC 2017*, Dortmund, Germany, 15 - 20. Print

L. Vasenko, H. Qu, 2018, Calcium phosphates recovery from digester supernatant by fast precipitation and recrystallization, *J Cryst Growth*, Vol 481, 1-6

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50276-7

A MINLP model to design desalinated water supply systems including solar energy as an energy source

Sebastián Herrera-León[a,b,c*], Andrzej Kraslawski[a,d], Luis A. Cisternas[b,c*]

[a]*School of Business and Management, Lappeenranta University of Technology, Skinnarilankatu 34, Lappeenranta FI-53854, Finland*
[b]*Departmento de Ingeniería Química y Procesos de Minerales,Universidad de Antofagasta, Avenida Angamos 601, Antofagasta 1240000, Chile*
[c]*Centro de Investigación Científico para la Minería,CICITEM, Calle Antonio José de Sucre 220, Antofagasta 1240000, Chile*
[d]*Department of Process and Environmental Engineering, Lodz University of Technology, 116 Żeromskiego Street, Lodz 90-924, Poland*
sebastian.herrera@lut.fi; luis.cisternas@uantof.cl

Abstract

A methodology to design desalinated water supply systems including solar energy as an energy source option is proposed. The methodology consists of two stages and uses a MINLP model. The objective of the first stage is to minimize the total cost of the whole system, while the objective of the second stage is to minimize the greenhouse gases (GHG) emissions related to desalinated water transport by considering the opportunity of installing photovoltaic panels. The MINLP model considers the installation and operation costs of the desalination plants, pumping stations, pipes, and photovoltaic panels. Moreover, it quantifies the GHG emissions produced by the construction and operation of these elements. The mathematical methodology was implemented in the GAMS environment, and the BARON solver was used to obtain the optimal global solution. The design of a desalinated water supply system in the Atacama Desert was considered as a case study to validate the applicability of the methodology proposed. Result of the case study showed that the incorporation of photovoltaic panels in the desalinated water supply system not only reduces the GHG emissions, but also the total costs of the whole system.

Keywords: water supply system; desalination; solar energy; design; MINLP model.

1. Introduction

Nowadays, four billion people in different regions of the world are facing a severe water scarcity (Mekonnen and Hoekstra, 2016). This phenomenon occurs, among other factors, when the water demand is higher than the existing supply in a specific area and it gets more complicated in arid and semi-arid regions, where the water sources are limited or even non-existent. In these regions, seawater (SW) has been seen as an opportunity to satisfy the increasing demand due to economic development and population growth. Although SW could satisfy the quantity of water required by the population, it does not usually meet the required quality standards. To carry out the transformation process from SW to water with high-quality standards, a desalination plant is required.

Desalination plants have been present for several decades providing water for municipal and/or industrial use, commonly in the regions with water scarcity. There are several desalination technologies, however, reverse osmosis (RO) is currently the most used,

representing 63% of the current market (Gude, 2016). SW is a nonconventional water source, whereby the incorporation of desalination plants brings new economic and environmental challenges that should be included in the common water supply management problems. According to Ostfeld (2015), these problems can be classified into three aspects: (1) The layout of the topology or connectivity of the system; (2) the designed size of the system; and (3) the operation of the system.

Several methodologies have been proposed in the literature to face water supply management problems considering SW as a source of supply. Atilhan et al. (2012) developed a methodology based on a linear mathematical model for an integrated layout and operation of water supply systems. The layout includes different water sources, desalination plants, users, wastewater treatment facilities, and water storages. The solution determines the optimal operation of the water supply system and the optimal monthly allocation and storage of water resources. González-Bravo et al. (2015) introduced a methodology based on a nonlinear mathematical model. The objectives were to integrate SW desalination plants with power plants in the layout of water supply systems to maximize the gross annual profit from water and energy sales. González-Bravo et al. (2016) proposed a multi-objective methodology based on a nonlinear mathematical model to integrate SW desalination with power plants in the optimal layout of water distribution systems. The objectives of the study were the maximization of the gross annual profit, the minimization of the overall GHG emissions, and the quantification of jobs generated by the system.

The majority of the studies previously mentioned were focused on the layout of the system. This means: The determination of the system topology or connectivity. However, they did not consider the design of the system that is concerned in its size, mainly to the size of the diameter of the pipes, which has influence on the costs and energy consumptions related to the water transportation. In this context, Herrera et al. (2015) proposed a methodology of two stages based on a MINLP model to design desalinated water supply systems with the objective to determine the location and size of desalination plants, pumping stations, and pipes. In addition, Araya et al. (2016) proposed a methodology based on a MILP model to design desalinated water supply systems with the same objectives including the option to install energy recovery devices. Both studies considered the Atacama Desert as case study mainly for its complex topography (users located at long distances from the ocean and at high altitudes), which makes it challenging to design desalinated water supply systems. In this study, we propose a methodology of two stages based on a MINLP model to design desalinated water supply systems. The objectives of this the study are the minimization of the total cost of the whole system, and the minimization of the GHG emissions related to desalinated water transport. Although this problem can be solved using a multi-objective optimization approach, here, the costs are minimized initially in order to prioritize economic aspects of the system, and then we seek to minimize the environmentally negative effects related to energy consumption. The main novelty of the methodology is the capability to select the option to install photovoltaic panels to supply the energy required for the water transport. The water supply system design in the Atacama Desert is considered as a case study to validate the applicability of the methodology proposed.

2. Mathematical methodology

The first step of the methodology is to define a network that contains all potential layouts of the water supply system. The elements of the networks are the users that determine the

demand, the desalination plants that determine the supply, and the pumping stations that deliver the energy to the system to transport the water. The links of the network represent the pipes with a determined diameter, where the water could be transported. As this study deals with the layout and design of the water supply system, the pipe diameter can be selected from a set of several pipe diameters available. Then, the network is described as a mathematical function that will be optimized in two stages. The objective of the first stage is to minimize the total cost of the whole system determining the location and size of the desalination plants, pumping stations, and pipes to satisfy the water demands of the users. The objective of the second stage is to minimize the GHG emissions considering the opportunity of installing photovoltaic panels to supply the energy related to desalinated water transport.

In the MINLP model four main sets are used: The set of desalination plants (SO), the set of pumping stations (N), the set of users (SI), and the set of discrete pipe diameters (D). The pipes between nodes i, j of the network are established in the set FIJ, with defined distances ($L_{i,j}$) and height differences ($\Delta Z_{i,j}$). The main constraints of the MINLP model are the flow conservation constraints based on the continuity equation, the maximum production capacity of desalination plants, the water requirements of the users, the selection of pipe diameter where a disjunctive constraint is used, and the maximum velocity of water flowing into the pipes. One fundamental equation used in the model is the energy equation expressed in Eq. (1), which results in the hydraulic head ($H_{i,j}$) that is necessary to transport desalinated from one node to another. In this equation, the parameter H_j represents the pressure required in the terminal node, and $h_{n,j}$ represents the pressure loss in the pipe due to friction.

$$H_{i,j} = \Delta Z_{i,j} + H_j + h_{i,j} \qquad \forall\, (n,j) \in FIJ \tag{1}$$

Another fundamental equation used in the model is the power required in the pumping stations to carry out the desalinated water transportation (Eq. 2). The power is a function of desalinated water transport $Q_{i,j}$ from node i to j and the hydraulic head.

$$P_n = \frac{\rho \cdot g}{\eta} \cdot \sum_{j \in output} Q_{i,j} \cdot H_{i,j} \qquad \forall\, n \in N \tag{2}$$

The parameters g, ρ and η represent the gravitational acceleration, the mass density of fluid, and the total efficiency of pumping stations respectively.

Eq. (3) represents the economic objective function that seeks to minimize the total costs of the whole system (ECT), which include the capital and operational costs of the desalination plants ($ECSO_{so}$), pumping stations (ECN_n), and pipes ($ECP_{i,j}$).

$$ECT = \sum_{so \in SO} ECSO_{so} + \sum_{n \in N} ECN_n + \sum_{i,j \in FIJ} ECP_{i,j} \tag{3}$$

The GHG emissions related to desalinated water transport ($OMENCN_n$) are minimized in the second objective function (Eq. 4), where $y1_n$ and $y2_n$ are binary variables used as decision variables to solve the problem. The GHG emissions without the installation of photovoltaic panels ($OMENCN1_n$) are shown in Eq. (5), where T represents the quantity of hours a year using conventional energy as an energy source, and UNE is the unit value of GHG emissions in $kg\ GHG/kWh$ related to the water transport. If the solution of the problem selects to install photovoltaic panels, the binary variable $y2_n$ is activated and Eq. (5) is used to calculate the GHG emissions ($OMENCN2_n$). In Eq. (5), t represents the

percentage of the hours supplied with solar energy. The variable t is equal to EHS/T, where EHS represents the effective hours of sun available in the area under study.

$$OMENCN_n = y1_n \cdot OMENCN1_n + y2_n \cdot OMENCN2_n \quad \forall\, n \in N \tag{4}$$

$$OMENCN1_n = T \cdot UNE \cdot P_n \quad \forall\, n \in N \tag{5}$$

$$OMENCN2_n = (1 - t) \cdot T \cdot UNE \cdot P_n \quad \forall\, n \in N \tag{6}$$

3. Case study

The case study considered the design of a water supply system in the Atacama Desert. This area has relevant water scarcity due to the limited quantity of available water, the population growth, and mainly the economic development pulled by minerals production. In the Atacama Desert, more than 40 mining companies are operating and several more are in the process of being installed. Due to water scarcity, a large part of these mining companies must obtain water from the ocean, which is a problem because several of its operations are located at high altitudes, even at 2,000 meters above sea level. This problem increases the costs of the system, because of an increase the energy required for the water transportation. While some mining operations have used SW directly in their operations, most of them prefer to use desalinated water to not interfere with some of their processes and to expand their facilities. Therefore, a desalination plant is required. Another important aspect is that the Atacama Desert has one of the greatest solar radiations on Earth, which makes it a perfect area to apply the methodology proposed.

The case study considered RO plants and mining operations as desalination plants and users respectively, as shown in Figure 1. The case study includes three desalination plants, 18 pumping stations, and three mining operations. The elevation and distances from the ocean are included in Figure 1, where the dotted lines represent the potential pipes where the desalinated water could be transported (the continuous lines are the optimal solution). The desalinated water demands of each mining operation are 0.6, 1.0, and 0.9 m^3/s, and the maximum capacity of each desalination plant is 2.5 m^3/s. The set of pipe diameters includes three sizes considering commercial availability: 42" (~1 m), 48" (~1.2 m), and 60" (~1.5 m). Monocrystalline silicon photovoltaic panel was considered, and the costs of panels were based on the work of Mariaud et al. (2017).

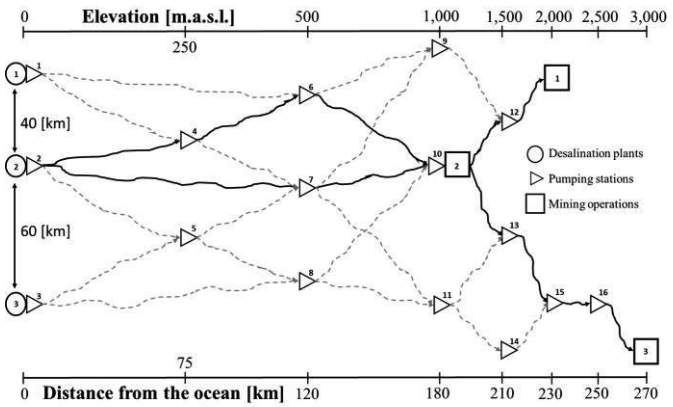


Figure 1: Conceptual representation of case study (dotted lines are potential pipes; the continuous lines represent the optimal solution).

4. Results and discussion

The mathematical methodology was implemented in the GAMS environment, and the BARON solver was used to obtain the optimal global solution. The whole model includes 1,443 single equations, 1,786 single variables, 116 discrete variables, and the solution of the first stage was reached in 2,451.83 seconds, while the solution of the second stage in only 1.81 seconds. The main results obtained are shown in Table 1. Stage 1 considered the minimization of the total costs of the system, while stage 2 considered the minimization of GHG emissions related to water transport. Three different effective hours of sun per day were considered in stage 2, and in each case, the solution selected the installation of photovoltaic panels to decrease the GHG emissions.

The results show that the total costs of the system decrease when the installation of photovoltaic panels is considered, even when the effective hours of sun are limited. Nevertheless, by decomposing the total costs we can observe that the capital costs of the system increase while the operational costs decrease.

Table 1: Results of the case study.

Stage		1	2		
Effective hours of sun			3	6	9
Total costs [million US$/y]	Whole system	361.87	353.59	345.31	337.02
Capital costs [million US$]	RO	1042.82	1042.82	1042.82	1042.82
	PS	261.62	261.62	261.62	261.62
	PIPES	462.48	462.48	462.48	462.48
	PP	0	30.99	61.98	92.98
Operational costs [million US$/y]	RO	73.33	73.33	73.33	73.33
	PS	101.11	88.47	75.83	63.19
	PP	0	1.07	2.14	3.21
GHG emissions by construction [Mt GHG]	RO	181.82	181.82	181.82	181.82
	PS	277.01	309.82	342.64	375.45
	PIPES	80.65	80.65	80.65	80.65
GHG emissions by operation [Mt GHG/y]	RO	261.78	261.78	261.78	261.78
	PS	652.48	570.92	489.36	407.80

RO: Reverse osmosis plants; PS: Pumping stations facilities; PP: Photovoltaic panels.

The increase in the capital costs is clearly due to the installation of photovoltaic panels. Whereas the decrease in the operational costs is because the operational costs of the photovoltaic panels are lower compared to the operational costs of the water transportation, when the energy is obtained from a conventional source. However, the cost to clean solar panels was not considered.

When the GHG emissions of water transportation are minimized, it a reduction of 81.56, 163.12, and 244.68 [Mt GHG/y] was obtained with 3, 6, and 9 effective hours of sun per day, respectively. However, the GHG emissions related to the construction of facilities increased due to photovoltaic panel installation, which is possible to appreciate in the quantities of GHG emissions due to the construction of the pumping stations, where the emissions related to photovoltaic panels were added. In this context, it is clear that the GHG emissions increase in the construction of the system, but the decrease in the operation is much more significant in quantitative terms.

5. Conclusions

A mathematical methodology to design desalinated water supply systems including solar energy as an energy source option was proposed. The methodology was applied to a case study in the Atacama Desert validating its applicability. The results of the case study showed that the incorporation of solar energy in the water supply systems reduces the GHG emissions and also the total costs of the whole system.

Acknowledgments

S.H.L. thanks CICITEM for funding through the project R15A20002 and L.A.C. thanks CONICYT for funding through the Fondecyt 1171341.

References

N. Araya, L. A. Cisternas, F. Lucay, E. D. Gálvez, 2017, Design of desalinated water distribution networks including energy recovery devices, Computer Aided Process Engineering, 40, 925-930.

S. Atilhan, A. Mahfouz, B. Batchelor, et. al., 2012, A system-integration approach to the optimization of macroscopic water desalination and distribution networks: a general framework applied to Qatar's water resources, Clean Technologies and Environmental Policy, 14, 161-171.

V. Gude, 2016, Desalination and sustainability-an appraisal and current perspective, Water research, 89, 87-106.

R. Gonzáles-Bravo, F. Nápoles-Rivera, J. Ponde-Ortega, M. El-Halwagi, 2015, Involving integrated seawater desalination-power plants in the optimal design of water distribution networks, Resources, Conservation and Recycling, 104, 181-193.

R. Gonzáles-Bravo, F. Nápoles-Rivera, J. Ponde-Ortega, M. El-Halwagi, 2016, Multiobjective Optimization of dual-purpose power plants and water distribution networks, ACS Sustainable Chemistry & Engineering, 4, 6852-6866.

S. Herrera, L. Cisternas, E. Galvez, 2015, Simultaneous design of desalination plants and distribution water network, Computer Aided Process Engineering, 37, 1193-1198.

A. Mariaud, S. Acha, N. Ekins-Daukes, N. Shah, C. Christos, 2017, Integrated optimisation of photovoltaic and battery storage systems for UK commercial buildings. Applied energy, 199, 466-478.

M. Mekonnen, A. Hoekstra, 2016, Four billion people facing severe water scarcity. Science advances, 2(2), e1500323.

A. Ostfeld, 2015, Water distribution networks, Intelligent Monitoring, Control, and Security of Critical Infrastructure Systems, 565, 101-124, Heidelberg-Germany.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50277-9

Life cycle assessment of bio-based sustainable polylimonene carbonate production processes

Dongda Zhang[a,*], Ehecatl Antonio del Rio-Chanona[a], Nilay Shah[1]

[a]*Centre for Process Systems Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, UK*
dongda.zhang11@ic.ac.uk

Abstract

Biomass derived polymers are considered as promising candidates to replace petroleum based polymers due to their potential environmental friendliness. To facilitate their application, in this study, a newly proposed biopolymer, polylimonene carbonate, was chosen as the representative to investigate the environmental impacts of the biopolymer production process. Different feedstocks (citrus waste and microalgae) were selected and a comprehensive process design from limonene oxidation to polymer synthesis was completed. Through life cycle assessment, effects of biomass treatment methods, energy integration, and use of solvents on the process environmental impacts were thoroughly discussed. It was found that for sustainable polylimonene carbonate synthesis, a more environmentally-friendly and energy-efficient limonene oxidation method should be developed. Based on the economic analysis, the polymer's cost was estimated to be around 1.36 to 1.51 $/kg, indicating its great potential as a substitute for petrol-based polystyrene. Moreover, this study found that both feedstock selection and biowaste treatment method significantly affect the process environmental impacts, and a carbon negative biopolymer can be achieved when the remaining waste is used for energy generation. Therefore, a new concept that considers CO_2 as an efficient solar energy carrier for future sustainable process design is proposed in this study.

Keywords: biomass, life cycle assessment, polylimonene carbonate, process simulation, economic analysis.

1. Introduction

Polymers are one of the most widely used commodities in both industry and daily life. With the rapid growth and development of advanced polymerisation technologies, a variety of polymers have been produced with substantial applications in different areas including electronics, healthcare products, toys, packaging materials, and coatings (Hottle et al. 2013; Zhu et al. 2016). To improve the process sustainability and mitigate environmental damages caused by the traditional polymer synthesis routes, more environmentally friendly raw materials, in particular biomass, have been chosen to replace fossil fuels for biopolymer production within the last two decades (Zhu et al. 2016; Miller 2013). For instance, a diversity of bio-based polymers such as polycarbonates, polyesters, and polyurethane (Hauenstein et al. 2016; Winkler et al. 2015) have been synthesised through different biomass feedstock including agricultural biomass, citrus wastes, and microorganisms.

Amongst different newly proposed biopolymers, polylimonene carbonate (PLC) has been given significant attention within the last five years due to its outstanding

characteristics compared to other candidates (Peña Carrodeguas et al. 2015; Hauenstein et al. 2016). For example, the precursor of PLC is limonene, a sustainable chemical that can be produced via different biomass feedstocks including orange peels and microalgae. Furthermore, CO_2 consumption is directly involved in the reaction steps for both limonene and PLC production, suggesting its promising potential as a carbon negative and environmentally friendly polymer. In addition, previous study has claimed that PLC has similar properties to polystyrene (Dalaeli et al. 2005), one of the most widely used fossil based polymers. However, the environmental impacts of the entire PLC production process ranging from feedstock selection to downstream PLC synthesis have not been explored so far, remaining the process sustainability, economic viability, and industrialisation feasibility an undetermined question. As a result, the current study aims to conduct a detailed PLC production process design and life cycle assessment to complete this open challenge and facilitate the industrialisation of PLC production.

2. Methodology

2.1. Process descriptions

The entire polylimonene carbonate (PLC) production process includes 4 operating sections (shown in Fig. 1, excluding limonene production): limonene production, tert-butyl hydroperoxide production, limonene oxide production, and PLC synthesis.

Limonene production from citrus and microalgae: Limonene is sold as a by-product from the orange juice industry because of its high content in orange peels (3.8 wt %) (Ciriminna et al. 2014). In a recent study (Roberts 2012), a citrus limonene production process was analysed from fertilising orange trees to final waste management. Therefore, this process is chosen as the standard route for limonene production from citrus in the current research (Roberts 2012). Meanwhile, microalgae been genetically modified to produce limonene (Davies et al. 2014; Jahandideh et al. 2017). In a very recent study (Johnson et al. 2016), a detailed process design for algal limonene and energy co-production was proposed. As a result, this process is selected to compare against the citrus limonene production process. In the algal process, limonene is considered as the major product, and algal biomass waste is sent to anaerobic digestion for hydrogen and methane production. The biogas is then used to generate electricity and heat. The detailed introduction of this process can be found in the recent study (Jahandideh et al. 2017). It is important to emphasise that in this study, the highest algal limonene productivity (Davies et al. 2014), 4.5 mg L^{-1} h^{-1}, is used to replace that used in the recent study, so that this process describes better the reality.

Limonene oxide and tert-butyl hydroperoxide production: Limonene oxide is the direct precursor for PLC synthesis. Tert-butyl hydroperoxide (TBHP) was recommended as a suitable oxidant for limonene oxidation because of its mild operating conditions (room temperature, 1 bar), high single-pass conversion (75 %) and selectivity (88 %) (Dalaeli et al. 2005). As this reaction has only been conducted under laboratory scale experiments, it is assumed that reaction selectivity and conversion efficiency in large scale systems are also identical. On the other hand, tert-butyl hydroperoxide (TBHP) is commercially derived from tert-butyl alcohol (TBA) oxidation by hydrogen peroxide. As TBA is the reduction product of TBHP in the limonene oxidation section, in practice only hydrogen peroxide needs to be provided to oxidise limonene, and TBA/TBHP actually acts as the electron carrier. The reaction happens at room temperature and 1 bar

in an acidic environment. The single-pass TBHP yield is high (90 %) (Bell & Vaughan 1953).

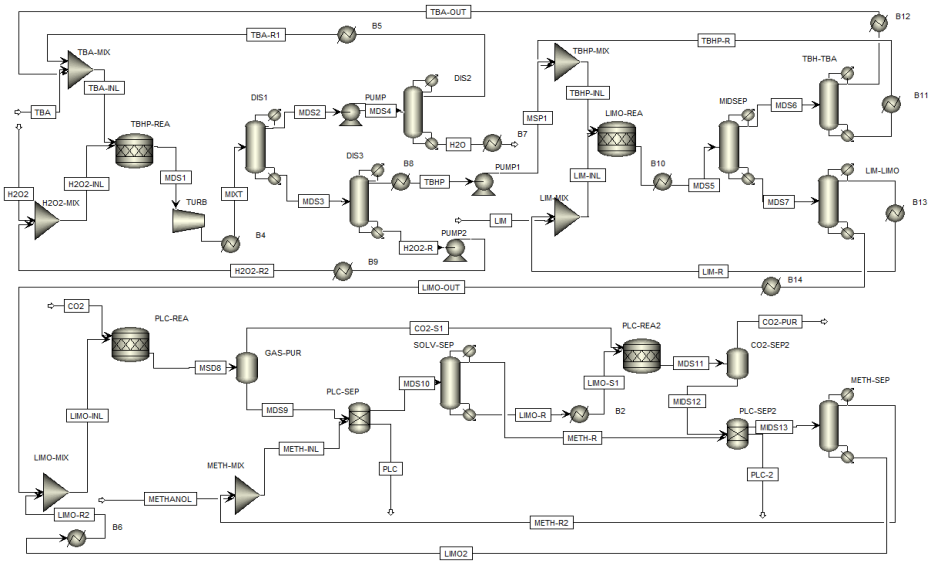


Figure 1: PLC production process (limonene oxidation, tert-butyl hydroperoxide production, and PLC synthesis)

Polylimonene carbonate production: Finally, PLC is synthesised by CO_2 and limonene oxide at a high pressure of 7 bar and ambient temperature (Dalaeli et al. 2005; Hauenstein et al. 2016). The single-pass conversion of this polymerisation process is 50 % (limonene oxide) (Dalaeli et al. 2005). The catalyst (a zinc based homogeneous catalyst) (Hauenstein et al. 2016) is dissolved in limonene oxide.

2.2. Life cycle assessment

A cradle-to-gate life cycle assessment is used to estimate the environmental impacts of PLC production processes (Yates & Barlow 2013). The primary goal of this LCA is to investigate whether the newly proposed biopolymer synthesis routes are environmentally friendly. In particular, to provide a realistic analysis for PLC production at its early research stage, two different biomass resources, microalgae and citrus wastes, are compared. More importantly, to explore the process carbon footprint and solar energy utilisation efficiency, different biowaste disposal scenarios are included in this study. The function unit of this LCA is set to be producing 21,600 tonnes PLC (eq. 15,000 tonnes limonene). In addition, when using citrus to produce limonene, only 1 % of environmental impacts is allocated to limonene, since it is a waste derived by-product and its production is much lower than orange juice (Roberts 2012). Whilst if the feedstock is algae, limonene should be regarded as the major product, since the process is particularly designed for limonene production (Jahandideh et al. 2017). As a result, an allocation factor of 66 % is given to limonene, and the remaining 34 % is allocated to the biogas derived through anaerobic digestion. From limonene oxide production to PLC production, the only by-product is water, which is assumed to be used inside the process. Thus, the environmental impacts are 100 % allocated to PLC.

2.3. Software

Process design and simulation (from limonene to PLC production) were executed in Aspen Plus V9. This aims to estimate energy consumption and mass balances in each process section. An analysis on the separation of limonene from orange peels has been completed in (Dalaeli et al. 2005), thus this is not included in the current simulation. Moreover, a preliminary evaluation of energy integration was conducted in this software to reduce the energy cost during the product separation sections (distillations), together with an approximation of process capital and operation costs. LCA was performed in Simapro 8.1, and the Ecoinvent 3.2 LCA database was used. The ReCiPe Midpoint method was used due to its wide applications (Roberts 2012; Keller et al. 2015).

3. Methodology

3.1. Effect of feedstock and biomass treatment method

To explore the importance of different feedstocks and biowaste treatment method on PLC production process environmental impacts, three scenarios are selected for comparison. These use citrus waste as cattle feed, algae biomass for energy production, and algae biomass as animal feed. Several conclusions are observed through LCA (Fig. 2). First, it is found that when biomass is sold as animal feed, no matter which feedstock is chosen for limonene synthesis, the PLC production process always emits CO_2 on a nett basis and has negative impacts to the environment (1.21 kg CO_2 eq./kg PLC and 11.99 kg CO_2 eq./kg PLC from the citrus and the algae process, respectively). Thus, biopolymer produced from these processes cannot achieve carbon negativity. On the contrary, when biomass is used for energy generation, it is found that not only is the biopolymer produced carbon negative, but also in most cases the process is environmentally friendly. This is because both algae and citrus are derived from solar energy and CO_2 via photosynthesis. Therefore, a significant amount of solar energy is stored inside biomass. For example, based on the recent work (Jahandideh et al. 2017), energy generated through anaerobic digestion (AD) in this PLC production process is 751.4 MJ kg^{-1} PLC (mainly electricity), over 30 times the energy required for the entire PLC production process (approximate 22.0 MJ kg^{-1} PLC, mainly heat). Hence, when algal biomass is converted back to CO_2, the stored solar energy is released and converted to electricity and heat. Thus, the obtained energy not only fulfils all the energy required for PLC synthesis, but can also be sent to electricity grids for other processes displacing other generation. However, if biomass is sold as animal feed, the stored solar energy will not be able utilised, and the process becomes less sustainable.

3.2. Effect of solvent, oxidant and energy integration

In this study, methanol is used as a solvent to precipitate PLC, and TBHP is used as the direct oxidant for limonene oxidation. As introduced before, both of them are recycled inside process without being emitted. Therefore, it is plausible to exclude the environmental impacts of these chemicals, since their purchase is one-off and their required amount is much lower than other chemicals. However, through LCA, it is found that although methanol only has negligible effect to the system, the use of TBHP shows a dramatic impact to the current process, in particular climate change, human toxicity, and ecotoxicity. For example, it causes a 156.7 % and 1603.4 % increase on climate change and human toxicity, respectively, in the citrus based PLC production process. This clearly suggests that although TBHP/TBA is recycled inside the process and only employed as the electron carrier to facilitate the oxidation-reduction reaction

between limonene and hydrogen peroxide, its use due to its current synthesis route could significantly aggravate the environmental impacts of PLC production process. As a result, more environmentally friendly routes should be developed for TBHP/TBA synthesis, and more effort should be given to research focusing on the direct reaction between limonene and hydrogen peroxide. In addition, it is found that through energy integration, in total 20.8 % of energy input can be saved within the PLC production process. Nonetheless, apart from the mild improvement on climate change (6.2 %) and ozone depletion (15.2 %), for most environmental categories, energy integration does not lead to distinguishable benefit to the process. This indicates that energy integration may not serve as the best option to improve the current process environmental-friendliness, and meanwhile could increase the process investment cost due to the addition of heat exchangers and other auxiliary equipment.

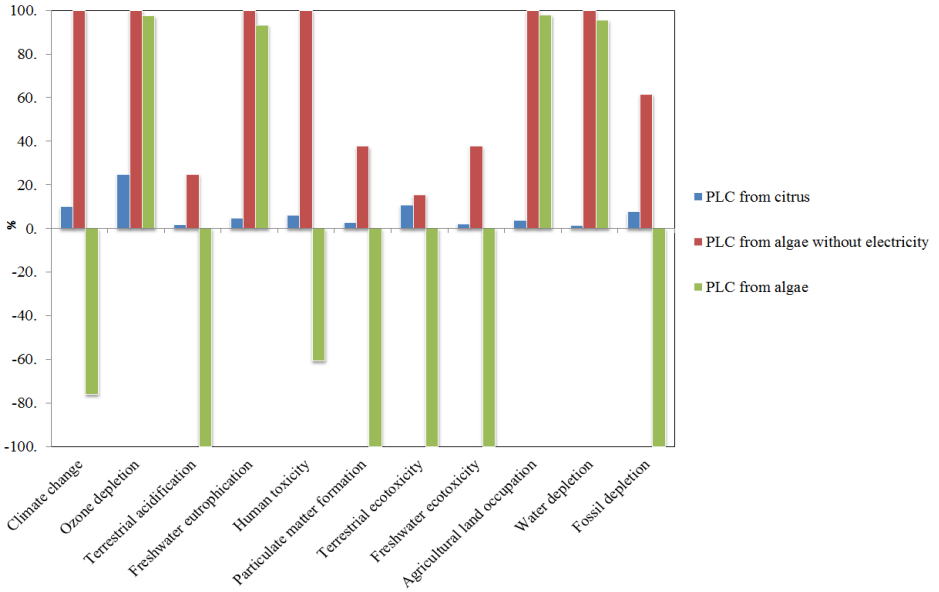


Figure 2: LCA of PLC production with different feedstocks and waste treatment methods. “PLC from algae” is the process where biomass is used to generate energy.

3.3. Process economic analysis

Finally, a brief analysis on the cost of different PLC production routes is carried out in this study. Process capital and operating costs are estimated from Aspen Plus V9, and prices of raw materials are obtained from manufacturer websites and recent studies (CO2 price is assumed to be zero) (Dalaeli et al. 2005; Johnson et al. 2016). From the study, it is estimated that the cost of PLC is around 1.36 to 1.51 $/kg, depending on the limonene feedstock. This result is higher than the price of polyethylene (0.91 S/kg), but falls within the price range of polystyrene (1.2 to 1.6 $/kg), a traditional commercial polymer that has similar properties to PLC but based on conventional fossil resources. Overall, based on the current process design, life cycle assessment, and process economic analysis, it is concluded that there is a great potential to replace fossil fuel based polystyrene by biomass derived PLC in the future.

4. Conclusion

Eventually, from the current research, a new concept is proposed for future sustainable process design, which is compared to exclusively focusing on the limited amount of biowaste derived sustainable chemicals as well as their downstream processing efficiencies, it is more crucial to identify the most effective way to maximise the utilisation efficiency of the huge amount of solar energy stored inside biowaste. In other words, this concept suggests that compared to only considering CO_2 derived biomass as the feedstock for sustainable chemicals synthesis, it is more important to also consider these biomass as an efficient solar energy storage to systematically optimise the process energy and materials utilisation efficiency.

References

E. Belland W. Vaughan, 1953. Preparation of organic hydroperoxides. p.6.

R. Ciriminna, et al., 2014. Limonene: a versatile chemical of the bioeconomy. Chem. Commun., 50(97), pp.15288–15296.

J. Dalaeli, C. Fogle, and J. Yang, 2005. Polymers from Oranges: Design and Feasibility of Polymer Production from Orange Oil Derivatives, Oklahoma.

F. K. Davies, et al., 2014. Engineering Limonene and Bisabolene Production in Wild Type and a Glycogen-Deficient Mutant of Synechococcus sp. PCC 7002. Frontiers in Bioengineering and Biotechnology, 2.

O. Hauenstein, et al., 2016. Bio-based polycarbonate from limonene oxide and CO 2 with high molecular weight, excellent thermal resistance, hardness and transparency. Green Chem., 18(3), pp.760–770.

T. A. Hottle, M.M. Bilec, and A. E. Landis, 2013. Sustainability assessments of bio-based polymers. Polymer Degradation and Stability, 98(9), pp.1898–1907.

A. Jahandideh, et al., 2017. Life cycle analysis of a large-scale limonene production facility utilizing filamentous N 2 -fixing cyanobacteria. Algal Research, 23, pp.1–11.

T. J. Johnson, et al., 2016. Producing next-generation biofuels from filamentous cyanobacteria: An economic feasibility analysis. Algal Research, 20, pp.218–228.

H. Keller, N. Rettenmaier, and G. A. Reinhardt, 2015. Integrated life cycle sustainability assessment – A practical approach applied to biorefineries. Applied Energy, 154, pp.1072–1081.

S. A. Miller, 2013. Sustainable Polymers: Opportunities for the Next Decade. ACS Macro Letters, 2(6), pp.550–554.

L. Peña Carrodeguas, et al., 2015. Al III -Catalysed Formation of Poly(limonene)carbonate: DFT Analysis of the Origin of Stereoregularity. Chemistry - A European Journal, 21(16), pp.6115–6122.

G. Roberts, 2012. Life Cycle Assessment - Renewable and Sustainable Citrus Oils Final Report, Chicago.

M. Winkler, et al., 2015. Renewable polycarbonates and polyesters from 1,4-cyclohexadiene. Green Chem., 17(1), pp.300–306.

M. R. Yates, and C. Y. Barlow, 2013. Life cycle assessments of biodegradable, commercial biopolymers—A critical review. Resources, Conservation and Recycling, 78, pp.54–66.

Y. Zhu, C. Romain, and C. K. Williams, 2016. Sustainable polymers from renewable resources. Nature, 540(7633), pp.354–362.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50278-0

Process Intensification of Algae Oil Extraction to Biodiesel

Geetanjali Yadav[a], Warren D. Seider[a*], Lindsay Soh[b], Julie Zimmerman[c]

[a]*University of Pennsylvania, Department of Chemical and Biomolecular Engineering, Philadelphia, Pennsylvania, 19104-6393, United States*
[b]*Lafayette University, Department of Chemical and Biomolecular Engineering, Easton, PA 18042, USA*
[c]*Yale University, Department of Chemical and Environmental Engineering, New Haven, CT 06511, USA*
seider@seas.upenn.edu

Abstract

Dilute algae are often grown in raceways or bioreactors by photosynthesis using carbon dioxide in flue gas. After being harvested, algae oil must be extracted before being transesterified to biodiesel. This process, which generates a carbon-neutral, renewable, liquid fuel, has been shown to be competitive with other liquid fuel sources (Silva et al., 2014). Herein, we introduce the use of carbon dioxide microbubbles to lyse algae cell walls, releasing triglyceride oils, combined with their transesterification using methanol in a single vessel. To obtain more reliable multiphase equilibrium calculations to predict the phase distribution (often vapor-liquid-liquid), we use the SAFT equation-of-state (the latest PC-SAFT in ASPEN PLUS and gSAFT in gPROMS) with new experimental data, where necessary. This permits the use of an extended kinetic model to compute biodiesel production rates. At lower temperatures and intermediate pressures, carbon dioxide causes methanol to dissolve in the triglyceride phase and triglyceride to dissolve in the methanol phase, increasing the transesterification reaction rate. This brief paper describes the status of our research without providing the detailed model and results.

Keywords: multi-phase equilibrium, SAFT EOS, algae biodiesel, transesterification, supercritical CO_2.

1. Introduction

1.1 Background

Increasingly, fossil fuels are being replaced with sustainable and renewable alternatives like microalgae, for the production of transportation fuels like biodiesel. Microalgae offer several advantages as compared to fossil-based counterparts including carbon-neutral re-use of CO_2. Biodiesel, comprised of fatty-acid methyl esters (FAMEs), is produced from the transesterification of triglycerides with methanol, yielding the glycerol byproduct.

In our earlier work (Silva et al., 2014), a process was designed to carry out transesterification using a solid-base catalyst at 1 atm with rate data regressed in a kinetic model. Using cost estimates from the Aspen Process Economic Analyzer (APEA) and industrial quotes, a cash flow analysis was performed for an entire carbon sequestration-to-biodiesel production train, yielding a competitive biodiesel selling price of $4.34/gal.

Subsequently, Silva et al. (2016) demonstrated that transesterification of algal oil (triglycerides) using methanol is accelerated using supercritical CO_2 at moderate temperatures and pressures. Recently, we have introduced the carbon dioxide in microbubbles to lyse the algae cell walls, releasing triglyceride oils. Our research focuses on the intensification/combination of these two processes to reduce the energy consumption and capital costs of processes to generate biodiesel. Initially, we are carrying out more reliable multiphase equilibrium calculations to predict the phase distribution (as simple as a single supercritical phase or as complex as V-L_I-L_{II}) using the SAFT equations-of-state. It has been shown experimentally in the laboratory (Soh et al., 2014, 2015) and using simulation software (ASPEN PLUS and gPROMS – Silva et al., 2016) that at intermediate pressures, CO_2 causes triglyceride to dissolve in the methanol phase, significantly increasing the transesterification reaction rate. But, additional data are needed to improve estimates of the rates. Also, smaller methanol/triglyceride ratios are needed, yielding a triglyceride phase that dissolves methanol.

They used an approximate single-reaction, kinetic mechanism, as shown in Figure 1. This model included just six chemical species: triolein, methanol, methyl oleate, glycerol, CO_2, and water. Their large methanol/triolein molar ratios simplified the experiments, but were commercially unrealistic. Note that Freedman et al. (1984) report the industrial use of methanol/triolein molar ratio = 6 and Maeda et al. (2011) report even smaller molar ratios for methanol/canola oil. Also, because algae grow profusely in dilute aqueous media, the presence of water cannot be avoided.

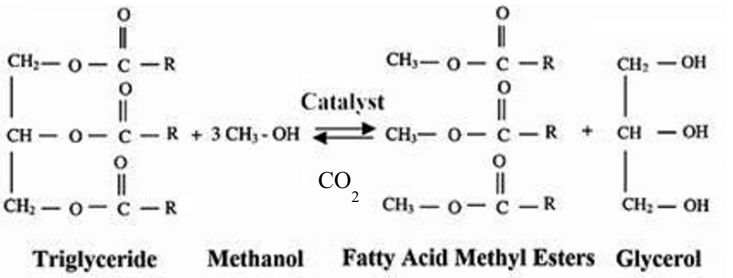


Figure 1. A single reversible transesterification reaction to convert triglycerides into biodiesel

In this paper, the mechanism is extended to include three reversible reactions (R_1, R_2, and R_3 in Figure 2) with eight chemical species, including diolein and monoolein. In our earlier work, a crude model was used, the Redlich-Kwong equation in ASPEN PLUS, which is not suitable in the critical region (Silva et al., 2016). To compute concentrations of diolein and monoolein in three phases at equilibrium, routine liquid density and liquid vapor pressure measurements are underway. These data are being regressed to obtain pure-species PC-SAFT and gSAFT parameters. Additionally, we are measuring LLE and VLE data for binary mixtures including diolein and monoolein. Their binary interaction parameters are being regressed to improve the multi-phase equilibrium calculations. Note that pure-species parameters for PC-SAFT were published by Perez-Cisneros et al. (2016), but have not yielded realistic LLE and VLE results in our ASPEN PLUS calculations.

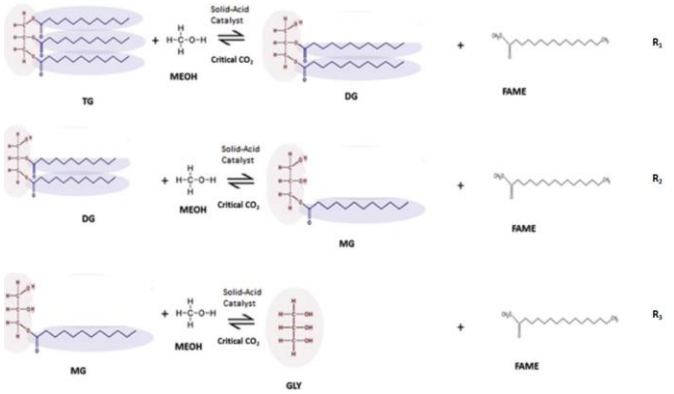


Figure 2. Kinetic mechanism involving three reversible reactions. TG is triglyceride (triolein), MEOH is methanol, DG is diglyceride (diolein), FAME is fatty-acid methyl-ester (methyl oleate – biodiesel), MG is monoglyceride (monoolein), GLY is glycerol, and critical CO_2 is in or near the critical region

2. Multiphase Kinetics Model

To model a batch reactor, involving up to three phases (V, L_I, and L_{II}), for each phase, the mass balances are expressed:

$$\frac{d[C_i]}{dt} = \nu_{R_{j,i}} * r_j \tag{1}$$

where i is the species counter in reaction (R_j), j = 1, 2, 3, and the intrinsic rates of reaction are expressed in:

$$r_1 = k_{1\mathrm{f}}[cat]^{n_{1\mathrm{f}}}[triolein][methanol] - k_{1\mathrm{r}}[cat]^{n_{1\mathrm{r}}}[diolein][methyl-oleate]$$

$$r_2 = k_{2\mathrm{f}}[cat]^{n_{2\mathrm{f}}}[diolein][methanol] - k_{2\mathrm{r}}[cat]^{n_{2\mathrm{r}}}[monoolein][methyl-oleate]$$

$$r_3 = k_{3\mathrm{f}}[cat]^{n_{3\mathrm{f}}}[monoolein][methanol] - k_{3\mathrm{r}}[cat]^{n_{3\mathrm{r}}}[glycerol][methyl-oleate] \tag{2}$$

where $k_{j\mathrm{f}}$ is the forward rate constant, $k_{j\mathrm{r}}$ is the reverse rate constant, [cat] is the catalyst concentration, $n_{j\mathrm{f}}$ is the exponent of the catalyst concentration in the forward direction, $n_{j\mathrm{r}}$ is the exponent of the catalyst concentration in the reverse direction, $[C_i]$ is the concentration of species i, and $\nu_{Rj,i}$ is the stoichiometric coefficient of species i in reaction R_j. In the absence of catalyst, the n coefficients are zero.

Prior to each time-step, when integrating mass balances (1) for each phase, the concentrations in the three phases and phase volumes are computed by minimizing G (Gibbs free energy). It is necessary to check for phase stability using the Gibbs tangent-plane distance criterion, as discussed by Silva et al. (2016). These calculations are completed using PC-SAFT in ASPEN PLUS and gSAFT in gPROMS, as summarized in Section 3. Experimental concentration measurements for the three phases (see Section 4) in the batch reactor are being taken to regress the rate constants in Eq. (2).

3. Equations of State (EOS)

3.1 PC-SAFT EOS

The EOS is formulated in terms of the Helmholtz free energy, as shown in:

$$\frac{A}{Nk_BT} = \frac{A^{ig}}{Nk_BT} + \frac{A^{hc}}{Nk_BT} + \frac{A^{disp}}{Nk_BT} + \frac{A^{assoc}}{Nk_BT} \quad (3)$$

where the ideal-gas contribution is augmented with hard-chain, dispersion, and association contributions. Although the PC-SAFT model is based on advanced molecular theory, only three pure-species parameters (m, σ, and ε) are required to model both phase equilibria and bulk properties of hydrocarbon mixtures.

3.2 SAFT-γ Mie

Like PC-SAFT, SAFT-γ Mie is formulated in terms of Helmholtz free energy. However, its formulation includes monomeric functional-group interactions, rather than a dispersion contribution:

$$\frac{A}{Nk_BT} = \frac{A^{ig}}{Nk_BT} + \frac{A^{mono}}{Nk_BT} + \frac{A^{chain}}{Nk_BT} + \frac{A^{assoc}}{Nk_BT} \quad (4)$$

In our current work, we examine the performance of the PC-SAFT and SAFT-γ Mie approach for the description of a broad variety of properties of pure fluids (such as liquid density and vapour pressure for diolein and monoolein) and binary mixtures including phase behaviour (VLL, VL, LL). Comparisons of the calculations are made with the experimental data in the development of the group parameters and for predictions of properties not used in the characterization of the groups.

4. Experimental

4.1. Vapor pressure measurement: Transpiration method

Vapor pressures of monoolein and diolein were determined using the method of transference in a saturated stream of inert gas and applying the Clausius-Clapeyron equation. A sample of approximately 0.5 g was mixed with glass beads (0.25 g) and placed in a U-tube 10 cm long and 0.5 cm in diameter having a thermostadt. A preheated helium stream was passed through the U-tube at different temperatures at the flow rate of 8 dm^3 h^{-1} which ensured that the carrier gas was in saturated equilibrium with the coexisting liquid phase in the saturation tube (Emel'yanenko et al. 2007). The material was transported in a cold trap upon bubbling through the isopropanol solvent and analyzed by supercritical fluid chromatography-mass spectroscopy (SFC-MS).

4.2. Liquid density measurement

Density was measured using a 5 mL pycnometer and a thermostatically controlled water bath was used to control the temperature (Ihmels and Lemmon, 2007).

4.3. Vapor-liquid equilibria

A 500 mL bulb connected to a condenser is being used to measure VLE. The bulb is heated from below with an electric heating jacket, and cooling water is circulated around the condenser at 7°C. The well-insulated apparatus is stirred by a magnetic stirrer in the liquid phase and heated to allow boiling and equilibration for 1 hour. Samples were collected and analyzed by SFC-MS.

4.4. Liquid-liquid equilibria

The binary pairs involving diolein and monoolein were added at equal mass loadings to a 10 mL separatory funnel that was used for equilibrium and sampling (Silva, et al., 2016). The top and bottom phases were sampled using a pipette and stopcock, respectively. The collected samples were analyzed by SFC-MS.

Recently, PC-SAFT calculations were repeated using ASPEN PLUS Version 10 for the triolein-water binary pair. As shown in Figure 3, good agreement is obtained with the experimental data (better than obtained by Silva et al. (2016).) Similar comparisons have been completed and, as needed, improved experimental measurements are planned for ternary LLE.

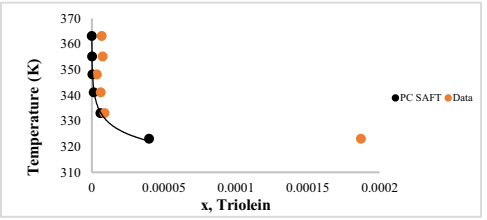


Figure 3. LLE – triolein in water at 1 atm

5. Microbubbles for Oil Extraction

Traditional methods for cell disruption include bead milling, microwave-assisted, sonication, hydrodynamic cavitation, and atomic force microscopy (AFM.) Recently, microbubbles have been proposed as promising candidates for improving cell-disruption efficiency and reducing energy requirements. In this regard, Krehbiel et al. (2014) showed that ultrasound cell disruption, with low-frequency waves (at ~10 kHz), increased by up to 58-% through the addition of microbubbles, with peak disruption occurring in the vicinity of 10^8 microbubbles/mL. The fluid motion due to the bubbles and shock waves that accompany bubble collapse lead to algal cell disruption (Krehbiel et al., 2014).

Methods for lipid extraction have energy requirements varying from $3.3 \times 10^7 – 5.3 \times 10^8$ J/kg of dry biomass. Furthermore, conventional cell disruption techniques often lead to negative net energy balances for biofuel production. To circumvent this, microbubbles are expected to advance significantly the efficiency of algal-cell disruption.

Of special note, supercritical CO_2 (SC-CO_2), for intensification of extraction and transesterification, shows great promise for replacing traditional organic solvents. SC-CO_2 accelerates the formation of the FAMEs in the transesterification reaction between two immiscible phases, oil and methanol, in a shorter extraction time compared with the use of organic solvents (Soh and Zimmerman, 2011). So, in this work, we describe the potential of CO_2 microbubbles for lipid extraction by cell disruption and as a co-solvent for the transesterification reaction that converts lipids into biodiesel. Experiments are planned to carry out extraction and transesterification in a single vessel. Also, calculations are underway to locate reactor temperatures and pressures that reduce the selling price of methyl-oleate (biodiesel) below $4.34/gal, as estimated using a basic-solid catalyst at 82°C and 1 atm (Silva et al., 2014).

6. Conclusions

To improve process design, a three-reversible reaction model involving diolein, monoolein, and triolein using PC-SAFT and gPROMS has been developed. It will improve kinetic modelling of the intensified extraction/transesterification process by enabling multiphase equilibria involving VLLE. This will permit an accurate determination of the phase distribution when triolein is present in sufficiently large quantities to create a second liquid phase. Binary VLE and LLE data for the eight species are used to obtain binary interaction parameters in PC-SAFT and cross parameters in SAFT-γ Mie. Herein, for the triolein-water binary, improved agreement of LLE calculations using PC-SAFT in ASPEN PLUS with experimental data has been achieved (compared with the calculations of Silva et al. (2016).) Additional comparisons are planned for other binary pairs, and where appropriate, additional experimental data will be measured. Furthermore, capital cost estimates and energy requirements for the intensified extraction/transesterification process are being computed.

References

B. Freedman, E. H. Pryde, T. L. Mounts, 1984. Variables affecting the yields of fatty esters from transesterified vegetable oils. J. Amer. Oil Chem. Soc. 61, 1638-1643.

V. Emel'yanenko, S. P. Verevkin, A Heintz, 2007. The Gaseous Enthalpy of Formation of the Ionic Liquid 1-Butyl-3-methylimidazolium Dicyanamide from Combustion Calorimetry, Vapor Pressure Measurements, and Ab Initio Calculations. J. Am. Chem. Soc. 129, 3930-3937

E.C. Ihmels, E.W. Lemmon, 2007. Exper. densities, vapor pressures, and critical point, and a fundamental eqn. of state for dimethyl ether. Fluid Phase Equil., 260, 36-38

J. D. Krehbiel, L. C. Schideman, D.A. King, J. B. Freund, 2014. Algal cell disruption using microbubbles to localize ultrasonic energy. Bioresource Tech., 173, 448–451.

Y. Maeda, L. T. Thanh, K. Imamura, K. Izutani, K. Okitsu, L. V. Boi, P. N. Lan, N. C. Tuan, Y. E. Yoo, N. Takenaka, 2011. New technology for the production of biodiesel fuel. Green Chem. 13, 1124-8.

E. S. Perez-Cisneros, X. Mena-Espino, V. Rodriguez-Lopez, M. Sales-Cruz, T. Viveros-Garcia, R. Lobo-Oehmichen, 2016. An integrated reactive distillation process for biodiesel production. Comput. Chem. Eng. 91, 233-246.

C. Silva, L. Soh, A. Barberio, J. Zimmerman, W.D. Seider, 2016. Phase equilibria of triolein to biodiesel reactor systems. Fluid Phase Equillibria, 409, 171-192.

C. Silva, E. Soliman, F. Cameron, L.A. Fabiano, W.D. Seider, E.H. Dunlop, A.K. Coaldrake, 2014. Commercial-scale biodiesel production from algae. Ind. Eng. Chem. Res., 53, 5311-5324.

L. Soh, J. Curry, E.J. Beckman, J. Zimmerman, 2014. Effect of system conditions for biodiesel production via transesterification using CO_2−methanol mixtures in the presence of a heterogeneous catalyst. ACS Sustain. Chem. and Eng., 2, 387-395.

L. Soh, C. Chen, T.A. Kwan , J. Zimmerman, 2015. Role of CO_2 in mass transfer, reaction kinetics, and interphase partitioning for the transesterification of triolein in an expanded methanol system with heterogeneous acid catalyst. ACS Sustain. Chem. and Eng., 3, 2669-2677.

L. Soh, and J. Zimmerman, 2011. Biodiesel production: the potential of algal lipids extracted with supercritical carbon dioxide. Green Chem., 13, 1422–9.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50279-2

Toward Optimal Synthesis of Renewable Ammonia and Methanol Processes (RAMP)

William W. Tso[a,b], C. Doga Demirhan[a,b], Joseph B. Powell[c], Efstratios N. Pistikopoulos[a,b,*]

[a]*Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX, USA*
[b]*Texas A&M Energy Institute, Texas A&M University, College Station, TX, USA*
[c]*Shell Technology Center, Shell Oil Company, Houston, TX*
stratos@tamu.edu

Abstract

The conversion and storage of renewables into useful and transportable energy vectors is an outstanding challenge in transitioning to a sustainable energy future. Ammonia and methanol are two promising vectors due to their proven production and transportation infrastructures. Individual processes consuming fossil fuels have been improved and optimized over the years, but there is no clear path forward for incorporating renewables or their coproduction. In this work, we introduce a deterministic global optimization-based process synthesis framework to determine the minimal cost for the coproduction of ammonia and methanol from hardwood biomass and natural gas. This will provide the guideline for future studies into solar and wind energy. A process superstructure with several competing technologies and integration routes is modeled and described. A customized branch-and-bound algorithm is used to solve the resulting mixed-integer nonlinear nonconvex optimization model. Results indicate that the biomass incorporated processes are cost-competitive with pure fossil fuel based processes, and the integration of ammonia and methanol production leads to 4-7% savings.

Keywords: Biomass, methanol, ammonia, process synthesis, global optimization

1. Introduction

Reducing mankind's dependency on fossil fuels in favor of an increasing share of renewables is an important shift for realizing a sustainable energy supply chain. While a transition to fully renewable energy production looks possible in the future, there will likely be a transition period, where renewables like solar, wind, or biomass are integrated alongside fossil fuels into the energy supply chain. In this respect, converting renewables into energy vectors that can efficiently store and transport the energy is a priority for creating an economically attractive sustainable energy supply chain (Floudas et al. 2016).

Although hydrogen is a promising energy vector to store renewables due to its high gravimetric energy storage density (33 kWh/kg), its low volumetric storage density (3 Wh/L under ambient conditions) makes storage and transportation problematic. Either extreme pressures or cryogenic temperatures are required for gaseous or liquid state storage, respectively. These two extreme storage conditions have motivated investigation into alternative energy vectors that can store hydrogen in their chemical structures, such as ammonia, methanol, liquid organic hydrogen carriers (LOHCs), and metal hydrides (Teichmann et al. 2012). Among these energy vectors, ammonia has a high capacity for hydrogen storage, 17.6 wt.% based on its molecular structure and a hydrogen volume

density of 105.0 kg/m^3, which is about 45 % higher than that of liquid hydrogen, which is reported as 71.2 kg/m^3. Similarly, methanol has 12.6 wt.% hydrogen based on its molecular structure and its liquid hydrogen volume density is 99.8 kg/m^3 (Thomas and Parks, 2006).

In addition to the high hydrogen content of ammonia and methanol, their storage and transportation infrastructures are also well-established. Therefore, we focus here on a transition-period ammonia and methanol coproduction plant in which both natural and biomass are used as feedstocks. Future studies will investigate solar and wind integration. Coproducing ammonia and methanol has received much attention. In 2017, Haldor Topsoe opened up the first such plant in Russia. Both production processes utilize syngas generated from steam reforming of methane, but the syngas quality in each is quite different. While methanol production desires a stoichiometric ratio of H_2/CO, ammonia production involves more complex gas purification steps to prevent oxide species from poisoning the catalyst in its synthesis converter. Typical ammonia processes also emit more CO_2 than typical methanol processes. However, ammonia contains more hydrogen per mass in its chemical structure than methanol does. Therefore, there are interesting tradeoffs to be investigated in the integrated process.

In order to determine the economic and technical feasibility of incorporating biomass into an integrated ammonia and methanol production, a global optimization-based process synthesis approach is used to determine the optimal process topologies while minimizing the annualized cost of energy carrier production. A superstructure of process alternatives is modeled as a large-scale mixed-integer nonlinear optimization (MINLP) model, with simultaneous heat, power, and water integration. The nonconvex MINLP model is solved to global optimality by using a branch-and-bound global optimization algorithm.

2. Biomass/Natural Gas to Ammonia and Methanol Process Superstructure

2.1 Natural Gas Reforming

Natural gas can be converted to raw syngas via a primary reformer, a secondary reformer or both. The primary reformer is a typical steam methane reformer, and the secondary reformer is an autothermal reformer. Both reforming routes are governed by the steam reforming and water-gas shift reactions shown in Eqs. (1) and (2). The primary reformer operates at 40 bar and three discrete temperatures (800 °C, 850 °C, and 900 °C). The secondary reformer operates at 39.5 bar and three distinct temperatures (900 °C, 950 °C, 1000 °C). In the secondary reformer, either highly pure oxygen from an air separation unit (ASU) or input air is added to partially oxidize the natural gas. Non-syngas and non-hydrocarbon species are assumed to be inert in the reformers. Syngas from both reformers are sent to the syngas purification section or methanol synthesis section.

$$CH_4 + H_2O \leftrightarrow CO + 3H_2 \quad (1)$$

$$CO + H_2O \leftrightarrow CO_2 + H_2 \quad (2)$$

2.2 Biomass Gasification and Bio-Syngas Cleaning

The hardwood-type biomass is delivered to the refinery as wood chips with a moisture content of 45 wt.%. Biomass is dried and lockhopped with compressed CO_2 at 31 bar. This mixture is then fed to the gasifier which operates at 30 bar and at either 900, 1000, 1100 °C. Oxygen and steam are also fed to the gasifier to facilitate char gasification. The effluent is sent to two cyclones to remove the ash. Unreacted char is either recycled to the

gasifier or sent to the tar cracker. The effluent of the tar cracker contains some sulfur and needs to be treated before being sent to ammonia or methanol synthesis sections. The bio-syngas stream is first cooled down to 35 °C and then sent to a NH_3/HCl scrubber. This is followed by a two-stage Rectisol acid gas removal unit and a Claus plant. The removed CO_2 is either recycled, sent to sequestration, or vented. Clean bio-syngas can either be sent to methanol synthesis section, the reformers to convert the light gases into syngas, to another WGS reactor, or the syngas purification section.

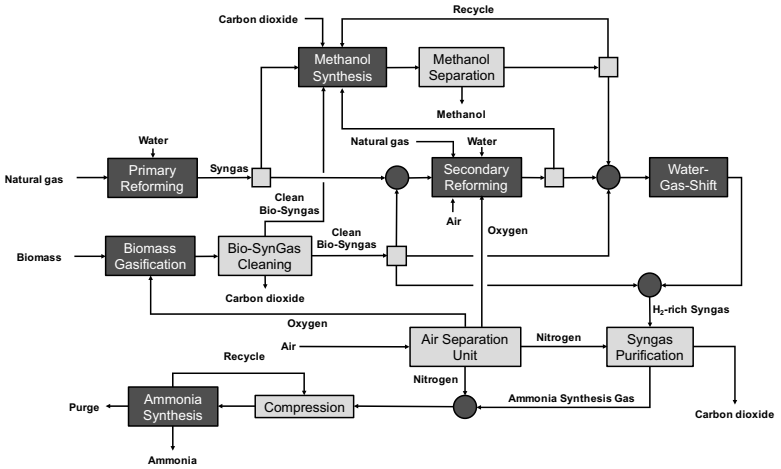


Figure 1. Process superstructure for Biomass/Natural Gas to Ammonia and Methanol

2.3 Methanol Synthesis

Clean syngas from biomass gasification and natural gas reforming is sent to the methanol synthesis section. Methanol synthesis section consists of two reactor configurations that are used together or separately. The first option is to use a once-through methanol reactor (OTMR) without recycle, and the second option is methanol synthesis reactor (MSWR) with recycle. The reactors serve two purposes: (i) methanol production and (ii) CO_x removal from syngas. The effluent of OTMR is flashed to separate methanol product from unreacted syngas. The unreacted syngas can either be sent to MSWR to produce more methanol or to the ammonia synthesis section. MSWR effluent is sent to a separation unit to separate methanol and unreacted syngas. A recycle loop configuration is used to increase the methanol conversion. Hydrogen-rich off-gases from MSWR can be purified and sent to ammonia synthesis section. Both reactors operate at 45 bar and 250 °C and reach equilibrium conversion according to the reactions shown in Eqs. (2) and (3).

$$CO + 2H_2 \leftrightarrow CH_3OH \tag{3}$$

2.4 Syngas Purification

Since oxygen containing species (e.g. CO_x, H_2O) are poisonous for the iron catalyst used in the ammonia converter, they need to be removed. Syngas coming from biomass gasification or natural gas reforming and H_2-rich streams from the methanol synthesis section are mixed and sent to a two-stage WGS reactor to further increase the H_2 yield (according to Eq. 2) and remove the CO_x species. After water is removed in a flash unit, an aMDEA process is used to remove 99.9 % of CO_2. This effluent can then be either washed with liquid N_2 to remove all the species but H_2 and N_2, sent to a methanator unit

where all CO_x species are converted to inert CH_4, or sent to a pressure swing adsorption (PSA) unit to remove impurities. Make-up nitrogen supply is available from an air separation unit to adjust the inert free ammonia syngas to the desired H_2/N_2 ratio.

2.5 Ammonia Synthesis

Five production trains are employed to have different converter options operating at distinct pressures (85.5, 130, 170, 210, and 250 bar). A single reaction shown in Eq. (4) takes place in the ammonia converter.

$$N_2 + 3H_2 \rightarrow 2NH_3 \tag{4}$$

The ammonia converters are not modeled as equilibrium converters, since industrial reactors do not reach equilibrium. Instead, the reactors are modeled through a data-driven approach using 25 industrial and experimental studies. Ammonia conversion is calculated as a function of operating temperature, pressure, and input compositions as shown in Eq. (5). The parameter estimation problem is formulated and solved in GAMS.

$$X_{NH_3} = A \cdot T + B \cdot P + C \cdot y_{H_2} + D \cdot y_{N_2} + E \cdot y_{NH_3} + F \cdot y_{inert} + G \tag{5}$$

The due to low single-pass conversion (20-35 %) multiple recycle options are used. The converter effluent is cooled to ambient conditions before being refrigerated to condense NH_3. A letdown vessel is used to flash out the liquid ammonia. The ammonia in the vapor effluent of the letdown unit is recovered with a water scrubber/distillation system. Inert gases are purged and the unreacted gases are recycled to the ammonia converter.

2.6 Utilities

The process superstructure includes alternatives for oxygen and hydrogen production. A comprehensive wastewater treatment network is also implemented. Simultaneous heat, power, and water integration ensures minimal utilization of external utility and maximum utilization of waste heat within the process (Elia et al., 2010).

3. Deterministic Global Optimization

3.1. Objective Function

The complete mathematical model representing the integrated ammonia and methanol process is a large-scale non-convex mixed-integer nonlinear program (MINLP). The objective function in Eq. (6) is to minimize the overall cost of hydrogen production, and all terms are normalized with respect to the total production of hydrogen in metric tons.

$$MIN \sum_{F \in Feed} Cost_F + Cost_{El} + Cost_{Seq} + \sum_{U \in Units} Cost_U \tag{6}$$

3.2. Branch-and-Bound Algorithm

Within each node in the branch-and-bound tree, a mixed-integer linear program (MILP) is first solved using CPLEX to determine a lower bound on the original MINLP by underestimating nonlinear terms with linear relaxations. Concave cost functions are approximated using piecewise linearization. Bilinear terms are relaxed using piecewise McCormick envelopes. The solution of the MILP provides initial starting points for the upper bound problem. By fixing binary variables at these starting points, a nonlinear program (NLP) that forms an upper bound on the original MINLP is created and then

solved using CONOPT (Baliban et al., 2012). More detailed discussions about global optimization theory and algorithms are given in textbooks by Floudas (1995, 2000).

4. Computational Studies

The process synthesis model is used to examine three case studies to observe the effect of coproduction on the optimal breakeven H_2 price and process topologies. The total output of the plant is set to 500 MT/day of H_2 containing energy vector. The cases are denoted as A-500 (ammonia-only), M-500 (methanol-only), and C-500 (ammonia-methanol coproduction). From mass conversion, 500 MT/day of H_2 translates to 2,830 MT/day of ammonia production in the A-500 plant and 4,000 MT/day of methanol production in the M-500 plant. For coproduction of ammonia and methanol, the product ratio is left as a decision variable. The greenhouse gas emissions (GHG) are restricted, such that they do not exceed more than 50% of the values produced by typical ammonia and methanol plants. The overall cost breakdown for all cases are shown in Table 1.

Table 1. Overall Cost Breakdown.

Contribution to Cost ($/kg of H_2 equivalent)	A-500	M-500	C-500
Biomass	0.57	0.95	0.61
Natural Gas	0.23	0.19	0.36
Water	0.01	0.01	0.00
Investment	0.99	0.91	0.90
CO_2 TS&M	0.00	0.02	0.02
O&M	0.26	0.24	0.24
Electricity Use	-0.03	-0.24	-0.20
Breakeven Price	2.03	2.08	1.94

The results indicate that coproduction of ammonia and methanol decreases the breakeven price of producing ammonia and methanol individually. The C-500 study shows that the optimal ratio in the product energy vector mix are 1,020 MT/day of ammonia and 2,560 MT/day of methanol. A mass balance indicates that 36% of the H_2 is stored in ammonia and 64% is stored in methanol. Converted breakeven prices to ammonia and methanol bases for the A-500 and M-500 plants are $360.33/MT and $254.80/MT, respectively.

It is observed that the A-500 and C-500 plants use a hybrid feedstock of biomass and natural gas, whereas the M-500 plant uses only biomass as feedstock. The M-500 and C-500 plants also use natural gas for electricity generation. Since the net electricity consumption values are negative, all the plants are producing electricity and this excess can be sold back to the grid.

The optimal process topologies are shown in Table 2. The A-500 plant uses both natural gas reformers and a biomass gasifier operating at 900 °C to produce syngas. The M-500 plant utilizes only biomass for its feedstock. It has a higher temperature gasifier, operating at 1100 °C to produce the methanol synthesis gas. OTMR and MSWR configurations are implemented together. The C-500 plant inputs both biomass and natural gas as feedstocks. Clean bio-syngas and primary reformer effluent are mixed and sent to OTMR, followed up by the MSWR. After methanol is produced, the unreacted and unrecycled synthesis gas is sent to the ammonia synthesis section. Liquid nitrogen wash technology is always selected over the methanator and PSA to obtain a pure ammonia synthesis gas, and a low-pressure ammonia synthesis loop is always preferred.

Table 2. Optimal Process Topologies.

Technology	A-500	M-500	C-500
Biomass Gasifier Temp.	900 °C	1100 °C	1100 °C
Primary Reformer Temp.	900 °C	-	900 °C
Secondary Reformer Temp.	1000 °C	-	-
OTMR	-	Y	Y
MSWR	-	Y	Y
ASU	Y	Y	Y
aMDEA	Y	-	Y
Liquid Nitrogen Wash	Y	-	Y
Methanator	-	-	-
PSA	-	-	-
Ammonia Converter Pres.	85.5 bar	-	85.5 bar
Fuel Combustor	Y	-	Y
CO_2 Sequestration	-	Y	Y

5. Conclusion

This paper introduced novel renewable processes for coproduction of ammonia and methanol as hydrogen containing energy vectors. A process superstructure consisting of several competing and commercial processes is modeled and solved to global optimality using a tailored deterministic global optimization branch-and-bound algorithm. Three examined case studies focused on plants that produce energy vectors in the form of ammonia and/or methanol that contain the same amount of H_2 that is equivalent to producing 500 MT/day of H_2. Coproduction of ammonia and methanol as energy vectors is found to lower the production cost of separate ammonia or methanol production plants between 4-7%. The reported breakeven prices for H_2 is cost-competitive.

Acknowledgments

The authors declare no competing financial interests and acknowledge financial support from the Shell Oil Company and the Texas A&M Energy Institute.

References

R.C. Baliban, J.A. Elia, R. Misener, C.A. Floudas, 2012, Global optimization of a MINLP process synthesis model for thermochemical based conversion of hybrid coal, biomass, and natural gas to liquid fuels, Comp. Chem. Eng. 42, 64-86.

J.A. Elia, R.C. Baliban, C.A. Floudas, 2010, Toward novel hybrid biomass: coal, and natural gas processes for satisfying current transportation fuel demands, 2: Simultaneous heat and power integration, Ind. Eng. Chem. Res. 49, 7371-7388.

C.A. Floudas, A.M. Niziolek, O. Onel, L.R. Matthews, 2016, Multi-scale systems engineering for energy and the environment: Challenges and opportunities, AIChE J. 62, 602-623.

C.A. Floudas, 1995, Nonlinear and Mixer-Integer Optimization, Oxford University Press.

C.A. Floudas, 2000, Deterministic Global Optimization: Theory, Methods, and Applications, Kluwer Academic Publishers.

D. Teichmann, W. Arlt, P. Wasserscheid, 2012, Liquid Organic Hydrogen Carriers as an efficient vector for the transport and storage of renewable energy, Int. J. Hydrogen Energy 37, 18118-18132.

G. Thomas, G. Parks, 2006, Potential Roles of Ammonia in a Hydrogen Economy, DOE.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50280-9

Framework for the optimal design of sustainable processes incorporating data envelopment analysis

Andres Gonzalez-Garay[a], Gonzalo Guillen-Gosalbez[a*]

[a] *Centre for Process Systems Engineering, Imperial College, South Kensington Campus, London, SW7 2AZ, UK*
g.guillen05@imperial.ac.uk

Abstract

The development of computational tools in the process industry is essential to achieve sustainable production systems. In this work, we present a framework which incorporates sustainability principles applied to the design and optimization of chemical processes. The methodology incorporates life cycle assessment, surrogate modeling, objective-reduction techniques, multi-objective optimization, and the use of Data Envelopment Analysis (DEA) as multi-criteria decision analysis tool. In particular, the use of DEA facilitates the post-optimal analysis of the Pareto front by filtering the number of designs and ranking them without the need to define explicit weights. Additionally, DEA also provides improvement targets for those designs identified as suboptimal. This characteristic is particularly appealing as it can guide retrofit efforts towards the most effective actions when benchmark processes are compared against the best technologies available. The capabilities of the framework are demonstrated in a case study based on the production of methanol from CO_2 and hydrogen.

Keywords: Sustainable process design, life cycle assessment, multi-objective optimization, objective-reduction, data envelopment analysis.

1. Introduction

The generation of renewable and cleaner production systems is essential to achieve sustainable development. In the process industry, this is observed in the increasing implementation of processes which consume less energy and resources while releasing lower amounts of waste and emissions. In other words, the development of new processes and technologies is now driven by sustainable criteria. However, the implementation of these new technologies will very likely take place in the coming decades creating also the need to improve the performance of current facilities.

Given the impact it has over the chemical industry and related industries, process systems engineering plays a key role to attain sustainable development (Reklaitis, 2014). In this context, we can identify three main stages to achieve a sustainable design: modeling and assessment of the process, optimization, and multi-criteria decision analysis. In the first stage, sustainability metrics are defined and incorporated within a modeling technique following either equation-oriented or sequential modular approaches. In the second stage, the design task is posed as a multi-objective optimization (MOO) problem that seeks to minimize or maximize the sustainability metrics previously selected. Typically, these metrics are in conflict with each other due to the existence of inherent trade-offs between them. As a result, their optimization results in a set of optimal solutions also known as Pareto frontier. In a third stage, this Pareto frontier is analyzed in order to select the best

design. Typically, this analysis is performed by stakeholders who are assisted by multi-criteria decision analysis (MCDA) tools.

In this work, we present a framework for the optimal design of sustainable chemical processes that combines life cycle assessment, surrogate modeling, objective-reduction techniques, multi-objective optimization, and MCDA tools within a unified framework. The main novelty of the framework, additional to the integration of approaches, is the use of Data Envelopment Analysis (DEA) to assess Pareto optimal designs. The use of DEA allows us to further screen and rank alternatives without the need of defining explicit weights. More precisely, in the context of process design DEA allows to: (i) filter and rank the Pareto solutions; and (ii) provide clear insight into how to enhance suboptimal solutions such that they achieve an optimal performance.

2. Methodology

The methodology proposed to generate sustainable processes is shown in Figure 1.

2.1 Flow sheet modeling and assessment

While the framework is general enough to accommodate any metric, the sustainability assessment proposed focuses on economic and environmental criteria. As standard practice, we propose to use the total annualized cost (TAC), as it is widely applied in process design. The environmental performance is assessed via LCA as described by the International Organization for Standardization (ISO-14040). Depending on the scope of the analysis, the process can be assessed from cradle-to-gate, cradle-to-grave or cradle-to-cradle, and allocation of environmental burdens can also be considered for those processes where more than one final product is generated.

As for the modeling, we primarily focus on commercial simulators based on the sequential modular approach. When complex or computationally expensive models are implemented in the simulator, the optimization task becomes challenging as convergence problems, numerical difficulties, or large use of computational resources may arise (e.g. modeling reactive distillation columns, reactors with complex kinetics, several recycling streams, etc.). In such cases, we propose the use of surrogate models to facilitate the solution of the model and enhance the optimization stage.

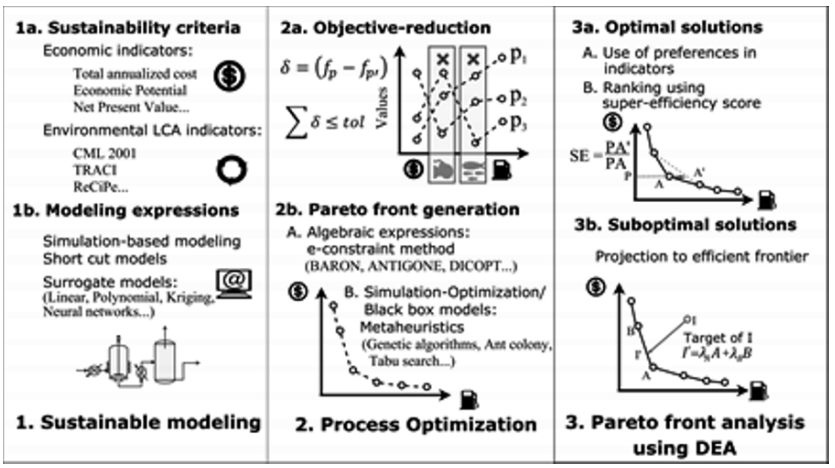


Figure 1. Framework for the optimal design of sustainable processes.

2.2 Process Optimization

To enhance the performance of the optimization solver and improve the quality of the Pareto frontier generated, we first apply an objective-reduction technique based on the MILP formulation introduced by Guillen-Gosalbez (2011). The advantage of this formulation is that the dominance structure of the Pareto set is preserved by maintaining a delta error δ below a given tolerance each time an objective is omitted.

During the optimization, we focus on the use of Multi-Objective Genetic Algorithms (MOGA's) given that they can be easily coupled not only with simulators/black box models but also with models based on algebraic expressions. However, other approaches can also be applied, such as the ε–constraint method.

2.3 Pareto frontier post-analysis using Data Envelopment Analysis (DEA)

DEA is a linear programming (LP) technique that quantifies the relative efficiency of a group of comparable decision-making units (DMUs), each consuming multiple inputs to produce multiple outputs. In the context of process design, a DMU refers to one process being assessed under multiple sustainability criteria.

The solution of the DEA model provides efficiency scores for all DMUs. Those units having an efficiency score of one are considered as efficient while those units having an efficiency score less than one are inefficient. Given the nature of DEA, the units identified as efficient form the convex envelope of the system, which can be considered as an efficient frontier. By projecting the inefficient units into the frontier, it is possible to determine improvement targets that would make them efficient. In this context, the most applied projections in DEA are the input and output oriented models. Herein, we present the input-oriented model (Banker et al., 1984), which aims to minimize the inputs (e.g. total annualized cost, global warming potential, etc.) while maintaining the same level of outputs (e.g. profit, amount of products and by-products produced, etc.). The input-oriented model is expressed as follows:

$$\max \sum_{r=1}^{s} u_r y_{ro} - u_o \tag{1}$$

$$\sum_{l=1}^{m} v_l x_{lo} = 1 \tag{2}$$

$$\sum_{r=1}^{s} u_r y_{rj} - u_o - \sum_{l=1}^{m} v_l x_{lj} \leq 0; \quad j = 1, \dots, n \tag{3}$$

$$u_r, v_l \geq 0; \quad \forall l, r; \quad \theta_o \; unconstrained$$

where multipliers v_l and u_r denote weights given to inputs x_l and outputs y_r, respectively. In the model, these multipliers are optimized in order to maximize the efficiency value of a given DMU_o rather than being defined by decision-makers. In this model, it is also possible to apply preferences in the multipliers by imposing additional constraints.

Enhancement of technologies. This characteristic is particularly appealing when we are comparing different processes or technologies. For instance, when we aim to compare a current production process against emerging technologies. In such cases, DEA not only identifies the process having the best practices ($\theta_o = 1$) but also provides improvement targets for the alternatives having the worst performance ($\theta_o < 1$). These targets are obtained by solving the following model, which is a reformulation of Eqs. 1-3:

$$\min \theta_o - \varphi \left(\sum_r s_r^+ + \sum_l s_l^- \right) \tag{4}$$

$$\sum_j \lambda_j x_{lj} + s_l^- = \theta_o x_{lo}; \quad l = 1, \dots, m \tag{5}$$

$$\sum_j \lambda_j y_{rj} - s_r^+ = y_{ro}; \quad r = 1, \dots, s \tag{6}$$

$$\sum_{j=1}^{n} \lambda_j = 1 \qquad j = 1, \dots, n \tag{7}$$

$$\lambda_j, s_l^-, s_r^+ \geq 0; \quad \forall i, j, r; \quad \theta_o \; unconstrained$$

where φ is a non-Archimedean infinitesimal value used to enforce the variables to be strictly positive; s_r^+ is a slack variable representing the additional amount of output r required by DMU_o to become efficient; s_l^- is a slack variable denoting the additional amount of input l to be reduced to become efficient; and λ_j are the linear weights used to project the inefficient units onto the efficient frontier. When $\theta_o = 1$ and $(s_r^+ + s_l^- = 0)$, the DMU is said to be strongly efficient. If $\theta_o = 1$ but $(s_r^+ + s_l^- \neq 0)$, the DMU is deemed as weakly efficient. If $\theta_o < 1$, the DMU is inefficient.

The improvement target for input l in a given design is expressed as:

$$\sum_{j=1}^{n} \lambda_j x_{lj} = \theta_o x_{lo} - s_l^- \tag{8}$$

Ranking of efficient (optimal) solutions. The ranking of optimal solutions is achieved by using a super-efficiency model, which is essentially the same as the one described in Eqs. 4-7, but in which the summation of lambdas excludes the efficient unit being assessed:

$$\sum_{j=1, j \neq j'}^{m} \lambda_j x_{lj} + s_l^- = \theta_j x_{ljj'}; \quad l = 1, \dots, m \tag{9}$$

$$\sum_{j=1, j \neq j'}^{s} \lambda_j y_{rj} - s_r^+ = y_{rjj'}; \quad r = 1, \dots, s \tag{10}$$

In an input-oriented model, the super-efficiency can be seen as the input savings achieved by an efficient DMU. These savings are measured by the extent to which the efficient frontier changes when such unit is removed. As a result, the model provides a super-efficiency score $\theta_{SE,j'}$ that is always greater or equal to one and can be used to further discriminate efficient units.

3. Case study: Methanol production from CO2 and hydrogen

The capabilities of the framework are shown in the assessment of the production of methanol from CO_2 and hydrogen. The base case is taken from Perez-Fortes et al. (2016).

Step 1: Flow sheet modeling and assessment. The sustainability of the process was assessed from cradle-to-gate defining 1 kg of methanol as functional unit. The indicators considered were the total annualized cost (TAC) and those included in the LCIA methodology CML 2001 (10 environmental indicators). The detailed process was first modeled in Aspen-HYSYS and then used to build a surrogate model using neural networks. A total of 10 continuous and 2 integer variables were considered for the

optimization of the process: CO_2 and hydrogen flow rates, temperature and volume of the reactor, recycling ratio of the flash gases, amount of purged being burned to recover heat, methanol recovery, inlet temperature of the feed to the column, distillation column pressure, reflux ratio, number of trays and feed tray.

Step 2: Process optimization.

The objective reduction was performed using an exhaustive exploration of the MILP formulation presented by Guillen-Gosalbez (2011) using a delta error $\delta = 0$. The results indicated that five out of the 11 objectives were required to keep the approximation error desired: total annualized cost (TAC), global warming potential (GWP), eutrophication potential (EP), human toxicity (HT), and photochemical oxidation (PO).

The optimization of the surrogate model was executed five times using a Multi-Objective Genetic Algorithm considering only the indicators included in the reduced space. A total of 137 non-repeated solutions were obtained, from which 126 points converged and were feasible in the original model in Aspen-HYSYS.

Step 3: Pareto front analysis using DEA. The Pareto front analysis included the 126 designs obtained from the MOO and the base case design taken from Perez-Fortes et al. (2016). After solving Eqs. 4-7, only ten out of the 127 designs were identified as efficient, including the base case (s_1 to s_9 and BC). These results show how a significant number of solutions were first ruled out by applying the concept of efficiency in DEA.

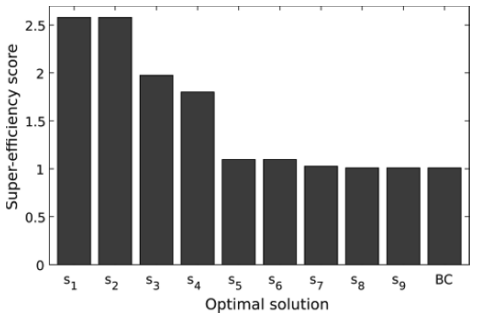


Figure 2. Super-efficiency score for the optimal designs identified as efficient.

To rank the efficient solutions, we calculated the super-efficiency score by solving Eqs.4, 7, 9 and 10. The ranking of the efficient solutions is shown in Figure 2. In this case, the super-efficiency model provided the same score for designs s_1 and s_2, indicating that a further analysis between those solutions has to be done to select the most suitable design.

In DEA, it is also possible to determine preferences in the sustainable indicators by imposing constraints to the multipliers u_r and v_l and including them in the model presented in Eqs. 1-3. To show this capability, we included the following constraint in our analysis:

$$v_{TAC} \geq v_{GWP} \geq v_{HT} \geq v_{EP} \geq v_{PO} \tag{11}$$

The efficient designs identified by solving Eqs. 1-3 and 11 are shown in Figure 3. In this case, only four designs were efficient and the base case was classified as inefficient. To identify the improvement targets for design BC, we solved Eqs. 4-8 and retrieved the values of λ_j. From the solution of the model, design BC had to be reduced to the level of solution s_1 in order to become optimal when constraint 11 is imposed.

Analyzing the best solutions identified by DEA, we found that hydrogen and CO_2 were the main contributors to the total value of the TAC and HT indicators (>90 %). In the case of GWP, hydrogen and CO_2 contributed by ≈80% while the remaining percentage was equally attributed to cooling water, heat recovery and emissions (≈6% in each case). EP

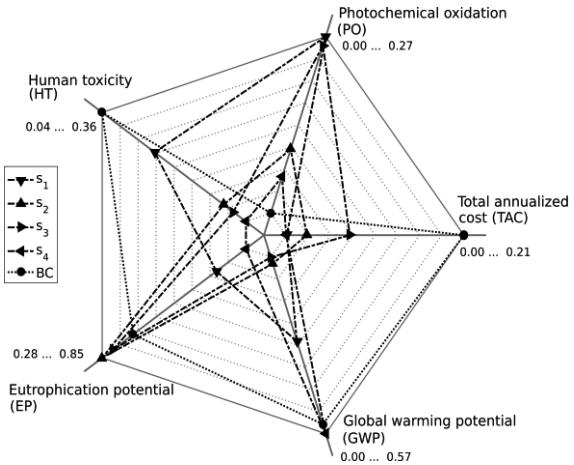


Figure 3. Normalized values for the efficient units and design BC in the reduced space when constraints in the multipliers are included.

and PO were mainly explained by the emissions generated in the process (>90% and >70%, respectively). Finally, the full methodology pointed designs s_1 and s_2 as the most sustainable options when no constraints were included in the indicators. The main difference between both designs was the higher consumption of hydrogen in design s_2, which also caused the TAC indicator to increase and was the reason to identify it as inefficient when constraints in the multipliers were included.

4. Conclusions

We introduced a framework for the optimal design of sustainable chemical process that combines life cycle assessment principles, surrogate modelling, objective reduction techniques, multi-objective optimization and DEA. The methodology was applied to the production of methanol from CO_2 and hydrogen, where all the stages of the methodology were evaluated. The application of DEA as multi-criteria analysis tool allowed us to narrow down the number of Pareto points and to establish improvement targets for those designs showing the worst performance. The same methodology allowed the ranking of optimal solutions, which facilitates the comparison between different designs. Finally, the achievement of sustainable development in the chemical industry can be facilitated by the development and application of computational tools such as the one presented in this work. This, not only by assessing new processes and technologies but also by identifying improvement targets for current suboptimal designs.

References

R.D. Banker, 1984. Estimating most productive scale size using data envelopment analysis. *European Journal of Operational Research*, 17(1), pp.35–44.

G. Guillén-Gosálbez, 2011. A novel MILP-based objective reduction method for multi-objective optimization: Application to environmental problems. *Computers and Chemical Engineering*, 35, pp.1469–1477.

M. Pérez-Fortes , J.C. Schoneberger, A. Boulamanti, and E. Tzimas, 2016. Methanol synthesis using captured CO 2 as raw material: Techno-economic and environmental assessment. *Applied Energy*, 161, pp.718–732.

G.V. Rex Reklaitis, 2014. Perspectives on Systems Engineering advances in Process and Product Design. *Computer Aided Chemical Engineering*, 34, pp.1–4.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50281-0

Synthesis of Water Distribution Networks through a Multi-Stakeholder Approach

Ramón González-Bravo[a], Luis Fabian Fuentes-Cortes[b], Mahmoud M. El-Halwagi[c], José María Ponce-Ortega[a*]

[a]*Chemical Engineering Department, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacán, 58060, México*
[b]*Tecnológico de Monterrey, Escuela de Ingeniería y Ciencias, Ave. Eugenio Garza Sada 2501, Monterrey, N.L., 64849, México.*
[c]*Chemical Engineering Department, Texas A&M University, College Station, TX, 77843, USA.*
jmponce@umich.mx

Abstract

This paper presents a systematic approach to synthesize optimal water distribution networks (WDN) taking into account the water requirement of consumers based on a multi-stakeholder approach. A structural representation is developed to satisfy the water requirements of all users without taking into account water concessions in the region, but considering the depletion of groundwater and surface water, to generate an integral water distribution network. A rigorous mathematical formulation is developed as a mixed-integer linear programming model to transform the water distribution problem into an optimization task that seeks to minimize the overall cost of the water supply network. The proposed optimization model was applied to the Sonoran water scarcity problem from Mexico, where the optimal water distribution network was obtained, based on a multi-criteria approach taking into account the industrial, domestic and agricultural users. The obtained results show that the proposed model can satisfy economic, environmental and social aspects, where the distribution of water is equal for all users without affecting natural resources. Also, the results show economic advantages due to the sales of water.

Keywords: Water management; Water distribution network; Optimization; Multi-Stakeholder Criteria.

1. Introduction

Water scarcity is becoming the major environmental issue affecting many regions of the world. Climate change and temporal variations between water recharge have intensified this problem, and combined with the increased water demand for domestic, agricultural, industrial and energy production, it has resulted in the overexploitation of water resources around the world (UN-Water, 2017). The water scarcity problem has affected large regions around the world, so that, the distribution of limited water resources for agricultural, domestic, industrial and environmental uses, requires the complete integration of demand and supply, taking into account sustainable considerations.

The synthesis of water distribution networks (WDN) based on optimization still requires the development of enhanced optimization models, efforts have been made to generate robust models that consider environmental, social and economic issues. For example,

Liu et al. (2011) proposed a mixed-integer linear programming (MILP) problem taking into account the production, distribution, and storage of desalinated water in a macroscopic case. Athilan et al. (2012) introduced a mathematical model for the synthesis of macroscopic WDN, where the authors considered fresh water, underground water, seawater, and desalinated water accounting for monthly demand fluctuations and wastewater treatment. Napoles-Rivera et al. (2015) developed a mathematical formulation for the optimal synthesis of WDN considering environmental and economic issues, the model considers dams, springs, deep wells and rainwater harvesting as main sources to satisfy industrial and domestic demands. Alnouri et al. (2015) considered the water-energy nexus by the development of a deterministic MILP problem used to obtain the optimal WDN and allocation of power-desalination technology taking into consideration seasonal variations through the year obtaining significant cost reductions. Gonzalez-Bravo et al. (2017) introduced a mathematical model for the synthesis of WDN involving power-desalination plants for water-stressed areas, in which the authors considered sustainability aspects in terms of energy and water preservation. Araya et al. (2017) presented a mixed-integer nonlinear programming (MINLP) problem for the synthesis of WDN taking into consideration the location of new desalination plants and energy recovery devices in regions with a complex topography.

This paper presents an optimization approach to perform a multi-objective optimization of integrated water distribution networks, the proposed optimization model is a decision-making tool based on multi-stakeholder theory, which takes into account the economic and environmental points of view, the characteristics of the water sources (availability of water, minimum and maximum water level, and over-exploitation status), as well as the characteristics of the consumers (seasonal variations on water demand), in order to generate a water distribution network that does not affect the sustainability of water resources. The model also considers the integration of energy issues related to the generation of water, based on power-desalination technology.

2. Proposed approach

It is important to differentiate between the minimum and the desired water supply. The first one represents the water needed to meet the basic requirements of the user and it is non-negotiable. The second one represents the water needed to cover all requirements of the user. Therefore, the social objective is to minimize the difference between the supply and the desired demand of the user, reducing the level of dissatisfaction. The proposed optimization model is based on stakeholder theory. The main idea of this model is to satisfy water demands of multiple users accounting for the dissatisfaction of each part, the water demands, as social objective functions, involve the water demands of domestic users, industrial users, and agricultural users, where the water demands can be satisfied by the existing water in dams, rivers, aquifers, deep wells, storage tanks and also the possible installation of water desalination plants. Each of the participants in the decision-making process expresses different level of preference over the objectives considered for the final configuration and operation of the system. The presented modelling approach allows defining an optimal solution that reaches a trade-off between the multiple criteria of the stakeholders and reducing the possible conflicts. The model also takes into account the water depletion of natural resources (dams and aquifers) based on a minimum and maximum amount of water to be extracted, these values depend on the natural recharge through the year. In addition, the model includes economic (profit), and environmental (CO_2 emissions) objective functions.

3. Mathematical formulation

The proposed equations include mass balances for aquifers, deep wells, dams, central stations (domestic, agricultural and industrial), storage tanks, new desalination plant, pumping cost, piping cost, solar collector equations, and water resource depletion equations. Additionally, the optimization model includes equations to account the existence of power desalination plants, capital and operating cost.

The water demands can be satisfied by the volume of water in central stations (h), plus the water in existing storage tanks (s^E) and new storage tanks (s^N).

For domestic users:

$$domdem_{r,t} = h_{r,t}^{dom} + \sum_p s_{p,r,t}^{E,dom} + \sum_q s_{q,r,t}^{N,dom}, \qquad \forall r \in R, \forall t \in T \tag{1}$$

The water demands for domestic users can be set using a minimum value, this value depends on the minimum amount of water that can be sent to the user to satisfy the domestic water demands, and it is stated as follows:

$$domdem_{r,t} \geq \mathrm{ddem}_{\mathrm{r,t}}^{\min} \qquad \forall r \in R, \forall t \in T \tag{2}$$

At the same time, the desired water demand that can be received by the domestic users can be stated as a maximum value, this value depends on the desired amount of water that can be sent to the user to satisfy the domestic water demands, as follows:

$$domdem_{r,t} \leq \mathrm{ddem}_{\mathrm{r,t}}^{\mathrm{desired}} \qquad \forall r \in R, \forall t \in T \tag{3}$$

Similarly, for industrial users:

$$inddem_{o,t} = h_{o,t}^{ind} + \sum_p s_{o,p,t}^{E,ind} + \sum_q s_{o,q,t}^{N,ind}, \qquad \forall o \in O, \forall t \in T \tag{4}$$

$$inddem_{o,t} \geq \mathrm{idem}_{\mathrm{o,t}}^{\min} \qquad \forall o \in O, \forall t \in T \tag{5}$$

$$inddem_{o,t} \leq \mathrm{idem}_{\mathrm{o,t}}^{\mathrm{desired}} \qquad \forall o \in O, \forall t \in T \tag{6}$$

For agricultural users:

$$agrdem_{g,t} = h_{g,t}^{agr} + \sum_p s_{g,p,t}^{E,agr} + \sum_q s_{g,q,t}^{N,agr}, \qquad \forall g \in G, \forall t \in T \tag{7}$$

$$agrdem_{g,t} \geq \mathrm{adem}_{\mathrm{g,t}}^{\min} \qquad \forall g \in G, \forall t \in T \tag{8}$$

$$agrdem_{g,t} \leq \mathrm{adem}_{\mathrm{g,t}}^{\mathrm{desired}} \qquad \forall g \in G, \forall t \in T \tag{9}$$

The economic objective function ($Profit$) seeks to maximize the total annual profit, which includes water sales and tax credit reduction, minus the total annual cost.

$$maxProfit = Water\ Sales + TCR - TAC \tag{10}$$

The environmental objective function ($OGHGE$) seeks to minimize the overall greenhouse gas emissions. In this case, we take into account the GHGE for fossil fuels and biofuels, the emissions for the solar collector is assumed to be zero, which is stated as follows:

$$Min\, OGHGE = \sum_{f}\sum_{u}\sum_{t}\left[GHGE_{f}^{fossil} \cdot Q_{f,u,t}^{fossil} \right] + \sum_{b}\sum_{u}\sum_{t}\left[GHGE_{b}^{biofuel} \cdot Q_{b,u,t}^{biofuel} \right] \tag{11}$$

4. Multi-stakeholder decision-making strategy

The optimal solution must be selected from the feasible solution space, which is obtained using the coordinates of the utopia point (UP), the UP point represents the optimal values of all objective functions (the UP is infeasible); the feasible solution also includes the Nadir solution (NS), which represents the opposite values of the UP, this region is obtained using lower bounds (LB) and upper bounds (UB) of the proposed objective functions. By scaling the objective functions using the UP and NS, the following equations are obtained:

Economic objective function:

$$\varphi = \frac{Profit^{UB} - Profit}{Profit^{UB} - Profit^{LB}} \tag{12}$$

Environmental objective function:

$$\tau = \frac{OGHGE - OGHGE^{LB}}{OGHGE^{UB} - OGHGE^{LB}} \tag{13}$$

Social objective function:

$$\psi_{e} = \frac{domdem_{e}^{UB} - \sum_{t} domdem_{e,t}}{domdem_{e}^{UB} - domdem_{e}^{LB}}, \qquad \forall e \in E \tag{14}$$

Where e represents any water consumer (either agricultural, domestic and industrial user). The social objective function allows selecting either one or e user demands to be prioritized, in order to evaluate the optimal solution (notice that the number of prioritized users increases the dimensions of the problem)

The compromise solution (CS) is to reduce the absolute distance between the utopia point (UP) and the objective functions, this solution will satisfy all parts and must be a feasible optimal solution, according to the following equation:

$$0 \le \varphi \le 1 \tag{15}$$

$$0 \le \tau \le 1 \tag{16}$$

$$0 \le \psi_{e} \le 1, \qquad \forall e \in E \tag{17}$$

By assigning weights to the CS function and taking into account k number of stakeholders, where each stakeholder has an individual solution (FSH_{k}), in this way these weights represent the level of individual priorities of each stakeholder:

$$FSH_{k} = w_{k}^{eco}\varphi_{k} + w_{k}^{envi}\tau_{k} + w_{k}^{soc}\sum\nolimits_{e}\psi_{k,e}, \qquad \forall k \in K, \tag{18}$$

where the average solution (AS) takes into account the contribution of each stakeholder as follows:

$$AS = \frac{\sum_{k=1}^{k} FSH_k}{CARD(k)} \tag{19}$$

Then the compensated solution of the problem seeks to minimize the average of the dissatisfactions and the main function, this solution represents the compensated solution of the involved stakeholders.

$$\min \left| CS - FSH \right| \tag{20}$$

5. Results

The water scarcity problem of Hermosillo Sonora in Mexico is presented to show the applicability of the proposed approach for designing a WDN, in this region the overexploitation of groundwater and surface water over the years has led to problems among the inhabitants of the driest regions by the dispute of this resource. In 2012, the inhabitants of one of the driest regions in Mexico (Hermosillo Sonora, near to the Sonoran Desert) faced a critical situation of water supply due to the low rainfall of the region. Later in 2014, a mining company of the region spilled 40,000 cubic meters of copper sulfate leachate in one of the most important rivers in the area, which deteriorated the water situation in the region.

The proposed model was coded in the software GAMS and it was solved using the solver BARON (Brooke et al., 2017). The model includes 10,724 equations and 10,910 continuous variables, and 764 binary variables, in this case, the average CPU time was 14,652 s in a computer with an i7 Intel core processor at 3.20 GHz and 32 GB of RAM. The model was solved taking into account environmental and economic points of view, then the most critical user of the case study was involved (domestic user from Hermosillo Region) as a priority to evaluate the differences between both scenarios. First, the evaluation of the net profit and the overall GHGE is presented in Figure 1, this figure presents an optimal economic solution in point A (1660 MM$/y and 1.45×10^7 tonCO_2eq/y) and the optimal environmental solution in point B (424 MM$/y and 1.36×10^7 tonCO_2eq/y). In the optimal economic solution, the power plant is installed at the Hermosillo Beach at the first period to supply the water deficit to the Hermosillo Region, (403×10^6 m^3/y). In this case, the prioritized users (domestic users from Hermosillo region) receive a total of 94.08×10^6 m^3/y, which represent the minimal requirement of water by the users. By applying the model prioritization through the multi-stakeholder optimization model for the domestic users from Hermosillo region, the total volume of water received by this user is 164.64, m^3/y, which represents 42% more water for this users, this increased volume inflicts directly over the net profit by reducing its value to 1493 MM$/y, which represents 11% less than the optimal economic solution. Also, the preservation of natural water resources is observed by maintaining the water levels through the year, and this is obtained because of the installation of the power-desalination plant.

6. Conclusions

This paper has presented a new optimization approach for the integrated synthesis of WDN, the proposed formulation takes into account economic and environmental impacts, as well as water demands as social objective, the problem was solved based on a multi-stakeholder theory which allows increasing the volume of water required by water-stressed users. This new formulation allows finding the optimal solution ensuring

the water demands of the most critical users. The results have shown important social, economic and environmental benefits that satisfy the involved stakeholders.

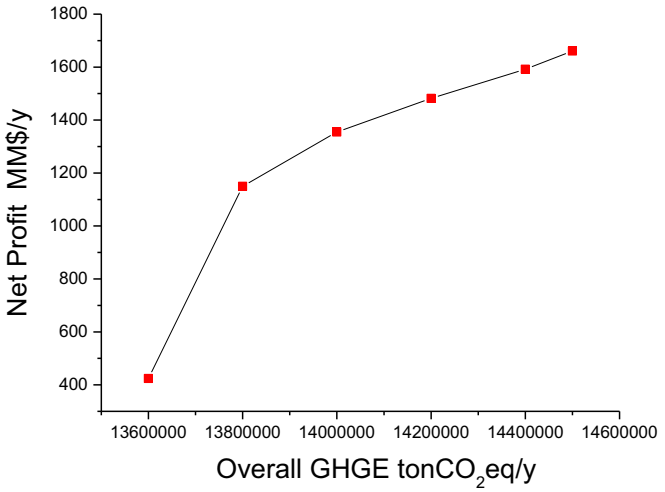


Figure 1. Pareto curve for case study

References

S. Y. Alnouri, P. Linke, M. M. El-Halwagi, 2015, A synthesis approach for industrial city water reuse networks considering central and distributed treatment systems, Journal of Cleaner Production, 89, 231-250.

N. Araya, L. A. Cisternas, F. Lucay, E. D. Galvez, 2017, Design of desalinated water distribution networks including energy recovery devices, Computer Aided Chemical Engineering, 40, 925-930.

S. Atilhan, A. Mahfouz, B. Batchelor, P. Linke, A. Abdel-Wahab, F. Nápoles-Rivera, A. Jiménez-Gutiérrez, M. M. El-Halwagi, 2012, A systems-integration approach to the optimization of macroscopic water desalination and distribution networks: a general framework applied to Qatar's water resources, Clean Technologies and Environmental Policy, 14, 2, 161-171.

A. Brooke, D. Kendrick, A. Meeruas, R. Raman, 2017, GAMS-language guide, Washington DC, GAMS Development Corporation.

R. González-Bravo, L. F. Fuentes-Cortés, J. M. Ponce-Ortega, 2017, Defining priorities in the design of power and water distribution networks, Energy, 137, 1026-1040.

S. Liu, F. Konstantopoulou, P. Gikas, L. G. Papageorgiou, 2011, A mixed integer optimization approach for integrated water resources management, Computers and Chemical Engineering, 35, 5, 858-875.

F. Nápoles-Rivera, M. G. Rojas-Torres, J. M. Ponce-Ortega, M. Serna-González, M. M. El-Halwagi, 2015, Optimal design of macroscopic water networks under parametric uncertainty, Journal of Cleaner Production, 88, 172-184.

UN-Water, WWAP (United Nations World Water Assessment Programme), 2014, The United Nations World Water Development Report 2017: Wastewater the untapped resource, UNESCO, Paris, ISBN 978-92-3-100201-4

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50282-2

Prediction of Environmental Properties Using a Hybrid Group Contribution Approach

Resul Al, Jérôme Frutiger, Alexandr Zubov, Gürkan Sin*

Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads 227, 2800 Kgs. Lyngby, Denmark
gsi@kt.dtu.dk

Abstract

Development of predictive environmental property models is increasingly becoming crucial as stringent regulations for substances with high global warming and ozone depletion potentials are being introduced. This contribution presents new environmental property models using two group contribution (GC) based approaches for the prediction of ozone depletion potential (ODP), global warming potential (GWP) and Daphnia magna lethal concentration (LC50, 48 hr), concentration of the test chemical in water (mg/L) that causes 50% of Daphnia magna to die after 48 hours. First, the classical group contribution approach, in which a property of interest is estimated from regression models that make use of available information about the chemical structure of a given compound (i.e. functional groups), is applied to develop models for selected properties using robust regression with outlier treatment. Second, a hybrid approach using only the first order GC-defined functional groups as predictors is presented to develop a number of data-driven models (a feedforward neural network (ANN) and a radial basis function network (RBFN), regression tree, etc.). Performance of the different models in predicting ODP, GWP, and LC50 is assessed for various industrially relevant chemicals and compared with results of the classical GC method. The experimental data for the selected properties is collected from the databases of Environmental Protection Agency (EPA) and the fifth assessment report of the Intergovernmental Panel on Climate Change (IPCC). The results showed that hybrid approach presents significant improvement on estimation accuracy for the considered environmental properties. This flexible approach builds on the basis of GC theory and extends it with nonlinear surrogate models to better describe the property of interest, which makes it a promising method to improve accuracy of property models in the wider domain of process systems engineering.

Keywords: predictive modeling, group contribution method, radial basis function network, artificial neural networks, environmental properties.

1. Introduction

The increasing legislative regulations as well as social awareness drive (bio)chemical, pharmaceutical and food industries to design sustainable processes with minimum environmental footprint. Performing sustainability assessment of manufacturing processes requires life cycle analysis, which among others rely on databases of environmental impact properties. There is a lack of sufficient data on the environmental properties such as ozone depletion potential, global warming potential and toxicity

endpoints like Daphnia magna LC50 (48 hr) for the majority of compounds. Development of these predictive property models is crucial for the process synthesis, design, and analysis of sustainable processes. GC based predictive property models are widely used in the literature. Frutiger et al. (2015) recently presented a methodology for the development of these models and applied it to the estimation of heat of combustion. Hukkerikar et al. (2012) estimated thermophysical and environment-related properties of pure components using GC methods. These methods have the advantage of being simple in model structure and capable of estimating new chemicals not used in training dataset. Approaches based on surrogate models like ANN using different pre-determined descriptors, also exist. Wernet et al. (2009), for example, developed ANN models to predict life cycle assessment related parameters including GWP using molecular weight and number of certain atoms in the molecular structure. Though these surrogate modelling approaches usually lack a close form expression of model structure, they are very effective at capturing nonlinearities that might be present in the datasets. In this contribution, by combining the best of the two modeling approaches, we introduce a hybrid approach building on the GC method by using its first order functional groups as predictors and utilizing the advanced model structure of surrogate models like ANNs to improve the predictive accuracy of the property models.

2. Theory

The classical group contribution method decomposes each chemical compound into first, second and, third order functional groups based on its molecular structure. Number of each different functional group present in a compound is used as predictors in a multiple linear regression model without any attempt to reduce the number of parameters in the model as the method tries to characterize the contributions from each different group. For instance, a certain function of the property of X to be predicted can be written as a function of contributions of first (i), second (j), and third (k) order groups as in the Eq.(1),

$$f(X) = \sum_i N_i C_i + \sum_j M_j D_j + \sum_k E_k O_k \quad (1)$$

where C_i, D_j, and O_k are fitted contribution factors and $N_i, M_j,$ and E_k are the number of times these groups appear in the compound. It was concluded in Frutiger et al.'s study that addition of higher order groups (second and the third) did not improve the accuracy of the predictions. Therefore, in this study, only the first order contributions are considered.

3. Methodology

3.1. Data collection

The experimental property data for GWP and ODP (100 years timeline for both) were collected from a recent paper from Huijbregts et al., and the data for LC_{50} were collected from the Ecotoxicology Database (ECOTOX) of U.S. Environmental Protection Agency. Available data for GWP covers chlorofluorocarbons, hydrochlorofluorocarbons, hydrofluorocarbons, chlorocarbons and hydrochlorocarbons, halons, fully fluorinated species, halogenated alcohols and ethers, whereas for ODP available data is much more limited, to those of chlorofluorocarbons, hydrochlorofluorocarbons, and halons. Before proceeding with model development, an

outlier treatment function is applied to the data, which trims the observations that are more than 1.5 interquartile ranges above the upper quartile or below the lower quartile. This ensures the developed model are robust to the outliers in the data. To determine the predictors for the multiple linear regression model, an occurrence matrix of each specific group present in the each different compound is generated using ICAS© software, which requires canonical SMILES expression of compounds. A robust regression using a nonlinear least squares solution algorithm (trust region reflective algorithm of MATLAB's lsqnonlin solver) is performed to calculate group contribution factors. The following functions are applied to the investigated properties.

$$f(GWP) = \log(GWP) = \sum_i N_i C_i \tag{2}$$

$$f(ODP) = \log(ODP) = \sum_i N_i C_i \tag{3}$$

$$f(LC_{50}) = -\log(LC_{50}) + LC_{50_0} = \sum_i N_i C_i \qquad with\ LC_{50_0} = 2.97 \tag{4}$$

Using the occurrence matrix and contribution factors as described in above equations, the first order GC predictions are obtained. An assessment on the goodness of fit of different models is made using the R^2 metric, which can be interpreted as the proportion of variability in property data that can be accounted for by the set of predictors in the model. A good estimate of the generalization potential of the developed models can be obtained using cross-validation. To account for that, one fifth of the data was held out as a test sample, and the model was trained on the remaining data. Performance measures, such as root mean squared error (RMSE) and mean absolute error (MAE), are calculated for the test set as shown in Eq. (5) and (6) respectively.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i^{exp} - y_i^{pred})^2} \tag{5}$$

$$MAE = \frac{1}{N}\sum_{j}^{N} |y_j^{exp} - y_j^{pred}| \tag{6}$$

When developing predictive models on datasets with strong clusterization, it is crucial to select a training set that is representative of the overall data trends. This problem can be addressed with the use of stratified sampling approach, which takes into account the clusters in the data and generates training and test samples preserving percentage of clusters for each set. This was the case with GWP data, and therefore a representative training set was created using this approach.

3.2. Description of neural networks based approach

Neural networks have gained a renewed interest among property modellers after their successful implementation in numerous other scientific and engineering problems such as function approximation, pattern recognition, etc. Recently, Mondejar et al. (2017) applied neural networks to predict thermophysical properties (critical temperature, critical pressure, normal boiling point, etc.) of a group of compounds using a GC based molecular groups as predictors. As the number of data available is so limited for the properties of interest in this study, our approach only uses the first order groups to keep the complexity of the models at a reasonable level.

Basically, a feedforward neural network (the kind of network where information flows only in one direction from inputs to targets) comprising of one hidden layer and one output layer (each with different number of neurons) was developed using the Neural Network toolbox of MATLAB software (Mathworks Inc.). Given the training data, each layer in the network adjusts its internal weights using a training algorithm and generates an output using a transfer function. The output of the output layer gives the network response. Development of an optimal network configuration for each property is achieved following a grid search approach. Each different combinations of three different transfer functions (i.e. logarithmic, tangential, and linear sigmoid functions) with three different training algorithms (i.e. Bayesian regularization backpropagation, Levenberg-Marquardt backpropagation, and scaled conjugate gradient backpropagation) was tested for different numbers of neurons in the hidden layer. As the initial assignment of weights and bias are random, each different network configuration is repeated 10 times. The configuration giving the lowest RMSE on the test set (the same set used for GC) was chosen. To prevent the network from overfitting, the number of network parameters is kept below the number of data points. Finally, the network performance is also compared with other popular machine learning algorithms such as regression trees, radial basis function networks using built-in MATLAB functions. Also applied is a piece-wise linear regression model using multivariate adaptive regression splines (MARS) method, for which the toolbox of Jekabsons (2016) is used.

4. Results and discussions

Performance results of first order group contribution based models trained on full data sets are summarised in Table 1. The predictors defined by GC models are able to explain the variations in the investigated properties with an acceptable accuracy (R^2 higher than 0.6) using a multiple linear regression model without the intercept term. However, cross-validation performance scores of GC models were not as satisfactory as the training performance. Table 2 shows performance metrics of holdout cross-validation results of GC based models in comparison with other methods. Compared to other methods, a significant reduction in the root mean squared error of GWP predictions (from around 2000 to 1196) was achieved using the hybrid approach with ANN. Similar reductions in RMSE were also observed for ODP and LC50 data.

Table 1. Summary of performances of first order GC-based models trained on full data sets.

Property	Number of data points used	Number of predictors	R2	MSE	RMSE	MAE	ARE
GWP	109	30	0.8544	1.72e+6	1311	689	13.92
ODP	30	16	0.9934	0.011	0.033	0.020	0.26
LC50	305	115	0.7983	0.4532	0.673	0.477	14.9

Figure 1 shows the comparison of the predictions attained using GC and neural networks based approaches with the experimental GWP and ODP data. GC method severely underestimates some of the compounds in GWP data while the neural network more closely follows the experimental data points.

Figure 2 shows the comparison of GC and ANN models for LC50 data prediction. Both models have some over and under estimates. The results suggest that the set of functional groups used in the model building are not able to explain the variations in LC50 data as well as they did for the other properties.

Table 2: Holdout cross-validation performance of GC models compared with other methods.

	GWP (kg CO2eq/ kg GHG)			ODP (kg CFC11-eq/ kg ODS)			LC50 (-Log10 mol/L)		
Model	R2	RMSE	MAE	R2	RMSE	MAE	R2	RMSE	MAE
GC	0.68	2002	940	0.35	0.90	0.50	0.39	1.37	0.95
MARS	0.60	2126	1835	0.32	0.40	0.33	0.34	1.34	1.05
RBFN	0.70	1968	1387	0.64	0.34	0.29	0.28	1.70	1.35
TREE	0.44	2553	1560	0.24	0.44	0.31	0.51	1.14	0.82
ANN	0.88	1196	741	0.97	0.06	0.03	0.74	0.77	0.55

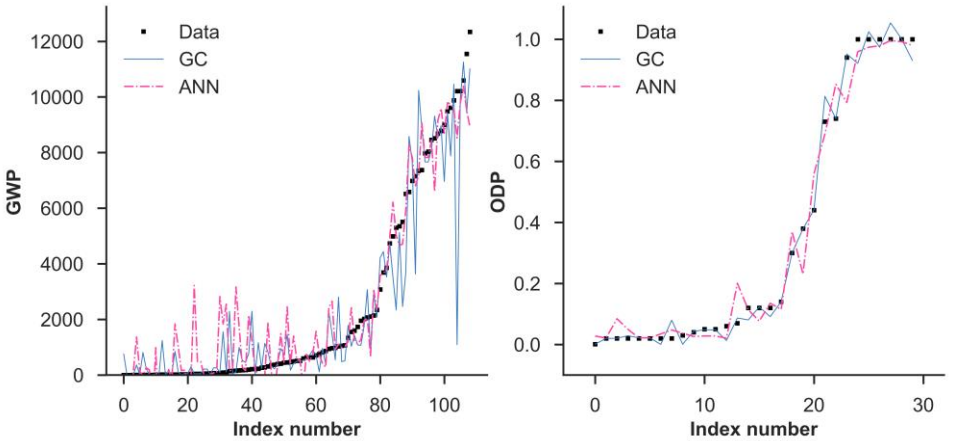


Figure 1: Comparison of group contribution method (line) and artificial neural network (dashed line) based method for predictions of the global warming potential data (left) and ozone depletion potential data (right). ANN models more closely follow the data points.

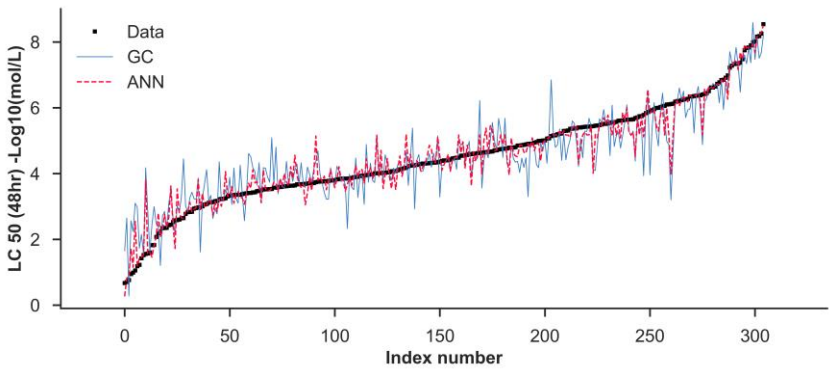


Figure 2: Comparison of group contribution model (line) and hybrid model using ANN (dashed line) for predictions of the full Daphnia magna LC50 dataset.

5. Conclusions

Predictive property models were developed for global warming potential, ozone depletion potential, and Daphnia magna lethal concentration 50 using the latest available data from the literature. First order functional groups defined by classical group contribution method are able to model the data, however, generalization potential of these models are found limited using holdout cross-validation. Compared to a multiple linear regression model as applied in GC methods, we have shown that a feedforward artificial neural network model using the same first order functional groups as inputs has better cross-validation performance for each of the investigated properties, also outperforming other tested surrogate models. This suggests that it is possible to describe relationships between first order functional groups and predicted environmental properties using an artificial neural network model with a better prediction accuracy.

Acknowledgements

The authors acknowledge funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no.675251.

References

A. Hukkerikar, S. Kalakul, B. Sarup, D. Young, G. Sin, R. Gani, 2012, Estimation of environment-related properties of chemicals for design of sustainable processes: Development of group-contribution+ (GC+) property models and uncertainty analysis, *Journal of Chemical Information and Modeling*, vol. 52, issue 11, pp. 2823-2839, doi:10.1021/ci300350r

G. Jekabsons, ARESLab (2016): Adaptive Regression Splines toolbox for Matlab/Octave, 2016, available at http://www.cs.rtu.lv/jekabsons/

M.A.J. Huijbregts, Z.J.N. Steinmann, P.M.F. Elshout, G. Stam, F. Verones, M. Vieira, M. Zijp, A. Hollander, R. van Zelm, 2017, ReCiPe2016: A harmonised life cycle impact assessment method at midpoint and endpoint level, *International Journal of Life Cycle Assessment*, 22(2), 138-147, doi:10.1007/s11367-016-1246-y

M. Mondejar, S. Cignitti, J. Abildskov, J. Woodley, F. Haglind, 2017, Prediction of properties of new halogenated olefins using two group contribution approaches, *Fluid Phase Equilibria*, vol. 433, pp. 79-96, doi:10.1016/j.fluid.2016.10.020

J. Frutiger, C. Marcarie, J. Abildskov, G. Sin, 2016, A comprehensive methodology for development, parameter estimation, and uncertainty analysis of group contribution based property models - an application to the heat of combustion, *Journal of Chemical and Engineering data*, vol. 61, pp. 602-613, doi: 10.1021/acs.jced.5b00750

G. Wernet, S. Papadokonstantakis, S. Hellweg, & K. Hungerbühler. (2009). Bridging data gaps in environmental assessments: Modeling impacts of fine and basic chemical production. Green Chemistry, 11(11), 1826. https://doi.org/10.1039/b905558d

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50283-4

Towards a low carbon economy via sorption-enhanced water gas shift and alcohol reforming

Diana Iruretagoyena[a,b], Nixon Sunny[a,b], Ehecatl A. del Rio-Chanona[a,b], David Chadwick[a], Niall Mac Dowell[b,c], Nilay Shah[a,b*]

[a]*Department of Chemical Engineering,Imperial College London, South Kensington Campus, London, SW7 2AZ, UK*
[b]*Centre for Process System Engineering, Imperial College London, South Kensington Campus, London, SW7 2AZ, UK*
[c]*Centre for Environmental Policy, Imperial College London, South Kensington Campus, London, SW7 2AZ, UK*
n.shah@imperial.ac.uk

Abstract

Hydrogen (H_2) is one of the most important raw materials in the chemical and refinery industries. H_2 is also regarded as a future "energy vector" due to its potential to act as an ultraclean fuel in the heat, power, and transport sectors. Therefore, development of efficient, low carbon routes to produce H_2 is essential to meet its current and growing demand. In particular, this has placed an imperative on improving the efficiency of steam reforming of hydrocarbons (SRH), this process being considered the most feasible and economic route to large scale H_2 production. The water gas shift reaction (WGS) is one of the most important stages of SRH, but is equilibrium limited and requires improvements in energy efficiency. In this study, we demonstrate that the overall efficiency of the WGS can be improved by removing CO_2 in situ and co-feeding alcohols such as methanol and ethanol. The feasibility of this novel concept is investigated by conducting thermodynamic analyses of the alcohol reforming/WGS (alcohol-to-shift) reactions for H_2 production alone and with simultaneous CO_2 adsorption (sorption-enhanced, SEalcohol-to-shift). To this end, a non-stoichiometric approach based on the minimisation of the Gibbs free energy is used. The results show that adding alcohols to the feed facilitates autothermal operation of the shift unit and significantly increases the amount of H_2 produced. The H_2 productivity can be further enhanced by adsorbing CO_2 in situ. The theoretical studies presented here are carried out under relevant operating conditions for SRH and aim to serve as a guideline for future work on alcohol-to-shift processes enhanced by adsorption.

Keywords: Hydrogen, WGS, alcohol reforming, sorption-enhancement, CO_2 adsorption

1. Introduction

Hydrogen (H_2) is considered to be a future "energy vector" playing a key role in the mitigation of climate change, which is one of the biggest energy challenges facing the world today (Committee on Climate Change, 2016). In particular, the use of H_2 as an ultraclean fuel in the transport, power, and heat sectors has the potential to provide deep carbon emission reductions. Contrary to fossil fuels that produce more than 180 gm/kWh CO_2 equivalent after combustion, H_2 has no emissions at the point of use. In addition, H_2 is an important raw material for many valuable chemicals (*e.g.* ammonia

and methanol), and it is widely used in hydrotreating processes to increase the yield of light fractions obtained from heavy oils and to meet increasingly strict environmental regulations (Iruretagoyena and Montesano, 2018). Hydrogen can be produced from a variety of sources and routes although the cheapest is steam reforming of methane and other hydrocarbons (SRH), [$CH_4 + H_2O \leftrightarrow 3H_2 + CO$]. A typical SRH plant includes a reformer followed by a water gas shift (WGS) stage, [$CO + H_2O \leftrightarrow H_2 + CO_2$] and a purification section. Steam reforming is strongly endothermic and the heat of reaction has to be supplied to the system, which accounts for a significant part of the production costs. Consequently, emphasis has been given to the development of processes that minimise the energy input, *e.g*. autothermal reforming (Zennaro et al. 2013) and to the improvement of reformer designs. However, while the energy efficiency of the reformer has increased significantly in recent years, in industry there has been little progress on improving the efficiency of the shift stage. One of the few attempts to enhance the performance of the WGS section is the methanol-to-shift (MtoS) process developed by Haldor Topsøe, which combines the exothermic WGS and the endothermic methanol decomposition [$CH_3OH \leftrightarrow CO + 2H_2$] and methanol reforming [$CH_3OH + H_2O \leftrightarrow CO_2 + 3H_2$] reactions, aiming to minimise the energy losses while boosting the H_2 productivity (up to 25 %). (Nielsen et al. 2009)

The efficiency of the WGS can also be significantly enhanced by the so-called sorption-enhanced water gas shift reaction (SEWGS), which combines the WGS with in situ removal of CO_2 using a solid adsorbent (Jang et al. 2012). This shifts the equilibrium towards the products thereby increasing the production of H_2. Additional advantages of SEWGS include lower consumption of steam, improved energy efficiency and capital cost reduction through process simplification. In this contribution, we investigate the feasibility of combining the methanol-to-shift process with CO_2 adsorption (i.e. SEMtoS). The analysis of the shift section is based on a typical composition of the effluent from a steam methane reformer (Rostrup-Nielsen, 2008). The effects of temperature, pressure, steam and methanol addition, fraction of CO_2 adsorption and energy efficiency of the shift systems are investigated. An advantage of using methanol as energy carrier in the process is that methanol is produced in very large scale at places where natural gas is cheap, and in the future is expected to serve as a vector for renewable hydrogen. The suitability of co-feeding ethanol (ethanol-to-shift, EtoS), a renewable fuel, in the WGS unit is also explored. This study extends the research on methanol-to-shift reported previously (Iruretagoyena, 2016)

2. Methods

2.1 Equilibrium calculations

An algorithm to calculate the chemical equilibrium of the WGS and alcohol-to-shift reactions with simultaneous CO_2 adsorption was developed following a non-stoichiometric approach based on the minimization of the Gibbs free energy (Iruretagoyena, 2016). The non-linear algebraic system, involving Lagrange multipliers, was solved in MATLAB (R2013a) using the in-built routine *fsolve* and the *code newtons* (Yang et al. 2005). A particular equation was written for carbon dioxide to account for the fraction, *f,* of CO_2 removed by adsorption, Eq. (1). In addition, a subroutine was developed to calculate the fugacity coefficients of each species in the gas mixture using the *Soave-Redlich-Kwong* equation of state. To validate the methodology used in this study, a comparison was made with published experimental

equilibrium data available for the forward and reverse water gas shift, methanol decomposition, methanol synthesis, MtoS and SEWGS (Nielsen et al. 2009, Graaf et al. 2016, Stevens Jr. et al. 2010).

$$\Delta G^{f}_{CO_2} + RTln\left(P\frac{n_{CO_2}(1-f)}{\sum_{i=1}^{N} n_i - n_{CO_2} f}\right) + RTln\phi_{CO_2} + \lambda_C + 2\lambda_O = 0 \quad (1)$$

2.2 Species in the alcohol-to-shift units

The free energy minimisation approach used here requires the specification of the species that are in chemical equilibrium in the system, i.e. methanol, ethanol. carbon monoxide, hydrogen and water. If a well-designed catalyst is used, the formation of by-products can be maintained at trace levels as reported by Nielsen et al (Nielsen et al. 2009). Therefore, the formation of by-products such as methane, dimethyl ether and formic acid is not considered in the present study.

2.3 Description of the simulated shift stage

The shift stage consists of an equilibrium reactor in which WGS alone or in combination with alcohol steam reforming takes place. To assess the energy efficiency of the section, besides the energy balance of the shift reactor, the heat that is removed from the effluent of the reformer and transferred to the water or alcohol streams if they are co-fed to the reactor are included. In the sorption-enhanced cases, a heat of adsorption of 17 kJ molCO_2^{-1} was assumed (Ding and Alpay, 2000). The equilibrium calculations presented in this study assume that CO_2 is the only species that is adsorbed on the surface, which is a good approximation to real adsorption systems using layered double oxide as adsorbents (Ding and Alpay, 2000, Iruretagoyena, 2014). The feed to the WGS reactor considered in our case study corresponds to the effluent from a heat exchange reformer operating at 20 barg and 1,173 K with a steam to carbon ratio 2.5:1, Table 1 (Rostrup-Nielsen, 2008).

Table 1. Feed composition to the shift section

Compounds	Composition (mole %)
H_2O	26.1
H_2	54.0
CO	12.3
CO_2	4.5
N_2	0.9
CH_4	2.2
Flow (Nm^3/h)	530

3. Results and discussion

3.1 Methanol-to-shift

The combination of WGS and steam reforming of methanol increases the productivity of H_2 in the range of temperatures relevant in the conventional WGS catalytic process (573-773 K), Figure 1a. Methanol is transformed to H_2 and CO, which is then shifted to produce even more H_2. The H_2 productivity shows a maximum between 523 K and 593 K, an operation window where both reactions, i.e. WGS and methanol steam reforming, are thermodynamically favourable. The operating pressure in the WGS alone has little impact on the equilibrium product distribution of the WGS since there is no change in the number of moles during the reaction. On the contrary, for the MtoS, the

thermodynamic constraints on methanol steam reforming at low temperatures can be compensated for by reducing the operating pressure due to the twofold increase in the number of moles, Figure 1b. In industry, it is more efficient to increase the pressure of natural gas rather than H_2, since smaller volumes need to be compressed. Consequently, the optimal operating pressure of a steam-methane reformer is typically around 20 barg (Rostrup-Nielsen, 2008).

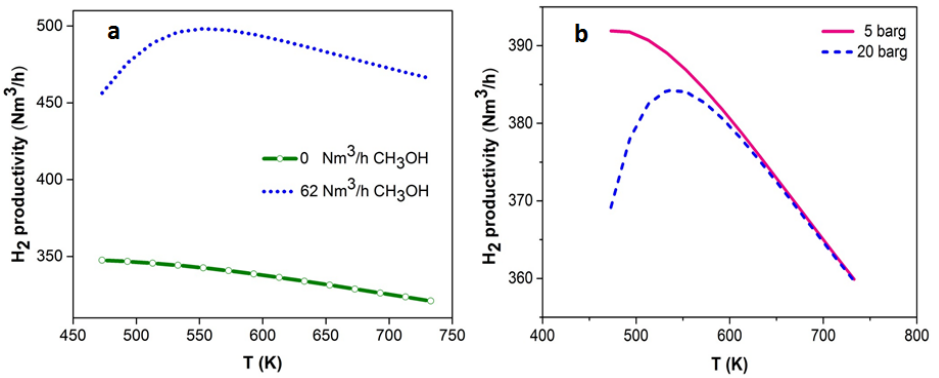


Figure 1. (a) Effect of methanol addition (20 barg). (b) Effect of temperature and pressure (CH_3OH: 16 Nm^3/h)

Besides the increase in the amount of H_2 produced, the main advantage of the MtoS process over WGS is to achieve autothermal operation of the shift unit which is not possible with the WGS alone. The heat released by WGS compensates the heat consumed by the steam reforming of methanol. It is worth nothing that for a given syngas composition, the amount of H_2 produced autothermally at equilibrium depends solely on the reactor temperature, Figure 2. This can be attributed to the fact that the partial pressures of the species have a negligible effect on the equilibrium conversion of WGS so the amount of methanol that needs to react does not depend on the pressure of the system. On the other hand, less methanol needs to be added at low pressure to achieve autothermal operation due to the more favourable thermodynamics for methanol reforming. Slightly higher H_2 productivities are obtained over the whole range of operating temperatures if the methanol feed to the MtoS reactor is replaced by a mixture of water and methanol. This stems from the decrease in the partial pressures of the species involved in the methanol decomposition, and from the shift of the WGS equilibrium towards CO_2 and H_2 due to an increased concentration of water.

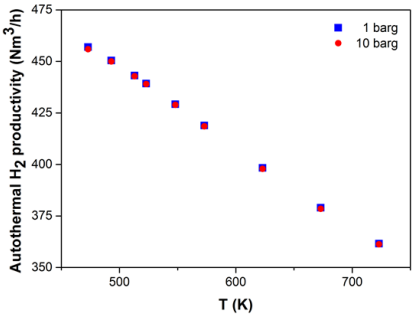


Figure 2. H_2 productivity under autothermal conditions (range of CH_3OH required for autothermal operation: ~20-60 Nm^3/h)

3.2 Sorption-enhanced methanol-to-shift

The maximum in the hydrogen productivity for the MtoS process between 523 K and 593 K, corresponds to the temperature window in which both WGS and methanol reforming are favourable. Therefore, adsorption of CO_2 becomes useful above ~523 K to overcome the WGS equilibrium limitations and to benefit from high reaction rates. When the MtoS process is conducted in the presence of a solid adsorbent, there is a marked increase in the hydrogen productivity as the fraction of CO_2 adsorbed, *f*, increases, and for different values of *f*, the maximum H_2 productivity is centered between 523 K and 593 K, Figure 3. Pressure has little effect on the H_2 productivity of SEMtoS at high temperatures, whereas low pressures favor the production of H_2 at temperatures below 623 K.

Under sorption-assisted conditions, the heat balance in the MtoS reactor must include the contribution of CO_2 adsorption. Considering a heat of adsorption of 17 kJ molCO_2^{-1} (Ding and Alpay, 2000) it is found that there is a methanol to carbon monoxide feed molar ratio that allows autothermal operation. This ratio decreases slightly in the temperature range relevant for SEMtoS. Due to the endothermicity of the steam reforming of methanol, the amount of methanol that needs to be added increases significantly at low temperatures.

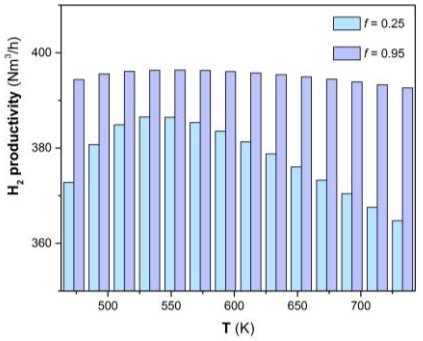


Figure 3. Effect of temperature on the H_2 productivity, SEMtoS (20 barg, 16 Nm^3/h CH_3OH)

3.3 Sorption-enhanced ethanol-to-shift

Analyses analogous to those described above for SEMtoS were conducted for sorption-enhanced ethanol-to-shift (SEEtoS). The overall results show a similar optimal operating window (i.e. ~523 K to 773 K). In this temperature range, the removal of CO_2 shifts the equilibrium towards the products, thereby increasing the production of H_2. A comparative study between SEMToS and SEEToS using the same carbon to CO feed molar ratio shows that the product distribution of both systems is approximately the same in the temperature range of interest. Moreover, the production of H_2 is only marginally higher in the methanol-system compared to that of the ethanol-system. Interestingly, an energy analysis shows that for both SEMtoS and SEEtoS, there is an amount of alcohol added which enables autothermal operation. However, if methane is produced at equilibrium as by-product, the system is not energy efficient at temperatures below 673K, i.e. the overall heat of reaction of SEWGS is lower than that of the sorption-enhanced alcohol-to-shift systems under the same operating conditions. Therefore, it is very important to use highly selective catalysts in the processes (Nielsen et al. 2009).

4. Conclusions

In this study a non-stoichiometric approach based on the minimization of the Gibbs free energy was used to assess the thermodynamic feasibility of sorption-enhanced alcohol-to-shift systems. SEMtoS and SEEtoS are promising strategies to enhance the productivity of H_2 while ensuring that the energy efficiency of the WGS is improved. The amount of H_2 produced in the SEalcohol-to-shift processes shows a maximum between 523 and 593 K, an operating window where the WGS and alcohol steam reforming reactions are thermodynamically favored. The adsorption of CO_2 in situ enhances the H_2 productivity more than other strategies to shift equilibrium (e.g. increasing the amount of steam in the feed or reducing the pressure of the shift units). A major advantage of the SEalcohol-to-shift systems compared to SEWGS is the ability to achieve autothermal operation of the shift unit, which is not possible with the WGS alone. The heat released by WGS and CO_2 adsorption is compensated by the heat consumed in the steam alcohol reforming reactions. The systems are not energy efficient if methane is produced at equilibrium as by-product.

References

Committee on Climate Change, 2016: Next steps for UK heat policy (https://www.theccc.org.uk/wp-content/uploads/2016/10/Next-steps-for-UK-heat-policy-Committee-on-Climate-Change-October 2016.pdf).

Ding, Y.; Alpay, E., 2000, Equilibria and kinetics of CO_2 adsorption on hydrotalcite adsorbent. *Chemical Engineering Science*, *55* (17), 3461-3474.

Graaf, G. H.; Winkelman, J. G. M., 2016, Chemical Equilibria in Methanol Synthesis Including the Water–Gas Shift Reaction: A Critical Reassessment. *Industrial & Engineering Chemistry Research, 55* (20), 5854-5864.

Iruretagoyena Ferrer, D., 2016, Sorption-Enhanced Methanol-to-Shift for H_2 Production: In *Supported Layered Double Hydroxides as CO_2 Adsorbents for Sorption-enhanced H_2 Production*, Springer International Publishing: Cham; pp 165-180.

Iruretagoyena, D.; Montesano, R., 2018, Selective Sulfur Removal from Liquid Fuels Using Nanostructured Adsorbents. In *Nanotechnology in Oil and Gas Industries: Principles and Applications*, Saleh, T. A., Ed. Springer International Publishing: Cham, pp 133-150.

Iruretagoyena, D.; Shaffer, M. P.; Chadwick, D., 2014, Adsorption of carbon dioxide on graphene oxide supported layered double oxides. *Adsorption*, *20* (2-3), 321-330.

Jang, H. M.; Lee, K. B.; Caram, H. S.; Sircar, S., 2012, High-purity hydrogen production through sorption enhanced water gas shift reaction using K_2CO_3-promoted hydrotalcite. *Chemical Engineering Science*, *73* (0), 431-438.

Nielsen, P. E. H.; Hansen, J. B.; Schiødt, N. C., 2009, Process for the preparation of a hydrogen-rich stream. US 7,527,781 B2.

Rostrup-Nielsen, J. R., 2008, Steam Reforming. In *Handbook of Heterogeneous Catalysis*, Wiley-VCH Verlag GmbH & Co. KGaA.

Stevens Jr, R. W.; Shamsi, A.; Carpenter, S.; Siriwardane, R., 2010, Sorption-enhanced water gas shift reaction by sodium-promoted calcium oxides. *Fuel, 89* (6), 1280-1286.

Yang, W. Y.; Cao, W.; Chung, T.; Morris, J., 2005, *Applied numerical methods using MATLAB*. John Wiley and Sons: NJ, USA.

Zennaro, R.; Ricci, M.; Bua, L.; Querci, C.; Carnelli, L.; d'Arminio Monforte, A., 2013, Syngas: The Basis of Fischer–Tropsch. In *Greener Fischer-Tropsch Processes for Fuels and Feedstocks*, Wiley-VCH Verlag GmbH & Co. KGaA, pp 17-51.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50284-6

Shear Tolerance and Lipid Content of *Botryococcus braunii* During and Post Non-Destructive Solvent Extraction

Brent A. Jackson[1], Parisa A. Bahri[1*], Navid R. Moheimani[2]

[1]School of Engineering and Information Technology, [2]Algae R & D Centre, School of Veterinary and Life Sciences, Murdoch University, 90 South Street, Murdoch, WA
P.Bahri@murdoch.edu.au

Abstract

The effect of applied shear stress on *Botryococcus braunii*, in the presence of solvent is important to further understanding the mechanisms behind in-situ solvent extraction and will be critical in the design and scale up of an effective extraction process. This paper looks at the effect of various mixing rates on colony structure and culture photosynthetic activity in the presence of heptane. Partial colony disruption was found to occur at shear rates of 335 s^{-1}, with the main mechanism of disruption believed to be microeddies generated by turbulence around the impeller. Photosynthetic activity was seen to drop by 55 % during the extraction period requiring a short recovery time of 3 days, reaching 66.5 % of its original value. These results provide valuable insight and data on the shear limits of *B. braunii* required in the design of a repeated milking process and will assist in the design and scale up of a viable milking process.

Keywords: Shear rate, Botryococcene, Microalgae, Lipid production, Solvent extraction.

1. Introduction

Botryococcus braunii is a fresh water, green microalgae, which forms colonies supported by an extracellular matrix (ECM) rich in long chain hydrocarbons of a nature which make them ideal as a source of biofuel (Banerjee, Sharma et al. 2002, Weiss, Roth et al. 2012). Historically the production of high value products from microalgae has followed a four-step process involving; biomass cultivation, harvesting/dewatering, extraction, and separation. This may work with the production of high value products; however, biofuels require larger volumes of product. Furthermore, while *B. braunii* may have a relatively high lipid content, it is a slow growing microalgae (Cabanelas, Marques et al. 2015). This is where methods such as in-situ solvent extraction become attractive. The ability of *B. braunii* to store these lipids extracellularly make them accessible for solvent extraction, with minimal damage to the cell wall, after which *B. braunii* cells can regenerate lipid stores for future extraction events (Frenz, Largeau et al. 1989).

Generally cell cultures require little agitation to achieve effective mixing, reducing the shear generated from impellers (Pol and Tramper 1998). However, higher mixing rates are required for effective solvent dispersion in a solvent extraction process. For large scale culture, the increased power input from can generate turbulent microeddies around the impeller. As average microeddy length drops below the cell diameter, cells can become damaged and rupture (Papoutsakis 1991, Pol and Tramper 1998, Mardikar and Niranjan 1999).

Solvent toxicity is likely to weaken colonies and increase the effect of these microeddies on culture viability, making this an important factor to consider in design of an effective non-destructive extraction process. There are five factors to consider in solvent selection; Biocompatibility, Downstream processing, Extraction efficiency, Cost and Safety (Jackson, Bahri et al. 2017). Solvents with high biocompatibility tend to have low extraction efficiencies resulting in the requirement of increased volumes and exposure periods (Miazek, Kratky et al. 2017). As the exposure period is increased the effect of solvent toxicity is increased resulting in cell mortality. Larger volumes and contact periods also result in increased process costs when considering scale up and economic viability. A solvent which has a reasonable extraction efficiency with an acceptable toxicity which does not result in irreversible damage to the culture is required for a viable large-scale process.

For this study heptane has been chosen due to a balance of biocompatibility (Frenz, Largeau et al. 1989, Moheimani, Cord-Ruwisch et al. 2013) and extraction efficiency, with ease of downstream separation and relatively low cost (Jackson, Bahri et al. 2017). The purpose of this paper is to observe and understand the effect of hydrodynamic shear on *Botryococcus braunii* during the solvent extraction process and the effect it has on culture activity and recovery.

2. Materials and Methods

2.1. Culture conditions:

Botryococcus braunii (BOT-22, B Race), originally obtained from The AOACC culture collection, Tsukuba, Japan, was used for this study. Culture conditions were identical to that of Moheimani, Matsuura et al. (2013). Culture density was maintained at 1.3 – 1.5 gL^{-1}.

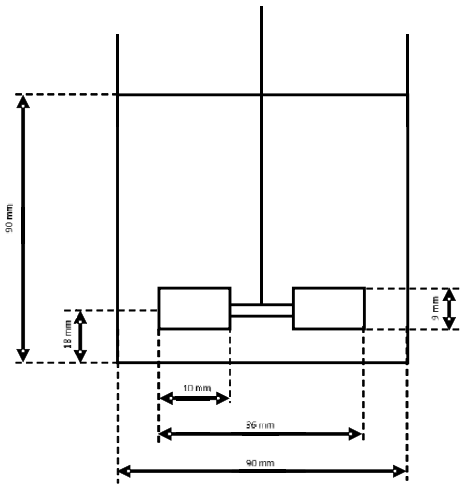


Figure 1 – Extraction vessel dimensions.

Figure 1 shows the dimensions of the vessel used for lipid extraction. For effective extraction it is important to keep turbulence to a minimum whilst still achieving sufficient solvent dispersion. To determine the minimum required mixing rate, tests were completed with water and heptane, increasing the mixing rate to a point where solvent dispersal is achieved. Eq. (1) was generated to select mixing rates for this work.

$$N_{min} = 21.5R + 289 \qquad 1.33 < R < 9 \qquad (1)$$

Where; R is the ratio of water to heptane and N_{min} is the minimum impeller speed necessary to achieve solvent dispersion (rev.min^{-1}).

Each 2-hour extraction was run in triplet. Photosynthetic activity was measured at 20-minute intervals for the duration of the experiments. Lipid and biomass content was measured directly before and after extraction. Heptane was added to the culture at time zero and topped up every 20-minutes throughout all extractions.

An R value of 2.33 and mixing rate of 500 rpm was used for the second round of experiments with a solvent contact period of 4 hours. Culture activity was measured half hourly during the extraction period and daily post extraction. CO_2 addition pre/post extraction was maintained at 1 Lmin^{-1} (1 % CO_2). Light was supplied to one side of the flask at 77 ± 0.4 μmol photons m^{-2}s^{-1}. Both light and CO_2 cycled on a 12 hr on: 12 hr off cycle.

Culture activity was measured using F_V/F_M as explained by Moheimani, Matsuura et al. (2013). 6 ml of sample was used for all activity samples. Samples taken during the extraction phase were replaced with fresh culture. All samples were diluted (6 ml sample : 9 ml fresh media) and dark adapted for 30 minutes prior to measurement.

2.2. CSTR shear rate correlation

Rushton, Costich et al. (1950) produced power curves for many stirred tank reactor setups correlating vessel power number (N_P) with Reynolds number (N_{Re}), one of these being an unbaffled reactor with a 6 bladed Rushton turbine, much the same as what was used in this study. Using these correlations the power draw generated by the agitator can be determined, which can then be used to calculate the shear rate imposed on the system (Coker 2001). Alternatively, the power requirement can be modelled using a complex set of equations taking into account; the vessel dimensions (height, diameter, number of baffles, width of baffles, height of baffles), impeller dimensions/type (blade length and width, number of blades, impeller diameter) and the characteristics of the fluid to be agitated (Furukawa, Kato et al. 2012). These equations can be used similar to that of Coker (2001) to obtain the power draw from the system and shear rate, leading to a correlation between shear rate (γ, s^{-1}) and mixing rate (N, rpm). For detailed equations and related correlations refer to Furukawa, Kato et al. (2012).

Eq. (2) (Coker 2001) and (3) (Furukawa, Kato et al. 2012) are the resulting correlations from the two methods described. Averaging these two equations gives Eq. (4), which was used for shear rate calculations on this system.

$$\gamma = 0.8603N - 81.677 \qquad 330 < N < 600 \qquad (2)$$
$$\gamma = 0.8440N - 101.280 \qquad 330 < N < 600 \qquad (3)$$
$$\gamma = 0.8522N - 91.479 \qquad 330 < N < 600 \qquad (4)$$

3. Results

In all extractions, once the heptane was added the culture activity steadily decreased over the extraction period, levelling out around 0.3 (approx. 40 % of the original value). At an R value of 4 and mixing rate of 380 rpm, very little loss in culture photosynthetic activity is seen with a reduction in activity of 9.7 %. At this mixing rate, a substantial portion of the solvent was still sitting on the surface, undispersed, resulting in reduced

culture/solvent contact. When increased to 400 rpm a significant drop in activity resulted in a final reduction of 52 %. As this was increased to 500 rpm, the initial drop in activity was much greater, reducing by 32 % after 40 minutes compared to 11 % at 400 rpm. There was no significant difference between the final activity at mixing rates of 500 and 600 rpm.

Table 1 – Reduction in culture activity after 2 hours of solvent contact.

Mixing rate	***380 rpm (232 s^{-1})***		***400 rpm (249 s^{-1})***		***500 rpm (335 s^{-1})***		
Culture/Heptane	400/100	350/150	500/0	400/100	500/0	400/100	350/150
R	4	2.33	-	4	-	4	2.33
Loss in activity	9.7 %	26.4 %	2.3 %	52.3 %	1.2 %	54.4 %	54.9 %

There was a clear initial drop in culture condition at 500 rpm at both R values of 4 and 2.33 which was not seen at 380 or 400 rpm. At a mixing rate of 380 rpm solvent dispersion was inefficient with an R value of 4, resulting in reduced solvent contact and a much lower reduction in activity. As R was reduced to 2.33 the culture was seen to be under more stress with the activity dropping by 26 % over the two-hour extraction period. Little difference could be seen in the culture performance at R values of 4 and 2.33 at a mixing rate of 500 rpm, indicating solvent dispersion is complete. In the absence of heptane, the culture was completely un-harmed.

Colony observations were made post extraction at a mixing rate of 500 rpm and culture/heptane ratio of 2.33. As to be expected, the effect of heptane on colony structure is clear. Large colonies post-extraction were less common with more single/double cells present in the culture. The colonies which were still intact were smaller and had a yellow/brown tinge to them indicating the production carotenoids in response to stress (Ambati, Ravi et al. 2010). With the extraction of lipids and hydrocarbons from the ECM, the strength of the colonies was compromised, increasing the cultures sensitivity to hydrodynamic shear. There are also what appear to be lipid globules surrounding the colony, indicating small tears in the ECM retaining sheath (Figure 2C).

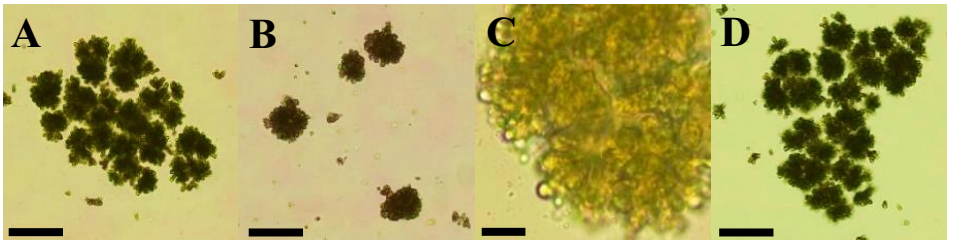


Figure 2 – Experiment 1.74 (500 rpm, R = 2.33); culture condition pre-extraction (A), directly after extraction (B, C) and 24 hours after extraction (D). Scale; A, B, D = 100 µm, C = 20 µm.

After just 24 hours of CO_2 exposure, the colonies returned to their original size/colour (Figure 2D). The extraction period at 500 rpm was extended to 4 hours with little further reduction in photosynthetic activity observed, showing the algae are biocompatible with heptane and can withstand prolonged exposure. The fact that the culture activity does not drop much in the second half of the extraction suggests the algae are stressed but are not dying, likely to be a result of colony disruption. Post extraction cultures recovered to an activity of 0.58 where it remained for the 6 days it was monitored.

4. Discussion

Botryococcus braunii has shown a strong tolerance to mixing rates of up to 500 rpm. At the lower mixing rates (N < 400 rpm) solvent dispersion was not complete with a portion of solvent remaining on the surface of the culture. This resulted in reduced contact and lower reductions in culture activity. At 500 rpm solvent dispersion was almost immediate, with culture activity dropping much quicker, stabilizing at 38 % of the original value after two hours. Solvent contact with the colony weakened the colony outer sheath, making colonies more susceptible to hydrodynamic shear. This resulted in the breaking up of colonies into smaller sub colonies and in some cases single/double cell aggregates (Figure 2B). While complete colony disruption can result in culture death (Hou, Park et al. 2014), the level of disruption in this case was not enough to do so.

As colonies are disrupted and split into more smaller colonies, the resistance to microeddies generated by shear forces around the impeller is expected to increase. The Kolmogorov scale can be used to model minimum microeddy length (η, m) (Eq. (5) and (6), (Papoutsakis 1991)).

$$\eta = (v^3/\varepsilon)^{1/4} \tag{5}$$

$$\varepsilon = P/\rho V \tag{6}$$

Where; v is the fluid kinematic viscosity (m^2s), ε is the rate of energy dissipation into the fluid body (m^2s^{-3}).

If the microeddy size is equal to, or less than that of the cells being mixed, damage is likely to occur (Papoutsakis 1991, Pol and Tramper 1998). At the mixing rates used in this work, the minimum microeddy sizes observed were 45-66 µm. At 500 rpm the eddy size calculated is 50 µm, which is much smaller than the large colonies seen pre-extraction (> 400 µm, Figure 2A), resulting in the breaking up of colonies into the smaller colonies seen post extraction (45-100 µm, Figure 2B). The size of these colonies suggests they would also be compromised, which is seen in the lipids leaking from the ECM (Fig. 3D), and a portion of the colonies which have been completely broken up, resulting in free floating single/double cells. Individual cells are much smaller than 45 µm and would be unaffected by the shear rates imposed in this work. The loss in activity must therefore be attributed to the break-down of colonies via mechanisms described by Papoutsakis (1991). Increasing the period of exposure is likely to lead to increased loss of lipids from the ECM and further colony disruption, which in this case would increase extraction efficiencies. Unfortunately, *B. braunii* cells have been seen to struggle once separated from the colony (Hou, Park et al. 2014), too much disruption is likely to result in increased cell mortality requiring an extended recovery phase.

5. Conclusions/Recommendations

The shear limits of *Botryococcus braunii* obtained in this work are vital in the design and scale up of an effective non-destructive extraction process. A mixing rate of 500 rpm (335 s^{-1}) at culture/solvent ratio of 2.33 proved to be optimal with a reduction in activity of 55%. At this mixing rate colonies were under enough stress to release lipids from the ECM while cellular activity was maintained for a quick recovery. The presence of solvent had a dramatic effect on the cultures shear tolerance which must be considered in the process design. The shear limits observed in this paper can be utilized in alternate extraction vessel design and modeling saving a great deal of time and resources in process

feasibility studies. Furthermore, this gives an initial starting point required for further work on long term repetitive extractions reducing the number of variables to be tested.

References

Ambati, R. R., S. Ravi and R. G. Aswathanarayana (2010). "Enhancement of Carotenoids in Green Alga-*Botryococcus braunii* in Various Autotrophic Media under Stress Conditions." International Journal of Biomedical Pharmaceutical Sciences **4**(2): 87-92.

Banerjee, A., R. Sharma, Y. Chisti and U. C. Banerjee (2002). "*Botryococcus braunii*: A Renewable Source of Hydrocarbons and Other Chemicals." Critical reviews in biotechnology **22**(3): 245 - 279.

Cabanelas, I. T. D., S. S. I. Marques, C. O. de Souza, J. I. Druzian and I. A. Nascimento (2015). "*Botryococcus*, what to do with it? Effect of nutrient concentration on biorefinery potential." Algal Research **11**: 43-49.

Coker, A. K. (2001). Modeling of chemical kinetics and reactor design. Boston, MA, Gulf Professional Pub.

Frenz, J., C. Largeau, E. Casadevall, F. Kollerup and A. J. Daugulis (1989). "Hydrocarbon recovery and biocompatability of solvents for extraction from cultures of *Botryococcus braunii*." Biotechnology and bioengineering **34**: 755-762.

Furukawa, H., Y. Kato, Y. Inoue, T. Kato, Y. Tada and S. Hashimoto (2012). "Correlation of Power Consumption for Several Kinds of Mixing Impellers." International Journal of Chemical Engineering **2012**: 1-6.

Hou, L., H. Park, S. Okada and T. Ohama (2014). "Release of single cells from the colonial oil-producing alga *Botryococcus braunii* by chemical treatments." Protoplasma **251**(1): 191-199.

Jackson, B. A., P. A. Bahri and N. R. Moheimani (2017). "Repetitive non-destructive milking of hydorcarbons from *Botryococcus braunii*." Renewable & sustainable energy reviews **79**: 1229-1240.

Mardikar, S. H. and K. Niranjan (1999). "Observations on the shear damage to different animal cells in a concentric cylinder viscometer." John Wiley & Sons.

Miazek, K., L. Kratky, R. Sulc, T. Jirout, M. Aguedo, A. Richel and D. Goffin (2017). "Effect of Organic Solvents on Microalgae Growth, Metabolism and Industrial Bioproduct Extraction: A Review." Int J Mol Sci **18**(7).

Moheimani, N. R., R. Cord-Ruwisch, E. Raes and M. A. Borowitzka (2013). "*Non-destructive oil extraction from Botryococcus braunii (Chlorophyta)*." Journal of Applied Phycology **25**(6).

Moheimani, N. R., H. Matsuura, M. M. Watanabe and M. A. Borowitzka (2013). "Non-destructive hydrocarbon extraction from *Botryococcus braunii* BOT-22 (race B)." Journal of Applied Phycology **25**(5).

Papoutsakis, E. T. (1991). "Fluid-mechanical damage of animal cells in bioreactors." Elsevier Science Publishers LTD **9**: 427 - 437.

Papoutsakis, E. T. (1991). "Media additives for protecting freely suspended animal cells against agitation and aeration damage." TIBTECH **Volume 9**: 316 - 324.

Pol, L. v. d. and J. Tramper (1998). "Shear sensitivity of animal cells from a culture medium perspective." Tibtech **16**: 323 - 328.

Rushton, J. H., E. W. Costich and H. J. Everett (1950). "Power characteristics of mixing impellers." Chemical Engineering Progress **46**: 467-476.

Weiss, T. L., R. Roth, C. Goodson, S. Vitha, I. Black, P. Azadi, J. Rusch, A. Holzenburg, T. P. Devarenne and U. Goodenough (2012). "Colony Organisation in the Green Alga *Botryococcus braunii* (Race B). Is Specified by a Complex Extracellular Matrix " Eukaryotic Cell **11**(12): 1424 - 1440.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50285-8

How can we solve systemic problems in plant-derived production based on simulation-based analysis?

Kotaro Ouchida[a], Tatsuya Okubo[a,b], Yasunori Kikuchi[a,b*]

[a]*Department of Chemical System Engineering, The University of Tokyo, 7–3–1 Hongo, Bunkyo-ku, Tokyo, 113–8656, Japan*
[b]*Presidential Endowed Chair for "Platinum Society", Organization for Interdisciplinary Research Project, The University of Tokyo 7–3–1 Hongo, Bunkyo-ku, Tokyo 113–8656, Japan*
kikuchi@platinum.u-tokyo.ac.jp

Abstract

While utilizing plant-derived resources has emerged as a key toward sustainability, systems should solve systemic problems that cannot be easily coped with because of their far-reaching and potentially unpredicted influence. In this study, we have tackled the systemic problems in plant-derived production through a case study on sugarcane-derived production. First, we developed a process model that can simulate mass and energy balance in a cane sugar mill considering agricultural and industrial technology options based on the site investigation, literature search and data analysis. Using the developed process model, a simulation-based analysis was conducted targeting Tanegashima, which is one of the sugarcane-farming areas in Japan. Some alternatives that a single decision maker or technology have never found were generated. Finally, the potential of this approach to generate scenarios for reaching the alternatives was confirmed through a scenario-planning meeting involving actual engineers in a cane sugar mill. It can be concluded that a simulation-based approach can be a strong tool to intensify plant-derived production through solving the systemic problems across agriculture and industry.

Keywords: Interdisciplinary approach, sugarcane, bioethanol, combined production

1. Introduction

Plant-derived resources have emerged as key alternatives to fossil resources for production of fuels and chemicals. There have been innumerable number of studies on the development of technology for converting plant-derived resources into the products within the industrial field (e.g., Vohra et al., 2014; Humbird et al., 2011). The benefit of these technologies, however, could be affected by the agricultural process as a raw material acquisition stage of plant-derived production. Unlike mining of fossil resources, agriculture has some unique characteristics that restrict the yield, composition and availability of plant-derived resources such as intervention of farmers and vulnerability to weather, insects and diseases. The method for the design of systems encompassing agriculture and industry is needed.

It is challenging to change agriculture and industry simultaneously because of the unclear relationship between them and existing limitations such as current land use, equipment in the plant, infrastructure or regulations. For example, in Japanese cane

sugar industry, it is not easy to transform from single sugar production to combined sugar and ethanol production even if the latter is acknowledged as better. While some high-yielding sugarcane cultivars that increase both sugar and ethanol production have been developed (Ohara et al., 2009), their adoption is challenging because of their high impurities that deteriorate the efficiency of sugar crystallization processes. In recent years, a technology has been developed to mitigate the unpleasant effect from the impurities and to support the introduction of the high-yielding cultivars (Ohara et al., 2012). However, it may be still hard for sugar mills to decide the technology introduction without building consensus with farmers in advance.

There are sometimes systemic problems, which cannot be clearly attributed to specific causes and, therefore, are difficult to be solved systematically in plant-derived production. The systems often become complex because of their lying on two different type of systems, i.e., agriculture and industry, which are different in decision makers, time scales, uncertainties and data availability. Because communication between them is insufficient, one player sometimes makes a decision that has far-reaching and potentially unpredicted influence on others' decisions and the performance of the systems. Although mere rational decision makings from the viewpoint of each player cannot always avoid systemic problems, appropriate communication within the systems may overcome them to adapt the new environment (Checkland and Scholes, 1990).

Over the decades, a simulation-based analysis has leveraged its expertise in synthesizing data to model real complex systems and to support decision makings (Klatt and Marquardt, 2009). Multi-scale modelling has been a recent challenge for the design of the systems having different temporal and spatial scales (Garcia and You, 2016). It should be a strong tool to design the systems of plant-derived production overcoming the differences between agriculture and industry, which have been obstacle to quantitative discussion.

In this study, we tackle systemic problems in plant-derived production through a case study on sugarcane-derived production with a simulation-based analysis. First, we develop a process model that can simulate mass and energy balance in a cane sugar mill considering agricultural and industrial technology options based on the site investigation, literature search and data analysis. Then, a scenario-planning meeting is conducted involving actual engineers in a cane sugar mill to verify the effectiveness of the model to generate scenarios for reaching new alternatives.

2. Materials and method

2.1. Requirements for systems design through solving systemic problems

Figure 1 shows the requirements for systems design in plant-derived production. To solve systemic problems, we propose two phases as described below: systematic approach through modelling and systemic approach with scenario-planning meeting with the model.

To avoid systemic problems, each player should recognize how his/her decision influence the whole system. It is necessary to develop a tool for analysing agricultural and industrial processes quantitatively through collecting and synthesizing data and knowledge that are dispersed across the systems. Since agriculture and industry are different in time scales, uncertainties and data availability, modelling them in an

integrated manner would be one of the challenges which process systems engineering should tackle with.

Just because a model is developed to help players to search alternatives considering agricultural and industrial options simultaneously, it does not follow that they are easily implementable for three reasons. First, the easiness of the transition highly depends on the order of options because each option would more or less have an impact on other options. Second, we should not overlook that some options need a period of time. For example, it takes several years for an unspecified large number of farmers to change the cultivar, while the constraints from externals such as regulations may change, encouraging or discouraging the player to take other options. Third, there may remain subjects that only simulation-based analysis cannot bring to light because the range of discussion with the model is limited by data availability and uncertainty from model structure. A method of planning the systems transition is needed to achieve the alternative.

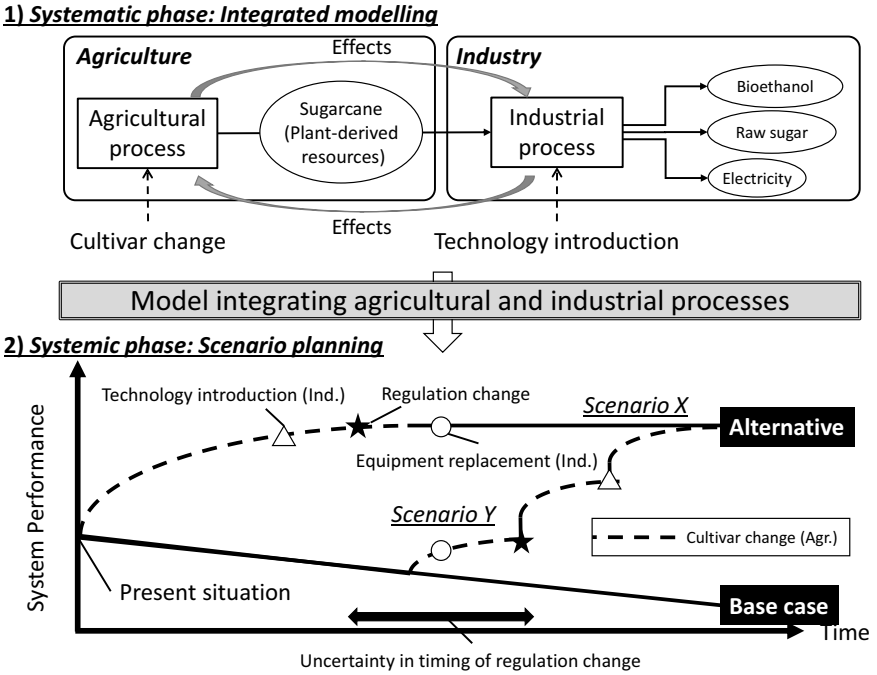


Figure 1 Requirements for systems design solving systemic problems in plant-derived production

2.2. Modelling and scenario-planning meeting

A process model that can simulate mass and energy balance in a cane sugar mill considering agricultural and industrial technology options has been developed (Ouchida et al., 2017a) and extended to analyse the effect of introducing a novel technology on the systems (Kikuchi et al., 2017). Data and information for modelling were collected through detailed investigations of actual sugar mills and literature search. This model could be a tool for promoting communication between agriculture and industry.

As a method of planning the systems transition, a scenario-planning meeting using the model was proposed in this study. A scenario-planning approach is a method for

clarifying the path to the goal and enhancing the ability to deal with uncertainties by considering the multiple possible future situations (Amer et al., 2013). A scenario-planning meeting was held on a sugar mill to generate the practical scenario to achieve the alternative.

3. Results and discussion

3.1. Development of model integrating agricultural and industrial processes

Figure 2 shows the examples of systems performances visualized by the model. Productivity, energy efficiency, lifecycle greenhouse gas emissions, economy and circulation of nitrogen, phosphorus and potassium can be evaluated considering both agricultural and industrial options simultaneously (Ouchida et al., 2016; 2017a). It also enables the implementation analysis of a technology on the systems (Kikuchi et al., 2017) and the operational decision support in the sugar mill considering the fluctuation in the quality of feedstock derived from uncertainties in agriculture (Ouchida et al., 2017b). This model has a potential of generating alternatives that a single decision maker or technology have never achieved.

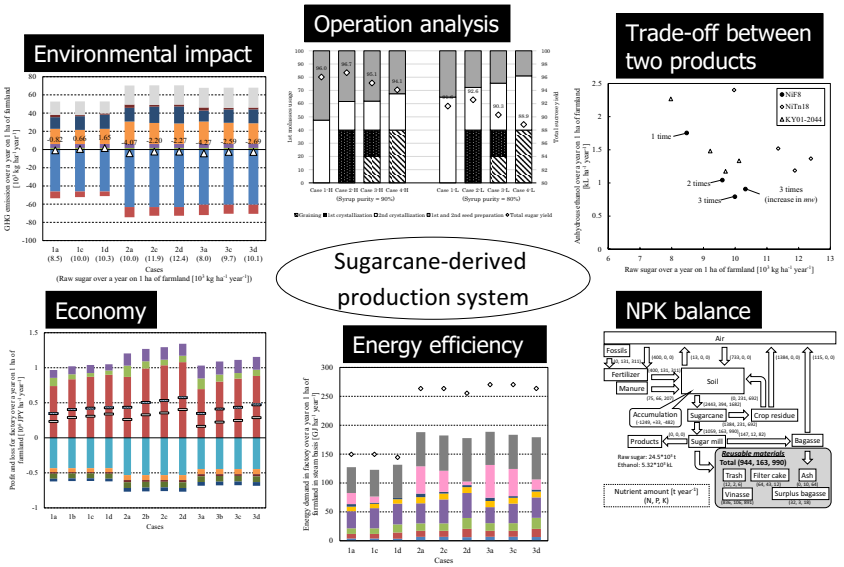


Figure 2 Visualization of systems performance in sugarcane-derived production by model.

3.2. Scenario-planning supported by model

Involving actual decision engineers in a cane sugar mill, a scenario-planning meeting was held on a cane sugar industry in Tanegashima, which is one of the islands in Japan. Tanegashima has 2,500 ha of sugarcane farmland and a sugar mill, which produces only raw sugar and the processing capacity of which is 1,500 t day^{-1}. Since the farmland area is shrinking year by year as the number of population decreases, some measures are needed to avoid the collapse of the cane sugar industry.

The model indicated that a set of key options to the systems transition are the cultivar change to the high-yielding cultivar, the replacement of the present back-pressure steam turbine with a condensing-extraction steam turbine to improve power conversion

efficiency, and starting to sell electricity. Based on the preliminary investigation on Tanegashima, it was found that the timings of the rise in the electricity price and the renovation of the turbine are crucial to the overall scenario plan. Because it is uncertain when the timing of the price rise comes, multiple scenario plans should be prepared.

One of the plans generated is shown in Figure 3. In this case, the timing of the renovation comes after the price rises and, consequently, electricity selling is highly motivated. Therefore, the mill should prepare for enhancing the benefit of electricity selling by cultivar change or energy saving before the renovation.

The feasibility of the scenario plans was discussed with actual engineers. Some obstacles to the implementation were extracted and the solutions for avoiding them were discussed. For example, an increase in the amount of bagasse by cultivar change will arise other problems such as the increase in labour loads of farmers in harvesting and the difficulty in the management of its storage. Such discussion has not appeared in the past. It can be expected that an iterative procedure of modelling and scenario planning will blush up the scenarios to make them more practicable.

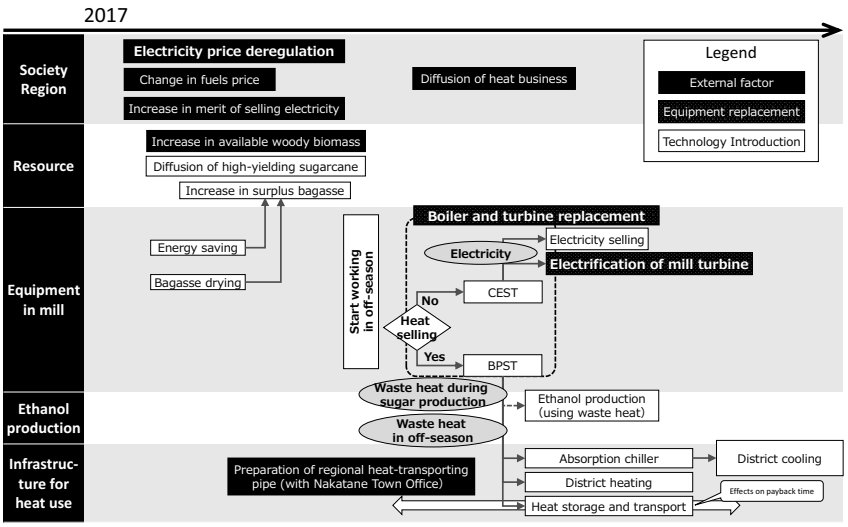


Figure 3 A scenario plan generated for the cane sugar mill and farmers in Tanegashima.

4. Conclusions

Through a case study on sugarcane-derived production, it was revealed that an integrated modelling of agricultural and industrial processes and a scenario-planning meeting using the model were effective to solve systemic problems in plant-derived production. In spite of the uncertainties, data availability and model limitations, a simulation-based approach can support alternative generations considering both agricultural and industrial options. In addition, actual decision makers can proceed to the next step through finding new challenges and generating alternative scenarios for avoiding them. In conclusion, modelling and scenario-planning meeting could deepen the discussion and contribute to solving systemic problems in plant-derived production. It can be expected that the development of the framework for conducting these procedure iteratively will intensify plant-derived production seamlessly.

Acknowledgement

The authors are grateful to Dr. Yasuhiro Fukushima, Dr. Satoshi Ohara, Dr. Akira Sugimoto, Dr. Yusuke Tarumoto, Dr. Taiichiro Hattori, Mr. Yoshifumi Terajima, Mr. Hidenori Hidaka, and Mr. Yosuke Hamada for their fruitful discussion and support. The authors also acknowledge the support given by the Japan International Research Center for Agricultural Sciences, and Shinko Sugar Mill Company. Part of this study is financially supported by grants from JSPS KAKENHI Grant Numbers 16H06126 (Grant-in-Aid for Young Scientists A). Activities of Presidential Endowed Chair for "Platinum Society" in the University of Tokyo are supported by the KAITEKI Institute Incorporated, Mitsui Fudosan Corporation, Shin-Etsu Chemical Co., ORIX Corporation, Sekisui House, Ltd., and East Japan Railway Company.

References

M. Amer, T.U. Daim, A. Jetter, 2013, A review of scenario planning. Futures, 46, 23–40.

P. Checkland, J. Scholes, 1990, Soft Systems Methodology in Action. John Wiley & Sons, Chichester, UK.

D.J. Garcia, F. You, 2016, The water-energy-food nexus and process systems engineering: A new focus. Computers and Chemical Engineering, 91, 49–67.

D. Humbird, R. Davis, L. Tao, C. Kinchin, D. Hsu, A. Aden, P. Schoen, J. Lukas, B. Olthof, M. Worley, D. Sexton, D. Dudgeon, 2011, Process Design and Economics for Biochemical Conversion of Lignocellulosic Biomass to Ethanol. Golden, Colorado: National Renewable Energy Laboratory.

Y. Kikuchi, K. Ouchida, Y. Kanemtsu, S. Ohara, Y. Fukushima, 2017, Retrofit energy integration for selective fermentation in cane sugar mills under hot/cold energy availability, Journal of Chemical Engineering of Japan, 50(4), 297–308.

K.U. Klatt, W. Marquardt, 2009, Perspectives for process systems engineering: Personal views from academia and industry, Computers and Chemical Engineering, 33, 536–550.

S. Ohara, Y. Fukushima, A. Sugimoto, Y. Terajima, T. Ishida, A. Sakoda, 2009, Reduction in greenhouse gas emissions from process retrofitting and cultivar improvement in combined sugar-ethanol production from sugarcane. Journal of Life Cycle Assessment, Japan, 5(4), 439–445.

S. Ohara, Y. Fukushima, A. Sugimoto, Y. Terajima, T. Ishida, A. Sakoda, 2012, Rethinking the cane sugar mill by using selective fermentation of reducing sugars by *Saccharomyces dairenensis*, prior to sugar crystallization. Biomass and Bioenergy, 42, 78–85.

K. Ouchida, Y. Hamada, T. Okubo, Y. Kikuchi, 2017, Simulation-based analysis for operational decision support on scheduling in sugar crystallization considering quality of molasses and syrup. Computer Aided Chemical Engineering, 40, 1807–1812.

K. Ouchida, Y. Kikuchi, T. Okubo, 2016, Integrated modeling of agricultural and industrial processes within life cycle design for environment. Computer Aided Chemical Engineering, 38, 1947–1952.

K. Ouchida, Y. Fukushima, S. Ohara, A. Sugimoto, M. Hirao, Y. Kikuchi, 2017, Integrated design of agricultural and industrial processes: A case study of combined sugar and ethanol production. AIChE Journal, 63(2), 560–581.

M. Vohra, J. Manwar, R. Manmode, S. Padgilwar, S. Patil, 2014, Bioethanol production: Feedstock and current technologies. Journal of Environmental Chemical Engineering, 2, 573–584.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50286-X

Optimization of the Thermal Efficiency of a Fixed-Bed Gasifier using Computational Fluid Dynamics

Nunurai Kapfunde[a], Cornelius Mduduzi Masuku[a*], Diane Hildebrandt[b]

[a]*Department of Civil and Chemical Engineering, University of South Africa, Private Bag X6, FLORIDA, 1710, South Africa.*
[b]*Materials and Process Synthesis, College of Science, Engineering and Technology, University of South Africa, Private Bag X6, FLORIDA, 1710, South Africa.*
masukcm@unisa.ac.za

Abstract

Coal or biomass gasification can be used to produce methane rich syngas and the heating value of this gas depends on methane concentration. In this research we use computational fluid dynamics (CFD) to optimize a fixed-bed gasifier to produce methane as a fuel. Only a limited amount of what is required to model gasification is available today in literature and most models fail to capture the intricate operations of a gasifier, but CFD can use this limited information to provide further information on parameters that cannot be easily and safely measured through experimentation.

Gasifier hydrodynamics, tar production, heat and mass transfer coefficients, and pressure drops are the principal areas requiring research for the modeling and optimization of a fixed-bed gasifier. A CFD model using the Euler–Euler model in ANSYS Fluent was developed to model the effect of composition and temperature on the production of tar and methane.

The results show that lower temperatures favor methane production. The devolatilization rate increases with increasing fluid solid interactions, heat and mass transfer rates. However, this lowers the methane production and hence the thermal efficiency of the gasifier. Thus, the use of CFD to optimize a fixed-bed gasifier does not only extend our knowledge of gasification and its components but moves us closer to providing the much-needed sustainable energy solutions that are easily affordable and accessible.

Keywords: Fixed-bed Gasifier Optimization, Coal and Biomass Devolatilization, Computational Fluid Dynamics Modeling, Energy Efficiency.

1. Introduction

Africa especially Sub-Saharan Africa still battles with a lack of basic service provision and a lack of basic infrastructure required to connect people to sustainable energy supplies (Butera et al, 2016; Musango, 2014). Substitute natural gas from biomass gasification could provide a solution but updraft gasifiers currently used in the gasification process are limited by their small processing throughput, and large production of waste water (Chen et al, 2011).

Tar is one of the major waste products in gasification with updraft gasifiers producing about 50mg/nm^3 compared to the required minimum of 0.1mg/nm^3. Tar production is determined by the reactor temperature, pressure, gasification medium and equivalence ratio. Passing the syngas through an external combustor before recycling it, was found to

reduce the amount of tar produced in a downdraft gasifier to minimum levels of about 0.1mg/nm^3 of tar (Susanto et al, 1996, Basu, 2010). This method has not been applied to an updraft gasifier. The purpose of this study was to use purely computational methods to simulate the impact of recycling syngas on tar and methane production in a 0.8x0.5 m^2 steel drum fixed-bed gasifier currently available at the University of South Africa.

2. Model Development

2.1. Literature Models

Most of the methane and the tar produced in a fixed-bed gasifier is assumed to come from devolatilization. Yoon (1978) studied the limits of operation of a moving-bed gasifier using a triangular diagram based on mass and energy balances. He assumed devolatilization to be thermally neutral and the inlet gas temperature to be equal to the outlet gas temperature to avoid tar condensation but focused on the combustion region of the gasifier. Hobbs et al (1992) attempted a comprehensive model of the fixed-bed gasifier removing most of the assumptions used in previous models. The devolatilization reaction was modeled using the functional group process. Tar was treated as a single species that has variable composition dependent on location in the reactor. Brundu et al (2009) considered a one-dimensional kinetic model of a fixed-bed gasifier and carried out a review of gasifier studies. They attempted using iron to reduce tar production during coal gasification. Their study modeled coal devolatilization using a global model. Murgia et al (2012) designed a comprehensive model of an air blown updraft gasifier using devolatilization data from the chemical percolation model.

2.2. Experimental Setup to be Modeled

The experimental gasifier currently stands on a raised palette and is fed compressed air from a compressor. Coal is charged from the top of the gasifier and gas is released through an outlet that is connected to a flare. The gasifier does not have a steam inlet. In our simulation, coal was charged into the gasifier forming a packed-bed. This result was achieved in ANSYS Fluent 18.2 using the region patch option and using the multiphase Euler–Euler method for the gas and solid phase. The simulation had two inlets, one for the compressed air as in the original gasifier and the other as a recycle inlet.

Two cases were run, with the first case involving zero input of syngas from the recycle stream. In ANSYS Fluent this was achieved by converting the feature type of the recycle inlet from a velocity inlet into a wall. The other case had the recycle inlet with a velocity specified together with the compressed air inlet. Syngas samples were used as the recycle inlet species composition. Changes within the gasifier in temperature, pressure, tar and methane concentrations were measured. The gasifier model was generated using design modeler and meshed using the parameters described in Table 1 below.

Table 1: Gasifier Model Parameters

Measurement	**Value**	**Source**
Density (kg/m^3)	Incompressible mixture	This study
Void fraction	0.465	Murgia et al, 2012
Particle diameter (m)	0.05-0.08	This study

Coal mass flow rate (kg/s)	Fixed bed	This study
Air velocity (m/s)	5	This study
Steam mass flow rate (kg/s)	0	This study
Wall temperature (°C)	42	Murgia et al, 2012
Bed height (m)	0.6	This study
Reactor length (m)	0.8	This study
Reactor width (m)	0.5	This study
Inlet and outlet pipe diameters (m)	0.02	This study

3. Hydrodynamics and Reaction Modeling

3.1. Hydrodynamic Model Selection

A purely theoretical computational fluid dynamics model was used to analyze the variation of methane production in a fixed-bed gasifier with temperature and pressure. The coal material used in this study was Jetson coal, with material parameters obtained from studies by Murgia et al (2012).

An Euler–Euler model was applied in this research and not the Lagrangian model to describe the behavior of coal, because the Lagrangian method was found to be more computationally expensive compared to the Euler–Euler method. The Lagrangian method is well suited for particles that are in motion such as found in an entrained bed gasifier but cannot adequately describe particles in a packed bed as in the case of the 0.8x0.5 m fixed-bed gasifier used in this research.

The Euler–Euler method applied in this model describes the phases as interpenetrating continua and this is ideal to describe fluidized bed gasifiers and in the case of a fixed bed gasifier it is more adequate for dense particle systems.

The Euler–Euler model in ANSYS does not provide an inbuilt model for the devolatilization process as the Lagrangian method does and so this study modelled devolatilization using coal as a mixture that contains three species namely char (solid carbon), volatiles and ash. All the volatiles in the coal are assumed to form methane, water, carbon dioxide, tar, carbon monoxide and char. The reaction kinetics data for the rest of the reactions within the gasifier are obtained from Hobbs et al (1992).

3.2. Computational Fluid Dynamics Modelling

ANSYS Fluent calculates the multiphase flow by solving the governing equations using numerical methods. The reactions that occur within the gasifier are included within the simulation through rate of reaction terms in the equations. Four main governing equations are solved. The mass conservation equation, the species mass conservation equation, the momentum conservation equation and the energy conservation equation.

The Euler–Euler method describes both the solid and gaseous phases as continuous phases. Therefore, the continuity equation is given by:

$$\frac{\partial(\rho\varepsilon)}{\partial t} + \Delta \cdot (\rho\varepsilon u) = \sum_{j=1}^{N} R_j \quad (1)$$

The conservation of mass of each species i is described by:

$$\frac{\partial(\rho\varepsilon Y_i)}{\partial t} + \Delta \cdot (\rho\varepsilon Y_i u) = \sum_{j=1}^{N} R_{ji} \quad (2)$$

The momentum balance equation is:

$$\frac{\partial(\rho\varepsilon u)}{\partial t} + \Delta \cdot (\rho\varepsilon uu) = \Delta \cdot S + \rho\varepsilon g - \sum_{k=1}^{N} I_{ki} \quad (3)$$

The energy balance is given by:

$$\rho\varepsilon Cp[\frac{\partial(T)}{\partial t} + u.\Delta T] = -\Delta q - Hk - \Delta Hr \quad (4)$$

3.3. Reaction Kinetics Modelling

The reactions that occur in the gasifier and the rate of reaction terms that are included in the simulation of the governing equations are listed in table 2 below:

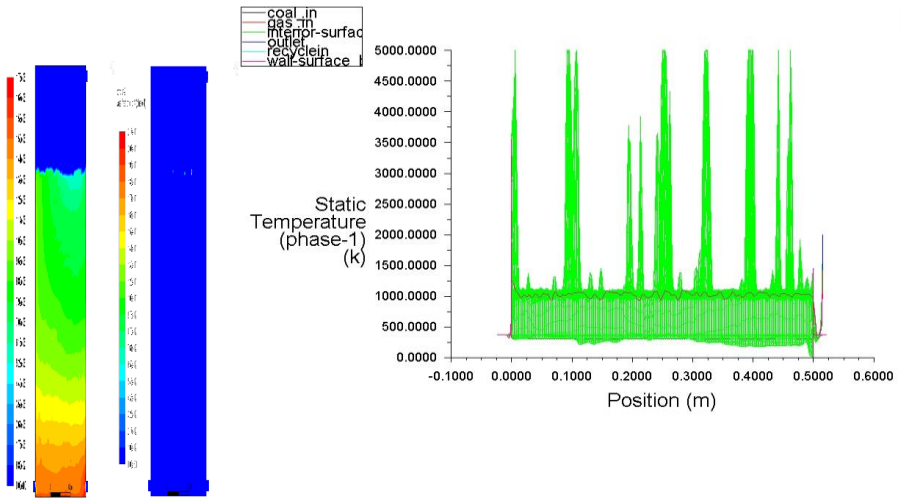
Table 2: Model Reactions and Parameters

Reaction equation	**Pre-exponential factor (m/sK)**	**Activation energy (E/R)**	**Source**
Char + 0.5 O_2 = CO	2.30	11,100	Hobbs et al, 1992
Char + CO_2 = 2CO	589	26,800	Hobbs et al, 1992
Char + H_2O = CO + H_2	589	26,800	Hobbs et al, 1992
Char + H_2 = CH_4	0.589	26,800	Hobbs et al, 1992
CO + 0.5 O_2 = CO_2	2.30	11,100	This study
CO + H_2O = CO_2 + H_2	589	26,800	This study
CH_4 + 0.5 O_2 = CO + 2H_2	2.30	11,100	This study
Coal = Char + Tar + H_2O + CO_2 + CH_4 + CO	$1.14x10^5$	8900	Murgia et al, 2012
Tar = Char +H_2O+CO_2+CH_4+CO	$1.14x10^5$	8900	this study

In this study we assumed that the reaction rate data for the oxidation of carbon monoxide, the water-gas shift and the oxidation of methane have reaction rates similar to those of the heterogeneous reactions. Therefore, literature values (Hobbs et al, 1992) were used for these reactions.

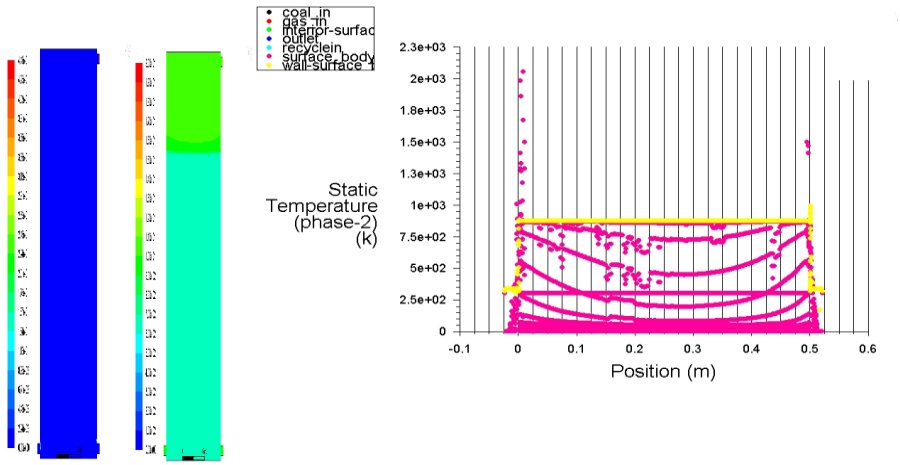
4. Results and Discussion

The CFD model results obtained for the two cases considered in this study are presented in Figures 1 and 2.



Tar methane temperature

Figure 1: Model results with zero input of syngas from the recycle stream



Tar methane temperature

Figure 2: Model results with input of syngas from the recycle stream

The results in Figures 1 and 2 should show that lower temperatures favor methane production. In the first case higher temperatures were achieved than in the second case. More methane was produced in the second case than in the first case with the opposite

occurring for tar. This suggests that the inclusion of syngas into the case reduced gas–solid interactions, heat and mass transfer within the reactor thus, reducing devolatilization and other gasification reactions. The devolatilization rate should increase with increasing fluid solid interactions, heat and mass transfer rates. However, as shown in the first case this lowers the methane production and hence the thermal efficiency of the gasifier.

5. Conclusions

The use of CFD to optimize a fixed-bed gasifier does not only extend our knowledge of gasification and its components but moves us closer to providing the much-needed sustainable energy solutions that are easily affordable and accessible. Recycling syngas into an updraft gasifier is limited by the gas–solid interactions which ultimately reduce the performance of the gasifier.

References

N.R. Amundsen, L.E. Arri, 1978. Char gasification in a countercurrent reactor. AIChE J., 24, 87–101.

M. Brundu, 2010. Gasification of solid fuels: Countercurrent reactor modeling and metallic iron use for tar removal. PhD Thesis, The University of Cagliari.

F. Butera, P. Caputo, R. Adhikari, A. Facchini, 2016. Urban development and energy access in informal settlements. A review for Latin America and Africa. Procedia Eng., 161, 2093–2099.

L. Chen, Q. Jiang, Z. Song, D. Posarac, 2011, Optimization of a methanol yield from a Lurgi gasifier, Chem. Eng. Technol., 34,817–822.

M.L. Hobbs, P.T. Radulovic, L.D. Smoot, 1992. Modelling fixed-bed coal gasifiers. AIChE J., 38, 681–702.

T.M. Ismail, M.A. El-Salam, 2017. Parametric studies on biomass gasification process on updraft gasifier high temperature air gasification. Appl. Therm. Eng., 112, 1460–1473.

S. Murgia, M. Vascellari, G. Cau, 2012. Comprehensive CFD model of an air-blown coal-fired updraft gasifier. Fuel, 101, 129–138.

J.K. Musango, 2014. Household electricity access and consumption behaviour in an urban environment: The case of Gauteng in South Africa. Energy Sustain. Dev., 23, 305–316.

N. Siefert, 2013. Experimental and thermo-economic analysis of catalytic gasification and fuel cell power systems. PhD Thesis, Carnegie Mellon University.

H.I. Somashekhar, S. Dasappa, N.H. Ravindranath, 2000. Rural bioenergy centres based on biomass gasifiers for decentralized power generation: case study of two villages in southern India. Energy Sustain. Dev., 4, 55–63.

H. Susanto, A.A.C.M. Beenackers, 1996, A moving bed gasifier with internal recycle of pyrolysis gas, Fuel, 75, 1339–1347.

A.A.P. Susastriawana, H. Saptoadi, Purnomo, 2017. Small-scale downdraft gasifiers for biomass gasification: A review. Renew. Sust. Energ. Rev., 76, 989–1003.

M. Syamlal, C. Guenther, A. Gel, S. Pannala, 2009. Advanced coal gasifier designs using large-scale simulations. J. Phys. Conf. Ser., 180, 012034.

H. Yoon, J. Wel, M.M. Denn, 1979. Feasible operating regions for moving bed coal gasification reactors. Ind. Eng. Chem. Process Des. Dev. 18, 306–312.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50287-1

Superstructure investigation for P-recovery technologies integration with macroalgae based hydrothermal liquefaction

Stavros Papadokonstantakis[a*], Andrea Gambardella[a], Johan Askmar[a], Yiyu Ding[a]

[a]*Chalmers University of Technology, Maskingränd 2, Göteborg 41258, Sweden*
stavros.papadokonstantakis@chalmers.se

Abstract

The aim of this work is to identify profitable and environmentally benign technological paths connecting phosphorus (P) recovery with macroalgae based hydrothermal liquefaction (HTL). For this purpose, a thorough literature review of relevant technologies in different fields of application was carried out. Together with a citation network analysis (CNA) a framework comprising qualitative and quantitative information about process significance, conditions, material and energy flows for a comprehensive list of potentially involved technologies was developed. On the basis of this information framework, processes were combined into a superstructure of options for P-recovery from macroalgae based HTL. Based on criteria such as technology maturity and severity of process conditions, two different but not mutually exclusive approaches for utilizing HTL waste streams were identified: exploiting the carbon and/or the nutrients potential of the streams. The process layout selected consists of hydrothermal liquefaction, catalytic hydrothermal gasification (CHG), incineration of solid residues, acidic leaching of incineration ashes, crystallization and precipitation of P in form of magnesium ammonium phosphate (struvite), methane steam reforming and biocrude upgrading through hydrotreatment. The selected layout's model was based on HTL reaction kinetics (kinetic constants extrapolated from experiment data and verified with laboratory results from HTL of macroalgae Ulva Lactuca collected from several sites of the Swedish coast) and on performances averaging (according to literature sources) for the other sub-processes sections. Net revenues as high as 21 $\$/t_{dry\ macroalgae}$ were predicted, coming mainly from upgraded oil and a small part from struvite, while including material and energy costs for operating the system. This revenue is highly affected by the cost of the macroalgae feedstock. Finally, environmental aspects and constraints related to the process, were addressed and evaluated with the cumulative energy demand (CED) indicator, which resulted stable around 34 $MJ_{eq}/kg_{products}$.

Keywords: Hydrothermal liquefaction; Phosphorus recovery; Macroalgae.

1. Introduction

Phosphorus (P) is a biologically indispensable nutrient and its largest use is by far in fertilizers (Desmidt et al., 2015). Today almost all P is produced from phosphate rock, but the mining process is not sustainable due to the negligible mineral P regeneration rate compared to the current and the future predicted exploitation. Moreover, world phosphate rock natural reserves are unevenly distributed, with more of 70 % situated in Morocco and the Western Sahara region (Desmidt et al., 2015), making P a geopolitically strategic

element. Phosphorus recovery and/or removal is both implemented at large scales and well documented for the treatment of municipal waste waters, but in the context of hydrothermal liquefaction (HTL) both large scale experience and literature is limited. Thus, the motivation for this work was to investigate such a possibility. In fact, the biomass feed of HTL can have considerable P contents, and only a small amount of it ends in the biocrude; the rest is distributed in the HTL aqueous and solid phases which are considered waste streams. In addition, phosphorus causes eutrophication of aquatic environments and therefore P-rich streams often must be treated prior to disposal.

2. Literature review and citation network analysis

P-recovery technologies can be categorized by the P source (aqueous, solid, sludge or ashes) or by the physical principle (mechanical, chemical, biological or thermal). These processes often consist of P concentration stages and/or heavy metal (HM) removal treatments, but in most cases, precipitation is required to effectively recover useful P-containing compounds. If not precipitated, P can be concentrated in different matrices such as ashes, adsorbents, metal slags and filters, or diluted in aqueous phases and used for cultivation of yeasts or microalgae. Commercial values of the recovered P can be associated to its potential as a fertilizer. Struvite is acknowledged as a slow releasing source of highly bioavailable-P (Desmidt et al., 2015) and therefore precipitation of struvite has been preferred to other phosphorus compounds in this work. The principles of P-recovery in wastewater treatment (WWT) can be applied to HTL even though the streams vary in composition. Moreover, besides nutrients recovery, HTL waste streams can be employed for energy recovery or further exploited for biofuels production. The set of material- or energy-recovery processes integrated to macroalgae based HTL outlines the superstructure shown in Figure 1. Citation network analysis (CNA) of the reviewed literature was carried out to prioritize information sources, assess the technological

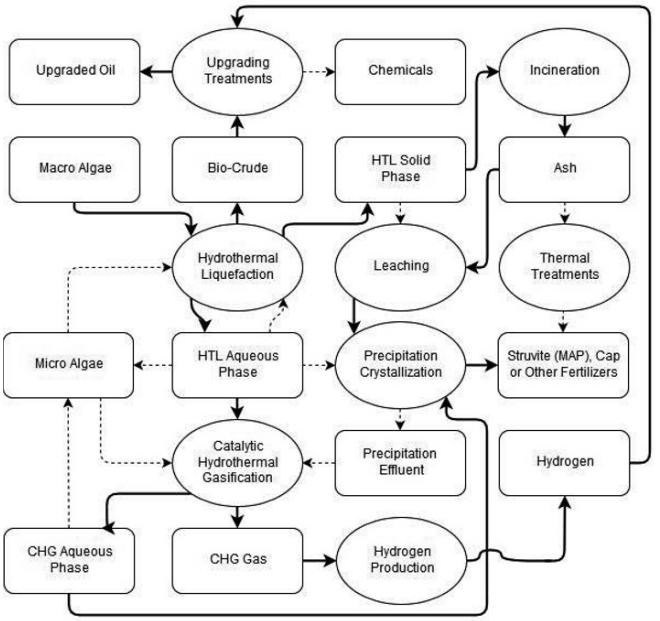


Figure 1: P-recovery for macroalgae HTL superstructure. Ovals and rectangles represent respectively sub-processes and materials. Bold arrows represent the pathway chosen in this work.

maturity of each process and highlight eventual technological affinities within the superstructure. An example of how to interpret the CNA is given in Figure 2: the temporal distribution of the sources was considered as an indicator of technological maturity, whereas the number of studies shared by two different technologies and the citations across sources of different fields were interpreted as indicators of technological affinity. Direct use of the HTL aqueous phase as a nutrients supplement for microalgae cultivation medium was initially taken into account as a valid option and successively discarded for its high costs and its poor affinity with P-recovery shown in the CNA.

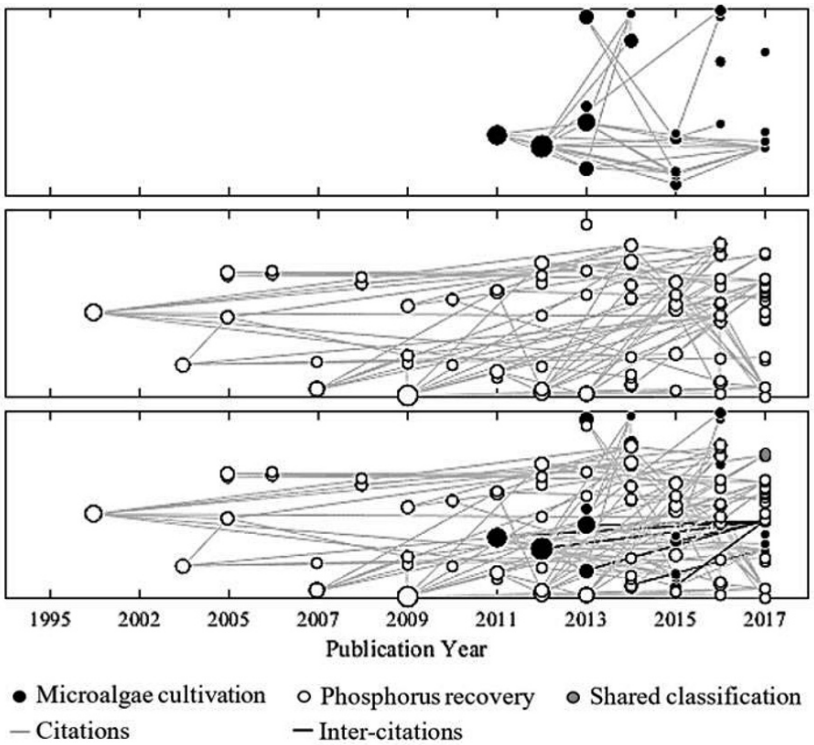
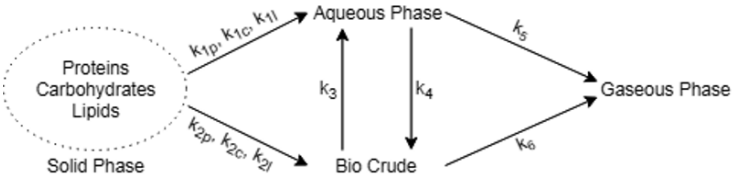


Figure 2: Citation network analysis for microalgae cultivation (top), P-recovery (center) and a combination of the two (bottom). Each circle represents a literature source, and its size is a measure of the number of times it has been cited (i.e. an indicator of its relevance for the analysis). Gray circles refer to literature sources combining microalgae cultivation and P-recovery. Gray lines represent citations within the same field, whereas black lines represent inter-citations. Gray circles and black lines are indicators of technological affinity.

3. Methodology

3.1. HTL kinetic model

To predict the HTL output yields, the general HTL kinetic model developed by Valdez et al. (2014), shown in Figure 3, was updated using a broader set of experimental data retrieved from different sources. The experimental set-ups presented subcritical water condition and a total solid content ranging between 10 % and 20 %, whereas they differed for feedstock biochemical composition, process temperature and residence time. Two different sets of kinetic constants were calculated: for residence times shorter than 15 min and from 15 min to 90 min. Then, the results "*Y*" were merged together according to Eq.(1).



$$Y = Y_{k,\theta<15} \cdot f(T,\theta) + Y_{k,\theta>15} \cdot \left(1 - f(T,\theta)\right) \quad (1)$$

Figure 3: Reactions framework for hydrothermal liquefaction proposed by Valdez et al. (2014).

Where Y_k is the result calculated from the respective set of kinetic constants, and f is a function of T and θ (respectively the HTL temperature and residence time expressed in °C and minutes) bounded between 0 and 1 used for weighting the results according to the pertinent HTL setting. The mathematical expression of $f(T,\theta)$ is given in Eq.(2).

$$f(T,\theta) = \left[1 - \frac{1}{1+\left(\frac{T}{400}\right)^{0,6}}\right] \cdot \frac{1}{1+\left(\frac{\theta}{90}\right)^{0,6}} \quad (2)$$

The reason of this is that using only one set of kinetic constants for the whole span of residence times produced unrealistic results for HTL residence times below 15 minutes.

3.2. P-recovery

HTL aqueous phase undergoes catalytic hydrothermal gasification (CHG) as described by Jones et al. (2014). The main product of CHG is a methane-rich gas which is used for hydrogen production through steam reforming as described by Barthe et al. (2015). Hydrogen is then used to upgrade the biocrude by the hydrodeoxygenation process as reported by Jones et al. (2014). HTL solid phase is incinerated and the P remaining in the ash is then released by acidic leaching at a pH below 2. The P dissolved in the leachate and in the CHG aqueous phase is then precipitated and recovered through pH adjustments with sodium hydroxide and addition of magnesium chloride. Part of the effluent is neutralized and recirculated as process water for the leaching, whereas the rest is sent to WWT. Energy recovery has been performed through heat integration and pinch analysis. Prices and environmental impacts for material and energy flows were retrieved from Ecoinvent database v3.3 except for catalysts prices which were extrapolated from rates from online suppliers. The system operating parameters were then optimized to maximize the net revenues. Furthermore, life cycle environmental aspects were assessed through the Cumulative Energy Demand (CED) indicator.

4. Results and discussion

The results of the HTL model for the macroalgae Ulva Intestinalis having the same biochemical composition given by Neveux et al. (2013) are shown in Figure 4. These results were validated by comparison with the HTL measurements provided by (Raikova et al., 2017). The costs and the maximized revenues for three different options are shown in Figure 5. Option 1 was previously described in section 3.2. In Option 2 and 3 no leaching is performed (ash is disposed of) and in Option 3 the CHG is not performed (HTL aqueous phase directly fed to the precipitation step). The cost of performing the ash leaching is equal or even higher than the value of the struvite recovered from it, but the

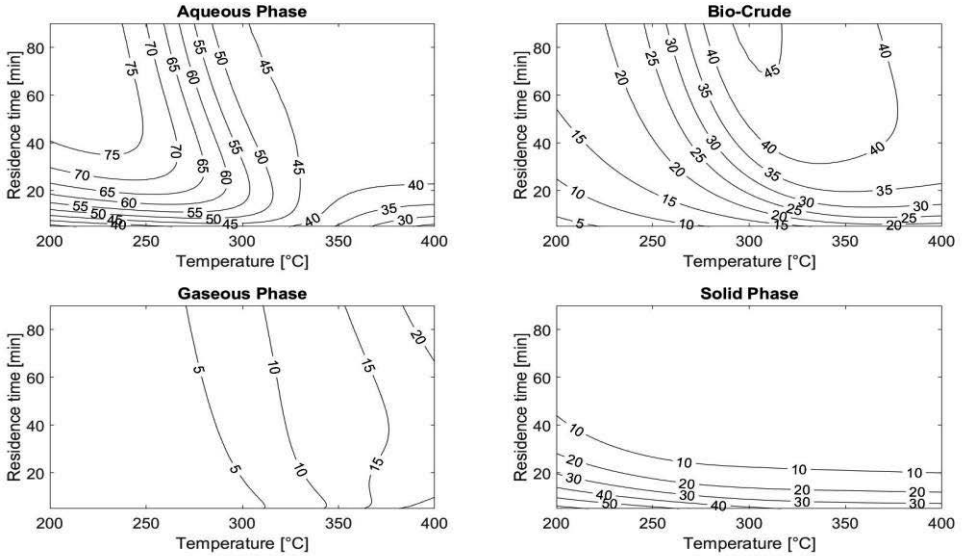


Figure 4: Predicted HTL yields as a percentage of the original dry, ash free feedstock mass.

P-content in the ash plays a key role in this assessment. Not performing the CHG leads to a substantial net revenues loss, as the aqueous phase total organic carbon (TOC) converted in the CHG becomes a cost for the waste water treatment, but might be counteracted by a reduced system capital cost. All the scenarios are heavily affected by the feedstock cost which accounts for more than 60 % of the total costs (the feedstock price being 93 $/t$_{dry\ macroalgae}$). The struvite production, peaking at 7.2 kg/t$_{dry\ macroalgae}$, has a marginal impact on the profits, but pays back its related operative costs. The CED of the three scenarios resulted stable around 34 MJ$_{eq}$/kg$_{products}$ (about 15 MJ$_{eq}$/kg$_{products}$ less

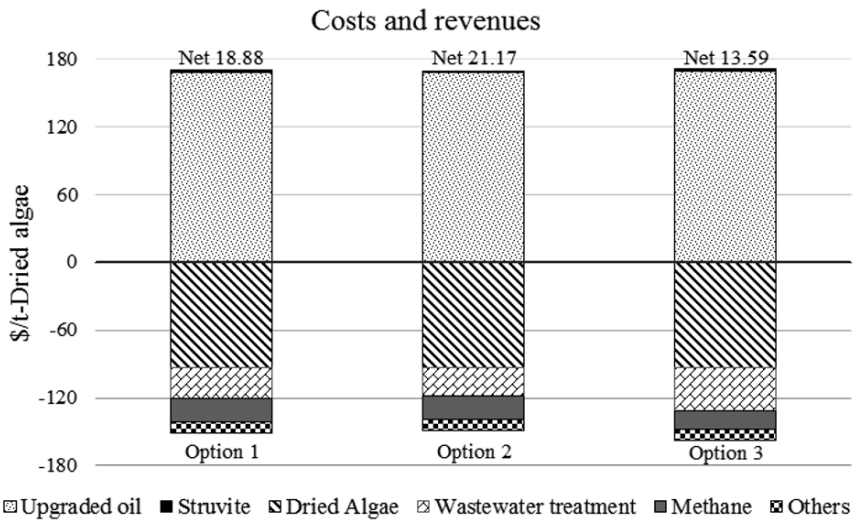


Figure 5: Operating costs and revenues for three different system scenarios.

than producing the same goods in a traditional way). Heavy metals contamination of struvite was not considered, but depending on the HM content of the feedstock a removal step could be necessary prior the precipitation (burdening on capital and operating costs and on the CED).

5. Conclusions

P-recovery could be effectively achieved from HTL waste streams treatments, diversifying the HTL product outputs. From a superstructure of process options supported by CNA and experimental data, cost effective process configurations were prioritized. Besides the phosphorous content and price of HTL feedstock, also the heavy metal content plays a key role in the economic and environmental assessment. To get more reliable results, it is important to include a capital cost analysis: for the same feedstock flow, a shorter HTL residence time would penalize the biocrude production but could substantially reduce the capital investment. Moreover, some of the most influential variables, such as the macroalgae price or the HTL yields, should be further verified. Some socio-economic factors could also affect the price setting. Main benefits include a reduced reliance on terrestrial biomass and enhanced energy security. Hindering factors are the early use of immature technology, which in combination with general negative public perception and conflicts with shipping lanes and anchorage space caused by the algae cultivation may pose significant risks for a smooth process learning period.

References

E. Desmidt, K. Ghyselbrecht, Y. Zhang, L. Pinoy, B. Van der Bruggen, W. Verstraete, K. Rabaey, B. Meesschaert, 2015, Global Phosphorus Scarcity and Full-Scale P-Recovery Techniques: A Review, Critical Reviews in Environmental Science and Technology, 45, 4, 336-384.

P.J. Valdez, V.J. Tocco, P.E. Savage, 2014 A general kinetic model for the hydrothermal liquefaction of microalgae, Bioresource Technology, 163, 123-127.

S.B. Jones, Y. Zhu, L.J. Snowden-Swan, D.B. Anderson, R.T. Hallen, A.J. Schmidt, K.A. Albrecht, D.C. Elliot, 2014, Whole Algae Hydrothermal Liquefaction: 2014 State of Technology, Pacific Northwest National Laboratory, United States of America.

P. Barthe, M. Chaugny, S. Roudier, L. Delgado Sancho, 2015, Applied Processes and Techniques, Best available techniques (BAT) reference document for the refining of mineral oil and gas industrial emissions, 88-91, Publications Office of the European Union, Luxembourg.

N. Neveux, A.K.L. Yuen, C. Jazrawi, M. Magnusson, B.S. Hayes, A.F. Masters, A. Montoya, N.A. Paul, T. Maschmeyer, R. de Nys, 2013, Biocrude yield and productivity from the hydrothermal liquefaction of marine and freshwater green macroalgae, Bioresource Technology, 155, 334-341.

S. Raikova, C.D. Le, T.A. Beacham, R.W. Jenkins, M.J. Allen, C.J. Chuck, 2017, Towards a marine biorefinery through the hydrothermal liquefaction of macroalgae native to the United Kingdom, Biomass and Bioenergy, 107, 244-253.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50288-3

A Data-Driven Approach for Design and Optimization of Energy Storage Systems

Lanyu Li, Xiaonan Wang*

Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore 117585 Singapore
chewxia@nus.edu.sg

Abstract

Energy storage systems are becoming an essential support for modern living. Since it is important but challenging to select the best energy storage technologies for the applications, this work aims to develop a data-driven optimization approach for energy storage selection as a decision-support tool for the decision makers. Technical suitability, economic feasibility and environmental impact were considered as objective criteria for selection and an augmented ε-constraint was used to find the efficient solutions for the multi-objective optimization problem of energy storage selection. A case study on arbitrage and voltage support applications was conducted. The optimal technologies for the applications was generally consistent with the common choices in literature. This model can be applied to the more complicated energy storage selection processes with more technologies, objectives and constrains taken into account.

Keywords: energy storage, technology selection, optimization, economic, environment

1. Introduction

Energy storage systems (ESS) play a critical role in the transformation from the current fossil fuel dependent energy systems to a low carbon-emitting, environmental-friendly and sustainable energy landscape, since they not only provide essential support for promoting the development of multiple evolving technologies in the energy sector such as renewable energy utilization, distributed micro-grid, and electric vehicles, but also offer a lot of benefits when integrated into the grid (Wang et al. 2015; Zhao 2016; Wang et al., 2017). Owning to the substantial research effort, a lot of energy storage technologies have been developed and their applications can be found across the power grid falling into the categories of customer energy management, transmission and distribution (T&D) support, and bulk power management (Huff et al., 2013). As it is shown in Figure 1, each of these technologies has its unique characteristics that make it only compatible with certain applications. As more and more energy storage technologies and applications are developed, it is necessary to develop a convenient tool for energy storage selection.

Although the development, characteristics and applications of energy storage technologies are widely discussed in review papers and technology roadmaps, only few publications have discussed the quantitative methods for energy storage selection. ES-select is the most comprehensive energy storage selection tool which is able to rank the energy storage technologies by their feasibility scores calculated by a weighted sum of technology feasibility, installation cost, maturity, and location feasibility (KEMA, 2012). Similarly, Raza et. al (2014) developed the sustainable index approach for energy storage

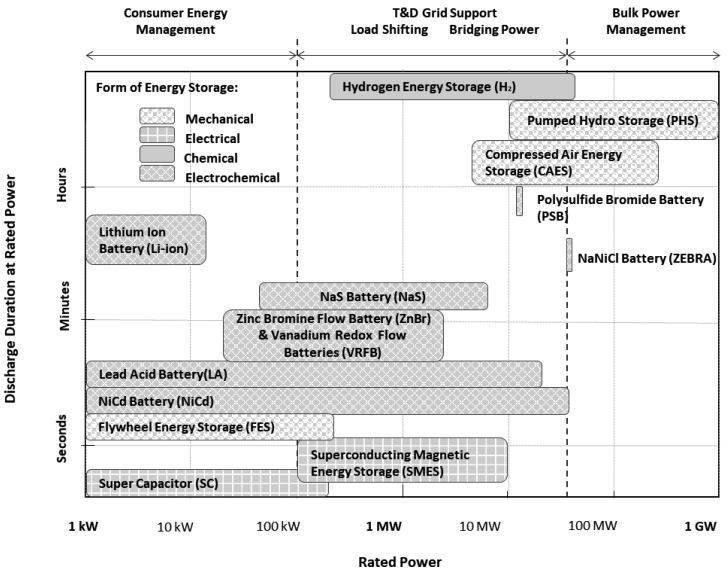


Figure 1 Characteristics of different energy storage technologies (adapted from Lawson, 2017)

technology selection for renewable energy application by ranking the technologies after assigning index to the evaluated factors and calculating a total score using the weighted sum of the normalized factors. Walker et al. (2016) used an AHP approach to study the feasibility of using different types of emerging chemical energy storage technologies for several key grid scale EES applications. However, the methods mentioned above belong to a priori methods where decision makers need to assign weighting factors and scores to evaluate the suitability of the technology for the applications, which is subjective and require a great deal of experience and expertise of the decision makers.

In this study, a data-driven multi-objective optimization approach, which is a posteriori method where efficient solutions of the problem are generated for decision makers, is proposed for the selection of energy storage technology. It is more convenient for decision-makers, especially those who is lack of expertise in this field, to select the technologies when the all possible optimal alternatives are generated by the optimization approach, compared with the previous a priori methods. In the following sections, the method and model for energy storage selection will be presented in Section 2. A case study and the results will be presented and discussed in Section 3, and conclusions and future work provided in Section 4.

2. Method

Three major criteria including technical suitability, economic feasibility, and environmental impact are considered for the selection of the energy storage technology. In this section, the detailed description of each of the criteria is given.

2.1. Technical Suitability

Generally, the two main characteristics that establish the potential set of suitable applications for a specific storage technology are rated power (P) and rated energy (E) (European Commission, 2017). In the current work, power and discharge duration are

selected to be the main factors for the technical suitability assessment because it is more common to use discharge duration (t_D) to represent rated energy as the relationship between them is:

$$E = P \cdot t_D \tag{1}$$

The technical feasibility (*TF*) for the matching between energy storage technology and application is evaluated by an overlapping coefficient (*OVL*). When the distribution density functions of the characteristics for both the technology (f_1) and the application (f_2) are known, the probability that an energy storage technology can be used for the application is the area under the two density functions, that is, the overlapping coefficient. It can be calculated by (Larson, 2014):

$$OVL = \begin{cases} \int_{R_n} \min[f_1(x), f_2(x)]dx, & \textit{Continuous distribution} \\ \sum_x \min[f_1(x), f_2(x)], & \textit{Descrete distribution} \end{cases} \tag{2}$$

Then the calculated *OVL* for the characteristic of the technology can be plug into Eq.(3) to calculate the total technical feasibility. w_P and w_D are weights for the factors and $w_P = w_D = 0.5$ since it was assumed that the power and discharge time which decides the total energy per charge cycle are equally important technical factors.

$$TF = w_P \times OVL_P + w_D \times OVL_{t_D} \tag{3}$$

2.2. Economic Feasibility

By assuming that different technologies bring the same revenue for a specific application, the economic performance can then be simplified to the cost of the energy storage system. A typical way of cost analysis for energy storage systems is using levelized cost of storage (*LCOS*). Recently, Zakeri and Syri (2015) conducted a thorough review and analysis on the levelized cost of energy storage system. Their method of cost analysis is commonly recognized and used in this model. The equations are shown below:

$$CRF = \frac{i(i+i)^T}{(i+i)^T - 1} \tag{4}$$

$$C_{OM,a} = C_{FOM,a} + C_{VOM} \cdot n \cdot t_D \tag{5}$$

$$C_{R,a} = CRF \cdot \sum_{k=1}^{r} (1+i)^{-kt} \cdot \frac{C_R \cdot t_D}{\eta} \tag{6}$$

$$C_{DR,a} = C_{DR} \cdot \frac{i}{(1+i)^T - 1} \tag{7}$$

$$C_{CAPEX,a} = CRF \cdot C_{CAPEX} = CRF \cdot (C_{PCS} + C_{ESS} \cdot t_D + C_{BoP}) \tag{8}$$

$$LCOS = \frac{C_{LCC,a}}{n \cdot t_D} = \frac{C_{CAPEX,a} + C_{OM,a} + C_{R,a} + C_{DR,a}}{n \cdot t_D} \tag{9}$$

where *CRF* is the recovery factor, *i* is the interest rate during the lifetime *T*, n is cycles per year, and η is cycle efficiency. C_{OM}, C_R, C_{DR}, C_{CAPEX}, C_{Lcc} are operation and maintenance cost (containing fixed and variable O&M C_{FOM} and C_{VOM}), replacement cost, discharge and recycle cost, capital cost, and life cycle cost, respectively. Subscript *a* denotes annual values.

2.3. Environmental Impact

The environmental impact of the system is not only the impact that is induced during the operation of the system, but also the amount resulted from its manufacturing, disposal and decommissioning processes. A lifecycle impact assessment (LCIA) involving the major environmental impact of the system throughout its lifecycle is a systematic and common way to evaluate the environmental impact of a system. In this study, ReCiPe point as a commonly used endpoint LCIA evaluation method, is adopted as indicator for the environmental performance of the technologies. The data and method is based on the energy storage system LCA study conducted by Oliveira et al. (2015), where the environmental impacts on climate change, human toxicity, particulate matter formation and fossil depletion were assessed and aggregated into a single ReCiPe endpoint in hierarchist version using European normalization and average weighting set.

2.4. Model Formulation

This work aims to identify one optimal energy storage technology that has high technical suitability, low cost and low negative environmental impact for a given application. To achieve this goal, a multi-objective optimization method is applied and the formulation of the problem is shown in Eq.(10), where z_i is a binary variable with 1 meaning selecting the technology i and 0 otherwise.

$$\begin{cases} \max\ T(z_i) \\ \min\ C(z_i) \\ \min\ E(z_i)) \end{cases} \quad \text{where} \quad \begin{cases} T(z_i) = \sum_{i=1}^{p} z_i \cdot TF_i \\ C(z_i) = \sum_{i=1}^{p} z_i \cdot LCOS_i \quad s.t.\ \sum_{i=1}^{p} z_i = 1 \\ E(z_i) = \sum_{i=1}^{p} z_i \cdot LCIA_i \end{cases} \tag{10}$$

The model is implemented in GAMS as a Multi-objective Mixed Integer Programming (MOMIP) problem using CPLEX solver. An augmented ε-constraint method proposed by Mavrotas (2007) is used to produce the efficient solutions to the problem.

2.5. Data Collection and Model Implementation

The considerable projects for different energy storage technologies generate a large amount of data describing the technical, economic and environmental features of the energy storage technologies and applications. In order to carry out an effective selection of the technologies for an application, descriptive statistics can be used to capture the main features of the technology and the applications from the abundant data. In this study, it is assumed that the distribution of the technologies and the applications for different characteristics follows normal distribution and the average value of the characteristic for each technology or application is used to represent the population. The data for economic and technical features of the energy storage technologies are mainly from Zakeri and Syri (2015) and the environmental impact data of the technologies are from Oliveira et al. (2015) and other data sources.

3. Case Study and Discussion

3.1. Technologies and Applications for the Case Study

Two bulk energy management applications, arbitrage and grid voltage stability, were chosen to be the application scenarios for this study. As is shown in Table 1, these two

applications have distinct technical features as arbitrage requires long discharge duration, while voltage stability application requires fast response and short discharge duration.

Table 1. Characteristics of the energy storage applications

	P (MW)	t_D	Cycles (per year)	Lifetime (y)
Arbitrage	10-300	2-10 h	350	25
Voltage stability	10-500	1 s	10	25

Figure 1 shows the basic information about the large variety of the energy storage technologies and their key technical characteristics (i.e. rated power and discharge duration) that are investigated in this case study.

3.2. Results for the Case Study

Figure 2 shows the evaluation of the key criteria, which are TF_i, $LCOE_i$ and $LCIA_i$, in the optimization process for the selection of energy storage technologies. The optimal set of efficient solutions of energy storage technologies for each application was returned by the multi-objective optimization as it is shown in Table 2. The most economically feasible, most technically suitable and least environmentally influential technologies for the applications highlighted in Table 2 can be obtained by individual optimization of each objective or interpreted from Figure2. The efficient solutions other than the above discussed cases are the one that has the advantage of compromised technical, economic and environmental performance, such as SC and VRFB for arbitrage application.

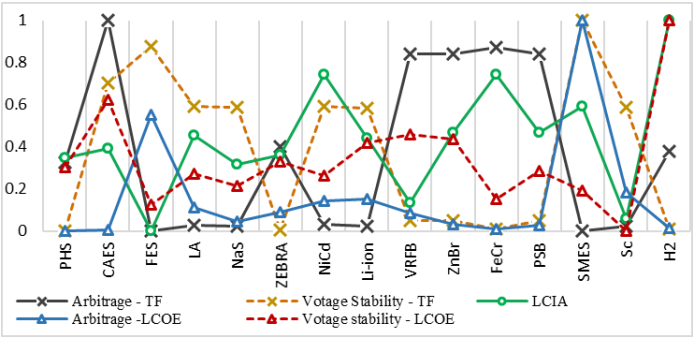


Figure 2. Trend for TF_i, $LCOE_i$ and $LCIA_i$ with varying technologies for arbitrage and voltage stability applications, where values are normalized.

Table 2. Efficient solutions returned by the optimization of energy storage technologies selection model for the arbitrage and voltage stability applications.

	Highest technical suitability	Lowest cost	Lowest environmental impact	Other efficient solutions	
Arbitrage	CAES	PHS	FES	SC	VRFB
Voltage stability	SMES	SC	FES	-	-

In general, CAES and PHS are energy storage technologies with large global deployment capacities and they are suitable for large-scale energy storage such as arbitrage, flow batteries are also good candidates for applications with discharge duration for more than 5 hours, and FES, SMES and SC are typical energy storage technologies with fast response and they are suitable technologies for short-duration energy storage. Therefore,

the results for voltage stability application are reasonable. But the FES and SC appear to be not technically suitable for the arbitrage application. They are included in the efficient solution set mainly because they are at the low end of environmental impact. In short, the result of the optimization is reliable and consistent with the recommended selections in the literature.

4. Conclusions and Future Work

This work developed an optimization approach to identify the optimal energy storage technology for a specific application. Several objectives including minimizing cost, maximizing technical suitability and minimizing environmental impact of the selected technology were considered in the selection process. An augmented ε-constraint method was used to find the pareto-optimum for the model. The result from the optimization for a case study on utility-scale applications was reasonable and this model can be applied to the larger system with more technologies and constraints. Although technical feasibility is modelled as an objective to be maximized in this conference paper, some technical requirements, such as the suitable range of the rated power and discharge duration for the applications, usually appear to be prerequisites for the selection. Therefore, as future work, several key technical requirements will be formulated as constraints in a journal article extended from on this paper.

References

European Commission, 2017. COMMISSION STAFF WORKING DOCUMENT Energy storage – the role of electricity EN.

Grid Energy Storage Systems [WWW Document], 2017. Electropaedia.

Huff, G., Currier, A.B., Kaun, B.C., Rastler, D.M., Chen, S.B., Bradshaw, D.T., Gauntlett, W.D., 2013. DOE/EPRI 2013 electricity storage handbook in collaboration with NRECA. Rep. SAND2013- … 340. https://doi.org/SAND2013-5131

KEMA, 2012. ES-Select™ Documentation and User's Manual.

Larson, B., 2014. Meet the Overlapping Coefficient.

Mavrotas, G., n.d. Generation of efficient solutions in Multiobjective Mathematical Programming problems using GAMS. Effective implementation of the ε-constraint method.

Oliveira, L., Messagie, M., Mertens, J., Laget, H., Coosemans, T., Van Mierlo, J., 2015. Environmental performance of electricity storage systems for grid applications, a life cycle approach. Energy Convers. Manag. https://doi.org/10.1016/j.enconman.2015.05.063

Raza, S.S., Janajreh, I., Ghenai, C., 2014. Sustainability index approach as a selection criteria for energy storage system of an intermittent renewable energy source. Appl. Energy. https://doi.org/10.1016/j.apenergy.2014.04.080

Walker, S.B., Mukherjee, U., Fowler, M., Elkamel, A., 2016. Benchmarking and selection of Power-to-Gas utilizing electrolytic hydrogen as an energy storage alternative. Int. J. Hydrogen Energy 41, 7717–7731. https://doi.org/10.1016/j.ijhydene.2015.09.008

Wang, X., El-Farra, N.H., Palazoglu, A., 2017. Optimal scheduling of demand responsive industrial production with hybrid renewable energy systems. Renew. Energy 100, 53–64. https://doi.org/10.1016/j.renene.2016.05.051

Wang, X., Palazoglu, A., El-Farra, N.H., 2015. Operational optimization and demand response of hybrid renewable energy systems. Appl. Energy 143, 324–335. https://doi.org/10.1016/j.apenergy.2015.01.004

Zakeri, B., Syri, S., 2015. Electrical energy storage systems: A comparative life cycle cost analysis. Renew. Sustain. Energy Rev. https://doi.org/10.1016/j.rser.2014.10.011

Zhao, Z., 2016. The Review of Energy Storage Technologies Selection. J. Electr. Electron. Syst. 5, 1–4. https://doi.org/10.4172/2332-0796.1000172

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50289-5

Comparative study on LCA co-product allocation method of bioethanol production in the development from sweet sorghum stem

Haoyu Liu, Tong Qiu*

Department of Chemical Engineering, Tsinghua University, 100084 Beijing, China
qiutong@tsinghua.edu.cn

Abstract

The biofuels technology in development stage have totally different separation and recycling of co-product compared with industry products, for which the greenhouse gases (GHGs) and energy consumption (EC) associated with different co-product allocation method of life cycle analysis (LCA) could apparently change. This study explores the impact of different co-product allocation method of fuel ethanol from sweet sorghum stalks using advanced solid state fermentation (ASSF) in china. The scenario analysis is utilized to predict the industrial application and recycling of co-product. The result reveal that apparent differences exist in different allocation method. Hybrid allocation is recommend to deal with different co-products. The performance of ASSF will be further improved by recycling the co-product and separation technology.

Keywords: Life-cycle analysis, Co-product allocation, Biofuels, Scenario analysis

1. Introduction

China is currently the world's largest oil consumer and GHGs emitter (EIA. 2014). Bio–ethanol derived from energy crop has been promoted by the Chinese government because of its capacity to reduce fossil consumption and GHGs emissions(Hao, Dunn et al. 2013). Ethanol from corn started being produced in 2002 and china has become the third largest producer and consumer of bio-ethanol, after the United States and Brazil(Ding, Yang et al. 2017). However, the concern about food safety promote the government to impose a ban on expanding maize and other grain-based bio-fuel production(Hong, Yuan et al. 2009). The study of second-generation biofuels has become an international trend to address these problems, while how to calculate the emission and reuse the non-energy co-product in the production process is the biggest controversy of non-food feedstock(Giarola, Shah et al. 2012).

Sweet sorghum stalks (SSS) have recently received attention as one of the most promising second-generation crop for bioethanol production in China because of its high photosynthetic efficiency, high biological yield, strong resistance, wide adaptability rich, as well as clean and relatively low production cost(Wang, Chen et al. 2014). The advanced solid state fermentation (ASSF) technology for fuel ethanol production from SSS was developed by the Chinese scientists. The pilot plant of this new technology is successfully applied in Inner Mongolia, located in northern China, where is estimated that there may be as many as 4.4 million hectares of land suitable for SSS planting, occupying 13% of this land(Han, Wang et al. 2010).

However, there are two controversy problems about ASSF: whether it is environment friendly and whether it is energy saving. Life cycle analysis is a tool to assess these problems. Tian, Y. S. established energy consumption and GHGs and other impact facts about ASSF, and the results showed SSS-based bioethanol making feasible from the view point of energy production(Tian, Li et al. 2011). Ding, N. reported that ASSF has advantages in terms of energy consumption but lack superiority in the GHGs(Ding, Yang et al. 2017). The different results are lead by the different co-product allocaiton of LCA they used. Because of the two kinds of co-product (the sweet sorghum grain and the dried distillers) existed in the production process, lack of systematic and suitable allocation method will strongly influence the GHGs and energy consumption results.

In recent years, a wide variety of co-product allocation method of biofuels has been proposed. Although the ISO14040 advocates the displacement method, and the mass, energy, market method are also adopted for different situations in studies(Wang, Hong et al. 2011). Jong compared the well-to-wake GHG emission performance of multiple renewable jet fuel conversion pathways and explored the impact of different co-product allocation method, and recommend using energy and economic allocation (for non-energy co-products)(Jong, Antonissen et al. 2017).Compared to the industry products, they discussed, the new biofuels technology, like ASSF, is deficient in separation and recycling of co-product, and the uncertainty of given process parameters is obvious. Error will be apparent in the result of LCA without solving these problems. The importance of this problem was also discussed by experts(Report, 2014).

Thus, the differences of ASSF laboratory data using multiple allocation methods is compare in this article. Four allocation methods are combined to analyze the uncertainty of the two kinds of non-energy co-product. Scenario analysis is utilized to predict the industrial application effect and reusage of biofuels in development stage, and the process replacement method is used to simulate the product heat transfer network and recycling. This life cycle process covers data from Chinese SSS-based bioethanol used ASSF. The energy consumption and GHG emissions are examined through the overall process. The result indicate that the scheme of co-product allocation has significant impact on the conversion period. Compared to the case of industrial applications, the effectiveness of the results is enhanced improved, the formation of a multi-scene analysis results is useful for policy makers.

2. Methods

2.1. LCA framework

A cradle-to-gate LCA was performed with GaBi ts Education software. The analysis focused on determine the energy consumption and GHGs of ASSF-based ethanol produced by the pilot plant in Inner Mongolia, to compare the results with different co-products allocation method and to find an eco-friendly heat transfer network and recycling.

2.2. System boundary and functional unit

The stages considered and the calculation logic of the energy consumption and the GHGs of the entire process is present in the Figure 1, including the plant cultivation, production and transport process. Effects of fertilizer emissions and land use change will not be considered for the moment due to the data given in the IPCC Guidelines have large deviations. The assessments were compared on the basis of their GHG emissions in g

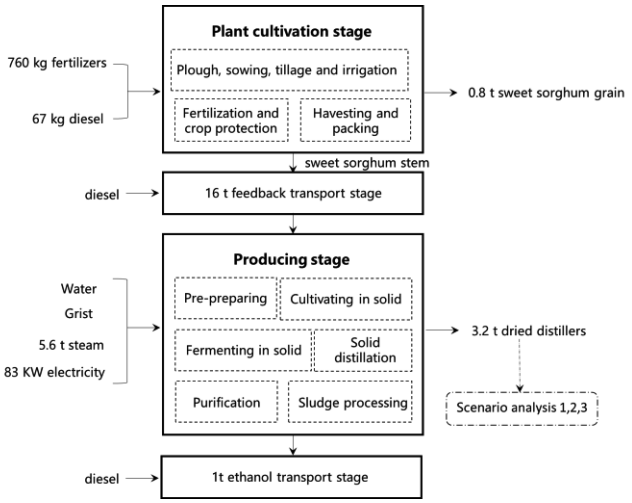


Figure 1. Product system boundary of this study.

CO2 per MJ_LHV_bioethanol and energy consumption MJ per MJ_LHV_bioethanol means of bioethanol produced from sweet sorghum stem in higher heating value.

2.3. Product system description

The ASSF technology in the lab mainly include pre-preparing, cultivating in solid, solid-state fermentation, solid distillation, purification, sludge processing. While there isn't sludge processing to deal with co-product in the pilot plant. An average of 1 t of ethanol produces 3.2 t of dried distillers, which is the biggest co-product produced in the pilot process.

2.4. Method to deal with two kinds of co-products

Three different means is suitable for this system: energy, mass and economic method. While benefits and disadvantages exist for each method. The non-energy coproduct is not suitable for energy allocation. For instance, the commercial crop, sweet sorghum grain, can't be valued by burning to get the energy. The mass allocation is limited to material products and cannot be used on immaterial products such as electricity. The third method economic allocation may be affected by the price fluctuations, geographical location or market distortions, especially for the non-commoditized products.

2.5. Scenarios for reusing the co-products

For the further applications, the distillers that can be used for forage or biogas will be comprehensive utilized. We gain the data from the pilot plant and the farmers. The possible systems for recycle the co-product, dried distillers, are divided into 3 scenarios: Scenario 1: Half of the dried distiller produces steam achieving self-supplied though burning, the other half is used to get biogas, biogas residue and biogas slurry turn to be fertilizer; Scenario 2: All the dried distiller are burned to achieve steam supply, Waste water is used to product biogas; Scenario 3: All dried distiller produce biogas, where need extra coal. The biogas is used to produce steam, while biogas residue and biogas slurry is used as fertilizer.

Table 1 Parameter about ASSF

Item Parameter	Value	Item Parameter	Value	Item Parameter	Value
Planting		**Pre-preparing**		**Cultivation**	
Nitrogen fertilizers (kg/ha)	600	Sweet sorghum stem (t/ha)	16	Electricity (kw)	15
Phosphorus fertilizers (kg/ha)	160	Electricity (kw)	41	Steam in 100℃ (t)	0.1
Diesel used in farming (kg/ha)	67	Steam in 100℃ (t)	0.1	Grist	0.17
Sweet sorghum stem (t/ha)	60	**Distillation and purification**		**Transport**	
Co-product: sweet sorghum grain (t/ha)	3	Electricity (kw)	6	Distance (farming and plant) (km)	25
LHV of SSS (MJ/kg)	4.56	Steam in 140℃ (t)	5.5	Distance (plant and market) (km)	300
LHV of grain (MJ/kg)	27.27	co-product: dried distillers (t)	3.2	Ethanol heat value	29.29
Fermentation		Water	264		
Electricity (kw)	21				

2.6. Life cycle inventory

Based on the investigation with the technology developer and farmers, open LCI databases and references, the LCA parameter is showed in the Table 1.

3. Result and discussion

3.1. Energy consumption and environment emissions

The contribution of different energy inputs and different processes to the energy consumption and the environment emissions are showed in the Figure 2. The total energy efficiency is 1.27MJ/MJ. The distillation and purification unit exhibits high contribution to EC due to the high steam input. Including the emission of all the co-product, the total GHGs is 116 g CO_2 /MJ and the ratio of stages is similar to the EC. To reduce the GHGs, we can consider enhancing the separation efficiency or using co-product to supply energy. Besides, the plant size is also to a factor that influence the consumption of coal.

3.2. Comparison between allocation methods

In this case, there are two process need to be allocated: the allocation of SSS and grain during plant cultivation stage, and the allocation of bioethanol and dried distillers during production stage. To illustrate the impact of two kinds of co-product, Figure 3 shows that the emissions for the SSS-based bioethanol range between 30 and 84 g CO_2 /MJ for different allocation methods. Huge sensitivity exists in the energy allocation for the grain and dried distillers, as is not the direct fuel. In the first stage, economic allocation is considered to be the best choice because the SSS will be waste if not used to fabricate

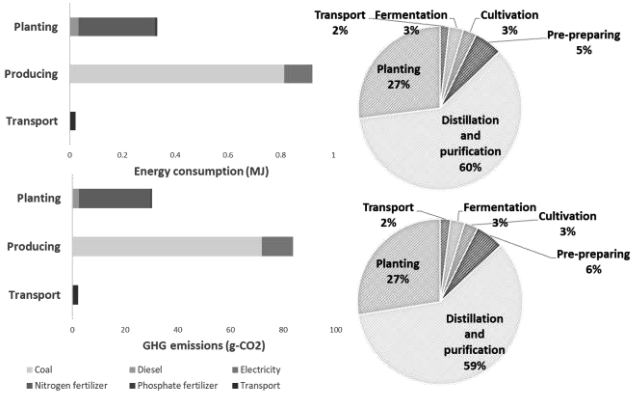


Figure 2. GHG emission and energy consumption in each production process of ASSF.

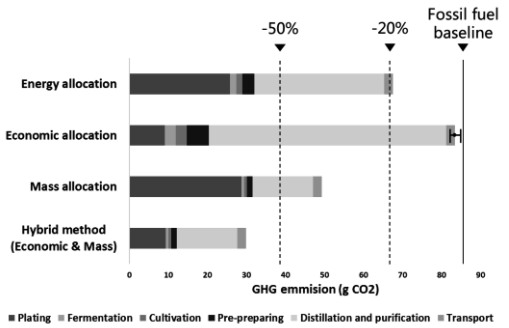


Figure 3. GHG emissions for ASSF using different allocation methods.

bioethanol. The factory can get SSS with a low price and the non-energy grain gain a big share. The mass allocation is most suitable for the ethanol production stage, in which the stem is not the main crop but a co-product and no one will buy distillers. So a hybrid allocation way, in which the first stage used economic allocation and the second stage used mass allocation, is utilized to prove the importance of allocation. To conclude, choose a suitable allocation way is important in LCA study.

3.3. Scenario analysis for the recycling of the co-product

Figure 4 shows that energy consumption and GHGs are greatly improved in each scenario, in which recycling the distillers will get the unexploited energy due to the low conversion ratio of the stem. The high energy efficiency in the scenario 2 by directly burning lead the negative emissions. Scenario 1 and 3 can be enhanced by using high-efficient biogas conversion method. From the scenario analysis we can draw the conclusion that comprehensive utilization about the co-product especially like dried distillers will be apparently positive to the environment impact.

4. Conclusion

The life cycle analysis about the ASSF in the Inner Mongolia is conducted in this study. Different allocation method (energy, economic, mass allocation) is utilized to illustrate the importance of how to calculate the co-product (especially the non-energy co-product). The effect and uncertainty of the biofuels in the development stage is compared by the scenario analysis, using process replacement method. Compared the single allocation and hybrid allocation, the biggest differences are found for non-energy product in the two process of production, which is more suitable for the mass allocation and economic allocation. With the idealized hybrid allocation, the result of GHGs in the ASSF is 30 g CO_2 /MJ. Scenario analysis showed the systematic and comprehensive utilization of

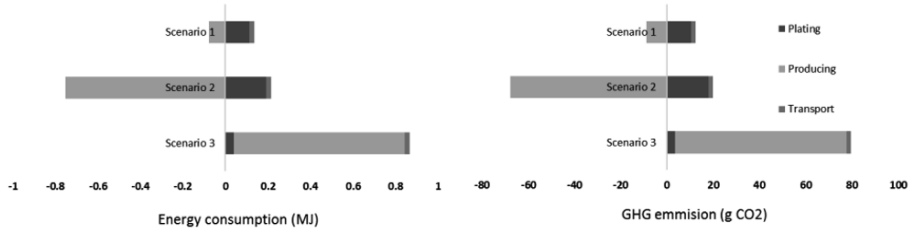


Figure 4. Three scenario analysis for reusing the co-product.

co-product will improve the environment impact. Reusing vinasse as the fuel and better separation of the product could significantly enhance the energy efficiency. Further study can focus on the scale of the heat transfer network in the pilot plant to recycle the co-product, the formation of a multi-scene analysis results is useful for producer and policy makers.

References

Ding, N., Y. Yang, H. Cai, J. Liu, L. Ren, J. Yang and G. H. Xie (2017). "Life cycle assessment of fuel ethanol produced from soluble sugar in sweet sorghum stalks in North China." Journal of Cleaner Production 161.

Giarola, S., N. Shah and F. Bezzo (2012). "A comprehensive approach to the design of ethanol supply chains including carbon trading effects." Bioresource Technology 107(2): 175-185.

Han, B., L. Wang, S. Li, E. Wang, L. Zhang and T. Li (2010). "[Ethanol production from sweet sorghum stalks by advanced solid state fermentation (ASSF) technology]." Chinese Journal of Biotechnology 26(7): 966.

Hao, C., J. B. Dunn, Z. Wang, J. Han and M. Q. Wang (2013). "Life-cycle energy use and greenhouse gas emissions of production of bioethanol from sorghum in the United States." Biotechnology for Biofuels 6(1): 141.

Hong, Y., Z. Yuan and J. Liu (2009). "Land and water requirements of biofuel and implications for food supply and the environment in China." Energy Policy 37(5): 1876-1885.

Jong, S. D., K. Antonissen, R. Hoefnagels, L. Lonza, M. Wang, A. Faaij and M. Junginger (2017). "Life-cycle analysis of greenhouse gas emissions from renewable jet fuel production." Biotechnology for Biofuels 10(1): 64.

Tian, Y. S., S. Z. Li, L. X. Zhao, H. B. Meng and L. L. Huo (2011). "Life cycle assessment on fuel ethanol producing from sweet sorghum stalks." Transactions of the Chinese Society for Agricultural Machinery 42(6): 132-137.

Wang, M., Y. Chen, X. Xia, J. Li and J. Liu (2014). "Energy efficiency and environmental performance of bioethanol production from sweet sorghum stem based on life cycle analysis." Bioresource Technology 163(163): 74-81.

Wang, M., H. Hong, S. Arora, R. Diazchavez and J. Woods (2011). "Methods of dealing with co-products of biofuels in life-cycle analysis and consequent results within the U.S. context." Energy Policy 39(10): 5726-5736.

EIA, The U.S. Energy Information Administration, 2014. China is Now the World's Largest Net Importer of Petroleum and Other Liquid Fuels.

http://www.eia.gov/todayinenergy/detail.cfm?id¼15531.

Consulting report of China Engineering Academy, 2014, Development, Demonstration and Application of Key Technology of Biofuels Industry in China.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50290-1

Network Visualization of Design Variables and Functions for Sustainable Packaging Design

Naoki Yokokawa[a*], Yutaro Masuda[a], Eri Amasawa[a], Hirokazu Sugiyama[a], Masahiko Hirao[a]

[a]*Department of Chemical System Engineering, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan*
n-yokokawa@pse.t.u-tokyo.ac.jp

Abstract

We visualized the network between design variables and product functions in a supply chain of product for sustainable packaging design. The multi-functional product design affects multiple processes in the supply chain including consumption process, while there are trade-offs between product functions because enhancing one functionality for sustainable supply chain could decrease another functionality. Visualizing the relationships between product functions and design variables as network structure supports to grasp the trade-offs in the design stage and to perform sustainable process and product design. Several case studies were performed to verify the proposed network structure, and design condition of the case studies was to decrease the content losses in consumption process. As a result, the proposed network structure reflected the difference of the product function and identified the loss reduction to decrease total environmental impacts by scenario analysis of consumption process.

Keywords: process and product design; food packaging; life cycle assessment; sustainable consumption and production; scenario analysis.

1. Introduction

Packaging has a lot of functions to provide benefits for stakeholders in a supply chain of a product, such as productivity, preservation of its content and usability. There are various requirements for product functions in the supply chain, and a packaging designer takes the requirements into account. To ensure different functions of the packaging, the packaging designer performs process and product design concurrently, and expressing product functions by packaging design affects multiple processes beyond production line. Linking packaging design with logistics is significant to develop a sustainable supply chain (García-Arca et al., 2017), which indicates that boundary expansion supports sustainable packaging design. In addition, consumption process needs to be included into the design boundary because there are high impacts from consumption process. For example, in a case of breakfast cereals, milk as an input of consumption process has a large contribution to CO_2 emission in a supply chain as well as the ingredients of the product (Juswani et al. 2015).

Packaging functions can prevent one of the crucial topics for global sustainability: food losses. Food losses are highly dependent on consumption process and cause environmental impacts (Wikström et al., 2014). We also presented the relationship between consumer behaviours and environmental impacts by scenario analysis (Yokokawa et al., 2018), and the effects of some of the packaging functions were

indicated. However, there remains a need to develop a support tool for the designer to obtain the effects of their decision-makings, i.e., decisions of product and process design variables for realizing product functions, on the supply chain.

The objective of this study is to visualize the network structure among process and product design variables and product functions in the supply chain of product for sustainable packaging design. The relations among variables and product functions were modelled, then they were structured by a matrix and visualized as a network structure. The proposed network structure was applied to milk in carton as a case study to show applicability of the network structure for packaging design.

2. Network visualization of design variables and product functions

2.1. Modelling the relationships

Design variables and product functions are regarded as attributes allocated to a supply chain. These attributes in the supply chain of packaging and its contents were identified, and the relationships were modelled by Eqs. (1)–(3):

$$I_l = g_l(\boldsymbol{V}, \boldsymbol{F}, \boldsymbol{P}) \tag{1}$$

$$F_m = h_m(\boldsymbol{V}, \boldsymbol{P}) \tag{2}$$

$$P_n = k_n(\boldsymbol{F}, \boldsymbol{P}) \tag{3}$$

$$\boldsymbol{I} = (I_1, I_2, \dots), \boldsymbol{V} = (V_1, V_2, \dots), \boldsymbol{F} = (F_1, F_2, \dots), \boldsymbol{P} = (P_1, P_2, \dots)$$

where I_l, V_j, F_m, P_n are the l-th objective function, the j-th product design variable, the m-th product function, and the n-th process design variable, respectively. Objective functions can be economic indicators such as cost, and environmental indicators such as life cycle CO_2 emission. From Eqs. (2) and (3), process design variables and product functions interact with one another. These equations prescribe the rules of visualization of relationships among attributes, and they represent a superstructure of the relationships among attributes and objective functions. The modelling reflects the multiple effects of the variable changes on the supply chain into the assessment of objective functions.

2.2. Structuring the relationships as a matrix

The relationships between one attribute to another modelled as Eqs. (1)–(3) were structured as a square matrix in which all attributes, i.e. $\boldsymbol{V}$, $\boldsymbol{P}$ and $\boldsymbol{F}$, are aligned in the row and column.

Figure 1 shows an excerpt of the matrix and the procedure to use the matrix for packaging design. In four steps shown in Figure 1, the spread of multiple effects on the supply chain by design change to fulfil one function is illustrated. As shown in the Step I, after deciding a product function to be enhanced (column F_5) to satisfy a request from the supply chain, the packaging designer seeks design variables for functionality enhancement. The shaded cells indicate the existence of relationships between the attributes. In this case, the rows V_1 and P_2 have relations with the column F_5. Then, as shown in the Step II, the designer can identify the effects of design change on the other functions. In this case, a design change of V_1 affects on F_9 as well as F_5. As shown in the Step III, the designer can identify

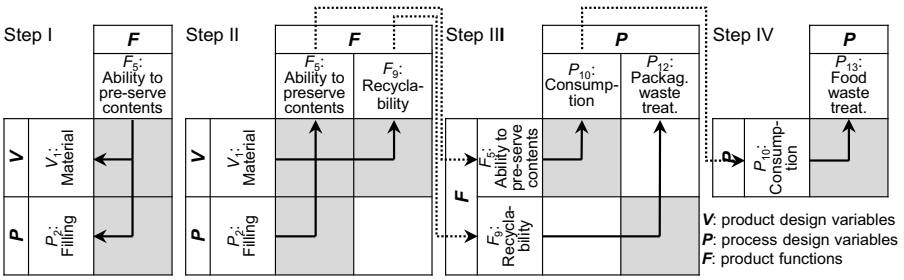


Figure 1 Excerpt matrix and one case of the multiple effects of changing product variables on other attributes.

the effects of functionality changes on process design variables. The changes of functionality, F_5 and F_9, affect on the consumption and end of life treatment P_{10} and P_{12}. Finally, as shown in the Step IV, the designer can identify the effects of the identified process design variables on other process design variables. The change of consumption P_{10} affects the end of life treatment P_{13} in the same manner as the Steps II and III.

2.3. Visualizing the relationships as a network

Network structure was developed to visualize all relationships between attributes and the places where the attributes are expressed. Figure 2 shows a part of the network structure based on the matrix in the case of Figure 1. The cross direction from left to right of Figure 2 indicates the stream of the supply chain, and the upper, lower and middle rows of Figure 2 indicate a packaging part, its content part and product part in the supply chain respectively. The relationships prescribed by Eqs. (1)–(3) are represented as nodes and arrows. Nodes surrounded by an oval, a square and two concentric ovals represent attributes ***V***, ***P*** and ***F*** respectively, and an arrow indicates the relationship, h_m and k_n. That is why an attribute at the start point of an arrow affects an attribute at the end point of the arrow. There are some double-headed arrows between process design variables and product functions, e.g. the arrow between "Filling" and "Ability to preserve contents" in Figure 2, because the decision order of these attributes differs among the cases. In Figure 2, the effects of decision-making, i.e. material selection, is shown as the arrows, while the requests are indicated by following the arrows conversely. The latter indicates the Step I shown in Figure1, and the former indicates the other Steps.

The relationships between attributes visualizes the experts' tacit knowledge, and the network structure contributes to concurrent design of process and product considering the

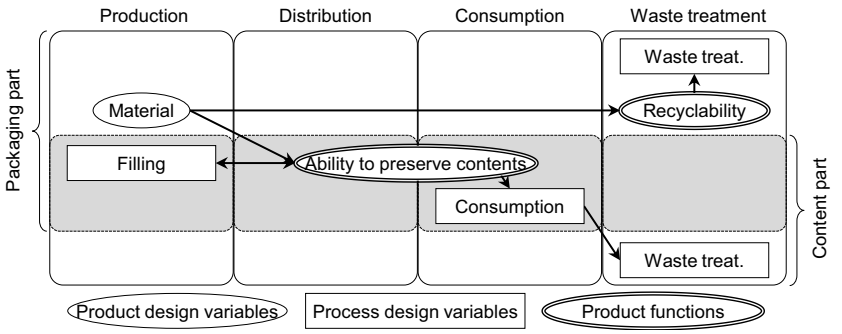


Figure 2 A part of network structure to visualize the relationships between the attributes.

trade-offs between product functions in the supply chain. For example, to enhance the ability to preserve contents F_5, packaging designers decide on packaging materials V_1 changing from packaging materials with high permeability to those with low permeability. It is certainly a decision to enhance the ability to preserve contents, but it decreases packaging recyclability F_9 as a consequent result of additional materials to packaging waste treatment process P_{12}. Environmental assessment based on the network structure can show the life cycle CO_2 emission of the product I_2 considering the trade-offs between product functions. The results support stakeholders to make a sustainable decision for sustainable process and product design.

3. Case study

We performed several case studies in different types of product design by using the network structure. In this paper, we present the case of milk products in carton to demonstrate the applicability of the proposed network visualization. To integrate various consumer behaviours, the design condition was set as food loss prevention in consumption process. Figures 1 and 2 represent the relationships focused on in the case of milk products. Life cycle assessment approach was adopted for evaluating environmental impacts, and life cycle CO_2 emission was set as an objective function in this case study because CO_2 emission is one of the reliable environmental indicators. The effects of the product functions were analysed by reflecting product characteristics in life cycle environmental analysis.

Table 1 shows product characteristics of milk products in this study. Refrigerated milk is a conventional product employing the carton packaging which consists of paper and polyethene. This type of the products is distributed with refrigeration through the supply chain till consumption, and the expiration date is one or two weeks. To prevent food losses, non-refrigerated milk has an extended expiration date by adding aluminium into the carton and increasing the sterilization temperature at the filling process. This product design is shown as a variable identification for functionality enhancement in the Step I. Figures 1 and 2 shows that packaging materials and filling process affect the ability to preserve food and the packaging recyclability at the same time. Finally, this decision on these changes affects the downstream of the supply chain including consumption process, and the non-refrigerated milk can be stored at normal temperature before opening.

The criteria for assessment of milk products was set as one litre of milk consumption by reflecting food characteristics. Food losses are derived from short/little consumption of food, and consumers do not take the same volume of nutrition as a result of the disposal of edible parts. When there are food losses, consumers purchase additional products to compensate for the losses. The model integrating food losses is represented by Eq. (4):

Table 1 Product characteristics of milk products in the case study.

	Refrigerated milk	Non-refrigerated milk
Packaging	Carton made by paper and polyethylene	Carton laminated with aluminum
Max temperature at sterilization [°C]	130	150
Storage condition before open	Refrigeration	Normal temperature
Expiration date [day]	7	60

$$E(x)^{product} = \left(E^{product} + xE^{\text{loss}}\right)/(1-x) \quad (4)$$

where x [–], $E(x)^{product}$ [kg CO_2eq/L-milk consumption], $E^{product}$ [kg CO_2eq/L-milk production], and E^{loss} [kg CO_2eq/L-milk] are the loss rate of food products which is defined as the rate of the losses to the provided food, the life cycle CO_2 emission of a product per consumption when the value of the loss rate is x [–], the CO_2 emission of a product per production, and the CO_2 emission of loss treatment process, respectively. This equation is unbounded when all contents are discarded (x=1) because unlimited purchase of products is needed for a certain amount of consumption.

Figure 3 shows the results of the environmental assessment comparing refrigerated milk containing one litre of milk (RF) and non-refrigerated milk containing one litre of milk (NR) based on Eq. (4). In the case of no food loss, the functionality enhancement increases the total CO_2 emission from the supply chain as a result of the additional use of aluminium for packaging. Meanwhile, because the CO_2 emission from the production of milk has a large contribution, food losses increase the total CO_2 emission. Therefore, a balance between the increased CO_2 emission by functionality enhancement and the effects of functions, i.e. food loss reduction, needs to be quantitatively evaluated and compared for sustainable packaging design. In other words, a break-even point on the increased CO_2 emission needs to be identified for decreasing the total CO_2 emission by food loss reduction. A previous study (Yokokawa et al., 2018) introduced an indicator, break-even rate of food loss, φ [–], to represent the break-even point. The break-even rate of food loss is given by Eq. (5):

$$\varphi = \left(E(0)^{\text{NR}} - E(0)^{\text{RF}}\right)/\left(E(0)^{\text{NR}} + E^{\text{loss}}\right) \quad (5)$$

The value of the break-even rate of food loss was 4.6 % in this study comparing NR with RF. The result shows NR has less CO_2 emission than RF has when enhancing the ability to preserve food reduces more than 4.6 % of the loss rate.

Food losses are dependent on consumer behaviour, and the value of the loss rate differs among consumer behaviours. Scenarios reflected various consumer behaviours to identify the food loss rate in the scenarios by setting the design variables of consumption process like consumption rate. For example, when the food losses were dependent on consumption rate r_{M} [L/day] and expiration date d_{expire} [day], the food loss rate was represented by Eq. (6)

$$x = \begin{cases} 1 - r_{\text{M}} d_{\text{expire}} / M_{\text{edible}} & \left(r_{\text{M}} d_{\text{expire}} \le M_{\text{edible}}\right) \\ 0 & \left(r_{\text{M}} d_{\text{expire}} \ge M_{\text{edible}}\right) \end{cases} \quad (6)$$

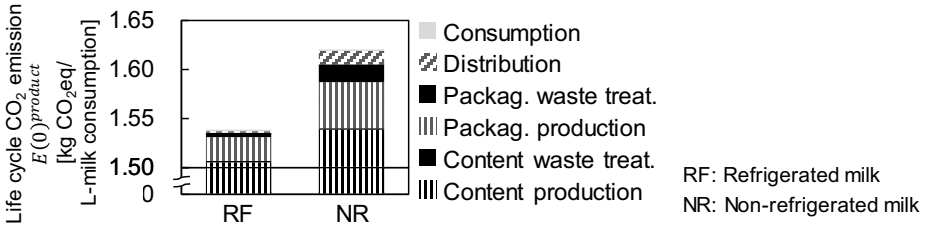


Figure 3 Environmental assessment results of refrigerated milk and non-refrigerated milk.

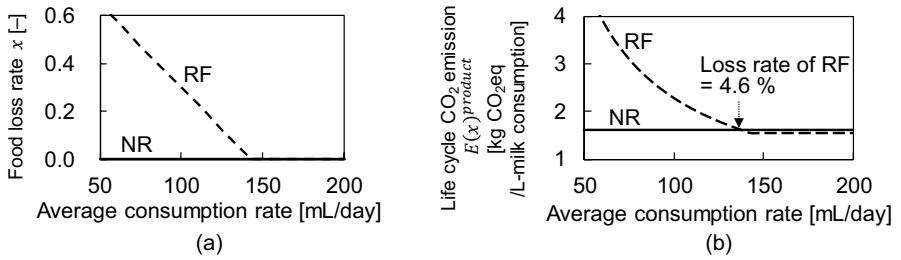


Figure 4 Scenario analysis results of consumer behavior comparing refrigerated milk (RF) and non-refrigerated milk (NR) in two indices (a) food loss rate (b) CO_2 emission

where M_{edible} [L] is the total edible parts of provided food. Figure 4 shows the scenario analysis results of milk products based on Eqs. (5) and (6). Figure 4(b) presents that the break-even rate of food loss is a useful indicator to identify the scenarios where NR is better for the environment than RF is. In this case, NR is effective for consumers who consume less than 136 mL/day of milk in decreasing the life cycle CO_2 emission.

4. Conclusions

We visualized the relationships between design variables and product functions in the supply chain of packaging and its contents for sustainable packaging design. By a case study, the applicability of the developed network structure of the relationships was presented. Including the losses of contents from consumption process into the evaluation shows the significance of integrating the effects of the product functions in the supply chain. This study can support packaging designers to make sustainable decisions by grasping the effects of their decisions on the supply chain.

5. Acknowledgement

We would like to thank the Packaging Working Group of Life Cycle Assessment Society of Japan for fruitful discussion. This study is supported by JSPS KAKENHI Grant Number JP17H01918 and the Environmental Research and Technology Development Fund (S-16) of the Environmental Restoration and Conservation Agency.

References

J. García-Arca, A. T. G. P. Garrido, J. C. Prado-Prado, 2017, "Sustainable packaging logistics". The link between sustainability and competitiveness in supply chains, Sustainability, 9, 7, 1098, DOI: 10.3390/su9071098

H. K. Jeswani, R. Burkinshaw, A. Azapagic, 2015, Environmental sustianability issues in the food-energy-water nexus: Breakfast cereals and snacks, Sustainable Production and Consumption, 2, 17–28

F. Wikström, H. Williams, K. Verghese, S. Clune, 2014, The influence of packaging attributes on consumer behaviour in food-packaging life cycle assessment studies - a neglected topic, Journal of Cleaner Production, 73, 100–108

N. Yokokawa, E. Kikuchi-Uehara, H. Sugiyama, M. Hirao, 2018, Framework for analyzing the effects of packaging on food loss reduction by considering consumer behavior, Journal of Cleaner Production, 174, 26–34

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50291-3

Semantically-enabled CAPS platform: Towards a generic and dynamic formulation for the synthesis and optimisation of value chains

Eirini Siougkrou*, Foteini Barla and Antonis C. Kokossis

Department of Chemical Engineering, National Technical University of Athens
9 Heroon Polytechniou Street, 157 80 Athens, Greece
esiougkrou@mail.ntua.gr

Abstract

A dynamic Computer-Aided Process Synthesis (CAPS) platform is developed able to offer high-throughput discovery and analysis of information for a wide range of users based on ontology engineering. The services that have been designed for the platform consider a range of user roles, including three main categories of services. (a) Basic services, such as search and upload of conversion pathways and models, (b) integration with external tools and databases, and (c) custom-made services that include synthesis and optimisation of value chains. The platform is further developed to enable a customised formulation for the optimisation of the value chains. A general formulation of the model equations is introduced, where each process unit is considered to be a black box described by inlets, outlets and generic functions that define the relationship between inlets and outlets. The user can define the form of the model equations in the ontology and then the model equations are transferred from the ontology to a mathematical form through a disjunctive programming formulation. Representative services of the semantic platform are illustrated for the case of the waste biorefinery.
Keywords: ontology engineering, repository, synthesis of value chains, optimisation of value chains

1. Introduction

The existing and developing resources in the CAPE (Computer-Aided Process Engineering) community can vary significantly in origin, type and scale; experimental data, algorithms, mathematical models, technology models, flowsheets, property models, executable files, etc.. A computational tool able to systematically store, retrieve and manipulate that knowledge and at the same time accessible to both experts and non-experts would be very beneficial. Ontology engineering can be the ideal base for the development of such a tool, as it can provide the means for high-throughput storing and processing of information.

Ontology engineering has been applied recently to several process engineering problems, such as process synthesis and optimisation (Kokossis et al., 2016; Barla et al., 2016; Magioglou et al., 2014) and industrial symbiosis (Cecelja et al., 2015). The current work anticipates the challenge to develop a dynamic platform able to host and integrate developments in the process engineering community. Ontologies are selected to model knowledge and process information services, such as synthesis and optimisation of value chains. The resources will be shared and will support dynamic updates and upgrades offering services to the users.

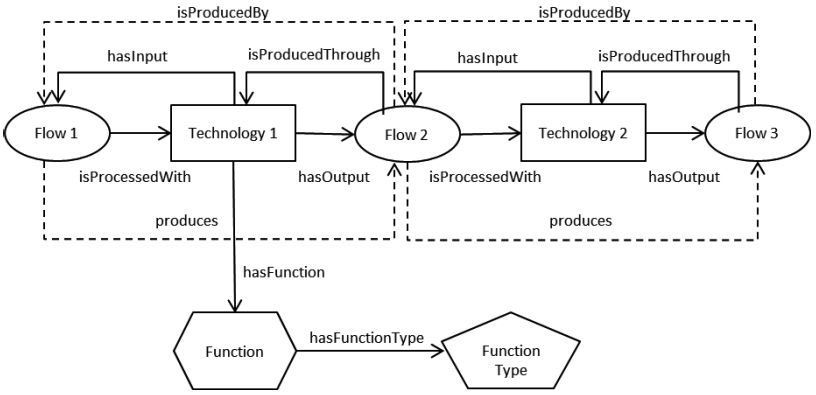


Figure 1. A representative pathway of the developed ontology and the defined object properties (relations). Different shapes correspond to different classes.

2. The CAPS platform

The platform is based on an ontological representation of the conversion pathways. The ontology comprises of four main classes; Flows, Technologies, Functions and Function types. Object properties are defined to describe the relations among the classes, as shown in Figure 1.

2.1 Services

Basic services

Basic services include Search and Upload of data, pathways and models. The users are able to search for existing pathways and technology models in the platform. In addition, new pathways and technology models can be added to the platform through a user-friendly interface. The new data that are uploaded are automatically integrated into the ontology and all possible connections with the existing data are made, possibly creating new knowledge. The user is also able to search for data to external databases that are linked via the CAPS platform.

Customised services

Customised services include the synthesis and optimisation of value chains, as well as integration with external databases and tools.

Synthesis of value chains

The Value Chain Builder is based on algorithms that have been developed to synthesise value chains from predefined feedstock/s or/and product/s. The algorithms make use of the object properties that connect the individuals of the ontology and are based on SPARQL commands. The corresponding steps are given in Figure 2. The steps that are followed for building value chains for preselected feedstock are the following:

Step 1. The Feedstock is selected.

Step 2. The flows that are contained in the value chains are found; all Intermediates and Products that *are produced by* the preselected Feedstock are found.

Step 3. The technologies that can connect the Feedstock and the Intermediates and Products that were found in Step 2 are found; the technologies that *have input* the

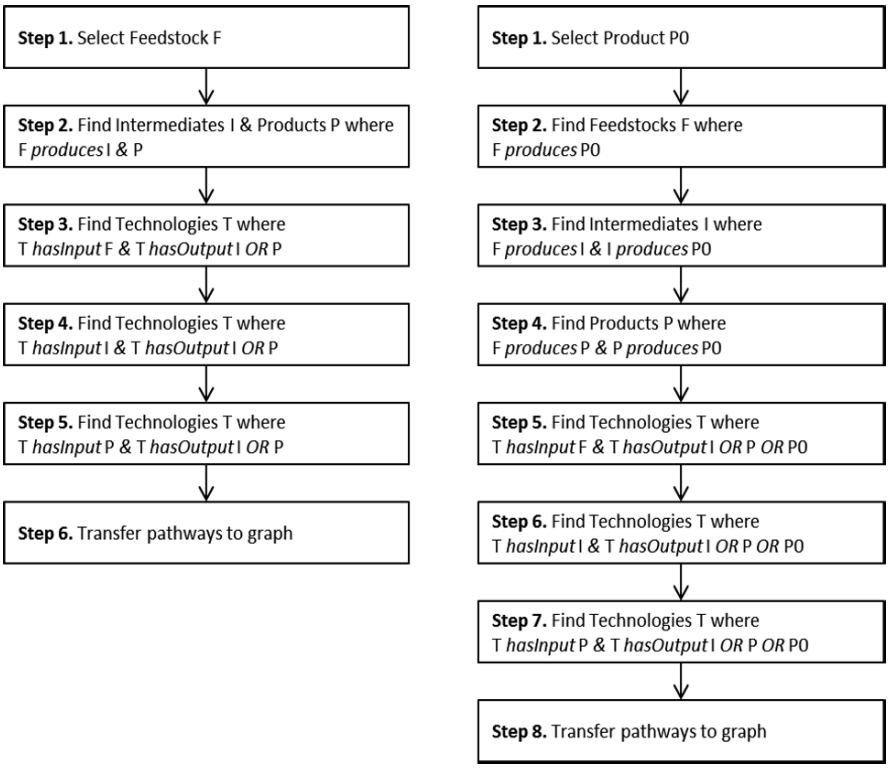


Figure 2. Algorithms for the building of value chains for selected Feedstock (left) and Product (right).

Feedstock *and have output* the Intermediates *or* Products found in the previous step are defined.

Step 4. Next, the technologies that can connect the Intermediates found in Step 1 and the other Intermediates and Products that were found in Step 2 are found. Thus, the technologies that *have input* the Intermediates *and have output* other Intermediates *or* Products are found.

Step 5. Similarly, the technologies that can connect the Products found in Step 1 and the Intermediates and other Products that were found in Step 2 are found. The technologies that *have input* the Products *and have output* other Intermediates *or* Products are defined.

Step 6. All pathways are transferred to a graph in the form of a superstructure.

The algorithm for preselected product has been developed in a similar way (Figure 2). The algorithms can be easily extended to consider also intermediates or combinations of flows, such as preselected feedstock and product, or feedstock and intermediate.

Optimisation of value chains

Once the desired value chain is obtained, the question that arises is "what is the optimal pathway?" This question can be answered by the Value Chain Optimiser, that provides

the optimal path of the value chain, for user-defined objective function and constraints. The Value Chain Optimiser, previously presented in Barla et al. (2016), has been further developed to consider a customised formulation of the optimisation problem. A generalized formulation of the optimisation problem is presented, enabling the use of user functions for selected technology models. Each technology block is described with inputs, outputs and functions that link the corresponding flows (f_i in Figure 3). These flows can be mass and energy flows, whereas the functions can be mass balances, energy requirement equations, environmental and cost constraints.

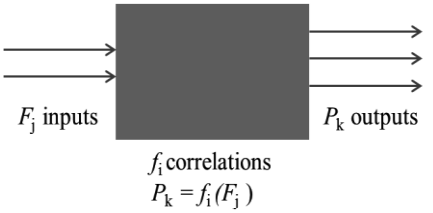


Figure 3. A technology unit represented by inputs (F_j), outputs (P_k) and relations f_i that connect them.

The optimisation problem is formulated as a Mixed Integer Non Linear Programming (MINLP) problem (equations 1-3), where the objective function (in this work, the total cost) is minimized subject to equality and inequality constraints. The equality constraints include model equations such as mass balances, energy requirements, cost and environmental models. The inequality constraints can be feedstock and product specifications or requirements.

$$\min \quad Obj\,(\boldsymbol{x}, \boldsymbol{y}) \tag{1}$$

$$\text{s.t.} \quad \boldsymbol{h}\,(\boldsymbol{x}, \boldsymbol{y}) = 0 \tag{2}$$

$$\boldsymbol{g}\,(\boldsymbol{x}, \boldsymbol{y}) \leq 0. \tag{3}$$

The optimization model is an extension of the formulation by Kokossis et al. (2015). The model equations are now defined using a disjunctive formulation, as

$$f_{l,i} = \begin{Bmatrix} \neg\, Z_{l.i} \\ \text{default eqn} \end{Bmatrix} \vee \begin{Bmatrix} Z_{l.i} \\ a_{l,i}x_l + b_{l,i}{x_l}^2 + c_{l,i}{x_l}^{d_{l,i}} + e_{l,i}y_l \end{Bmatrix} \tag{4}$$

where subscript l defines the technology unit and i defines the equation type, that is i={conversion, energy, cost, environmental, etc.}. $Z_{l.i}$ is a Boolean variable that is true, if the user has defined equation i for technology l. Thus, when an equation is not defined by the user, the corresponding equation by Kokossis et al. (2015) (default eqn in equation (4)) is used, while when a user-defined equation exists, the generalised formulation is used as shown in the second part of equation (4). Parameters a, b, c, d and e are values defined by the user, and x is the depending variable. y_l is a binary variable that represents the existence of technology l. As it is obvious from the equation, the user can enter linear, polynomial and exponential model equations. Non-zero parameters define the function type. Naturally, the equation is easily extended to include other forms of equations. The platform offers the possibility to use the default optimisation formulation (Kokossis et al., 2015), customised forms of equations, and any combinations of default and customised equations.

Integration with external tools and databases

The CAPS platform supports the integration of external tools, such as property or data models (e.g. LCA models, cost models). In addition, the platform is able to search for information to external databases and integrate internal and external data.

2.2 Implementation

The application platform has been developed upon Yii framework, a high-performance PHP framework using the Model-View-Controller (MVC) software design pattern. This web-based system lies on a private Linux Web Server provided by GRNET NOC's ViMa service ("ViMa," n.d.) and is part of IPSEN Tools, a collection of tools developed by the Industrial and Process Systems Engineering (IPSEN) group at the National Technical University of Athens. IPSEN tools are available in the URL *http://tools.ipsen.ntua.gr*. The architecture of the platform, as shown in Figure 4, contains three main separated layers, the presentation layer (the front end), the logic layer and the data access layer (the back end). The presentation layer provides the functionality for the user interface in order to interact with the data layer for both input and output.

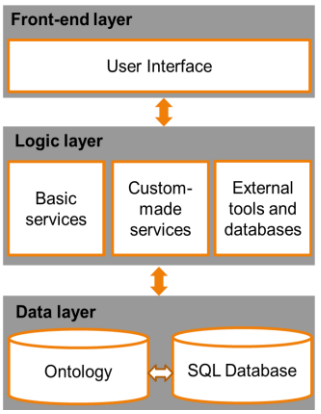


Figure 4. The architecture of the developed platform.

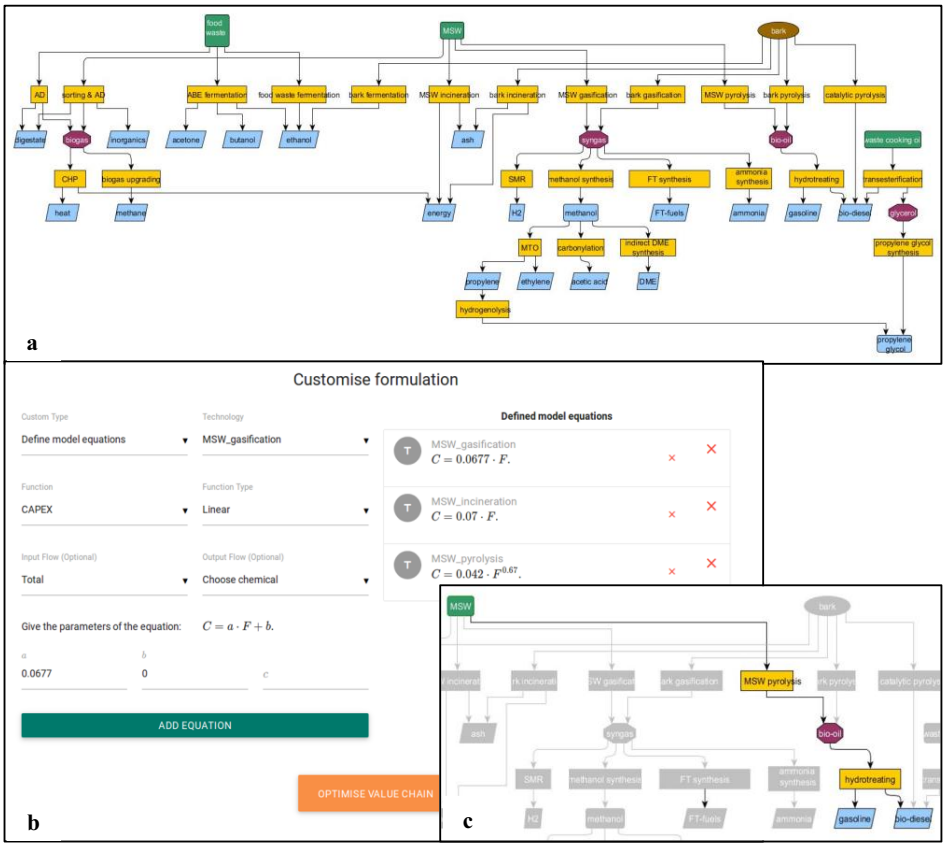


Figure 5. (a) The value chain produced for MSW, food waste, waste cooking oil and bark. (b) The interface of CAPS platform for customised optimisation formulation. (c) The optimised path.

3. The case of the waste biorefinery

The case of the waste biorefinery has been selected to demonstrate some of the functionalities of the platform. In this example, the user is looking for the optimised path for a defined set of feedstocks. Thus, the user firstly selects feedstocks to be included in the value chain; municipal solid waste (MSW), food waste, waste cooking oil and bark. The Value Chain Builder synthesises the value chain for the aforementioned feedstocks, as shown in Figure 5a. Next, the user chooses to "Optimise the Value Chain" for an amount of 500 Mt/year available total feedstock. In the next step, the user is called to choose between default and customised formulation. The default formulation corresponds to the use of existing models in the platform's database. The user selects customised formulation and defines the cost equations (Mountraki et al., 2016) for the technologies MSW_fast_pyrolysis, MSW_incineration and MSW_gasification (Figure 5b). The resulting optimal path is shown in Figure 5c, where two products are selected, gasoline and bio-diesel.

4. Conclusions

A dynamic platform has been presented that supports the integration and manipulation of knowledge in process systems engineering community. The platform stands as a repository of existing data, as well as a tool for the synthesis and optimisation of value chains. The development of the CAPS platform has been based on ontology engineering and its services include basic services, such as search and upload of data, and customised services for building and optimising value chains. A generalised formulation of the model equations for the optimisation has been presented, where the optimisation model can be customised according to user definitions. The user can partly define the model equations or data without needing to create a new optimisation model from scratch. The functionality of the developed tool has been demonstrated to a waste biorefinery example.

References

F. Barla, F. Lykokanellos, A.C. Kokossis, 2016. Discovering Valorisation Paths in Waste Biorefineries using an Ontology Engineering Approach, in: Computer Aided Chemical Engineering. Elsevier, pp. 2079–2084.

F. Cecelja, T. Raafat, N. Trokanas, S. Innes, M. Smith, A. Yang, Y. Zorgios, A. Korkofygas, A. C. Kokossis, 2015. e-Symbiosis: technology-enabled support for Industrial Symbiosis targeting Small and Medium Enterprises and innovation. J. Clean. Prod. 98, 336–352.

A.C. Kokossis, C. Labrador-Darder, F. Cecelja, 2016. Semantically enabled process synthesis and optimisation. Comput. Chem. Eng. 93, 64–86.

A.C. Kokossis, M. Tsakalova, K. Pyrgakis, 2015. Design of integrated biorefineries. Comput. Chem. Eng. 81, 40–56.

V. Magioglou, M. Tsakalova, A.C. Kokossis, 2014. Natural Language Modelling in Process Synthesis and Optimization, in: Computer Aided Chemical Engineering. Elsevier, pp. 877–882.

A. Mountraki, M. Tsakalova, A. Panteli, A. I. Papoutsi, A.C. Kokossis. Integrated Waste Management in Multiproduct Biorefineries: Systems Optimization and Analysis of a Real-Life Industrial Plant. Ind. Eng. Chem. Res. 55, 3478-3492.

E. Siougkrou, A.C. Kokossis, 2016. Development of semantically-enabled community hubs in biorefineries and biorenewables, in: Computer Aided Chemical Engineering. Elsevier, pp. 2013-2018.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50292-5

Computer Aided Design of Green Thermomorphic Solvent Systems for Homogeneous Catalyst Recovery

Kevin McBride[a*], Steffen Linke[a], Shuang Xu[b], Kai Sundmacher[a,b]

[a]*Max Planck Institute for Dynamics of Complex Technical Systems, Department Process Systems Engineering, 1 Sandtorstraße, D-39106 Magdeburg, Germany*
[b]*Otto-von-Guericke University Magdeburg, Process Systems Engineering, Universitätsplatz 2, D-39106 Magdeburg, Germany*
mcbride@mpi-magdeburg.mpg.de

Abstract

This work presents a new computer-aided thermomorphic solvent system (TMS) design methodology for the recovery of homogeneous catalysts that incorporates quantitative structure-activity relationships (QSAR) to predict various environmental, health, and safety (EHS) criteria using modern software packages. The quantum chemical method Conductor-like Screening Model for Real Solvents (COSMO-RS) (Klamt, 1995) is used for predicting catalyst solubility and the liquid-liquid equilibrium behavior of potential TMS designs. This methodology is then exemplified on the hydroformylation of 1-decene using the rhodium-Biphephos (Rh-BPP) transition metal catalyst and several green TMS designs were identified. Two of these TMS were then selected for experimental validation of both their LLE behavior and reaction performance.

Keywords: computer-aided molecular design (CAMD), COSMO-RS, green solvents, homogeneous catalysis

1. Introduction

The use of TMS has been shown to be a simple yet effective separation strategy in the recovery of homogeneous catalysts. Recently, a computational framework for a priori TMS design that incorporated predictions of catalyst solubility was developed (McBride and Sundmacher, 2015a) and experimentally validated (McBride et al., 2016) based on the quantum-chemical method COSMO-RS (Klamt, 1995). This design strategy was exemplified on the hydroformylation of 1-dodecene using the rhodium-Biphephos (Rh-BPP) homogeneous transition metal catalyst. Several potential TMS systems were identified and promising candidate systems were experimentally validated. Three of the considered TMS were very successful in recovering the catalyst and facilitating the reaction. Generally, the best performing TMS designs found consisted of mixtures using dimethylformamide (DMF) and alkanes in the range of C_8 – C_{14} as the polar and nonpolar component solvents, respectively.

However, a TMS consisting of DMF and decane has already been used in long-chain olefin hydroformylation research for several years. Since catalyst leaching remains economically high using this TMS (McBride and Sundmacher, 2015b), an alternative recycling strategy using a series of counter-current extraction stages using additional DMF was investigated. An evaluation of this separation strategy as part of an overall process optimization predicted that improved recovery of the catalyst at more

economical conditions is possible than previously (McBride et al, 2017). Nevertheless, due to its developmental toxicity, DMF is a solvent that should be substituted with a safer, more benign alternative. Continued use of this solvent is also in direct opposition to the idea of green and sustainable chemistry. Therefore a new computational strategy has been developed to find replacement solvents for DMF that are considered green and that would be suitable for recovering the catalyst as part of a multi-stage extraction.

2. Screening Method

This new method is based on the previous TMS design strategy first presented by McBride and Sundmacher (2015a), but with two major differences. Since it is known that an extraction process can be used to recover leached catalyst lost after the first decantation step (McBride et al., 2017), a solvent with lower catalyst recovery performance than DMF can be used. This allows one to explore a larger solvent search space than in the previous framework, increasing the potential to find a green solvent to replace DMF. Additionally, the simple prediction of the presence of a miscibility gap using binary LLE predictions has been replaced by a complete prediction of the ternary LLE for the reaction and separation stages at several temperatures. Although there are still inherent inaccuracies in the LLE predictions, they should provide an improved approximation of phase separation after the reaction which is also strongly correlated to catalyst leaching.

Since it is only of interest to replace DMF and that it was found previously that alkanes containing eight to 14 carbon atoms were the best choice for nonpolar solvents, it was decided that decane should remain fixed as the nonpolar solvent in the present work. This also has the benefit of reducing the amount of decane formed as a by-product due to the hydrogenation of 1-decene during the reaction. Thus, the focus here is to find a suitable and green solvent to replace DMF as the solvent used to recover the rhodium-Biphephos catalyst, which also simplifies the TMS design procedure by limiting our search to a single component solvent in the TMS. One of the challenges in green chemistry is finding feasible substitutes for currently used solvents. Fortunately, there are several methods that allow one to predict different EHS characteristics, such as persistency and carcinogenicity. The modern software packages VEGA (Benefenati et al, 2013) and EPI Suite (US EPA, 2017) are two such tools that enable one to make computationally based predictions of the EHS criteria for molecules simply by using their SMILES structures as the basis for QSAR-based models.

To start, two different lists of potential solvents are considered: those molecules included in COSMObase (ver. 1301 COSMOlogic GmbH) and a list containing known green solvents collected from Moity et al. (2012).

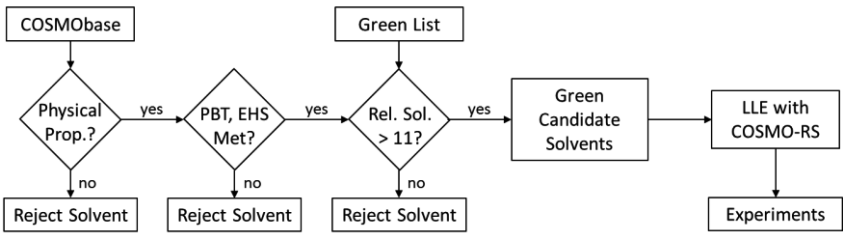


Figure 1. Screening methodology for green TMS design.

Table 1. Models used for each EHS characteristic and constraints on PBT criteria taken from US EPA TRI for substances of low concern.

Property	**Model (Software)**	**Limit**
Persistence	Biowin3 (Vega)	< 60 days
Bioaccumulation	CAESAR (Vega)	< 1000
Toxicity	ECOSAR (EPISuite)	< 10 mg/L
Carcinogenicity	CAESAR (Vega)	No
Mutagenicity	CAESAR (Vega)	No
Developmental Toxicity	CAESAR (Vega)	No

Those molecules provided by COSMObase are to be screened using physical property constraints and their predicted EHS criteria.The solvents included in the green solvent list are used directly in the LLE calculation step. A flowchart depicting this screening process is shown in Figure 1.

2.1. Predicted EHS Properties

In this framework, six molecular properties concerning EHS are considered: the persistence, bioaccumulation, and toxicity criteria (PBT), as well as predictions about whether each molecule is carcinogenic, mutagenic, or developmentally toxic. The screening is conducted using the values provided in Table 1 as constraints on the PBT, which have been set by the US EPA Toxics Release Inventory (TRI) Program. Predictions made for carcinogenicity, mutagenicity, and developmental toxicity are either yes or no decisions. The individual models for each property used in this work are also provided in Table 1 along with the software package used.

2.2. Relative Solubility

The second aspect considered is the relative solubility of the catalyst ligand in each potential green solvent. This is possible using COSMO-RS as implemented in the COSMOtherm software (Eckert and Klamt, 2016). Relative solubility has been shown to be a qualitatively acceptable measure for comparing the potential solvation power of a solvent for a specific solute and this was corroborated for the present catalyst in our previous work (McBride et al., 2016). The relative solubility compares the predicted chemical potential of a solute in a solvent at infinite dilution, which is an acceptable assumption for the catalyst ligand due to its very low concentration in the reaction mixture. To remove those solvents predicted to have poor solubility for the catalyst, a minimum value of 11 is selected as the cut-off value for the relative solubility (higher is better). For reference, DMF is predicted as having a relative solubility of 12.49. All calculations for relative solubility and the following LLE predictions were made using the BP_TZVPD_FINE_C30_1601 parameterization.

2.3. Liquid Phase Equilibrium

A requirement of the TMS is that a homogeneous mixture exists during the reaction and that a biphasic mixture forms upon subsequent cooling. For those solvents with suitable EHS characteristics and good relative solubility, ternary LLE calculations are predicted using COSMOtherm at 25 °C to obtain information about the potential ability of the mixture to separate. Care must be taken here as predictions of LLE behavior are

notoriously inaccurate for some substances, especially for mixtures containing more than two components. A similar calculation is performed for the ternary mixture consisting of the TMS and 1-decene at 100 °C to predict the degree of homogeneity in the reactor. These predictions were found to be unreliable and quite restrictive and therefore left as a consideration for future improvements to the TMS design method.

3. Results

3.1. Screening Results

First, those solvents that are contained within the COSMObase (ver. 1301, COSMOlogic GmbH) are considered. The only physical property constraint used in this example is a restriction on molecular weight (< 200 g/mol), which leaves 6261 candidates. After screening the molecules using the criteria outlined in Section 2.1 and 2.2 and evaluating the phase equilibrium behavior of each resulting ternary system as outlined in Section 2.3, only three molecules remain. These are listed in Table 2 on the left-hand side. Similarly, only three solvents included in the green solvent list were found to have acceptable LLE predictions. These are also presented in Table 2. Two solvents were then selected for experimental validation primarily based on price and accessibility: dimethylsuccinate (DMS), which was found in both lists, and tetrahydro-4h-pyran-4-one (THPO).

Table 2. List of potential solvents to replace DMF in the TMS using COSMObase (left) and from the green solvent list (right) taken from Moity et al (2012).

Molecule (COSMObase)	**Rel. Sol.**	**Molecule (Moity et al.)**	**Rel. Sol.**
2,7-Octanedione	13.49	Dimethylglutarate	12.05
Tetrahydro-4h-pyran-4-one	12.88	Dimethylsuccinate	11.87
Dimethylsuccinate	11.87	Glycerol-1-methyl monoether	11.33

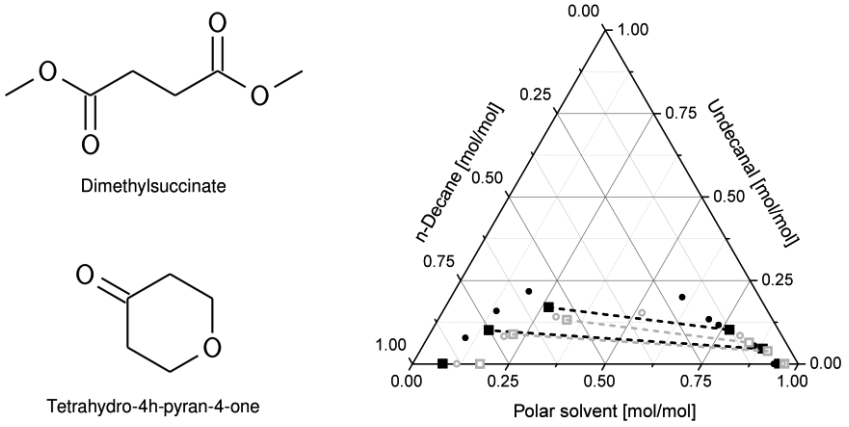


Figure 2. LLE (■) and cloud point measurements (●) for decane, undecanal, and DMS and LLE (□) and cloud point measurements (○) for decane, undecanal, and THPO at 25 °C. The molecular structures for both solvents are shown on the left.

3.2. LLE Measurements

Due to the uncertainty in predictions of LLE, they are measured directly. This ensures that one knows whether the mixture leaving the reactor forms a biphasic mixture upon cooling. This was performed on each mixture at 25 °C using a series of predefined compositions as well as by performing cloud point experiments. The results of these experiments are displayed in Figure 1. Additionally, several typical reaction mixtures of 1-decene, decane and the two polar solvents were heated to 100 °C and no phase separation was observed. Both solvents show acceptable LLE behavior for use in a TMS for the hydroformylation of 1-decene.

3.3. Reaction Experiments

The hydroformylation of 1-decene was performed in both TMS mixtures to confirm that they can facilitate the reaction. Each reaction was performed in a 1 L semi-batch reactor filled with 600 mL liquid. An equimolar synthesis gas feed was designed to maintain the pressure at 19 bar, which failed to do so due to very high reaction activity. This caused the pressure to fall to 6 bar using the TMS with DMS. A reduction in the catalyst amount by two-thirds was then used for THPO as a direct result of the high pressure loss during the first reaction. Despite this, the pressure still fell to 11 bar with THPO. Since our concern was in proving the feasibility of using these solvents in the reaction, further optimization of the reaction conditions was not pursued at this time.

The reaction profiles for both TMS are shown in Figure 3. In both cases, despite the differences in catalyst amounts used, the reaction proceeds rapidly towards a high conversion to the undecanal product. These preliminary investigations show that the hydroformylation can be efficiently performed in both TMS compositions. However, when using the TMS with THPO, a yellow substance formed during the reaction which may be caused by catalyst instability or degradation in the presence of THPO.

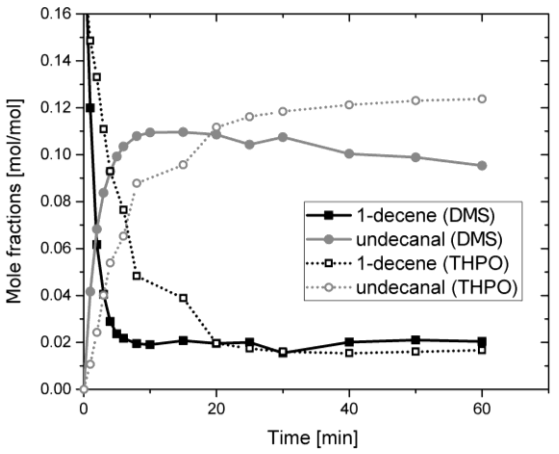


Figure 3. Mole fractions of 1-decene and undecanal during the reaction. The reactions were carried out at 105°C and 19 bar with 1-decene/decane/polar solvent = 0.20/0.48/0.32 mass ratios, H_2:CO = 1:1, catalyst/1-decene = 1:2000 for DMS, catalyst/1-decene = 1:6000 for THPO, and Rh/ligand = 1:5. All ratios given in mol/mol except for reactants and solvents.

4. Conclusions

This work presents a new framework for green solvent selection in TMS design used in replacing DMF as the currently used solvent in a TMS for catalyst recovery. The method was applied to the hydroformylation of 1-decene using the homogeneous rhodium-Biphephos catalyst. By considering the EHS criteria, catalyst ligand solubility, and LLE behaviors, several green solvents were identified. The LLE and reaction performance in two of these TMS designs were then experimentally validated. It was shown that performing the hydroformylation of 1-decene in a green TMS is not just feasible but functional. Future work will focus on a more detailed analysis of the catalyst recovery in a single separation stage and using a series of extractions. The effect of the solvents on the reaction performance is also to be considered in more detail, especially the reason why by-products formed when using THPO.

5. Acknowledgement

This work was conducted in part in cooperation with the Collaborative Research Centre "Integrated Chemical Processes in Liquid Multiphase Systems". The financial support from the German Research Foundation (DFG) under the grant SFB/TRR 63 is gratefully acknowledged.

References

E. Benefenati, A. Manganaro, G. Gini, 2013, VEGA-QSAR: AI inside a platform for predictive toxicology, CEUR Workshop Proceedings, 1107, pp. 21-28

F. Eckert and A. Klamt, COSMOtherm, version C16, COSMOlogic GmbH & Co. KG, Leverkusen, Germany, 2016

A. Klamt, Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena, 1995, J. Phys. Chem.-US, 99, 7, pp. 2224-2235

K. McBride, T. Gaide, A. Vorholt, A. Behr, K. Sundmacher, 2016, Thermomorphic solvent selection for homogeneous catalyst recovery based on COSMO-RS, Chem. Eng. Process., 99, pp. 97-106

K. McBride, N.M. Kaiser, K. Sundmacher, 2017, Integrated reaction-extraction process for the hydroformylation of long-chain alkenes with a homogeneous catalyst, Comput. Chem. Eng., 105, pp. 212-223

K. McBride and K. Sundmacher, 2015a, Computer-aided design of solvents for the recovery of a homogeneous catalyst used for alkene hydroformylation, Comput. Aided Chem. Eng., 37, pp. 2075-2080

K. McBride and K. Sundmacher, 2015b, Data driven conceptual process design for the hydroformylation of 1-dodecene in a thermomorphic solvent system, Ind. Eng. Chem. Res., 54, 26, pp. 6761-6771

L. Moity, M. Durand, A. Benazzouz, C. Pierlot, V, Molinier, J. Aubry, 2012, Panorama of sustainable solvents using the COSMO-RS approach, Green Chem., 14, pp. 1132-1145

US EPA, 2017, Estitmation Programs Interface Suite for Microsoft Windows, v 4.11

US EPA, https://www.epa.gov/toxics-release-inventory-tri-program/persistent-bioaccumulative-toxic-pbt-chemicals-rules-under-tri

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50293-7

Life Cycle Analysis of an Industrial Water Supply System

José O. N. de Jesus[a]*, Karla P. S. O. Esquerre[a], Asher Kiperstok[a], Albérico R. P. da Motta[a], Diego L. Medeiros[a], Carla P. de B. Lemos[a]

[a]*Federal University of Bahia (UFBA), Growing with Applied Modeling and Multivariate Analysis (GAMMA), Industrial Engineering Program (PEI), Federal University of Bahia (UFBA), Rua Professor Aristides Novis, 02, Federação, 40210-630, Salvador, Bahia, Brazil.*
jose.oduque@hotmail.com

Abstract

The life cycle assessment method was applied to identify the stages of the industrial water production system in an oil refinery and analyze the energy demand, carbon footprint and water footprint in order to propose scenarios for improvements. Based on the data collection carried out in the field, the inventory of the product system was elaborated in Simapro software with the ecoinvent database. Impact assessment methods applied were Cumulative Energy Demand (CED), Intergovernmental Panel on Climate Change (IPCC) (2013) and Available Water Remaining (AWARE). It was noticed that water losses represent 59% and electricity consumption 0.92 kWh per m^3 of distributed water. The proposed scenarios to use more efficient frequency inverters in water uptake, water capture in the nearest place and water losses reduction combined resulted in 28% water losses and 0.33 kWh of electricity consumption per m^3 of distributed water. The proposed scenarios combined also reduced CED by 51%, carbon footprint by 47% and water footprint by 39%. The proposed actions are part of the rationalization and eco-efficiency strategies of water use in oil refineries.

Keywords: Life Cycle Assessment, Energy Demand, Carbon Footprint, Water Footprint, Water Supply.

1. Introduction

Investigations on the IPCC climate scenarios from 2070 to 2100 for the State of Bahia in Brazil observed the largest reduction in precipitation on the coast, approximately 70%, and the highest temperature increase in the Northwest and North of the State by 5◦C (Tanajura et al., 2010). Water and energy rationalization are fundamental to achieve Sustainable Development Goals by 2030. About 2% to 3% of the world's energy consumption is used to capture, treat and distribute water to supply industries and homes (ASE, 2002). Water losses in uptake, treatment or distribution stages of water supply chain can occur by inadequate operation, lack of maintenance, depleted pipes and process management with low efficiency. Iwanek et al. (2016) argue that deteriorated infrastructure in the water supply chain causes leaks that reduce water carrying capacity, increase energy demand, cause system component failures, increase maintenance and operating costs, constant disruptions supply and reduce reliability. Therefore, the goal of this study was to assess the energy demand, carbon footprint and

water footprint of the industrial water supply chain to support the environmental management and operational costs reduction.

2. Methodology

The application of Life Cycle Analysis (LCA) followed the methodological structure recommended by ISO 14044 (2006). Data were collected from the oil refinery's water supply system in the foreground inventory (gate to gate) and the ecoinvent database version 3.3 was used in the background inventory (cradle to gate) in Simapro software version 8.4. The impact assessment methods were Cumulated Energy Demand (CED) (Jungbluth and Frischknecht, 2010), Carbon Footprint (IPCC 2013 - 100 years) and Water Footprint (Boulay et al., 2017).

The water supply system is the conventional type, composed of surface water uptake, treatment, storage and distribution. The water uptake uses three sets of motor-pump with a flow rate of 700 $m^3\ h^{-1}$ that pumps water to the Water Treatment Plant (WTP). There are three water catchment points, Catu river (245 $m^3\ h^{-1}$ or 35% of the total), Paraguaçu river (175 $m^3\ h^{-1}$ or 25% of the total) and São Paulo river (280 $m^3\ h^{-1}$ or 40% of the total). The electricity consumption of each set of motor-pump is 0.20 kWh m^{-3} for Paraguaçu river, 0.40 kWh m^{-3} for Catu river and 0.07 kWh m^{-3} for São Paulo river. In 2015 the total volume of water uptake was 6,117,984 m^3. The water supply system operates 24 h dia^{-1}. The WTP is composed of flocculation, decantation and filtration processes. The distribution of water flows by gravity. The reference flow adopted was 1 m^3 of distributed water. In this way, the amount of auxiliary materials and electricity consumption of the water supply system are presented in Table 1.

Table 1
Foreground inventory for 1 m^3 of distributed water

Stage	Amount	Unit	*GSD^2
Uptake			
Chlorine	0.030	kg	1.05
Electricity	0.54	kWh	1.07
Treatment			
Aluminum Sulfate	0.26	kg	1.05
Sodium Carbonate	0.12	kg	1.05
Polyelectrolyte	0.0026	kg	1.05
Electricity	0.38	kWh	1.07
Chlorine	0.025	kg	1.05
Sludge	0.020	m^3	1.09
Infrastructure	3.07E-9	unit	3.28
Distribution			
Electricity	0	kWh	-

*GSD^2: square geometric standard deviation.

Electricity consumption and water losses were identified as the largest contributors to the impact categories considered in this study. Therefore, improvement scenarios together with scenarios of water uptake per source were analyzed in order to identify

opportunities for environmental improvements. Seven additional scenarios were set for comparison with the Base Scenario.

- Inverter Scenario: considered the use of modern frequency inverters to reduce electricity consumption in uptake by 20%;
- Loss Scenario: considered the reduction of water losses in uptake to 5% and treatment to 20%;
- Combined Scenario: refers to the combination of Inverter Scenario and Loss Scenario;
- Paraguaçu Scenario: considered water uptake only from Paraguaçu river;
- Catu Scenario: considered water uptake only from Catu river;
- São Paulo Scenario: considered water uptake only from São Paulo river;
- São Paulo Combined Scenario: refers to the combination of Combined Scenario and São Paulo Scenario.

3. Results and discussion

The Base Scenario of the water supply system consumes 0.92 kWh of electricity per m^3 of distributed water. Fig. 1 shows the water balance of the oil refinery water supply system per stage.

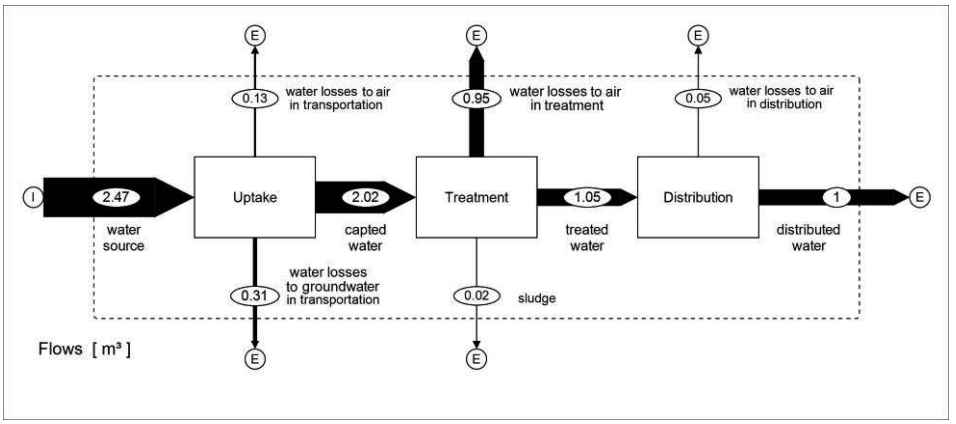


Fig. 1 Water balance along the water supply system (Base Scenario)

It was identified that each m^3 of distributed water requires 2.47 m^3 of water in uptake stage, 2.02 m^3 of water in treatment stage and 1.05 m^3 of water in distribution stage. Therefore, water losses represent 59% along the supply chain, considering 18% in the uptake stage, 48% in the treatment stage and 5% in the distribution stage. Water losses along the water supply system occur due to pipeline leakage, evaporation and sludge generation. The impact contributions of the water supply system are presented per life cycle stage for each scenario (Fig. 2, Fig. 3 and Fig. 4).

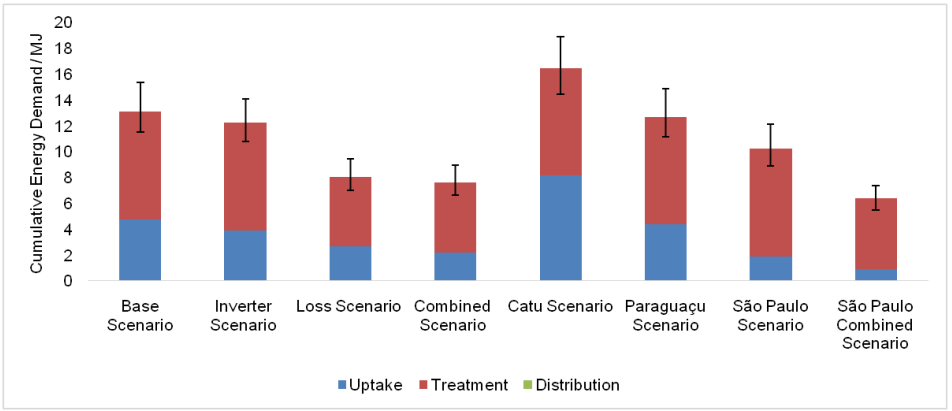


Fig. 2 Impact contribution of Cumulative Energy Demand with uncertainty (background and foreground) of 1 m³ of distributed water per stage along the life cycle for each scenario

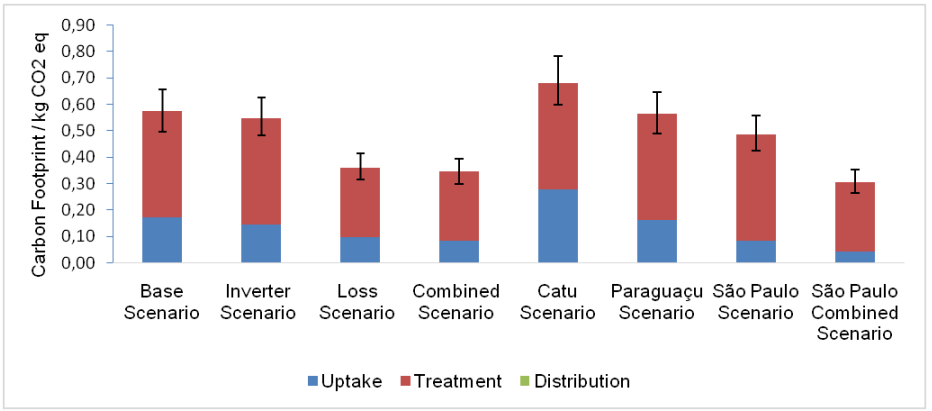


Fig. 3 Impact contribution of Carbon Footprint with uncertainty (background and foreground) of 1 m³ of distributed water per stage along the life cycle for each scenario

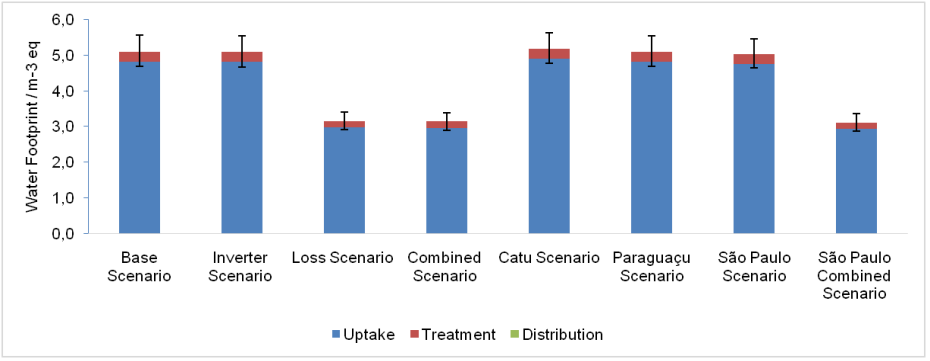


Fig. 4 Impact contribution of Water Footprint with uncertainty* (foreground) of 1 m³ of distributed water per stage along the life cycle for each scenario

*Uncertainty was only considered from foreground data due to inconsistencies presented in water flows from background data.

The largest contributions per stage of the water supply system were treatment stage for CED and carbon footprint, and uptake stage for water footprint. The distribution stage did not present significant impact contributions as infrastructure was considered in treatment stage and no energy was consumed at this stage. The main contribution to CED in the Base Scenario was electricity consumption (54%) of water uptake and treatment. The Inverter Scenario reduced CED by 6% and carbon footprint by 5% with the insertion of frequency inverters in the motor-pumps. On the other hand, Loss Scenario reduced the amount of water uptake from 2.47 m^3 to 1.39 m^3 that also reduced the electricity and chemical inputs. In this sense, the Loss Scenario reduced CED, carbon footprint and water footprint by 38% compared to Base Scenario. The Combined Scenario presented impacts slightly smaller than Loss Scenario, as the Inverter Scenario had a small contribution to impact reduction. The scenarios per water catchment points presented larger impacts for a distant source (Catu Scenario), average impacts for a medium distance source (Paraguaçu Scenario) and smaller impacts for a nearer source (São Paulo Scenario). São Paulo Scenario presented as an optimal point of water uptake with smaller energy consumption in relation to other catchment points. Therefore, the São Paulo Combined Scenario reduced CED by 51%, carbon footprint 47% and water footprint by 39%. Water management requires knowledge of the availability and demand of water resources to establish priority actions for water conservation. If the refinery plant reduces its water demand significantly it could be supplied by nearer water sources, reducing environmental impacts and operational costs. The climate change may also pressure water availability, as mentioned by Tanajura et al. (2010), therefore water rationalization along with alternative water sources should be considered.

4. Conclusion

The study analyzed the water balance, energy demand and environmental impacts of water supply system in an oil refinery with LCA methodology. The largest contributions in the energy demand, carbon footprint and water footprint were related to uptake and treatment stages, such as water pumping, electricity consumed in the water treatment and water losses. Some opportunities to improve energy efficiency and reduce water losses were identified, as well as water uptake from nearer sources. Energy and water efficiency along the supply chain is key to reduce environmental impacts and increase water availability for multiple uses. Thus, this study supports decision making aimed at reducing costs and improving environmental aspects of water supply systems.

References

A.-M. Boulay, J. Bare, L. Benini, M. Berger, M. Lathuillière, A. Manzardo, M. Margni, M. Motoshita, M. Núñez, A. Pastor, B. Ridoutt, T. Oki, S. Worbe, S. Pfister, 2017, The WULCA consensus characterization model for water scarcity footprints: assessing impacts of water consumption based on available water remaining (AWARE). Int. J. Life Cycle Assess. doi:10.1007/s11367-017-1333-8

Alliance to Save Energy (ASE), 2002, Watergy: taking advantage of untapped energy and water efficiency opportunities in municipal water systems. Washington.

C. Tanajura, F. Genz, H. Araujo, 2010, Mudanças climáticas e recursos hídricos na Bahia: Validação da simulação do clima presente do HadRM3P e comparação com os cenários A2 e B2 para 2070-2100, Revista Brasileira de Meteorologia (Impresso), 25, 345-358p.

International Organization for Standardization (ISO), 2006, ISO 14044: Environmental managements - life cycle assessments – requirements and guidelines, International Organization for Standardization, Geneva.

IPCC, 2013, Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535p.

M. Iwanek, B. Kowalska, E. Hawryluk, K. Kondraciuk, 2016, Distance and time of water effluence on soil surface after failure of buried water pipe, Laboratory investigations and statistical analysis, Eksploatacja i Niezawodnosc - Maintenance and Reliability, 18 (2), 278–284p.

N. Jungbluth, R. Frischknecht, 2010, Part II: 2. Cumulative energy demand, in: Hischier, R., Weidema, B., Althaus, H.-J., Bauer, C., Doka, G., Dones, R., Frischknecht, R., Hellweg, S., Humbert, S., Jungbluth, N., Köllner, T., Loerincik, Y., Margni, M., Nemecek, T. (Eds.), Implementation of Life Cycle Impact Assessment Methods, ecoinvent report no 3 v2.2, Dübendorf, 33-40p.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50294-9

Optimal Design of an Ambient Air Vaporizer using Numerical Model and DIRECT Algorithm

Yongkyu Lee [a], Jonggeol Na [a], Wonbo Lee [a*]

[a]*School of Chemical and Biological Engineering, Seoul National University, Build. 302, 1, Gwanak-ro, Gwanak-gu, Seoul (08826), Republic of Korea*
wblee@snu.ac.kr

Abstract

The ambient air vaporizer (AAV) using ambient air as heat source was environment-friendly method to vaporize LNG. The main problem of the vaporizer is performance degradation caused by frost formation on surface of tube. Determining the appropriate value of design variables under given climate is a critical issue. In this paper, design optimization in the vaporizer is performed using DIRECT algorithm whose result is compared with the result obtained by GA. Time series dataset including temperature and relative humidity of air for a year is categorized using k-means clustering. Finally, the optimal design shows 23.4 % improved performance of AAV and defrosting time is drastically reduced.

Keywords: Ambient air vaporizer, DIRECT algorithm, K-means clustering

1. Introduction

Ambient air vaporizer (AAV) is the most simple type of liquefied natural gas (LNG) vaporizer. The vaporizer do not require external power because it utilizes ambient air to vaporize LNG. Because cryogenic LNG is flowing through the internal section of finned tubes, water-vapour in the air is condensation and frost layer is accumulated at the surface of the finned tubes.[1, 2] The heat transfer efficiency of the vaporizer is reduced by thick frost layer, so defrosting time is typically required repeatedly.

Frost growth on the tube surface is generally known to be highly influenced by atmospheric temperature and humidity, but there is little research into the effects of AAV structures.[3-5] Frost growth is determined by the water-vapour density difference between air stream and frost surface. Weather condition such as relative humidity and temperature is related to the water-vapour density of air stream. But, water-vapour density on the frost surface is affected by the factors like cryogenic LNG, finned tube, etc. Hence, proper design of the AAV structure assists in maximizing heat transfer efficiency of the vaporizer.

Global optimization problems can be generally classified into stochastic and deterministic algorithms. Stochastic algorithms are based on the random generation of points, while deterministic optimization algorithms use a predetermined sequence of points to search for the global optimum.[6, 7] DIRECT, which is shorthand for Dividing Rectangles, is a deterministic global optimization algorithm. Given enough function evaluations and a highly sensitive stopping criteria, DIRECT will always find the global optimum of a function within a given domain.

Numerical AAV model simulate heat transfer from air to cryogenic LNG, mass transfer from air to frost layer and frost growth. The behaviour of the heat and mass transfer phenomena depends on the initially set structure and operation condition. The numerical

model which was validated by pilot-scale experimental data was used to simulate frost growth and temperature profile of LNG. The thickness and density of frost and temperature change of LNG in a vaporizer tube with time and position are predicted using simulation results.[5]

This paper proposes an optimal design of AAV that can vaporize LNG while minimizing the formation of the frost layer and defrosting time in a given climate condition. Climate conditions including temperature and relative humidity of specific region were categorized under k-means clustering. K-means clustering algorithm partitions the data set into k-clusters with the dual objective of making each cluster as compact as possible and the k-clusters as separated as possible. [8, 9] With the clustered data set, DIRECT algorithm and GA (genetic algorithm) were also used to find optimal design of AAV, which shows the difference in convergence speed and optimal point of design structure. Finally, the simulated results which were calculated under optimal design variables were compared with the results from original design variables. The comparison results shows proposed optimal design structure is more efficient than original design with a year's climate data.

2. Ambient air vaporizer model

The numerical model of the vaporizer calculate LNG temperature profile in fin tube of AAV, the thickness and density of frost layer on the tube surface and water-vapor in the air stream which flows outside. Input parameters includes climate data of a year at Incheon, Korea and structure related variables such as thickness and height of fin, tube radius and fan capacity. The climate data of a year was categorized under K-means clustering and centroid values of temperature and relative humidity represents each cluster. Hence, each cluster has its own representative climate condition and weight which was determined by ratio of the number of samples contained in the cluster to total data samples. For optimizing design variables, operating time until the temperature of outlet LNG becomes below -20 °C is defined as duration of AAV. The sum of duration at each cluster is defined as objective function which should be maximized at given design variables.

2.1. Numerical model of the vaporizer

The flowchart of numerical model is shown Fig. 1. First, the number of control volumes and time steps was defined, which was essential procedure for iterative techniques. The operation condition including temperature and humidity of air and temperature and flowrate of LNG is introduced. These parameters used to determine the properties like thermal diffusivity, convective heat transfer coefficient and density. Heat transfer model which calculates heat flux from air stream to LNG and mass transfer model which calculates growth of frost layer are solved based on previously obtained properties. In heat transfer model, heat fluxes and overall heat transfer coefficient affects the enthalpies of air and LNG. In mass transfer model, temperature of frost surface and mass flux of water-vapor determine the frost thickness and density. These two types of model are simultaneously solved and the results from the models are saved for current time step and control volume. The procedure is iteratively performed until the target control volume and time step reaches given values. The detailed explanation including equations and validation is contained at the paper of Lee et al. 2017. [5]

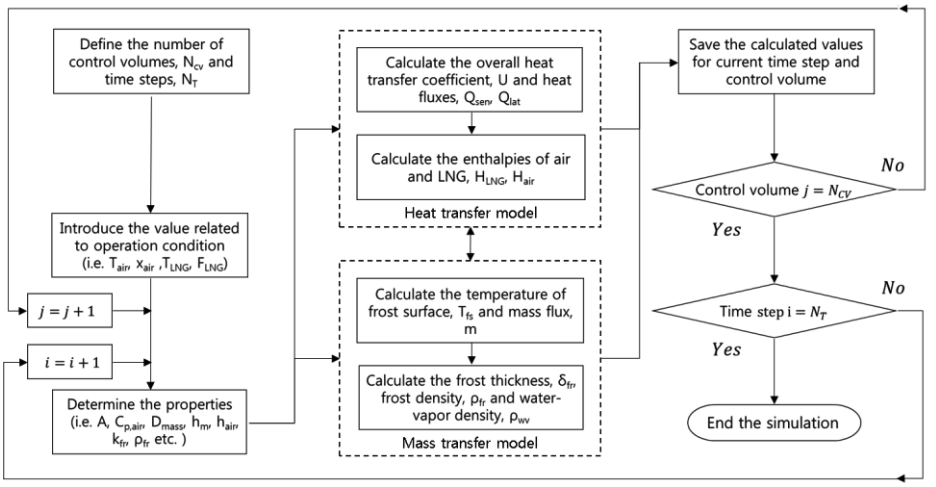


Figure 1. Flowchart of the vaporizer model

2.2. K-means clustering for categorized climate data

Climate condition of the region where the vaporizer was installed was a critical factor. The duration of AAV operation is mainly affected by the factor and the design of vaporizer structure should be considered under the factor. But temperature and relative humidity depend on location and season and are different at every hour of the day. Hence, the climate data with similar values should be grouped and the representative value of the group are selected for evaluation the design variables. K-means clustering is method for finding group which has similar features. Each group has its own centroid which represents the group. So it is simple method for grouping unlabeled big data. The data of a year have information of temperature and humidity every minute and the number of data points is 527,040. Figure 2. shows climate condition of a year in Incheon, Korea.

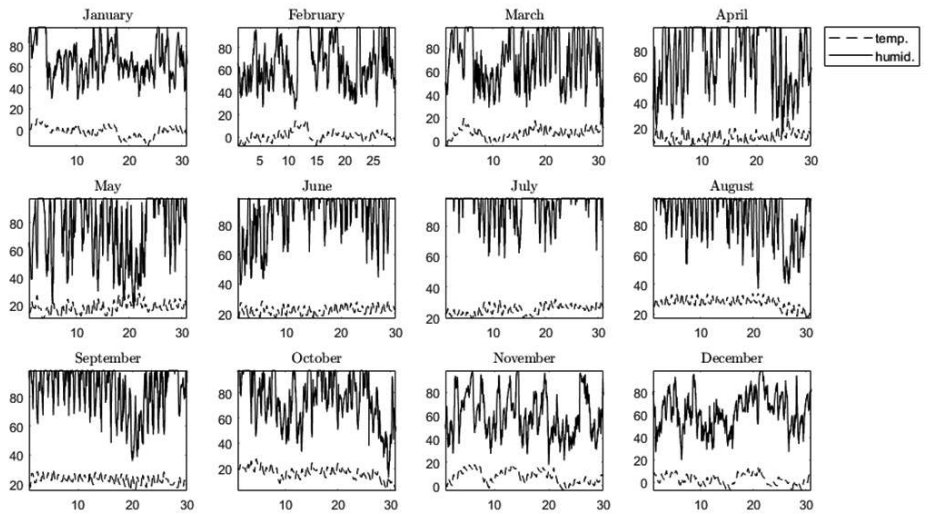


Figure 2. The data of a year in Incheon, Korea in 2016.

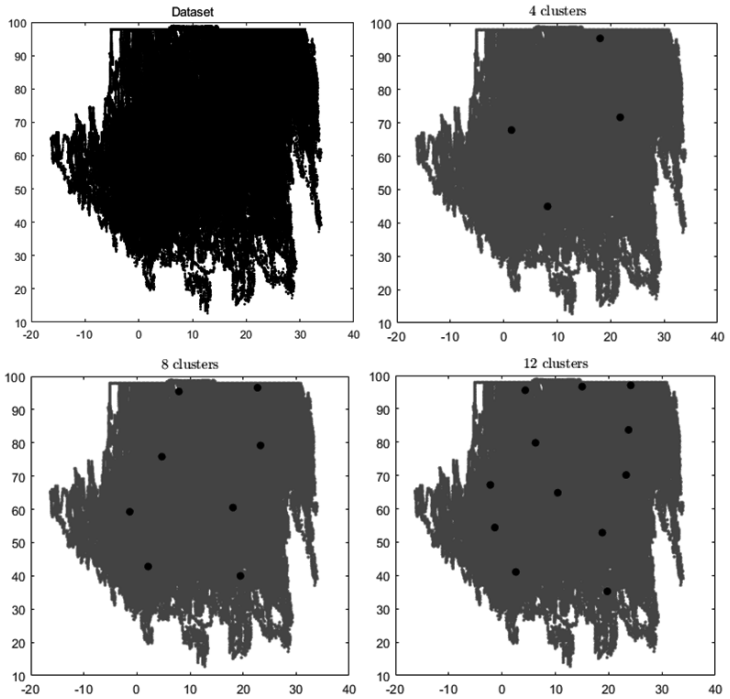


Figure 3. The dataset and centroids by k-means clustering (left-top: dataset, right-top: 4 clusters, left-bottom: 8 clusters, right-bottom: 12 clusters)

4 clusters, 8 clusters and 12 clusters are obtained by k-means clustering algorithm. Figure 3. shows the climate dataset and centroids of clusters.

2.3. Optimization formulation

There are no constraints in optimization problem. But the problem has multiple local optimum because of the complexity of numerical model. The Eq.(1) shows the optimization problem. $f(\vec{x}, w, C)$ is the objective function at which $\vec{x}$ means the design variables, w is weight which was ratio of the number of target cluster's data samples to the number of total data samples and C is temperature and relative humidity of air of clusters centroids. $g(\vec{x}, C_i)$ indicates the duration of validated model operation under $\vec{x}$ and C_i condition. Optimal design variables vector $\vec{x}$ was searched to maximize the sum of product of weight and duration. The boundary conditions of design variables $\vec{x}$ are [2000, 0.005, 0.05, 0.001] for lower bound and [4000, 0.02, 0.15, 0.005] for upper bound. The values in vectors indicate fan capacity, tube radius, fin height and fin thickness in order.

$$\underset{\vec{x}}{maximize} \quad f(\vec{x}, w, C) = \sum_{i=1}^{N_c} w_i g(\vec{x}, C_i) \tag{1}$$

$$s.t.: \quad LB_i \leq x_i \leq UB_i, \quad i = 1, 2, 3, 4$$

3. Optimization result and discussion

3.1. Optimal design of the vaporizer

Proposed optimization problem was solved by GA and DIRECT algorithm. The convergence rates of both algorithms are shown Figure 4. DIRECT algorithm approaches to the optimal point faster than GA at first. But the total iteration steps to optimal point is almost same. The weight of the objective function makes the optimal values almost same in different cluster size.

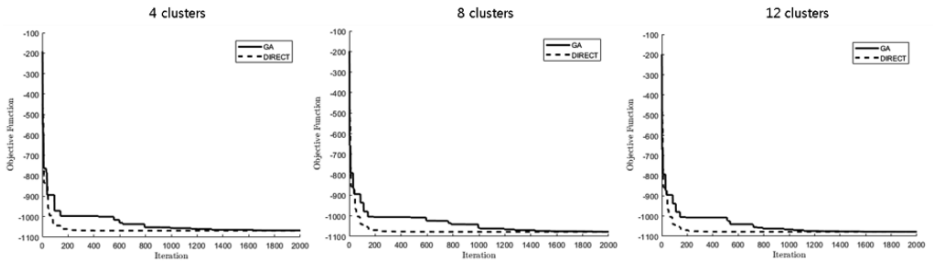


Figure 4. The convergence rate of GA and DIRECT algorithm

Table 1. shows the design variables of optimal solution of DIRECT algorithm at various clusters because DIRECT gives better solution than GA. Tube radius and fin thickness are converged at upper bound. Fan capacity at optimal solution is reduced as the number of clusters increases and fin height is converged near 0.1 m. The results show that proper design of fan capacity and fin height for specific climate is the critical for performance of AAV.

Table 1. The design variables of various conditions

	Fan capacity [m^3/min]	Tube radius [m]	Fin height [m]	Fin thickness [m]
Original	4000	0.0105	0.085992	0.0025
4 clusters	3738	0.0200	0.099826	0.0050
8 clusters	3492	0.0200	0.104100	0.0050
12 clusters	3357	0.0200	0.099588	0.0050

3.2. Comparison of performance in different designs

Figure 5. shows the ratio of operation time to a year. The number of clusters don't affect the results much but comparison of results between original design and optimal design shows 23.4 %, 22.4 %, and 22.6 % increase in operating time respectively. Although 12 clusters reflects better the overall climate than 4 clusters, optimal solution from 4 clusters deliver the best performance.

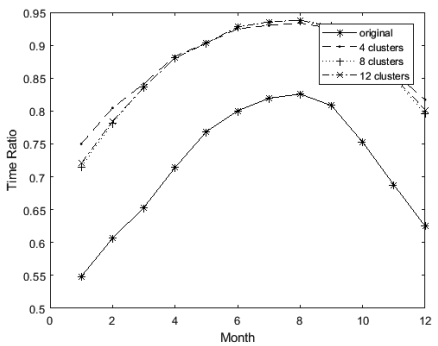


Figure 5. The ratio of operation time to a year

It shows the limitation of the clustering method when dealing with time series data. Figure 6. shows that optimal design provides outstanding performance and vaporizes LNG while minimizing the frost growth and defrosting time.

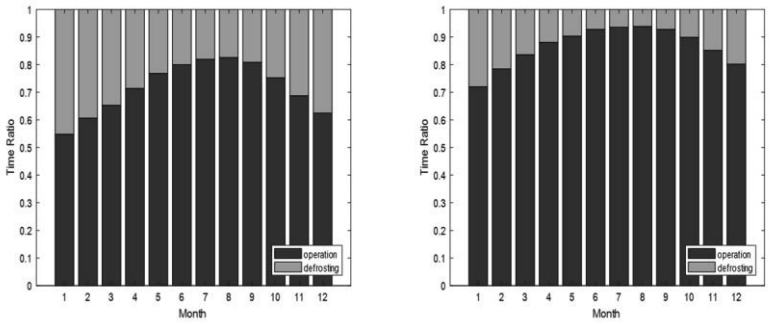


Figure 6. The ratio of operation time to defrosting time (left: original design, right: optimal design – 4 clusters)

4. Conclusions

The optimal design of AAV is proposed using K-means clustering method for data processing and GA and DIRECT algorithm for solving optimization problem. As the number of clusters increases, each cluster better reflects the overall climate. But many clusters does not guarantee optimal design of the vaporizer because clustering method is not best option for treating time series dataset. However, all optimal design shows better performance than original design under climate for a year. The ratio of operation time for clusters increases by 23.4 %, 22.4 %, and 22.6 % respectively. It means the vaporizer can be operated long time with short defrosting time.

References

T. Dendy, R. Nanda, 2008, Utilization of Atmospheric Heat Exchangers in LNG vaporization processes: a comparison of systems and methods, SPX Corporation, Houston, Texas.

D. Patel, J. Mak, D. Rivera, J. Angtuaco, 2013, LNG vaporizer selection based on site ambient conditions, Proceedings of the LNG, 17, 16-19.

S. Liu, W. Jiao, L. Ren, H. Wang, P. Zhang, 2017, Dynamic heat transfer analysis of liquefied natural gas ambient air vaporizer under frost conditions, Applied Thermal Engineering, 110, 999-1006.

Y. Kuang, C. Yi, W. Wang, 2015, Numerical simulation of frosting behavior and its effect on a direct-contact ambient air vaporizer, Journal of Natural Gas Science and Engineering, 27, 55-63.

Y. Lee, J. Park, C. Han, 2017, Modeling and Analysis of Frost Growth in Pilot-Scale Ambient Air Vaporizer, Industrial & Engineering Chemistry Research.

A. Younis, Z. Dong, 2010, Trends, features, and tests of common and recently introduced global optimization methods, Engineering Optimization, 42, 691-718.

H. Liu, S. Xu, X. Wang, J. Wu, Y. Song, 2015, A global optimization algorithm for simulation-based problems via the extended DIRECT scheme, Engineering Optimization, 47, 1441-1458.

J. MacQueen, 1967, Some methods for classification and analysis of multivariate observations, in: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Oakland, CA, USA., pp. 281-297.

A. Ahmad, L. Dey, 2007, A k-mean clustering algorithm for mixed numeric and categorical data, Data & Knowledge Engineering, 63, 503-527.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50295-0

Thermoeconomic design of biomass biochemical conversion technologies for advanced fuel, heat and power production

Maxime Alexandre Vigot[a], Theodoros Damartzis[a*], François Maréchal[a]

[a]*Industrial Process and Energy Systems Engineering (IPESE), École Polytechnique Fédérale de Lausanne (EPFL), 1951, Sion, Switzerland*
theodoros.damartzisl@epfl.ch

Abstract

The current work addresses the design and assessment of different biomass conversion pathways based on anaerobic digestion for the efficient transformation of the raw materials to a variety of products, including bio-SNG, heat and power. Using a generic superstructure-based approach, various conversion pathways can be synthesized in the context of a MILP optimization problem that aims to minimize the total process cost. A comprehensive methodology is utilized, able to produce a large variety of process alternatives. Heat integration is employed among all designed cases to best exploit the process thermal streams and minimize the energetic needs. Four indicative flowsheet designs are presented and assessed here, regarding their energetic consumption as well as their associated revenues.

Keywords: anaerobic digestion, biogas, bio-SNG, heat and power production, fruit and vegetable residues.

1. Introduction

Recent fluctuations in the prices of fossil fuels as well as the overall shift towards renewable energy have led to the intensification in the design and employment of biomass-based conversion technologies. The latter, together with the high available biomass potential constitute the enhancement of the role of biomass in the global energy mix as well as the development of efficient conversion technologies to support the chemical transformation to energy and fuels a necessity.

Biomass conversion technologies allow the transformation of raw biomass to a variety of gaseous, liquid or solid fuels as well as the simultaneous production of heat and power for energy services supply. Biomass-produced fuels can substitute nearly any kind of conventional fuel, decreasing the associated greenhouse gas emissions. However, the current high investment costs, together with the relatively low conversion efficiencies hamper the quick propagation of these technologies and limit the utilization of biomass to direct combustion in order to maintain a profitable operation (Codina Girones et al., 2017). On the other hand, the drawbacks associated with the efficiency limitations during the chemical transformation of biomass can be easily compensated by an effective reformulation of the conversion path, including the employment of a variety of sub-processes that increase the overall efficiency.

Anaerobic digestion (AD) is a biochemical process for the conversion of a wide variety of wet biomass (i.e. animal manure, sludge from wastewater treatment plants (WWTP),

household or industrial residues containing significant levels of humidity etc.) into biogas, a mixture of CH_4 and CO_2 (Ward et al., 2008). Biogas can be then upgraded to bio-SNG or used as fuel for heat and power production. Anaerobic digestion also produces a sludge residue, which includes the undigested part of the biomass. The latter can be directly used as soil amendment of further treated for methane or energy production as well as the extraction of the contained nutrients in the form of phosphorus, nitrogen and potassium (Abubaker et al., 2012).

A large variety of possible pre-treatment steps can be used enhance the digestibility of the biomass by either chemically intervening to the system's biochemistry (Passos et al., 2017) or simply increasing the effective area for digestion with simple size reduction of the raw material. Similarly, the upgrading of biogas into a stream of pure CH_4 is realized with the employment of a number of techniques that include the separation and removal of CO_2 from the gaseous mix, the conversion of CO_2 into CH_4 or even a combination of the two (Singhal et al., 2017). The solid digestate can also be processed in a number of ways including drying and reuse as fertilizer (Alburquerque et al., 2012) or further treated in order to recover the contained carbon material (Tambone et al., 2010). Finally, depending on the employed conversion and upgrading technology, the produced stream of either biogas or bio-SNG can be further treated for energy recovery in gas boilers, cogeneration units of even injected into the natural gas grid, provided that the specifications are met.

The large number of possible options in all processing stages between the raw materials and the end products creates a plethora of conversion chains that differentiate in terms of conversion efficiency as well as total process cost. This work aims towards the systematic identification, development and assessment of efficient pathways through which biomass is effectively converted to a number of potential products based on anaerobic digestion. This complex problem is handled within a generic framework that utilizes a process superstructure approach for the representation of the possible conversion pathways as well as their synthesis from a list of available technologies that act as building blocks. The efficient design of the arising conversion pathways is established through the use of a MILP optimization formulation that aims to minimize the total capital and operating expenses of the process. Process integration considerations are further employed to ensure the minimization of the energetic demands in the designed plants and further enhance the process economics through the efficient design of the heat exchanger network. Thus, the emerging solutions can be easily assessed for their economic performance and decisions regarding the selection of the most suitable conversion pathway based on the type of raw biomass used and intended product can be facilitated.

2. Methodology

To address the problem of designing an economic and energy efficient process for the treatment of biomass through anaerobic digestion, a process superstructure was formulated consisting of the possible options representing the employed technologies. The arising superstructure is schematically depicted in Figure 1. The considered technologies revolve around the processes of pre-treatment, conversion to biogas, upgrading of biogas and provision of services.

Pre-treatment of biomass involves the mechanical, thermal or chemical processing of the raw material in order to facilitate the digestion process. Mechanical pre-treatment

methods usually resort to size reduction in order to expose a larger area to the methanogenic bacteria and increase the rate of conversion. On the other hand, thermal pre-treatment methods involve the processing of the biomass at elevated temperatures for a specified amount of time in order to enhance the solubility of organic waste constituents and subsequently promote the hydrolysis step. Finally, chemical methods aim towards the initial breakdown of the organic matter via the addition of alkaline or acid solutions.

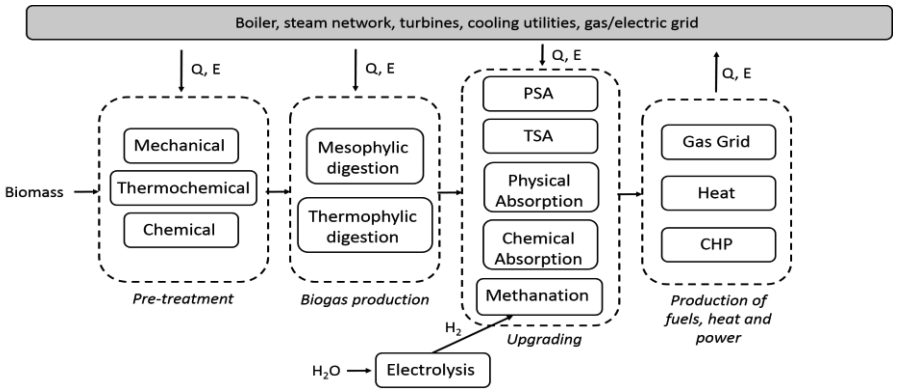


Figure 1. Biomass conversion superstructure.

The considered conversion of biomass to biogas involves two scenarios; utilizing either mesophilic or thermophilic bacteria. The major differences between the two involve the operating conditions such as temperature and retention time but also affect the CH_4 content of the biogas.

A number of different biogas upgrading methods are considered as shown in Figure 1. PSA and TSA exploit the pressure and temperature effects on the CO_2 adsorption equilibrium, respectively, creating an alternating cycle of adsorption and desorption in suitable solid materials. Membrane separation is based on the gas dissolution and diffusion into polymeric materials, responsible for selectively allowing the passing of certain gases, thus achieving the separation of a gaseous mixture. Solvent-based CO_2 capture processes involve the physical or chemical entrapment of CO_2 in usually packed absorption / reactive absorption columns. Depending on the employed solvent, various regeneration and operating costs apply for these processes. Finally, direct methanation involves the reaction of CO_2 with H_2 in moderate temperature and pressure for the production of additional CH_4. Subsequently, the product of upgrading can either be directly subjected into the natural gas grid of further converted to heat and/or power using turbines, fuel cells or advanced cogeneration units.

The addressed problem is formulated as a MILP optimization focusing on the minimization of the total process expenses that account for the annualized capital and operating costs of each designed process, further implementing heat integration considerations to account for the minimization of the heating/cooling requirements. The methodology is implemented in the OSMOSE process integration platform (Bolliger, 2010), where the mathematical programming is defined using AMPL. The employed optimization scheme can be seen in Figure 2, showing the information exchange between the various modeling layers.

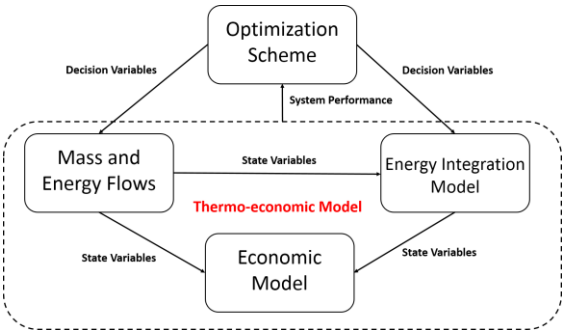


Figure 2. Optimization scheme

3. Case study

Fruit and vegetable residues with a total solids (TS) content of 10 % wt. have been selected as the raw material. The volatile solids (VS) content of the biomass is 86 % wt (Gunaseelan, 1997) and the plant capacity was set to 12500 kg/day.

Four indicative scenarios are presented here as part of the attained solutions. The first two scenarios are depicted in Figure 3 and differentiate by the end services. Specifically, in FS-1, the methane from the upgrading step is directly inserted to the natural gas grid, while a SOFC is used in FS-2 to convert it to electricity. Both flowsheets neglect the use of a pre-treatment step as is was not deemed economically viable, while they employ a biofiltration step for the cleaning of the biogas impurities as well as PSA for the upgrading step. In both cases, the reactor is a handled as a thermophilic digester with an operating temperature of 55 ºC and a retention time of 20 days, a typical value for the thermophilic digestion of food waste (Mir et al., 2016). The PSA utilizes a pressure of 8 bars to achieve the desirable separation.

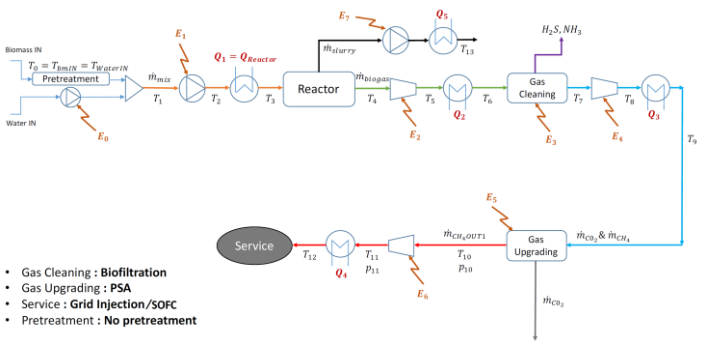


Figure 3. Schematic representation of scenarios FS-1 and FS-2.

Figure 4 shows the designed flowsheets FS-3 and FS-4. In these, a methanation reactor operating at 220 ºC and 20 bars is used to convert the captured CO_2 into extra methane and enrich the final methane yield prior to grid injection or further treatment. The required H_2 for the reaction is obtained from water electrolysis using renewable energy sources, in the form of solar energy. However, in this study, the electrolyzer design and economics were not taken into consideration. Like the first two flowsheets, FS-3 and

FS-4 differ in the end services, with FS-3 producing a bio-SNG stream that is injected into the gas grid while FS-4 employs a fuel cell for power generation. The operating principles of the digester and PSA are the same as the respective ones in FS-1 and FS-2.

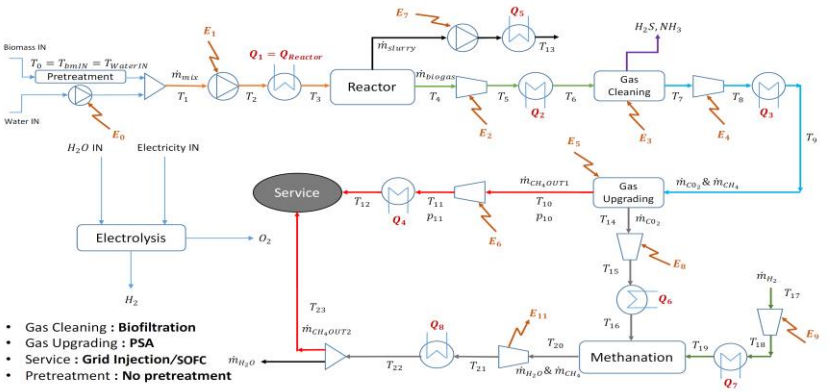


Figure 4. Schematic representation of scenarios FS-3 and FS-4.

In all four cases, the obtained heat from the heat exchangers is sufficient to fulfill the energetic needs of the process regarding heating, eliminating the need for an extra boiler, while in the cases of power generation the latter is partly used to supply the necessary power in the pumps, motors and compressors of the process. Cooling, on the other hand, is still necessary to reduce the temperature of the intermediate process streams.

Table 1. Operating expenses and revenues for the four reported flowsheets.

	FS-1	FS-2	FS-3	FS-4
Expenses (in €)				
Process Water	1966.27	1966.27	1879.39	1879.39
Electricity	25352.85	0	37066.63	0
Revenues (in €)				
SNG	59270.6	0	93854.96	0
Electricity	0	70010.27	0	112626.71

Table 1 summarizes the energy consumption in the four designed flowsheets as well as the obtained revenues from the distribution of bio-SNG or electricity into the respective grids. The process water used to create the slurry fed into the digester is smaller in the cases where methanation is used. This is because the water produced by the reaction is assumed to be used in the process after a condensation step. It can be seen that the power demands in the cases where a SOFC is used are fully satisfied by the product of the process itself, disregarding the need for grid imported power. These include the energy demands for the compressors, motors, pumps as well as the operation of the cooling tower. Finally, the option of using a methanation reactor significantly increases the production of bio-SNG through the recovery of extra methane leading to increased revenues, both in the case of grid injection as well as in the case of power generation.

4. Conclusions

A comprehensive methodology for the design and thermoeconomic assessment of biochemical biomass conversion technologies was presented. A generalized superstructure approach, encompassing the various intermediate processes that lead from the raw materials to the end products, coupled with a MILP optimization problem, has led to the design of a variety of process alternatives. A case study is shown here, accounting for the transformation of fruit and vegetable residues to bio-SNG, heat and power and four indicative solutions are presented. Heat integration has allowed the minimization of the heating demands in the processes and a variance is observed among the final revenues depending on the type of the end services as well as the conversion path used. The employment of a methanation reactor for the enrichment of the produced bio-SNG has led to improved attained revenues in the range of 58 % for the case of bio-SNG production and 61 % for the case of power generation.

Acknowledgements

The authors gratefully acknowledge the financial support of this work by the Swiss Competence Center for Energy Research, BIOSWEET - Biomass for Swiss Energy Future.

References

J.A. Alburquerque, C. de la Fuente, A. Ferrer-Costa, L. Carrasco, J. Cegarra, M. Abad, M.P. Bernal, 2012, Assessment of the fertiliser potential of digestates from farm and agroindustrial residues Biomass and Bioenergy, 40, 181-189.

R. Bolliger, Methodologie de la syntheese des systeemes energeetiques industriels, 2010, Ph.D. Thesis: Available online: https://infoscience.epfl.ch/record/151501?ln=en.

V. Codina Girones, S. Moret, E. Peduzzi, M. Nasato, F. Maréchal, 2017, Optimal use of biomass in large-scale energy systems: Insights for energy policy, Energy, In press, doi: 10.1016/j.energy.2017.05.027.

V. N. Gunaseelan, 1997, Anaerobic digestion of biomass for methane production: A review, Biomass and Bioenergy, 13,(1–2), 83-114.

M. A. Mir, A., Hussain, C. Verma, 2016, Design considerations and operational performance of anaerobic digester: A review, Cogent Engineering, 3, 1181696.

F. Passos, V. Ortega, A. Donoso-Bravo, 2017, Thermochemical pretreatment and anaerobic digestion of dairy cow manure: Experimental and economic evaluation, Bioresource Technology, 227, 239–246.

S. Singhal, S. Agarwal, S. Arora, P. Sharma N. Singhal, 2017, Upgrading techniques for transformation of biogas to bio-CNG: a review, International Journal of Energy Research, 41, 1657–1669.

F. Tambone, B. Scaglia, G. D'Imporzano, A. Schievano, V. Orzi, S. Salati, 2010, Assessing amendment and fertilizing properties of digestates from anaerobic digestion through a comparative study with digested sludge and compost, Chemosphere, 81, 577-583.

A.J. Ward, P.J. Hobbs, P.J. Holliman, D.L. Jones, 2008, Optimisation of the anaerobic digestion of agricultural resources Bioresource Technololy, 99, 7928-7940.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50296-2

Prediction of Dissolved Methane Loss in Anaerobically Treated Effluent Based on ADM 1 and Equilibrium Conditions

Júlia Carolina B. F. Bijos[a]*, Robson W. S. Pessoa[a], Karla P. S. O. R. Esquerre[a],

[a]*Federal University of Bahia, Professor Aristides Novis St., Salvador 40210-630, Brazil*
juliabijos@outlook.com

Abstract

The following work aims to estimate the production of biogas in the liquid and gaseous phases in an anaerobic reactor by means of simulation. For this, the ADM 1 model modified by Mendes et al. (2016) is used to represent the simultaneous removal of carbon and nitrogen (SRCN) from a fishery effluent. Henry equation, originally used to represent thermodynamic equilibrium, is hereby replaced by a method that aggregates more robust equations, available in the simulator Aspen Plus. Therefore, the validity of Henry's law is tested by applying an equilibrium condition that considerate liquid phase modeling by means of eNRTL method. The results show that when applying the thermodynamic method, the gaseous concentrations for CH_4, N_2, H_2 and acetic acid are lower than those foreseen by ADM 1, and the presence of these dissolved components in the liquid phase is indicated. That way, it is verified that the consideration of equilibrium condition is fundamentally important for anaerobic digestion process, besides indicating the need for more in-depth studies about the influence of this consideration in the process.

Keywords: anaerobic digestion, equilibrium conditions, ADM 1, modeling.

1. Introduction

Anaerobic processes allow treatment of effluents with low sludge generation and production of biogas (renewable energy). Less space is needed for operation and it's applied to the treatment of both diluted and concentrated effluents. In this process, about 70 - 90 % of the influent COD is converted to biogas, leaving the reactor in the gaseous phase. A portion ranging from 5 – 15 % is converted to biomass, and about 10 – 30 % of the influent COD is converted neither into biogas or biomass, so this portion composes the liquid phase leaving the reactor as non-degraded material (Chernicharo, 2007). In this portion, some dissolved gases such as N_2, CO_2 and especially methane gas (CH_4) are found.

In this sense, anaerobic digestion presents itself as a potential source of greenhouse gas emissions, since CH_4 and CO_2 have high polluting potential and may be emitted by the process. Particularly in relation to CH_4 gas, besides the pollution potential, its loss in the liquid phase leads to the reduction of energy potential of the system (Lobato, 2011).

This concern motivated several authors to investigate the presence of CH_4 dissolved in both diluted and concentrated wastewater. The results of the studies indicate a large variation in the dissolved amount. Souza et al. (2011), observed values around 36-41% for domestic water treatment, and other authors reported by Crone et al. (2016) found

values of 11-100% at different temperatures, concluding that the concentrations of dissolved methane may be many times higher than those values predicted by Henry's law, in a supersaturation state. In this sense, it's necessary an accurate modeling of this process, based on biochemical fundamentals in order to predict the biogas production. Also, it's important to do a reliable consideration of the liquid-vapor phase equilibrium, aiming a correct calculation of the gaseous solubility. This approach allows a closer representation of the real situation.

In light of this, the present work uses the gas production data predicted by the Anaerobic Digestion Model N.1 (ADM 1) for the wastewater treatment of salmon effluent in a CSTR proposed by Mendes et al. (2016) and implemented in Matlab/Simulink. Together with ADM 1, it is proposed to improve the thermodynamic representation of the process by using a simulation tool, Aspen Plus. Therefore, it is possible to verify the prediction of biogas formation, and the presence of dissolved gases (mainly CH_4) in the liquid phase, after the application of a thermodynamic method.

2. Anaerobic Biodigestion Modeling and the Representation of Equilibrium

Knowing the complexity of the anaerobic digestion process, Batstone et al. (2002) proposed a generic model to control the operation and prediction of biogas, the ADM 1. It is composed of biochemical and physical-chemical reactions, represented by acid-base and gas-liquid transfer reactions. The gas liquid equilibrium is represented by Henry's Law. Despite being used, this law does not present great relevance in relation to gas transfer, indicating how much the liquid-gas system is far from equilibrium. (Souza, 2010). Mendes et al. (2016) proposes a modification to represent the simultaneous removal of carbon and nitrogen (SRCN) in anaerobic-anoxic reactor.

Batstone et al. (2015) ponders the need to consider the solubility of the gases, including methane. Since this gas has a low mass transfer coefficient, it ends up being released with the final effluent. It is also necessary to consider the influence of electrolytes, since these ions interact with stronger and far-reaching forces than those of the neutral molecules. In addition, because of the different functional groups, the effluent will have a high non-ideality. Consequently, Henry's Law will be significantly different for this case over those with very low electrolyte concentrations. (Sandler, 2006).

Since the equilibrium conditions control the entire reaction mechanism that occurs between species in reactor, it becomes essential to apply a more robust thermodynamic method to the model in order to obtain results closer to reality.

3. Methodology

For integration between the ADM1 model proposed by Mendes et al. (2016) and the external simulator, ActiveX tool was used. Due to consideration of acid-base pairs, the chosen method on the thermodynamic simulation was "Electrolyte-NRTL" (eNRTL), implemented as ELECNRTL. This is a comprehensive electrolytic method since it handles solutions of very low concentrations up to high concentrations. It applies a modified Debye-Rueckel equation, able to represent long and middle range interactions. Furthermore, it takes Redlich-Kwong equation for predict the vapor phase properties and Henry's Law for modeling the gases solubility. This method also handles short range interactions. (Aspen Technology, 2013; Biernacki, 2014).

This way, ADM1 for SRCN is integrated from an initial instant t to a final instant t+1, step by step, until it reaches the steady-state. In each instant of time, Matlab communicates with the simulator estimating the compounds distributions between gaseous and liquid phase, the equilibrium conditions. Figure 1 shows a schematic representation of the integrated model developed, as well as the flow of required information and data for the simulation.

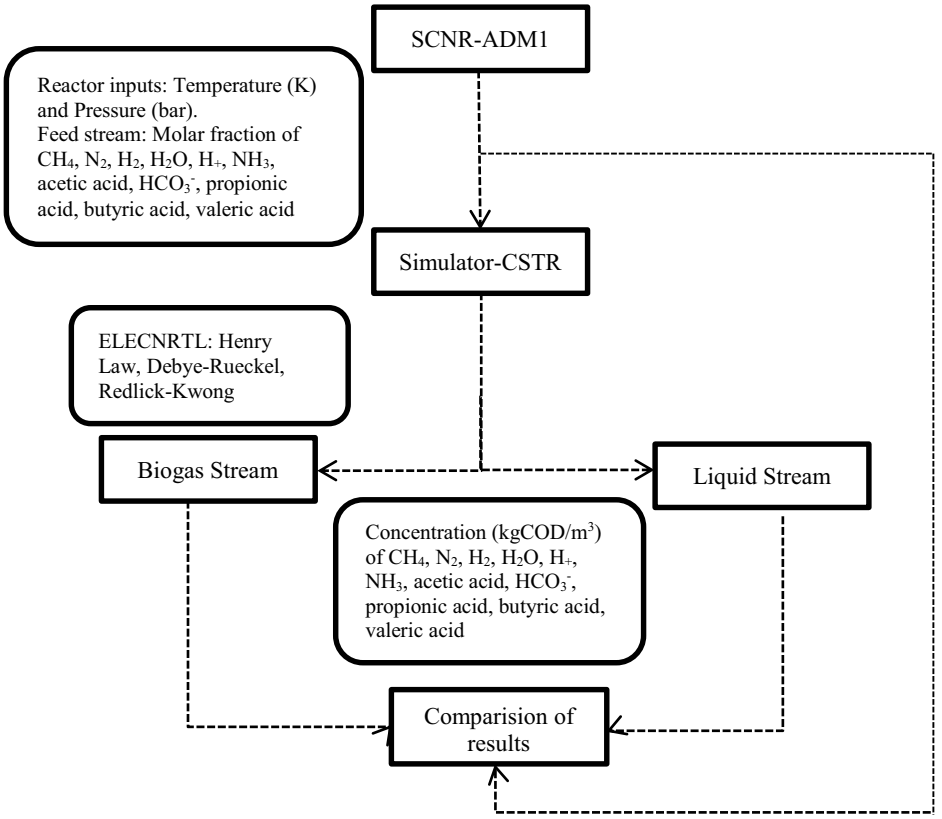


Figure 1- Representative scheme

4. Initial results

Figure 2 shows the prediction results using the proposed method for equilibrium. The concentration is given in kgCOD.m^{-3} and time is given in hours. In letter a, the predicted concentration of produced CH_4 from ADM 1 is shown. It may be noted that ADM 1 indicates a biogas production with a concentration of 0.1897 kg COD m^{-3}. When the thermodynamic equilibrium is represented by the simulation tool, a lower concentration is obtained that makes up the gaseous phase, 0.0953 kgCOD.m^{-3}, with a portion remaining in the liquid phase 0.0073 kg COD.m^{-3}. Letter b shows the behavior for Nitrogen, the final product of the denitrification process, occurring from the oxidation of ammonia to nitrate, and subsequent reduction from nitrate to nitrogen. It is interesting to note that nitrogen oxides have an inhibitory effect on methanogenesis, since there is a substrate competition between fermentative, methanogenic and denitrifying bacteria.

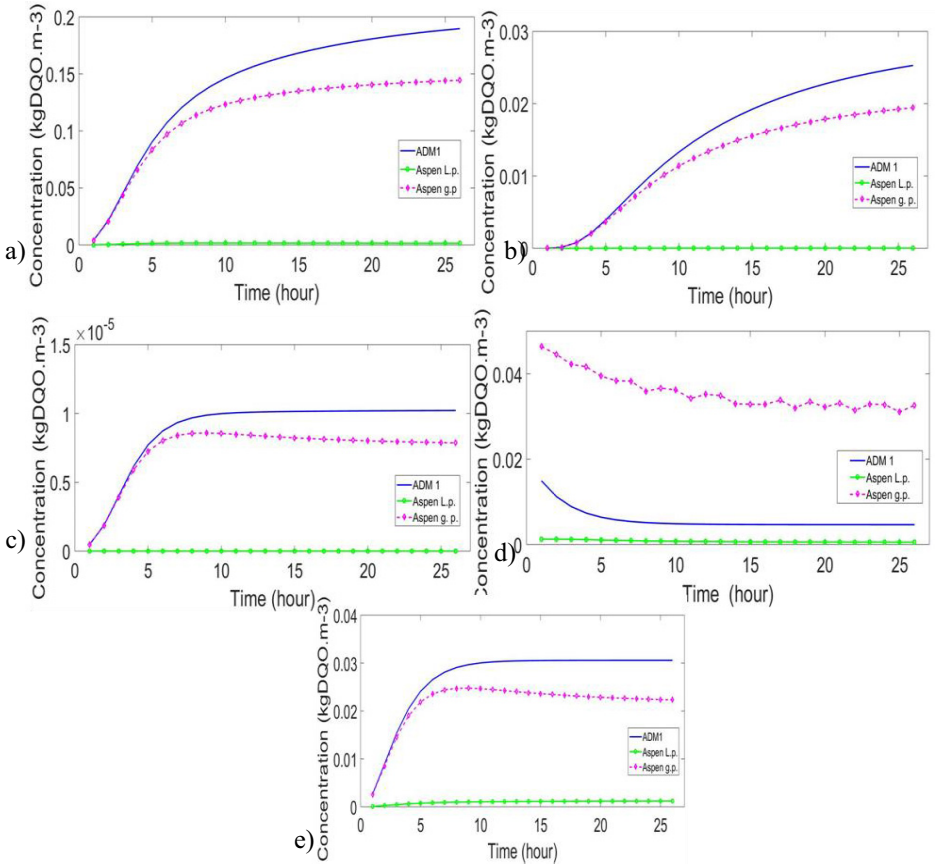


Figure 2- Results for prediction of concentration in each instant t: a) CH_4 b) N_2 c) H_2 d) CO_2 e) Acetic acid

By the figure, it is possible to notice that the behavior for gaseous phase remained the same, but the simulation with the thermodynamic model eNRTL once again found a lower gas concentration (0.0136 kgCOD.m^{-3}) and a small concentration in the liquid phase (1.1414e^{-4} kgCOD.m^{-3}). For Hydrogen, figure c, the same behavior as the previous compounds is observed. It is noted in the beginning of the process, the acidogenesis of the products resulted from hydrolysis, noticing a rapid increase in the hydrogen concentration. H_2 is also formed in acetogenesis from organic acids, along with CO_2 and acetic acid. The concentration becomes constant while the hydrogen is consumed by the homoacetogenic to form acetate and later by the methanogenic for the formation of CH_4, where the hydrogenotrophic bacteria will use it as energy source.

For CO_2, letter d, it is verified that the prediction using the thermodynamic model eNRTL did not progress with the same tendency of the ADM 1. Initially, this gas is formed and consumed in acetogenesis and after it is a product of methanogenesis from acetate, whereas it is consumed in methanogenesis from H_2. The trend simulated using the new thermodynamic model exhibits a worrying behavior since it confirms a

concentration higher than expected for the gaseous phase and a considerable concentration for the liquid phase. This could be due to numeric method or the influence of consider it as an important integral acid-base pair. This dilution in liquid phase can cause a drop in pH, since in water CO_2 is maintained as carbonic acid. With decreasing pH, CO_2 is dissolved as a neutral molecule, but with increasing pH it forms carbonic acid that ionizes. (Serrano, 2011).

The acetic acid, letter e, follows the same trend proposed by ADM1. Since it is one of the main components of the process, once after being formed in the acidogenesis it is used by the methanogenic bacteria to produce CH_4. Ammonia, propionic acid, valeric acid, butyric acid and carbonic acid had much lower concentrations than those predicted by ADM 1, besides the fact that they did not present behavior following the trend of the original model. Except for ammonia and carbonic acid, the acid-base reactions for volatile acids were not specified by the thermodynamic model because their ions are not present in the simulator databases and, for the implementation, a great number of information is necessary. In any case, they have a low dissociation constant and therefore, their higher or lower concentration does not strongly affect the result, and may have a negligible acid-base effect. (Serrano 2010)

Pauss et. al. (1990) considers the importance to considerate steady-state (as considered in the simulation), when no variation occurs in flow, temperature or pressure. In this case, normally, the concentration of dissolved gases is close to equilibrium values. About the highly soluble gases as CO_2, H_2S and NH_3, they are likely to be in thermodynamic equilibrium, even with low mass transfer coefficient. About the poorly soluble gases, as CH_4, N_2 and H_2, they are far from the equilibrium. In fact, the effluent here considered presents many compounds that affirm its non-ideality.

This discussion raises important considerations involving evidences cited for previous authors and possibly confirm that effluent may being eliminated before reaching the equilibrium, once high values of dissolved methane was measured. Even if this effluent goes to others steps of post treatment (as indicated) and if the methane is transferred from liquid to gaseous phase, it would be emitted to atmosphere. This investigation needs more steps to confirm how much this evidences and purposes is correlated.

5. Conclusion

This work allowed to visualize the influence of consideration about thermodynamic equilibrium in the model, using an electrolytic method. It is a significant approach, especially for CH_4, since its loss also characterizes energy loss of the system, as well as environmental damages. For further steps, new investigations must be carried out by adding all the acid-base reactions in the simulator, considering each functional group and investigation if in this case, methane is in supersaturation condition, as previously reported. The work ends as a finding that thermodynamic equilibrium does indeed provide a reliable representation of the results, indicating a deeper investigation is needed to validate this observation. Besides, it is an advance in ADM1 usage by improving its results with a rigorous thermodynamic approach, since it may allow a precise improvement in ADM1 values, ensuring a strict representation of this process.

References

AspenTech, 2013, Aspen physical property system, Physical property methods.

A. Pauss, G. Andre, M. Perrier, S. Guiot, 1990, Liquid-to-gas mass transfer in anaerobic processes: Inevitable transfer limitations of methane and hydrogen in the biomethanation process. Applied Environmental and Microbiology, 56, 6, 1636, (1990).

B. Crone, J. Garland, G. Sorial, L. Vane, 2016, Corrigendum to "Significance of dissolved methane in effluents of anaerobically treated low strength wastewater and potential for recovery as an energy product: A review", Water Research, 104, 520.

C. Chernicharo, 2007, Biological wastewater treatment series. Anaerobic reactores, 4, 188.

C. Mendes, K. Esquerre, L. Queiroz, 2016, Modeling simoutaneous carbon and nitrogen removal (SCNR) in anaerobic/anoxic reactor treating domestic wastewater, Journal of Enviromental Management,177, 119-128.

C. Souza, C. Chernicharo, S. Aquino, 2011, Quantification of dissolved methane in UASB reactors treating domestic wastewater under different operating conditions, Water Science &Technology, 64, 11, 2259-2264.

C. Souza, 2010, Estudo das rotas de formação, transporte e consumo dos gases metano e sulfeto de hidrogênio resultantes do tratamento de esgoto doméstico em reatores UASB, Tese, Universidade Federal e Minas Gerais, 147.

D. Batstone, J. Keller, I. Angelidaki, S. Kalyuzhnyi, S. Pavlostathis, A. Rozzi, W. Sanders, H. Siegrist, V. Vavilin, 2002, The IWA Anaerobic Digestion Model No 1 (ADM1), Water Science and Technology, 45, 10, 65.

D. Batstone, D. Puyol, X. Flores-Alsina, J. Rodríguez, 2015, Mathematical Modelling of Anaerobic Digestion Processes: Applications and Future Needs, Environmental Science and Biotechnology, 14, 595.

L. Lobato, 2011, Aproveitamento energético de biogás gerado em reatores uasb tratando esgoto doméstico, Tese, Universidade Federal de Minas Gerais, 184.

P. Biernacki, 2014, Model based sustainable production of biomethane, Tese, Universidade de Oldenburg, 196.

R. Serrano, 2011, Biogas Process Simulation using Aspen Plus, Final Master Thesis, Syddansk Universitet, 88.

S. Sandler, 2006, Chemical, Biochemical and Engineering Thermodynamics,4 ed.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50297-4

Optimal Configuration of Work-Heat Exchanger Networks (WHEN) in the Presence of Demand Response Objectives

Mats Rademacher[a,b], Yu Liu[a], Nael H. El-Farra[a] , Ahmet Palazoglu[a,*]

[a]*Department of Chemical Engineering, University of California, One Shields Avenue, 95616 Davis, USA*

[b]*Process Dynamics and Operations Group, TU Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany*

anpalazoglu@ucdavis.edu

Abstract

Demand responsive operation of chemical processes has recently become the focus of significant research. While various operational scheduling solutions have been proposed to realize demand response (DR) objectives, the possibility of plant flowsheet reconfiguration in real-time has rarely been addressed. In this work, we compare the different deterministic design solutions of a work-heat exchanger network under varying electricity prices. We present a heuristic design that allows for reconfigurability during DR events using a superstructure formulation to conceptualize a DR-enabled process design.

Keywords: Demand Response, Work-heat Exchange Network, MINLP, Flowsheet Reconfiguration.

1. Introduction

The fluctuation of electricity prices, caused by changing demand and generation profiles during a specific period, is already playing a significant role in industrial energy management strategies, and this is likely to continue with the growing share of renewables in future energy systems. The challenge of decreasing the impact of volatile electricity price profiles evoked concepts like demand response (DR), where electricity consumption is adjusted depending on its price during a specific period. While a higher electricity price encourages consumers to lower their demand, low prices are an incentive to increase it. As a result, fluctuating electricity price profiles are smoothed benefiting both the energy providers and the consumers.

DR has been discussed recently in the context of chemical process systems (e.g., Chmielewski 2014; Harijunkoski et al. 2014). Primarily, the focus has been on the scheduling of process operations, typically by increasing or decreasing production rates during the day or by shifting production load among specific units (Kwag and Kim, 2012; Mitra et al., 2012; Wang and Li, 2013; Wang et al., 2014). Wang et al. (2015) also presented a proactive approach for operational reconfiguration of heat exchanger networks (HENs) under varying electricity prices and renewable energy resource availability to satisfy DR objectives. Compared with operational DR solutions, the

design of DR-enabled chemical processes has received limited attention, and has focused mainly on air separation units (Pattison and Baldea, 2014), where variable capacity operations are accounted for during the process design step. The design of a process network, allowing varied operating levels as well as enabling the optimal process flowsheet reconfiguration under different DR scenarios is a novel concept. Intuitively, reconfiguration suggests the purchase of more (redundant) process equipment which generally results in a higher capital investment. Whether the economic benefits from DR operation can compensate for the higher investment and thereafter reduce the overall cost remains an open research question.

In this work, we propose a heuristic study of an existing work-heat exchanger network (WHEN) model (Huang and Karimi, 2016) to demonstrate the concept of a reconfigurable flowsheet subject to DR objectives. We first obtain the optimal design solution under different electricity prices; then an ad-hoc method is introduced to explore the generation of a design that can switch flowsheet configurations to accommodate DR objectives.

2. WHEN Modeling and Optimization

While air separation processes are significant energy consumers, several other processes have also been identified as potential candidates due to their high electricity consumption, including pulp and paper, aluminum, chlor-alkali, cement, and steel (Paulus and Borggrefe, 2011). Another promising technology is the combination of work exchanger networks (WENs) with heat-exchanger networks (HENs), which also utilize energy-intensive equipment. Similar to HENs in which heat-exchangers are used to have process stream meet temperature targets, a WEN is equipped with single-shaft-turbine-compressor (SSTC) units for transferring work from high- to low-pressure process streams. Razib et al. (2012) first introduced a superstructure-based mixed-integer nonlinear programming problem (MINLP) model to include the SSTC into the work recovery. Based on their study, Onishi et al. (2014) further developed a MINLP formulation to include heat integration. However, their model includes highly nonlinear equipment cost functions and a number of big-M constraints. Huang and Karimi (2016) proposed an alternative formulation to address these issues, yielding a more efficient model for WHEN synthesis.

The WHEN model consists of three parts, including the set of equations describing the HEN and WEN modules and their connections. In this work, the WEN model is taken from Huang and Karimi (2016) and the HEN model is based on a formulation proposed by Yee and Grossmann (1990). The connection of HEN and WEN is achieved by considering the outlet temperature of a HEN stage as the inlet temperature of the subsequent WEN stage, while the pressure drop in the HEN is neglected. The objective is to minimize the total annualized cost (TAC), including the capital (CAPEX) and operational (OPEX) costs.

The WHEN model results in a MINLP problem. Its highly nonlinear characteristics preclude the guarantee of obtaining a global optimum, and in this work, a heuristic, ad-hoc method was used for finding the local optima with GAMS/SBB. Various scenarios are developed and following the adjustment of the initialization and electricity prices, the optimizations are run consecutively. While other solvers, i.e. DICOPT, are also useable for MINLPs, SBB showed the best performance in terms of computation time and obtaining feasible results.

3. Case Study

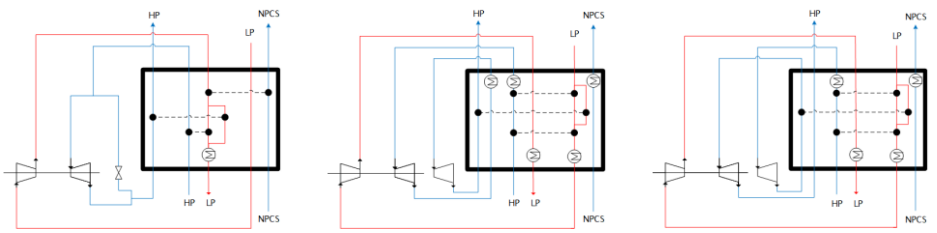
3.1 Process description

To test the WHEN model formulation and confirm the scenario assumptions, the data of example 3 in Huang and Karimi (2016) are used for evaluating a LNG process. This case study deals with the work and heat integration between two gaseous and one liquid process streams. For the liquefaction of natural gas (NG), the feed stream has to be compressed and thus is considered as a low-pressure (LP) stream in the WEN, while it is regarded as a hot stream in the HEN since it is more energy efficient to compress cold streams. In contrast, the liquid nitrogen used to cool down NG is supposed to be a high-pressure (HP) and cold stream. Liquid CO_2, a side-stream of the LNG process, is considered to be a no-pressure-change stream (NPCS) to be heated, and is not involved in the WEN. All process data are taken from (Huang and Karimi, 2016), while this paper focuses on a varying electricity price structure to demonstrate how DR objectives are met. Furthermore, since selling back electricity to the grid is allowed and would generate a revenue, the electricity selling price is assumed to be 50 % of the purchase price. For evaluating off- / mid- / and on-peak scenarios, three different electricity price levels are assumed. As time of use (TOU) electricity price, the following distribution was assumed: \$0.07/kWh for 11 h/d (midnight to 11 am), \$0.15/kWh for 7 h/d (11 am to 6 pm) and \$0.30/kWh for 6 h/d (6 pm to midnight).

4. Results and Discussion

First, the MINLP problem is solved using the TOU electricity price structure. As a result, the optimal configurations for the different periods are obtained as shown in Table 1. While an off-peak electricity price leads to the use of a valve for expanding HP, a utility turbine is chosen for higher prices. Since the utility turbine is assumed to be much more expensive than a valve, its use has to be economically compensated by the generated revenue, which is the case when the electricity selling price is high.

Although the configurations for mid- and on-peak process look similar (Figure 1), their loads and process stream connections are slightly different. Specifically, in the WEN the utility and SSTC turbine are in reversed order, and in the HEN the number of heaters is higher in the mid-peak scenario.



(a) (b) (c)

Figure 1: Optimal configurations during (a) off-peak, (b) mid-peak and (c) on-peak time periods.

Table 1: Stream loads for the optimal configurations in Figure 1.

Equipment	Stream(s), Stage	Load [kW]		
		Off-Peak	Mid-Peak	On-Peak
SSTC turbine	HP1, s1	216.1	71.2	-
	HP1, s2	-	-	71.2
SSTC compressor	LP1, s1	216.1	71.2	71.2
Valve	HP1,s1	0*	-	-
Utility turbine	HP1, s1	-	-	621.9
	HP1, s2	-	629.1	-
Heat exchanger	LP1, HP1, s1	-	338.1	238.0
	LP1, HP1, s2	-	46.6	80.9
	LP1, HP2, s1	-	-	353.0
	LP1, HP3, s1	-	288.9	-
	LP3, HP1, s2	211.4	-	-
	LP3, HP2, s2	363.8	-	-
	LP3, HP4, s1	131.6	-	-
Heater	HP1	-	190.2	364.2
	HP2	-	179.4	-
	HP4	-	131.6	131.6
Cooler	LP1	-	118.3	120.0
	LP2	-	-	31.3
	LP3	261.2	31.3	-
TAC [$/y]		124,699	116,842	-251,328

*0: Used but no load; - : not used

To gain further insight into the DR behavior of the WHEN, we consider and evaluate the following scenarios:

1. Use one of the optimal configurations (Fig. 1) throughout the entire day, recognizing that each configuration would be suboptimal during the other two periods.
2. Commit to the purchase of all equipment under all three configurations and allow for the reconfiguration of the flowsheet among the available equipment, switching configurations as electricity price changes. From the CAPEX standpoint, this is the worst case scenario. We will refer to this as the WHEN superstructure.
3. Create a WHEN *suprastructure* that modifies each optimal configuration in an ad-hoc fashion so that the modified configurations have more common equipment. When all such equipment are combined, one still has a reconfigurable flowsheet but much less capital commitment than the WHEN superstructure. This comes at the expense of likely increased OPEX during the realization of various electricity price levels.

For evaluating the total cost of using a single configuration for the entire time (on a daily basis), the binary variables of the WHEN are fixed as suggested by the solutions and run under the other two electricity purchasing and selling prices.

The WHEN suprastructure is achieved by fixing the HEN equipment, while the binary variables of the WEN remain as decision variables. This scenario was used since a single WEN configuration under a varying electricity price showed high TAC, so one HEN was assumed to lower the cost. Though a slight decrease of the cost was achieved (see Table 2), it could be different for other electricity price profiles due to the heuristic nature of this approach. Furthermore, although only the HEN equipment was fixed under different electricity prices, the WEN configuration is almost the same in mid- and on-peak periods.

The results of the above scenarios are summarized in Table 2. One can see that when using a single configuration for the entire day, the mid-peak configuration has lowest cost, while the WHEN suprastructure is slightly cheaper compared to the WHEN superstructure. Although the CAPEX and OPEX for the off-peak configuration are the lowest, the total cost is higher than the mid-peak and on-peak configurations since no revenue is generated due to the low electricity selling price.

Table 2: Daily cost of all configurations.

Cost [$/d]	Off-Peak Configu-ration	Mid-Peak Configu-ration	On-Peak Configu-ration	WHEN Superstructure	WHEN Suprastructure
CAPEX	350	1,040	1,041	1,066	1,037
OPEX	6	347	347	295	269
Revenue	0	1,069	1,063	887	836
Total	356	318	325	474	470

Table 3: Stream loads of WHEN suprastructure.

Equipment	Stream(s), Stage	Load [kW]		
		Off-Peak	Mid-Peak	On-Peak
SSTC turbine	HP1, s1	125.7	71.2	71.2
SSTC compressor	LP1, s1	125.7	71.2	71.2
Valve	HP1,s1	0	-	-
Utility turbine	HP1, s1	-	584.4	584.4
Heat exchanger	LP1, HP1, s1	-	356.2	356.2
	LP1, HP1, s2	130.4	21.8	21.8
	LP1, HP2, s1	363.8	315.5	315.5
Heater	HP1	-	305.1	305.1
	HP4	131.6	131.6	131.6
Cooler	LP1	-	98.4	98.4
	LP2	383.4	31.3	31.3

4. Conclusions and Future Work

In this study, the concept of a reconfigurable design for process flowsheets under varying electricity profiles was introduced and demonstrated using a WHEN case study. The preliminary results show that different electricity purchasing and selling prices highly influence the WHEN configuration while the use of the single mid-peak configuration for the entire time yields the lowest TAC. We note that the current results are ad-hoc solutions, and no global optima is presumed to be obtained, meaning that under other assumptions the WHEN superstructure could have lower TAC. Further research needs to be conducted, focusing on other electricity price profiles, cost data and scenarios using a broader variety of process streams. While the MINLP problem of work-heat integration between two gaseous and one liquid process stream is a relatively simple example, a higher number of streams would result in more complex models, which are difficult to solve. More importantly, a rigorous solution of the optimal network reconfiguration needs to be formulated.

References

D. J. Chmielewski, 2014, Smart grid: the basics – What? Why? Who? How?, *Chem. Eng. Prog.*, 110:28–33.

I. Harjunkoski, E. Scholtz, and X. Feng, 2014, Industry meets the smart grid, *Chem. Eng. Prog*., 110:45–50.

K. Huang and I. A. Karimi, 2015, Work-heat exchanger network synthesis (WHENS), *Energy*, 113, 1006–1017.

H. G. Kwag and J. O. Kim, 2012, Optimal combined scheduling of generation and demand response with demand resource constraints, *Appl. Energy*, 96:161–170.

B. Linnhoff and J. R. Flower, 1978, Synthesis of heat exchanger networks: I. Systematic generation of energy optimal networks, *AIChE J.*, 24(4), 633–642.

S. Mitra, I. E. Grossmann, M. Pinto, and N. Arora, 2012, Optimal production planning under time-sensitive electricity prices for continuous power-intensive processes, *Comp. Chem. Eng.*, 38:171–184.

V. C. Onishi, M. A. S. S. Ravagnani, and J. A. Caballero, 2014, Simultaneous synthesis of work exchange networks with heat integration, *Chem. Eng.*, 112, 87–107.

M. S. Razib, M. Hasan, and I. A. Karimi, 2012, Preliminary synthesis of work exchange networks, *Comp. Chem. Eng.*, 37, 262-277.

X. Wang, N. H. El-Farra, and A. Palazoglu, 2015, Proactive Reconfiguration of Heat-Exchanger Supernetworks, *Ind. Eng. Chem. Res.*, 54(37), 9178–9190.

X. Wang, H. Teichgraber, A. Palazoglu, and N.H. El-Farra, 2014, An economic receding horizon optimization approach for energy management in the chlor-alkali process with hybrid renewable energy generation, *J. Proc. Contr.*, 24:1318–1327.

Y. Wang and L. Li, 2013, Time-of-use based electricity demand response for sustainable manufacturing systems, *Energy*, 63:233–244.

T. F. Yee and I. E. Grossmann, 1990, Simultaneous optimization models for heat integration—II. Heat exchanger network synthesis, *Comp. Chem. Eng.*, 14(10), 1165–1184.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50298-6

Open Sugarcane Process Simulation Platform

Rubens E. N. de Castro[a], Rita M. B. Alves[a], Adam Hawkes[b], Claudio A. O. Nascimento[a*]

[a] *Department of Chemical Engineering, Escola Politécnica, Universidade de São Paulo, Av. Prof Luciano Gualberto 380, tr 3, São Paulo, SP, 05508-900, Brazil*
[b] *Department of Chemical Engineering, Imperial College London, South Kensington Campus, London SW72AZ, UK*
oller@usp.br

Abstract

Software tools for process simulation and optimization have increasingly been used in industry to design and operate complex, highly interconnected plants. This allows the design of industrial plants to be profitable and to meet quality, safety, environmental and other standards. The aim of this work is to create a platform to simulate industrial sugarcane first generation process. Brazilian sugarcane industry is a well known process with many parameters available from industrial and literature data. The current first generation process seems to have reached the state of art and not great improvements seams to emerge nowadays. However engineering research has dedicated great efforts recently to improve efficiency through the use of industrial and agricultural residues. Most of these studies are related to the use of lignocellulosic material and vinasse. In this context an easy and simple platform has been developed to provide reliable outputs that could provide data for the studies of viability, social and environmental impacts when an additional process are interconnected to the first generation plant. The model has been validated using industrial data in order to attain the most realistic outputs.

Keywords: Sugarcane, biofuel, biogas, process simulation.

1. Introduction

In Brazilian sugarcane industries, several studies have been carried out as an attempt to use agricultural residues like straw (Dias, et al., 2012) or industrial residues like vinasse (Longati, et al., 2017) and bagasse (Furlan, et al., 2012) to sum up with new co-products or improve the whole sector more efficiently (Junqueira, et al., 2011) These new processes can be understood as an interconnected plant to the current first generation process where materials and utilities are shared. (Fonseca, et al., 2017). Indeed, there is a synergetic interaction between the new and the main process which could result in better overall energy balance. (Furlan, et al., 2012) The objective of this work is to develop a model for quantitative simulation of a feasible ethanol and sugar conversion process that uses these residues as raw materials and is interconnected to the first generation process.

2. Process Overview

The development of this computational platform follows the process configuration presented in Fig 1, which is a typical sugarcane mill. Each step represents an area at the sugar mill industry and each area has a number of unit operation processes. In Fig 1, "cane reception and extraction" is the process where the sugarcane feed is reduced into

small pieces and the sugar bearing cells is ruptured to facilitate the subsequent extraction process. The size reduction is generally achieved through the use of rotating knives and swing hammer shredder in the cane conveying system. In the case of billeted cane, mechanically harvested, it can be fed directly into a shredder without any knifing. For sugarcane juice extraction, two systems are used: milling tandem or diffuser. In Brazil, the preference has been for the mill extraction process which consists of a set of four to six mills. This step produces two streams: bagasse and juice.

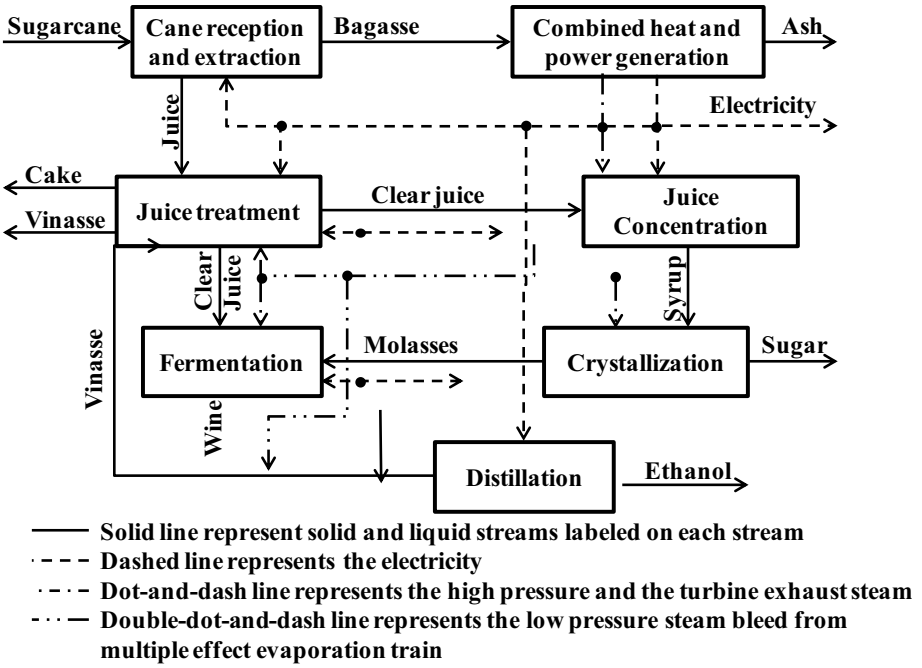


Figure 1 - Schematic flowchart of a sugarcane industry

Juice after mill contains several impurities which must be removed prior to fermentation or concentration. This process takes place at "juice treatment" area where these impurities are removed using a set of unit operations which basically consist in heating, pH adjust, settling the precipitate formed in the body juice and filtering.

Clear juice must be concentrated to achieve a suitable brix to start the crystallization process, about 60 % (w/w) sucrose. This concentration is obtained using multiple effect evaporators, which use the exhaust steam from the turbines and produce lower pressure steam used as utility in the process.

Concentrated juice, called syrup, is further concentrated in a vacuum boiling pan until it becomes supersaturated. Finely ground sugar crystals suspended in alcohol are introduced into the vacuum pan as seed crystals around which sucrose is deposited. These crystals then grow until they are ready to be discharged. Maximum recovery of sucrose by crystallization cannot be achieved in one step. Various crystallization schemes can be employed. In Brazil, two-boiling scheme is the most used, because it is not necessary to completely exhaust molasses since ethanol is produced using it.

Must for alcoholic fermentation is a mixture containing molasses and clear juice. This mixture uses a biological process that converts sugars such as glucose, fructose, and

sucrose into cellular energy, producing ethanol and carbon dioxide as a side-effect. The overall chemical reaction for alcoholic fermentation is:

$$C_{12}H_{22}O_{11} + H_2O + enzyme \rightarrow 2\ C_6H_{12}O_6 \rightarrow 4\ C_2H_5OH + 4\ CO_2$$

In the first step of alcoholic fermentation, the enzyme invertase cleaves the glycolic linkage between the glucose and fructose molecules. This process usually takes place in a fed batch fermentation process. When fermentation finishes the fermented wine is centrifuged to remove yeast (*Saccharomyces cerevisiae*) which is recycled to the next fed batch while wine with ethanol content about 8 ºGL is sent to next process.

Ethanol is recovered from wine in the distillation process. Fuel ethanol or hydrous ethanol 93 wt% is produced by a simple distillation process while anhydrous ethanol 99.3 wt% for mixing with gasoline is produced by azeotropic distillation with cyclohexane or monoethyleneglycol or by absorption on molecular sieves. Vinasse rich in nutrients is used in fertirrigation of the sugarcane fields.

Sugarcane bagasse is burnt in the boiler, producing steam that is expanded in turbines coupled with electric generators. Steam exhaust from the turbines is used as thermal energy at the evaporation area. The surplus steam (not consumed in the process) is used in condensing steam turbines to maximize the electricity production. Therefore sugarcane mill is self-sufficient in electricity and can sell surplus to the grid.

3. Model

The model-based simulation framework involved two main parts: (1) mass and energy balances; (2) constitutive equations.. The in-house model was implemented in MatLab®.

Sugarcane, biomass feedstock for the first generation process, can be represented by 3 main parts: (1) the insoluble part includes fibrous and other insoluble materials like soil, which in this text is called 'Fibre'; (2) water which is about 70 % of the total sugarcane; and (3) soluble solids like sucrose, other sugars, some proteins, some salts, etc., which is usually called 'Brix'. Consequently mass balance is applied for the three parts (Eqs 1 to 3).

$$\sum_{i=1}^{n} Fibre_{in} - \sum_{i=1}^{n} Fibre_{out} = 0 \quad (1)$$

$$\sum_{i=1}^{n} Brix_{in} - \sum_{i=1}^{n} Brix_{out} = 0 \quad (2)$$

$$\sum_{i=1}^{n} Flow_{in} - \sum_{i=1}^{n} Flow_{out} = 0 \quad (3)$$

For the energy balance, Eq. 4, thermodynamic properties of water were obtained adjoining packages from IPAWS. Specific heat of juice is given by Eq. 5 (Hugot, 1986) and specific heat of fibre is given by Eq. 6 adapted from Hatakeyama et al (1982)

The parameters for the constitutive equations come from industrial or literature data, or from equipment modelling (Rein, 2007). This equation and its parameters depend on the type of operation, for example, overall coefficient and area is the needed parameters for

heat exchangers; mud concentration is the parameter for the juice clarification process; and so one.

$$\sum_{i-stream}^{n-stream} (Flow_i \cdot cp_i \cdot T_i)_{in} - \sum_{i-stream}^{n-stream} (Flow_i \cdot cp_i \cdot T_i)_{out} = 0 \quad (4)$$

$$cp_j = \left(1 - \left(0.6 - 0.0018 \cdot T + 0.08 \cdot (1 - P)\right) \cdot Brix\right) \cdot 4.187 \quad (5)$$

$$cp_b = 1.364 + 5.06 \cdot 10^{-3} \cdot (T - 76.85) \quad (6)$$

The model is then comprised by three independent equations from mass balance are obtained: Fibre balance, Brix balance and overall balance or water balance and the energy balance. Consequently, the number of parameters and constitutive equations depends on the number of unknown variables. Usually feed stream is known and the outflow is unknown, consequently, the number of parametric equation needed will depend on the number of streams out of the 'unit operation'.

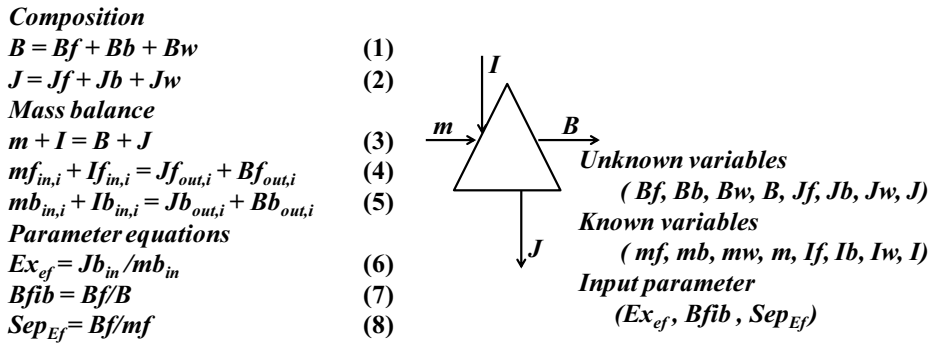


Figure 2 - Schematic flowchart of a sugarcane industry

Several configurations have been proposed (Rein, 2007). This platform aim to allow the user to customize the configuration in many ways one requires. Consequently this model deals with each unit operation as a separate algorithm. Fig 2 shows an example of one unit operation.

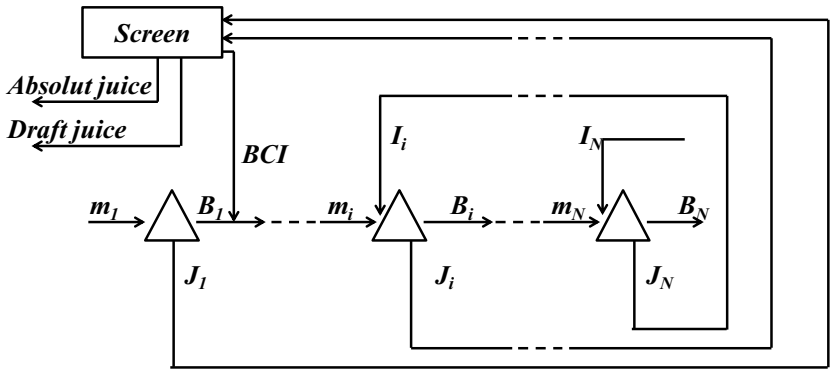


Figure 3 - Schematic flowchart of a sugarcane industry

Another script is used to call a given unit operations by identifying recycling loops to solve each area individually as shown in Fig 3. The result from one unit operation becomes the input for the next unit operation.

Flowsheet analysis or simulation is complicated because the mass and energy balances are tightly coupled between different unit operations and consequently between areas. This problem were stated in a standard form of a "Direct Substitution Method", which uses a fixed point relation $x^{k+1} = g(x^k)$ (Biegler, et al., 1999, p. 268), where x is for instance the stream flow rate and $g(x)$ is the corresponding algorithm to calculate this stream and x_0 is the initial guess. The convergence of this method depends on the relation:

$$|\lambda| = \frac{|x^{k+1} - x^k|}{|x^k - x^{k-1}|} \tag{7}$$

Where $|\lambda|^{max}$ is the Euclidean norm and a necessary and sufficient condition for convergence is that $|\lambda|^{max} < 1$. The fixed point methods developed for recycle convergence are strongly influenced by the structure of the flowsheet and the choice of the tear streams. The algorithm proposed by Biegler et al. (1999) was used to find the partitions and precedence ordering of resolution.

4. Results

Fig 4 shows the synergetic effect of producing sugar and ethanol simultaneously and Table 1 shows specifications and operating conditions used in this simulation. When both sugar and ethanol are produced, the steam consumption in first generation process is reduced while surplus power generation increase against industrial mix. Industrial mix is the step to decide which product is going to be produced using the sucrose from juice: ethanol or sugar. In this configuration, steam form the first effect of the multiple effect evaporation train is bleed and directly injected in a distillation column. It is observed that when at about 60 % of cane is used to produce sugar, the steam consumption is the lowest and consequently the energy production is the highest. This occurs because great amount of water has to be taken off the sugar juice in order to concentrate it for the crystallization process and the water withdrawn as steam from the sugar juice is used in the distillation process. For higher sugar production, the steam used to concentrate the juice is predominant while for a lower sugar production

Table 1 - Specifications and operations conditions used in this simulation

Sugarcane	1000 t/h
Fibre % Cane	13.0%
Sucrose wt%	15.0%
Bagasse	265 t/h
Extraction efficiency	97.7
Boiler steam pressure and temperature	67 bar / 520ºC
Boiler efficiency	79%
Back pressures turbine overall efficiency	82%
Condensation turbine overall efficiency	73%

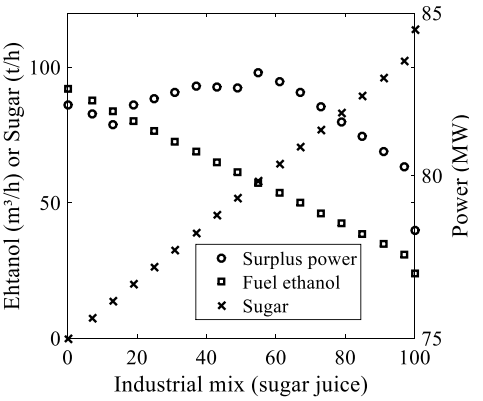


Figure 4- Ethanol Sugar and Power production

distillation it is major responsible for the overall heat consumption and it is necessary to concentrate ethanol juice using one effect evaporator which increases the overall steam consumption when compared to multiple effect evaporation.

It is important to say that, at this scenario, even when the whole sugarcane juice stream is sent to the sugar process, ethanol is still being produced using molasses. Nevertheless the behaviour showed at Fig 4 is valid for this configuration of equipments and operation streams. Consequently, changing the configuration would change the result even keeping the same parameters and specifications or vice-versa.

5. Conclusion

A typical sugar and ethanol process has been presented. Assumptions and results of the simulation have been carefully validated using industrial data to reflect the reality. Consequently a platform simulation has been developed to provide reliable outputs that could provide data for studies of technical and economic viability and social and environmental impacts when any new process is interconnected to the first generation plant, by sharing materials, utilities and products. Research will continue using agricultural residues (straw) or industrial residues (vinasse and bagasse) to sum up with new co-products or improve the whole sector more efficiently.

6. Acknowledgements

The authors kindly acknowledge FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) process number 2015/50684-9 for its financial support.

7. References

L. Biegler, I. Grossmann, A. Westerberg, 1999. *Systematic Methods of Chemical Process Design.* New Jersey: Prentice Hall PTR.

Dias, et al., 2012. Evaluation of different cogeneration systems in first and second generation ethanol production from sugarcane. *Computer Aided Chemical Engineering*, pp. 172-176.

G. Fonseca, C. Costa, A. J. Cruz, 2017. Superstructural economic optimization of sugarcane bagasse exploitation in an ethanol distillery connected to Rankine cycle, BIGCC system and second generation ethanol process. *Computer Aided Chemical Engineering*, pp. 889-894.

F. Furlan, et al., 2012. Assessing the production of first and second generation bioethanol from sugarcane through the integration of global optimization and process detailed modeling. *Computer and Chemical Engineering*, pp. 1 - 9.

T. Hatakeyama, K. Nakamura, H. Hatakeyama, 1982. Studies on heat capacity of cellulose and lignin by differential scanning calorimetry. *Polymer*, pp. 1801 - 1804.

E. Hugot, 1986. *Handbook of cane sugar engineering.* Amsterdam - Oxford - New York - Tokyo: Elsevier.

T. Junqueira, et al., 2011. Simulation and evaluation of autonomous and annexed sugarcane distilleries. *Chemical Engineering Transactions*, pp. 941-946.

A. Longati, O. Cavalett, A. Cruz, 2017. Life Cycle Assessment of vinasse biogas production in sugarcane biorefineries. *Computer Aided Chemical Engineering*, pp. 2017 - 2022.

P. Rein, 2007. *Cane sugar engineering.* Berlin, Germany: Verlag Dr. Albert Bartens.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50299-8

Design of energy-saving CO_2 capture process using circulating fluidized bed for power generation

Yasuki Kansha[a*], Masanori Ishizuka[a], Atsushi Tsutsumi[a]

[a]*Collaborative Research Center for Energy Engineering, Institute of Industrial Science, The University of Tokyo, 4-6-1, Komaba, Meguro-ku Tokyo 153-8505, Japan*
kansha@iis.u-tokyo.ac.jp

Abstract

In this research an innovative energy-saving process using chemical absorption by a metal solid in a circulating fluidized bed has been proposed. To recirculate of all process heat, the reaction heat that accompanies absorption is successfully supplied through a heat pump to achieve thermal decomposition in a regenerator, leading to large energy saving of the process. Absorption and fluidization experiments were conducted to determine the process conditions and requirements of the proposed process. An energy balance of the proposed process was simulated using the experimental data by a commercial process simulator. The simulation results indicate that the proposed CO_2 separation process has large energy-saving potential as compared with conventional counterparts.

Keywords: CO_2 capture, Energy saving, Process design

1. Introduction

Carbon capture and storage (CCS) is considered to be the most significant strategy to control the greenhouse gas effect and mitigate climate change. Among CCS, CO_2 separation is the most energy-intensive component: 75%–80% of the total cost of CCS is consumed for CO_2 separation (Goto et al., 2009). It is therefore necessary to reduce the energy consumption of the CO_2 separation process to promote development of CCS technology (Figueroa et al., 2008).

Mainly three CO_2 capture ways are applicable to the power plants, namely post-combustion, pre-combustion and oxy-fuel combustion. The most widely used for CO_2 separation among these three ways is chemical absorption using amine-based solutions, because of the reversible and continuous nature of the chemical reaction (Ali et al., 2016). In fact, an aqueous alkaline solution has been commercially used for absorption of acidic or sour gases (Arias et al., 2016). In the conventional absorption process using amine solutions, thermal energy is supplied from the bottom of the regenerator, which utilizes reaction heat for regeneration of the solutions and heat of vaporization for stripping.

Concurrently, the authors have developed an energy-saving process design: that of self-heat recuperation (Kansha et al., 2009). By applying self-heat recuperation to thermal processes, whole process heat can be circulated in the process with equivalent work of minimum exergy destruction of heat transfer (Kansha et al., 2013). Following this design, CO_2 gas separation processes based on self-heat recuperation for both post- and pre-combustion power generation systems have been developed (Kishimoto et al., 2011, 2012). In these processes, all of the process heat circulates in the process by using heat

pump, resulting in large energy savings. Although this technology contributes large energy savings, absorption using aqueous solutions requires additional energy for vaporization for the stripping step. To avoid the need to supply vaporization heat, some researchers have investigated the use of solid absorbents. However, the reaction temperatures of these solid absorbents are relatively high to apply heat-pump systems (Kato et al., 2002). It is therefore difficult to recirculate all process heat. In fact, many of these solid absorbents are used for chemical looping combustions integrated into the power generation (Cormos et al., 2015).

In this research, a new energy-saving process using chemical absorption by metal oxide in a circulating fluidized bed with reaction heat circulation by a heat pump is proposed. The energy-saving potential was evaluated by simulation based on experimental data.

2. Proposed CO_2 Separation Process

Figure 1 shows the conceptual design of the proposed process using chemical absorption by metal oxide particles (Kansha et al. 2017). The feed mixture containing CO_2 and other gases is fed as stream 1 and supplied to an absorber through a blower (stream 2). This absorber is a fluidized bed, in which the fluidized particles are selected zinc oxide (ZnO) fluidized by the feed gas in consideration of the reaction temperatures. CO_2 reacts with the ZnO particles to zinc carbonate ($ZnCO_3$) particles. This reaction heat is transferred to a heat-pump system where the temperature of the heat is elevated. The unreacted gases are exhausted from the absorber as stream 3. $ZnCO_3$ particles from the absorber are fed to a regenerator with a carrier gas (stream 4). In the regenerator, the reverse reaction, thermal decomposition ($ZnCO_3 \rightarrow CO_2 + ZnO$), takes place by supplying absorption heat through the heat-pump system. ZnO particles are resupplied to the absorber as stream 5. CO_2 gas is separated from the particles (stream 6 → stream 7). Stream 7 is the separated CO_2 gas. A portion of the CO_2 produced is supplied to the moving-bed regenerator through a blower to reduce friction between the particles and wall (stream 8 → stream 9). Therefore, this process does not require the stripping step and all process heat including reaction heat can be recirculated by heat pump and fluidized particles.

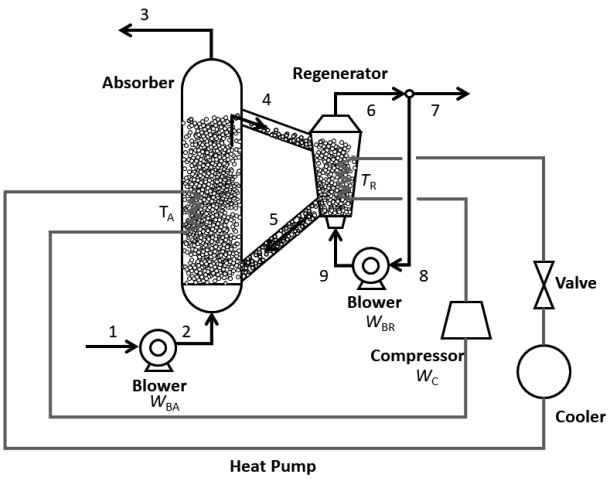


Fig. 1. Flow diagram of the proposed process

3. Experimental Test work and Results

To determine suitable reaction conditions for the proposed process, the CO_2 absorption reaction experiment was conducted using experimental setup shown in Figure 2. ZnO powder (Wako Pure Chemical JIS Special Grade, Wako, 5 g) was pre-treated by heating at 250 °C in a test tube under pure nitrogen (N_2) flow to ensure the desorption of any CO_2 from the atmosphere. The heater was set to the selected reaction temperature under pure N_2 flow. After stabilizing the temperature, the CO_2 flow valve was opened and CO_2 was mixed with the N_2 flow. The feed compositions were ensured by means of flow controllers (FCST 1005LC-4F2-F10, Fujikin). The exhaust gas composition was analysed by Fourier Transform infrared spectroscopy (Nicolet™ iS™ 5 FT-IR Spectrometer, Thermo Fisher Scientific) each minute.

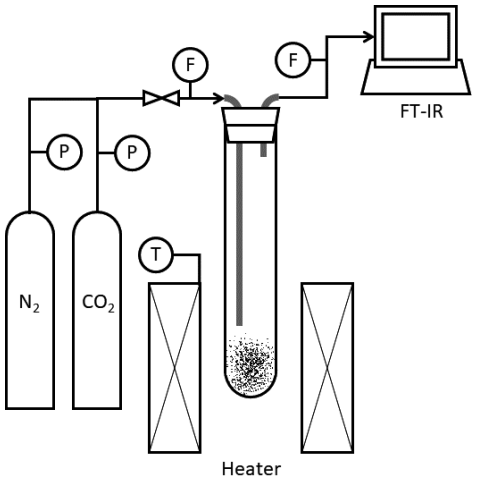


Fig. 2 Experimental setup for absorption reaction

Figure 3 shows one of the results of this experiment to see the extent of reaction for absorption as a function of time at the different reaction temperatures (80 °C, 100 °C, 120 °C), using the heater to ensure isothermal reaction conditions. The N_2 flow rate was set to 10 mL/min and the CO_2 flow rate was set to 4.0 mL/min (29%). It can be seen from Figure 3 that these reaction temperature changes did not significantly affect the reaction rate and all of reaction was completed within 25 mins. It is well known that the conversion ratio appeared to be affected according to Le Chatelier's principle. This absorption reaction is exothermic reaction. Thus, it was expected to reduce the reaction rates with temperature increase. However, the effect was too small to accurately observe in this experiment. It can be said from these results that the absorption reaction temperature can be set lower than 120 °C. By the way, only less than 2.5% of the ZnO was converted to $ZnCO_3$ due to the limited contact area between ZnO powder and CO_2 gas in the test tube. It was therefore necessary to expand the contact area in order to absorb CO_2 effectively.

From these experiments, it appeared appropriate to set the reaction temperature of the absorber to 120 °C; however, it was necessary to increase the contact area of ZnO with CO_2 to increase the conversion ratio of absorption. In addition, we considered that the rate of thermal decomposition in the regenerator was relatively fast compared with that of the absorption reaction.

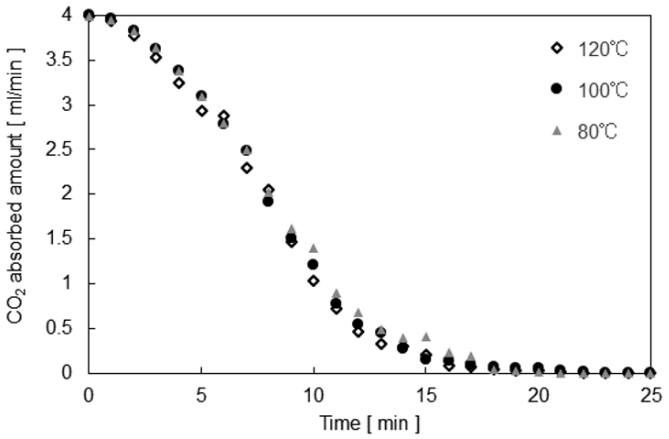


Fig. 3 Rate of CO_2 absorption as a function of time for different reaction temperature

A fluidized bed was therefore selected for the absorber and a moving bed for the regenerator. Although a fluidized bed requires additional energy for the blower that supplies the fluidizing gas, it is well known that heat transfer inside the bed is enhanced and a uniform bed temperature is achieved, leading to good heat transfer from the absorption reaction to the heat-pump system.

To investigate the additional energy required for the blower (W_{BA}), the minimum fluidizing gas velocity was determined using an experimental fluidized bed. The bed was made of glass and had a cylindrical shape with an inner diameter of 50 mm and height of 500 mm. ZnO particles with an average diameter of 40 μm were used as the fluidized medium (457 g, bed height of unfluidized condition was 120 mm). According to the pressure drop changes for different fluidizing gas velocities and observation of this experiment, the minimum fluidizing gas velocity (U_{mf}) was determined to be about 2.6 L/min.

4. Simulations

To examine the energy required for the proposed process, a simulation at steady-state conditions was conducted using PRO/II Ver. 10.0 software (Schneider Electric). The Soave–Redlich–Kwong equation was selected for the thermodynamic data and an adiabatic efficiency of 80% was initially assumed for the heat-pump compressor. The same simulations were also conducted assuming 60% adiabatic efficiency to be close to actual adiabatic efficiency of the compressors. Water was selected as a representative working fluid in the heat-pump system. The reaction temperature (T_A) in the absorber was assumed to be 120 °C, from the experimental test work; the temperature of the thermal decomposition in the regenerator (T_R) was assumed to be 200 °C, from the equilibrium of the reaction between ZnO and CO_2 using the Materials-Oriented Little Thermodynamic Database for PC (MALT, Kagaku Gijutsu-Sha). The bed temperature was assumed to be uniform due to the properties of a fluidized bed and the reaction was assumed to take place isothermally. We therefore assumed that the heat-pump system successfully received the reaction heat of adsorption and provided reaction heat for thermal decomposition in the regenerator. The minimum temperature differences between the heat-pump system and each beds were assumed to be 10 and 20 K. The blower duty for the fluidized bed was calculated from the assumptions of $2.0U_{mf}$ (=5.2

L/min) for the fluidized gas velocity, 0.6 kPa of pressure drop in the fluidized bed acquired from previous experiments. In addition, 10% of conversion ratio in the absorber was assumed. The blower duty (W_{BR}) of the moving bed was neglected due to its negligible influence and heat loss was neglected. The energy required for the process at each temperature for 80 and 60% adiabatic efficiencies are summarized in Tables 1 and 2.

Table 1 Energy required for the proposed process assuming 80% adiabatic efficiency of the heat-pump compressor and blower

Temperature of heat pump at absorber [°C]	110	100
Temperature of heat pump at regenerator [°C]	210	220
Work for heat pump (W_C) [MJ/kg-CO_2]	0.56	0.70
Work for blower (W_{BA}) [MJ/kg-CO_2]	0.04	0.04
Total Work (W_C+ W_{BA}) [MJ/kg-CO_2]	0.60	0.74

Adiabatic efficiency: 80%

Table 2 Energy required for the proposed process assuming 60% adiabatic efficiency of the heat-pump compressor and blower

Temperature of heat pump at absorber [°C]	110	100
Temperature of heat pump at regenerator [°C]	210	220
Work for heat pump (W_C) [MJ/kg-CO_2]	0.75	0.93
Work for blower (W_{BA}) [MJ/kg-CO_2]	0.05	0.05
Total Work (W_C+ W_{BA}) [MJ/kg-CO_2]	0.80	0.98

Adiabatic efficiency: 60%

From the simulation results, the energy required for the proposed process was from 0.60 to 0.98 MJ/kg-CO_2 with 60 or 80% adiabatic efficiency of the compressors at steady state. These values are less than 1/4 of that required for conventional CO_2 gas separation processes using amine-based absorption (4.1 MJ/kg-CO_2) (Kishimoto et al., 2011, 2012). According to the temperature range of this process, it can be expected to apply for post-combustion CO_2 capture. Although it is necessary to conduct further investigations for installing the proposed process to an actual power plant such as integrating with a scrubber system for separation of other gas mixtures or other processes in a power plant, and optimization of process conditions, it can be said that this process has a great potential for energy saving CO_2 separation for power generation.

5. Conclusion

An innovative energy-saving CO_2 chemical absorption process, integrating a circulating fluidized bed and heat pump, is proposed. The energy-saving performance of the proposed process was evaluated both experimentally and by simulation. In fact, this process required for less than 1/6 of conventional CO_2 separation. Thus, it can be said that the proposed process has a great potential for CO_2 separation with further investigations.

6. Acknowledgements

This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number 15H05554.

References

U. Ali, E.O. Agbonghae, K.J. Hughes, D.B. Ingham, L. Ma, M. Pourkashanian, 2016, Techno-economic process design of a commercial-scale amine-based CO_2 capture system for natural gas combined cycle power plant with exhaust gas recirculation, Appl. Therm. Eng., 103, 747–758

A.M. Arias, P.L. Mores, N.J. Scenna, S.F. Mussati, 2016, Optimal design and sensitivity analysis of post-combustion CO_2 capture process by chemical absorption with amines, J. Clean. Prod., 115, 315–331

A-M. Cormos, A. Simon, 2015, Assessment of CO_2 capture by calcium looping (CaL) process in a flexible power plant operation scenario, Appl. Therm. Eng., 80, 319–327

J.D. Figueroa, T. Fout, S. Plasynski, H. Mcllvried, R.D. Srivastava, 2008, Advances in CO_2 capture technology—The U.S. Department of Energy's Carbon Sequestration Program, Int. J. Greenh. Gas Control, 2, 9–20

K. Goto, H. Okabe, S. Shimizu, M. Onoda, Y. Fujioka, 2009, Evaluation method of novel solutions for CO_2 capture, Energy Procedia, 1, 1083–1089

Y. Kansha, N. Tsuru, K. Sato, C. Fushimi, A. Tsutsumi, 2009, Self-heat recuperation technology for energy saving in chemical processes, Ind. Eng. Chem. Res., 48, 7682–7686

Y. Kansha, M. Aziz, A. Kishimoto, A. Tsutsumi, 2013, Evaluation of a self-heat recuperative thermal process based on thermodynamic irreversibility and exergy, J. Chem. Eng. Japan, 46, 87–91

Y. Kansha, M. Ishizuka, H. Mizuno, A. Tsutsumi, 2017, Design of energy-saving carbon dioxide separation process using fluidized bed, Appl. Therm. Eng., 126, 134–138

M. Kato, S. Yoshikawa, K. Nakagawa, 2002, Carbon dioxide absorption by lithium orthosilicate in a wide range of temperature and carbon dioxide concentrations, J. Mater. Sci. Lett., 21, 485–487

A. Kishimoto, Y. Kansha, C. Fushimi, A. Tsutsumi, 2011, Exergy recuperative CO_2 gas separation in post-combustion capture, Ind. Eng. Chem. Res., 50, 10128–10135

A. Kishimoto, Y. Kansha, C. Fushimi, A. Tsutsumi, 2012, Exergy recuperative CO_2 gas separation in pre-combustion capture, Clean Technol. Environ. Policy 14, 465–474

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50300-1

Insertion of Renewable Sources in the Brazilian Electricity Matrix: an Analysis through Portfolio Theory

Lucas Lyrio de Oliveira[a*], Celma de Oliveira Ribeiro[a], Julia Tomei[b], Erik Eduardo Rego[a]

[a]*Polytechnic School of the University of Sao Paulo, Av. Prof. Luciano Gualberto, 380, Sao Paulo 05508-010, Brazil*
[b]*UCL Institute for Sustainable Resources, University College London, 14 Upper Woburn Place, London WC1H 0NN, UK*
lucas.lyrio.oliveira@usp.br

Abstract

The insertion of renewable sources in the world energy matrix is a promising pathway to reduce greenhouse gas (GHG) emissions in the future. Brazil is one of the leaders in the use of renewable energy, pioneering new biofuel technologies and presenting an electricity matrix based on hydroelectric power. Due to its vast territory, Brazil presents a high heterogeneity regarding natural resources availability and socioeconomic conditions, resulting in regions with different electricity supply and demand profiles. Hydropower, biomass and wind power account respectively for 61.40%, 8.77% and 6.71% of the Brazilian total installed capacity. However, the high dependence on hydroelectricity exposes the country to risks associated with variations on the hydrological regimes. The renewable alternatives to hydroelectricity in the country - biomass and wind power - can increase the reliability of the system and the choice of the optimal power mix is a challenge.

This study analyzes the optimal Brazilian electricity mix based on a portfolio model considering regional supply and demand characteristics. The main drivers to analyze the problem are the generation and transmission costs and the regional supply and demand. At an aggregate level, the results show that bioelectricity is the most economical option among the thermal sources. The results motivate the analysis of land use for biomass crops, the development of new technologies to increase the harvest productivity and establishment of policies to attract investments on bioelectricity.

Keywords: electricity generation, portfolio theory, renewable sources, Brazilian electricity sector.

1. Introduction

The global electrical energy demand may increase by a factor of seven over the next five decades (Siirola, 2012). As a developing country, Brazil presents a high potential for electricity consumption growth. Forecasts from the Empresa de Pesquisa Energética (EPE) – shows that in 2026 the country's demand for electricity will be 43.6% higher than in 2016 (EPE, 2017). The Brazilian electricity mix is predominantly renewable but due to recent droughts there has been an expressive increase in the demand for fossil fuels for thermopower generation. Thus, an important question is: how to expand the

Brazilian electricity capacity to attend the demand growth considering renewable sources.

The Brazilian grid is made up of four interconnected electricity subsystems with more than 99% of the electricity generated in the country being delivered through the national interconnected system (Operador Nacional do Sistema (ONS), 2017). Hydropower is the main generation source, corresponding to 61.40% of the total installed capacity (Agência Nacional de Energia Elétrica (Aneel), 2017). Other important renewable sources are biomass and wind, representing 8.77% and 6.71% of the country's installed capacity. Fossil fuels account for 16.76%, nuclear 1.24% and importation from Paraguay, Argentina, Venezuela and Uruguay 5.04% (Aneel, 2017). Natural gas plays an important role to guarantee the system reliability. Brazil uses fossil fuel power plants to overcome the variability on the hydro, wind and biomass supply, with natural gas representing more than 48% of the fossil installed capacity (Aneel, 2017).

As the hydro potential in many regions is already saturated and most of the country's remaining potential is situated far from the centres with the highest demand, high investments on transmission lines would be required to attend the demand growth. Moreover, the operation of hydropower plants depends on the hydrological conditions, exposing the system to the risk of electricity shortfalls during drought periods. Consequently, the expansion of other sources to meet the growing demand is needed and the choice of such sources is a challenge.

One approach to determine the amount of expansion for each energy source is through Modern Portfolio Theory (Markowitz, 1952). Originally designed to find portfolios of financial assets with minimum risk, this approach has been adapted to many applications, including electricity portfolios (Awerbuch and Berger, 2003).

Expressive regional differences with respect to natural resources availability, level of industrialization, socioeconomic conditions and climate influence the operation of the whole Brazilian system. Therefore, it is possible that transmission constraints and limits on the amount of generation in different regions of the country may affect the overall participation of each generation source. Such issues are considered in the present study through the development of optimization models. The objective is to find electricity matrices with minimum cost risk, to meet the demand forecasted by EPE for the year 2026 (EPE, 2017).

The models consider two risk measures: the variance of costs, as in the original work of Markowitz (1952), and the β-CVaR, a tail risk measure proposed by Rockafellar and Uryasev (2000) defined as the expected cost, above a fixed β-percentile. The decision variables - x_{ij}^{k} - are the amount of electricity generated through source k, in the supply subsystem i, to be sent to the consumption subsystem j.

Recent discoveries of natural gas reserves in the Brazilian pre-salt, the effort to maintain high shares of renewable sources and improvements on bioelectricity generation technologies suggest four different scenarios to be analysed regarding the costs of electricity generation: Scenario 1 considers industry experts costs estimation, Scenario 2 and 3 assume respectively that natural gas and biomass generation costs maintains 20% lower than in Scenario 1, and Scenario 4 considers that both natural gas and biomass generation costs maintain 20% lower than in Scenario 1.

Section 2 presents an overview of the Brazilian power sector. The portfolio models are described in section 3 and the results are presented in section 4. Section 5 concludes this study.

2. Brazilian Power Sector

The Brazilian power sector comprises four interconnected subsystems allowing electricity exchanges between regions. As regions with electricity surplus can supply regions with deficit of electricity, it increases the reliability of the whole system. Moreover, it is possible to take economic advantages of this structure by generating electricity where the generation costs are lower, even when the consumption centre is far from the supply source. The national grid accounts with more than 150,000 kilometres of transmission lines (ONS, 2017).

To maintain and expand the system, generation and distribution companies pay a tariff for the use of the transmission system. This tariff is calculated by the government and the total amount must be enough to cover expenses and to provide profits to the transmission companies. Moreover, generators located farther from the consumption centres pay higher tariff values to compensate their need for extensive lines.

The great size of the Brazilian system makes the management of both generation and transmission costs important to provide affordable electricity prices to the population. Therefore, the costs considered in the models are the sum of the levelized cost of electricity, which is the present value of operation and maintenance (O&M), investment and fuel costs incurred during the lifetime of a powerplant divided by the amount of electricity generated during this period, and the tariff for the use of transmission system.

3. Model Description

The objective is to find minimum risk electricity mixes under two risk measures: the variance and the 95%-CVaR of the total costs. The generation sources considered are hydro, nuclear, wind, biomass, natural gas, oil and coal. Due to investment costs already incurred in hydro and nuclear powerplants currently in operation (sunk costs), these sources were grouped in two categories: old and new plants. It is assumed that old hydro and old nuclear plants have their effective generation capacities fully used. Due to the investment costs, it is also assumed that the only nuclear plant under construction in Brazil, Angra III, will use the estimated effective capacity. As to guarantee supply security, the government contracts a minimum amount of thermoelectricity. The model constraints are linear and described as follows:

a) After transmission and distribution losses are discounted, demand must be attended;

b) Generation through each source for each region is bounded to lower and upper limits;

c) The expected cost is bounded from above;

d) Each region must generate a minimum amount of thermal electricity, which is given by the sum of generation through biomass, natural gas, coal and fuel oil;

e) The amount of electricity sent from one region to another cannot exceed the transmission capacity.

The total cost – T - of the system is the sum of the costs - $Cost_i^k$ - over all sources, supply regions and consumption regions (Eq.(1)).

$$T = \sum_i \sum_j \sum_k Cost_i^k x_{ij}^k = \sum_i \sum_j \sum_k (invest^k + O\&M^k + fuel^k + transm_i) x_{ij}^k \quad (1)$$

Investment, O&M, fuel and transmission costs were considered independent from each other. Different sources are independent with respect to investment and O&M costs. The fuel costs cross-correlation matrix estimated by Awerbuch and Berger (2003) was adopted. The regions were considered independent with respect to the transmission costs. Investment and O&M costs for each source, and transmission costs for each region were assumed to be random variables uniformly distributed. Fuel costs were assumed to be normally distributed.

The model formulation is:

$$\min_{\boldsymbol{x} \in X} \big(risk(\boldsymbol{x}) \big) \quad (2)$$

The set X considers linear constraints. The risk measure is quadratic if the model adopts variance as risk measure, and it is linear in case CVaR is adopted (Rockafellar and Uryasev, 2000).

4. Results

The optimal solutions when transmission constraints are considered are presented in Figure 1 for Scenarios 1 and 2. Except for the solutions with lowest expected costs, there was no significant difference between the optimal solution regarding the use of transmission constraints. The optimal generation mix for Scenarios 1, 3 and 4 were similar, therefore Scenarios 3 and 4 are not shown in Figure 1. For each expected cost level (x-axis) the optimal composition (the y-axis) are shown. Although the amount of generation from each source varies according to the region, indicating the importance of regional public policies, only the aggregated results are presented in this study.

The results obtained through CVaR and variance minimization are respectively shown in the top and bottom of Figure 1. The main differences between the optimal solutions result from the risk measure considered. Due to fuel costs, the expected costs of thermopower sources are higher when compared to other sources. Even when the costs obtained by wind and new hydro power are higher than their respective expected values, they are still lower than the expected costs of thermopower sources. Therefore, as CVaR minimizes the expected cost of the matrix, given that the cost is above a fixed percentile, it does not favour the insertion of thermoelectricity, which is included in the matrix at the lower bound level to assure supply security. On the other hand, the minimization of variance concerns to minimize deviations from the matrix expected cost value and does not consider how high is the resulted expected cost, which results on the insertion of thermoelectricity above the lower bound level.

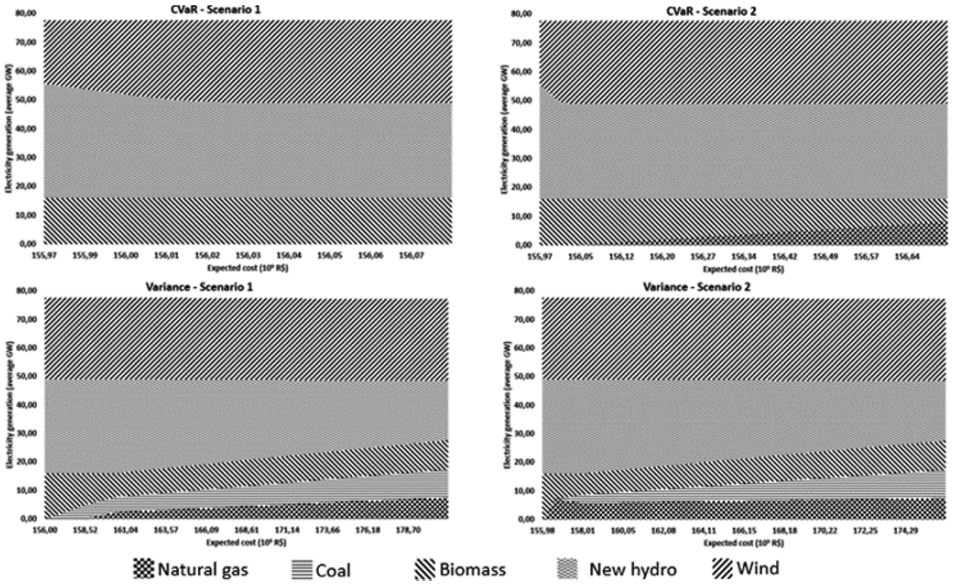


Figure 1: Efficient electricity matrices compositions.
Source: Own elaboration.

All the solutions indicate a high share of wind power in the matrix. This result is aligned with the increase from 2,501 average megawatts in 2015 to 3,809 average megawatts in the electricity generated through wind in 2016 (ONS, 2017). The trade-off expected cost versus cost risk favours the use of wind even compared to hydropower. As a result, for higher expected costs and lower risk solutions, the amount of wind generation grows as substitution to hydropower.

In Scenario 1, under CVaR risk measure, the only thermopower source included in the matrix is biomass. However, if the cost of electricity from natural gas is reduced by 20% and its variance is maintained constant, its expected cost becomes just 2% higher than biomass and it is included in the matrix. When considering variance as the risk measure, Scenario 2 also presents a higher participation of natural gas when compared to Scenario 1.

5. Conclusions

Based on portfolio theory, the four models were applied to find possible efficient electricity matrices to meet the estimated Brazilian demand for 2026. Operational constraints such as limited generation potential for each source in each region, transmission capacities and distribution and transmission losses were considered. Regional differences on the demand were also considered.

Variance and CVaR were adopted as risk measures. With the probability density functions adopted for the costs, it was shown that efficient solutions in terms of variance differ from efficient solutions in terms of CVaR. However, they are similar with respect to indication of high participation of wind, hydropower and biomass.

Biomass was predominant among the thermopower sources. The solutions obtained were very similar regardless transmission constraints were considered or not. Also,

there is no significant change in the solutions when biomass cost is 20% lower than the costs estimated by industry experts. This reduction would only maintain the predominance of electricity generated through biomass among other thermopower sources.

As there was no upper bound for bioelectricity, its predominance raises issues regarding the debate food versus energy. It also indicates that the creation of government policies to incentive investment on bioelectricity is an adequate path to be followed. Moreover, the development of new technologies to generate bioelectricity more efficiently may be crucial to keep renewable sources predominant in the Brazilian power sector.

Under the scenario where the discovery of natural gas reserves in the Brazilian pre-salt layer would cause a 20% reduction on the electricity cost generated by this source, natural gas has an augmented participation on the matrix. As a result, it substitutes part of the bioelectricity generation. Therefore, it can be concluded that in case of land unavailability or lack of bioelectricity investors, natural gas would be the most economical thermopower source to be used instead of biomass.

As a continuity of this research, a multi-period optimization model can be developed to incorporate temporal fluctuations on energy availability.

Funding and Acknowledgments

This work was supported in part by Project FAPESP (São Paulo Research Foundation)/BG Brasil, through the Research Center for Gas Innovation, grant 2014/50279-4 as well as FAPESP/RCUK-NERC Grant Proc. 2015/50684-9.

L.L. Oliveira is grateful to CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, for a PhD scholarship.

References

Agência Nacional de Energia Elétrica (Aneel), 2017. Matriz de Energia Elétrica. http://www2.aneel.gov.br/aplicacoes/capacidadebrasil/OperacaoCapacidadeBrasil.cfm (accessed 23 November 2017).

Awerbuch, S., Berger, M., 2003. Applying portfolio theory to EU electricity planning and policy making. IAEA/EET Working Paper No. 03, EET.

Empresa de Pesquisa Energética (EPE), 2017. Plano Decenal de Expansão de Energia 2026. http://antigo.epe.gov.br/PDE/Documents/Arquivos/PDE2026/PDE2026_versao_para_ConsultaPublica.pdf (accessed 29 January 2018).

Markowitz, H., 1952. Portfolio selection. The journal of finance 7: 77-91.

Operador Nacional do Sistema (ONS), 2017. http://www.ons.org.br (accessed 23 November 2017).

Rockafellar, R.T., Uryasev, S., 2000. Optimization of conditional value-at-risk. Journal of risk 2: 21-42.

Siirola, J.J., 2012. A Perspective on Energy and Sustainability. 11th International Symposium on Process Systems Engineering – PSE 2012. Computer Aided Chemical Engineering 31: 1-7.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50301-3

Including Ecosystem Services in Sustainable Process Design across Multiple Spatial Scales

Tapajyoti Ghosh[a], Xinyu Liu[a], Bhavik R. Bakshi[a*]

[a]*William G Lowrie Department of Chemical and Biomolecular Engineering,The Ohio State University,Columbus, Ohio,USA.*

bakshi.2@osu.edu

Abstract

This work aims to combine process design, supply chain design (SPD), technology choices, hybrid life cycle assessment (LCA) as well as incorporation of ecosystem services in a comprehensive optimization model. While most work on sustainable supply chains couples process life cycle assessment as an addendum to the optimization model, the Process to Planet – Techno Ecological Synergy (P2P-TES) framework seamlessly integrates the engineering process as well as the supply chain design module into the hybrid LCA matrix structure. Inclusion of process design results in simultaneous optimization of both the production process as well as the supply chain network. Choice of technology options in the life cycle model provide the freedom to supplier to choose between pathway alternatives for promoting sustainability in the supply chain. Additionally, this framework includes the role of ecosystem services at the local process level as well as in all the other scales of the hybrid life cycle model. Inclusion of ecosystem services in the optimization framework results in opportunities where technological systems can harness ecological systems to offset the environmental impacts and verifies if the manufacturing process as well as its entire supply chain network is within the ecological carrying capacity of nature and satisfies a necessary condition for sustainability.

Keywords: Supply Chain design, sustainability, optimization, process, ecosystem services.

1. Introduction

Sustainable process design has made enormous advances in the last decade and industries, governments as well as consumers are leaning towards technologies and products whose production and use have low environmental impact. Major chemical industries are aiming for eliminating toxic chemical use, environmental impacts, reducing human risks and improving social conditions not only locally but over the entire supply chain. While LCA methods have been popular for gauging environmental impacts, their application in the field of supply chain design has been limited. Studies which deal with "green supply chains" keep LCA separated from the design problem and apply it as an analytical tool for environmental impacts (Borghi et al. 2014). There have been few studies where supply chains for biofuels have been designed while satisfying environmental objectives. Mele et al. (2009) studied optimal planning of bioethanol and sugar production supply chains while You et al. (2012) considered the design of sustainable cellulosic biofuel supply chains. Aguirre (2009) studied the design of maleic anhydride production supply chain. Both these studies coupled LCA within the optimization framework. To the best of our knowledge, none of the existing work

considers ecosystem services in the purview of the design problem. In this study, we present a novel P2P-TES optimization framework that extends the boundaries of supply chain design by integrating process design, supply chain design and ecological services within the scope of the design problem. An additional advantage of P2P-TES is that it preserves the matrix structure of conventional LCA, making its application simpler and more generalized. Individually, the Process to Planet (P2P) framework (Hanes and Bakshi, 2015) brings in the capability of process design, SPD, LCA and choice of technology while the Techno-Ecological Synergy (TES) framework (Bakshi et al. 2015) provides the foundation for coupling of process, life cycle and economy with the ecosystem services. Using a corn ethanol biorefinery network, an application of P2P-TES framework for process design, supply chain design, and choice of technology is demonstrated. Changes in the decision variables on consideration of ecosystem services within the optimization problem is also demonstrated.

2. Methodology

The framework is built to be analogous with the matrix form of conventional life cycle assessment to promote ease of usability, integration with environmental impact indicators and generalization for applicability to a wide range of problems.

2.1 Background

The mathematical formulation of Conventional LCA is presented as

$$Xm = f \tag{1}$$

where X is the technology matrix that contains information about the life cycle network, flows between activities and linear models of the activities. m is the scaling factor matrix that decides the sizes of the activities to satisfy a certain final demand f of a product on the entire system by society. The environmental impact is given by

$$g = Bm \tag{2}$$

where B is the interventions matrix with information regarding environmental impact of each activity within the system. The P2P framework, while preserving the structure of conventional LCA allows incorporation of economy scale as well as process design in an optimization framework as shown in Fig. 1. It has three separate scales – economy, value chain (for life cycle models) and equipment (for process engineering design model). The framework can be represented mathematically as

$$\overline{\underline{X}}\,(\{z\})\,\overline{\underline{m}} = \overline{\underline{f}} \tag{3}$$

where $\overline{\underline{X}}$ represents a multiscale technology matrix that includes life cycle model information, economy scale as well as engineering design variables z for the fundamental engineering models. Similarly, $\overline{\underline{m}}$ and $\overline{\underline{f}}$ represent the multiscale scaling vector and multiscale final demand that contain the sizes of the three separate scales and their separate final demands. The overbars and underbars are used to represent its multiscale nature. Environmental impact from this multiscale framework is given by

$$\overline{\underline{g}} = \overline{\underline{B}}(\{\mathrm{z}\})\,\overline{\underline{m}} \tag{4}$$

where $\overline{\underline{B}}$ is the multiscale environmental interventions matrix containing impact information for activities at all three scales.

2.2 P2P-TES mathematical framework

The P2P framework is modified to include ecosystem services information resulting in the P2P-TES framework which is written as

$$\begin{bmatrix} \underline{\overline{X}}(\underline{\overline{z}}) & \underline{\overline{C}}(\underline{\overline{z}}) \\ \underline{\overline{B}}(\underline{\overline{z}}) & \underline{\overline{S}}(\underline{\overline{z}}) \end{bmatrix} \begin{bmatrix} \underline{\overline{m}} \\ \underline{\overline{m}}_e \end{bmatrix} = \begin{bmatrix} \underline{\overline{f}} \\ \underline{\overline{f}}_e \end{bmatrix} \tag{5}$$

In contrast to P2P framework, engineering process variables z has been replaced by $\underline{\overline{z}}$ which includes $\underline{z}$ supply chain design/choice of technology and $\overline{z}$ economic policy design variables. $\underline{\overline{S}}(\underline{\overline{z}})$ is ecosystem services matrix that holds data pertaining to ecosystem activities modelled as unit operations at different scales. $\underline{\overline{m}}_e$ is the ecosystem scaling matrix that determines ecosystem sizes whereas $\underline{\overline{f}}_e$ is the multiscale ecosystem final demand matrix. The final demand of ecosystems can be envisaged as the requirement of technological systems to reduce environmental impact, such as carbon dioxide sequestration and is obtained by subtracting the demand of ecosystem services from supply. Ideally, if $\underline{\overline{f}}_e$ is 0, it means no environmental impact as ecosystem service supply becomes equal to the demand. $\underline{\overline{C}}(\{\underline{\overline{z}}\})$ is the management matrix that models flows from technological systems to ecosystems for maintenance. The net environmental impact is obtained from

$$\underline{\overline{g}} = \sum_i \underline{\overline{f}}_e^i \tag{6}$$

which represents the total final demand of ecosystem services. It is different from the previous frameworks since we are interested in not only the direct environmental impact but after ecosystem services have been considered and resulted in impact remediation. The dotted arrows in Fig. 1 represent the flow to and from ecological systems.

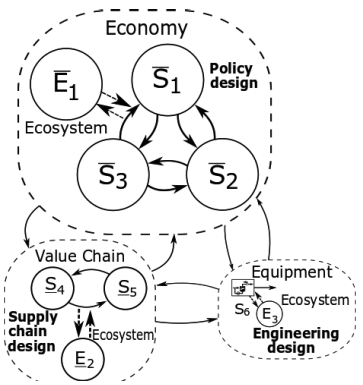


Figure 1: P2P-TES Framework showing economy, value chain and equipment scale with ecosystems.

As seen from Fig. 1 different modules of the P2P-TES framework entail different design problems. Variables at equipment scale constitute engineering design problem, at the value chain scale establish a supply chain design problem, while variables at economy scale forge a policy design study. With environmental impact assessment and ecosystem services included at all three scales, P2P-TES creates the foundation of a comprehensive, yet coherent and systematic tool which can undertake various challenges as mentioned earlier.

2.3 Optimization Formulation

The optimization problem based on the P2P-TES framework can be multifaceted with multiple objectives based on user preference. For the purposes of this paper, we define a multi-objective scenario, with minimization of environmental impact and maximization of economic profitability of the engineering process through

$$Minimize \begin{cases} \Sigma_i \overline{\underline{f}}_e^i \\ -NPV(\{z\}) \end{cases} \quad (7)$$

subject to constraints,

Eq. 3, $\mathrm{H}(\{z\}) \geq 0$, $\overline{\underline{\mathrm{m}}}_{\mathrm{e}} \geq 0$, $\overline{\underline{m}} \geq 0$, $S(\{z\}) \geq 0$

$\mathrm{H}(\{z\})$ represents the engineering design, comprising process variables, mass, energy balances etc. $S(\{z\})$ represents the ecosystems modelled as unit operations and should also comply with physical mass flow laws. Eq. 3 forms the core equation for P2P-TES framework while the sizes of technological and ecological systems should be non-negative. Solving the optimization problem determines the engineering process design variables z, supply chain design and/or choice of technology variables $\underline{z}$ and policy design variables $\overline{z}$.

3. Case Study

In this article, we demonstrate P2P-TES application on a corn ethanol distribution network. The goal of the study is to design the ethanol manufacturing process, the supply chain as well as choose between farming objectives. Major constraints imposed on this network are demand of food and energy by the society. As seen from Fig. 2, we consider a land area divided into 18 grid regions. Each of these regions are attributed to their own available agricultural area, population dependant food and biofuel demand. The demand and production of materials are concentrated at the centroid of grids. Each of the 18 grid regions can be a possible location of a biorefinery whose total number is constrained at 6 as because we estimated 6 refineries would be enough to satisfy the entire corn ethanol demand of this area. The white regions are assumed to practice no till corn farming while the grey regions practice farming with conventional till. The biorefineries as well as consumers have the option of choosing between the different farms available at the 18 grids for meeting their fuel and food demand respectively. Combination of rail and truck transportation modes was considered from centroid to centroid distances for material movement. Environmental impact considered in this study is nitrate loading in water from farming and refining activities. The optimization problem was set up to minimize the environmental impact as described by

$$Minimize \ \Sigma_i \overline{\underline{f}}_e^i \ or \ \overline{\underline{B}}(\overline{\underline{z}})\overline{\underline{m}} + \overline{\underline{S}}(\overline{\underline{z}})\overline{\underline{m}}_e \quad (8)$$

Design variables included process parameters for corn to ethanol conversion in biorefinery, quantity of corn supply from the different farms to the individual biorefineries and consumer, binary variables to determine the location of biorefineries and scaling variables to determine the total consumption from individual farms. The corn to ethanol production technology assumed a wet milling process with 15 projected design variables which included temperature of steam jets, flow concentrations and recovery rates in distillation columns. Ecosystems considered in this case study are wetlands which are modelled as bioreactors and can take up the nitrates in the run-off water from technological activities. Limited amount of areas in each grid region is available for conversion to a wetland ecosystem. The final optimization problem is an MINLP consisting of 8541 equations, 10915 single variables and 36 discrete variables. The non-linearities in the model mostly come from the engineering superstructure module of the biorefineries. It is solved using BARON on GAMS optimization platform in an 8300 core Intel Xeon Cluster.

3.1 No-ecosystem scenario

In this scenario, ecosystems are not considered in the optimization problem. The objective function minimizes the nitrate runoff from the farms and the biorefineries. Solving this problem resulted in placement of biorefineries in the grids as shown in Fig. 2 (top). From the network plot in Fig. 2 (top), it is observed that all the grid regions practicing no till agriculture is chosen for satisfying food and fuel demand. The network plot depicts the quantity (width of interconnecting curves) of flow of corn from farms to biorefineries and consumers (interconnecting curves between farms and refineries, farms and consumers) along with the flow of ethanol from refineries to consumers (interconnecting curves between refineries and consumers). However, the total corn production of these 6 regions are not enough for satisfying the total food and fuel demand. Thus, additional farming regions are selected for corn production. However, the optimization solution expands no till practicing farms to their largest possible production bounds before sourcing corn from farms practicing conventional till.

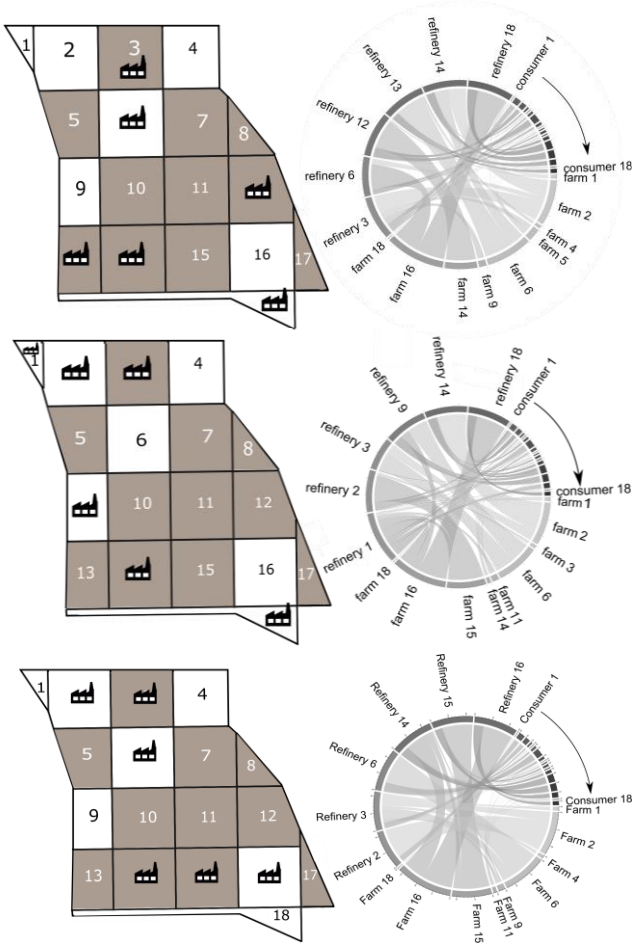


Figure 2: Distribution of biorefineries in the grids and network plot for product flow connections for (top) no-ecosystem scenario, (middle) with-ecosystem scenario and (bottom) transportation cost reduction scenario. Consumers 2-18 have not been labelled due to space constraints.

3.2 With-ecosystem scenario

In this scenario, land area is available at the farms for conversion into wetlands. These wetlands, modelled as flow reactors can consume nitrate loaded water, treat it naturally and supply clean water to the watershed. The amount of land area available is barren or shrubs land and is determined randomly for the 18 grids. Solving this optimization problem, the distribution of refineries is obtained as shown in Fig. 2 (middle). Total nitrate loading in water was reduced by 22%. From the network plot in Fig. 2 (middle), it is observed that for this scenario, along with some no till faming grids, conventional till practicing regions

3,11,14,15 are chosen for production of corn. This is due to the construction and maintenance of wetlands in these regions which are large enough to take up a huge amount of nitrate in the water. Thus, the net nitrate emissions for these regions are lower than obtained from till agriculture. Farming in regions 4 and 9 are not longer viable because they have relatively less area available for wetlands and results in larger nitrate emissions compared to the selected regions.

3.3 Transportation cost reduction scenario

In this scenario, the objective function is modified to reduce cost of transportation of products between different sectors. Ecosystems are considered in this analysis and the nitrate loading in water is fixed to the amount obtained from the previous with-ecosystem scenario. The distribution of biorefineries for this scenario as shown in Fig. 2 (bottom) is such that both transportation of corn from farms to biorefineries and farms to consumers as well ethanol transportation cost from biorefineries to consumers is reduced. The network plot in Fig. 2 (bottom) clearly depicts this change. More interconnections are seen between activities in the same grid (Farm 2 – Refinery 2, Farm 6 – Refinery 6 etc.). Biorefineries also tend to be located near high ethanol demand centres. (Grids 14,15,16.)

4. Conclusion

In this article, we present a novel framework that can solve a process design, supply chain design, technology choice optimization problem and considers ecosystem services within its analysis. The inclusion of ecosystem services provides a metric for sustainability where we may claim that the supply chain does not exceed the carrying capacity of ecological sources and may be environmentally sustainable. The case study effectively demonstrates how this framework can be applied for determination of process parameters and facility locations of the biorefineries, choice of farming technologies and consideration of wetland ecosystems.

References

Del Borghi, A., Gallo, M., Strazza, C. and Del Borghi, M., 2014. An evaluation of environmental sustainability in the food industry through Life Cycle Assessment: the case study of tomato products supply chain. *Journal of cleaner production*, *78*, pp.121-130.

Mele, F.D., Guillén-Gosálbez, G. and Jiménez, L., 2009. Optimal planning of supply chains for bioethanol and sugar production with economic and environmental concerns. *Computer aided chemical engineering*, *26*, pp.997-1002.

You, F., Tao, L., Graziano, D.J. and Snyder, S.W., 2012. Optimal design of sustainable cellulosic biofuel supply chains: multiobjective optimization coupled with life cycle assessment and input–output analysis. *AIChE Journal, 58*(4), pp.1157-1180.

Laınez-Aguirre, J. M., 2009. "Towards Integrated Supply Chain Management for the Enterprise Sustainability." *PhD Thesis.*

Hanes, R.J. and Bakshi, B.R., 2015. Process to planet: A multiscale modeling framework toward sustainable engineering. *AIChE Journal*, *61*(10), pp.3332-3352.

Bakshi, B.R., Ziv, G. and Lepech, M.D., 2015. Techno-ecological synergy: A framework for sustainable engineering. *Environmental science & technology*, *49*(3), pp.1752-1760.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50302-5

Long-Term Energy Storage: What is the Need and is Ammonia a Solution?

Richard M. Nayak-Luke[a], René Bañares-Alcántara[a*]

[a]*University of Oxford, Dept of Engineering Science, Parks Road, Oxford OX1 3PJ, UK*
rene.banares@eng.ox.ac.uk

Abstract

Ammonia production contributes 1.3% of global carbon dioxide emissions (IEA 2016, Philibert 2017). In the context of global emission targets and growing demand, decarbonisation of this process is environmentally imperative. Ammonia can also have a significant additional role as a seasonal electrical energy storage vector, and the potential requirement for this could be substantial. We present a method to calculate a first estimate for the optimum size of an electrified ammonia production plant (at the process level), the required renewable energy (RE) supply, and the levelised cost of ammonia (LCOA) for islanded operation with a hydrogen buffer. A model was developed to quantitatively identify the key variables that impact the LCOA (relative to a ±10 GBP/t change in LCOA): levelised cost of electricity (±0.89 GBP/MWh), electrolyser CAPEX (±65 GBP/kW), minimum Haber-Bosch (HB) load (±12% of rated power), maximum rate of HB load ramping, and RE supply mix. Using 2025/2030 estimates results in a LCOA of 588 GBP/t for Lerwick, Scotland. The application of the model can guide and improve the production of carbon-free ammonia in the future.

Keywords: Ammonia; Electrification; Haber-Bosch; LCOA; Seasonal storage.

1. Introduction

On a global scale, carbon dioxide emissions are increasing at a rate of 2.4% per annum (IEA 2016). These emissions originate from four main sectors: electricity, heat, transport and industry, the last being the most commonly overlooked. With chemical production predicted to account for 14% of industrial emissions by 2050 (Philibert 2017), it is an environmental imperative to identify and decarbonise those chemicals produced in significant quantities and with high dependency on fossil fuels (for both energy and feedstock). Ammonia is a prominent example, with a current global production of c.180M t/y and, due to the 95% of its hydrogen feedstock coming from fossil fuels, and generation of 420M t/y of carbon dioxide emissions (1.3% of the global amount) (IEA 2016, Philibert 2017). Primarily driven by population growth, this is predicted to increase to c.270M t/y of ammonia by 2050 (Philibert 2017, UN 2017). Currently about 85% of ammonia produced is used as fertiliser.

In recent years ammonia has been re-considered as a potential energy vector for use in three other sectors: electricity, transport and heating. Ammonia can provide a grid service for seasonal energy storage. There has been significant integration of renewable energy (RE) sources onto existing electrical networks. To date this has mainly been in response to binding national/regional targets (e.g. EU Renewable Energy Directive), but the substantial falling cost and significant availability of RE resources also drives it (IPCC 2011). However, as calculated in (Denholm and Hand 2011) and proven in practice by

the application of Germany's Energiewende (Ehlers 2015), increasing RE penetration on networks requires an increased network flexibility for successful integration. To gain quantitative understanding of the supply-demand balancing problem and to facilitate identification of the impact that key variables (such as RE penetration) have on the storage required we have defined the storage duration index (SDI) (Eq.1).

$$\text{SDI} = \left(\begin{matrix} \text{Fraction of demand met by} \\ \text{energy that has been stored} \end{matrix} \Big/ \begin{matrix} \text{Normalised variance from} \\ \text{uniform duration storage} \end{matrix} \right) \quad (1)$$

The SDI condenses the storage requirement information from the storage profile (Figure 1) into two non-dimensional numbers that relate to the magnitude of storage use and duration of storage use respectively. This enables the quick identification of the impact that variables have on the storage requirements. Using this tool and data from the Met Office (Office 2006) and ELEXON (ELEXON 2017) we have investigated the impact that RE mix and penetration would have on an electricity network in Lerwick, Scotland. From these results, shown in Figure 2, we have three conclusions for this location: increased RE penetration increases the requirement for storage, increased RE penetration leads to a reduction in the short-term storage bias, and the RE mix has a large impact on storage requirements which is exacerbated by RE penetration.

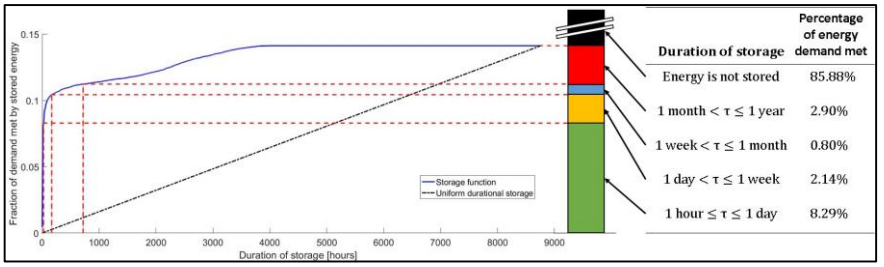


Figure 1: Storage profile against the duration of storage for Lerwick with 60% RES penetration and 50% Wind:50% Solar power supply

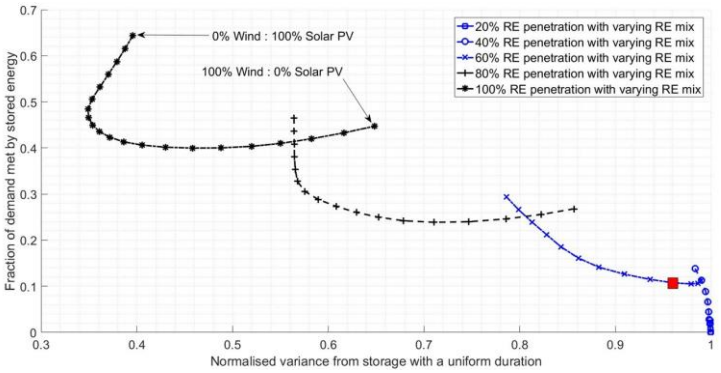


Figure 2: The impact of RES ratio (Wind:Solar PV energy supplied) and RES penetration on the system's storage duration index (SDI)

We also investigated the viability of an islanded 'green' ammonia production process, i.e. using a Haber-Bosch (HB) process fed with hydrogen produced by water electrolysis and nitrogen from air separation powered entirely by RE. We modelled the production process in MATLAB at the process level accounting for the intermittency of RE (i.e. fluctuations

≥ 1 hour), a gap in the knowledge base recently identified by the IEA (Philibert 2017). Few studies on process analysis of electrified ammonia production have been published. The four most important RE studies dealt with intermittency differently so LCOA estimates vary (in GBP/t NH_3, Eq.2): 447-942 (Morgan 2013), 264-528 (Beerbühl, Fröhling et al. 2015), 504-609 (Bañares-Alcántara, Dericks III et al. 2015) and 524-1771 (Tunå, Hulteberg et al. 2014). To stabilise the supply Morgan and Tunå et al. used dispatchable sources such as the grid or biomass, whereas Beerbühl et al. and Bañares-Alcántara et al. assumed islanded systems that had hydrogen and ammonia "buffers" respectively. Our analysis presented here goes beyond: it provides a method applicacable to any location and quantitatively identifies the key variables affecting the LCOA.

Section 2 outlines the modelling assumptions and methodology, Section 3 describes the model's calculations, Section 4 presents the results and their significance. Section 5 summarises the key conclusions and identifies areas for future development.

2. Methods (modelling methodology)

The model was written in MATLAB®. The model, when provided with RE supply profiles, outputs the optimum size of the electrolyser, air separation unit (ASU), HB process and hydrogen storage (see Figure 3). Optimisation of additional variables such as the RE mix and HB operating variables (maximum rate of load ramping and minimum process load) is also possible. The model uses RE supply data with thirty minute or hourly granularity to accommodate electrical settlement periods in the UK (the most commonly available data granularity). This limits the accuracy of the final results unless sub-minute variations are taken into account using dynamic modelling. The ASU and HB processes are lumped together (ASU/HB) for power allocation to simplify the problem, a reasonable assumption due to the small power demand and capital cost (CAPEX) of the ASU relative to the HB process, and particularly the electrolyser.

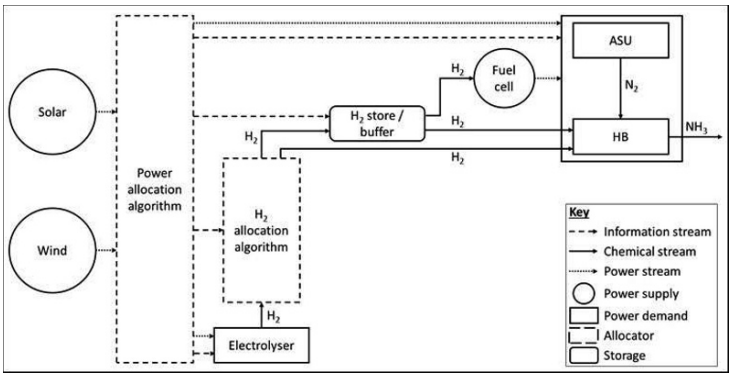


Figure 3: Process block diagram including power, chemical and information flows

3. Calculation

In order to identify the key variables that impact LCOA (Eq.2) and its sensitivity to them, the model was run with (2000-2016) Met Office wind and solar MIDAS data which have a granularity of one hour (Office 2006). The use of historical data, instead of an artificial statistical dataset, and its long time coverage enabled the model to provide more realistic results. We decided to perform our investigation using data for Lerwick, Scotland for two

reasons: plentiful wind power (thus enabling a high electrolyser average load factor (Philibert 2017)), and a predicted increase of curtailed RE energy in the regional network (see ITM's BIG HIT project for further details (Hyde 2017)).

$$\text{LCOA} = \frac{\text{Cost of production}}{\text{Mass of ammonia produced}} \quad (2)$$

To facilitate the comparison of different cases, we constrained the supply power to an average of 100 MW rather than fix the plant production rate (in t/day of ammonia); the simulation then finds the optimal size plant that minimises the LCOA. For Lerwick the LCOE from renewables was taken as 67 and 80 GBP/MWh for wind (BEIS 2016) and solar photovoltaic (BEIS 2016) sources respectively. The solar data is provided in power per unit area, so it does not require pre-conditioning. However, the wind speed provided in the MIDAS data is converted first to a hub height of 80 m using the wind profile power law (and roughness level taken as 0.03), then converted to power using the power profile of the Vestas V90 3.0 MW turbine with an air density of 1.225 kg/m^3. The cut-in and cut-out speeds were taken as 4 and 25 m/s respectively.

The identification of the key variables that determine the LCOA value was achieved by performing sensitivity analyses on numerous variables, prioritised when possible (if they were cost of production variables), by their fraction of total production cost. Having identified five key variables (discussed further in Section 4) that significantly impact LCOA, we performed more detailed sensitivity analyses. Analysis of production cost variables held the size of the plant (all component processes) constant and changed linearly the value of the variable. However, analysis of production process variables allowed the size of the plant (each individual component process) to change, allowing the model to find a better plant size (and its corresponding RE supply mix) if one is available, thus providing an impact on LCOA together with individual process sizing.

4. Results

The five key variables have the most significant impact on LCOA:

- Production cost variables (those that dominate OPEX and CAPEX values)
  i. Levelised cost of electricity (LCOE)
  ii. Electrolyser CAPEX per kW of rated power
- Production process variables (those that influence the mass of ammonia produced)
  iii. RE sources ratio (dependent on RE profiles, i.e. geographical location)
  iv. ASU/HB process minimum power consumption (P_{MIN})
  v. ASU/HB process maximum ramping rate (maximum difference in power for contiguous time intervals)

The production cost variables, as shown in Figure 4 for LCOE, have significant impact on the LCOA: a change in LCOE ±0.89 GBP/MWh results in a change of ±10 GBP/t NH_3. In turn, a change in the electrolyser CAPEX per kW rated power of ±65 GBP/kW has the same effect on the LCOA, i.e. ±10 GBP/t NH_3. The significance of these variables is widely acknowledged, but has only been previously quantified in (Beerbühl, Kolbe et al. 2014) (using a dispatchable electrical supply) and in (ISPT 2017) respectively. The results presented here corroborate Beerbühl's result of ±0.97 GBP/MWh resulting in ±10 GBP/t NH_3, and correlate well with ISPT's finding that a significant reduction in

electrolyser CAPEX (to 300 GBP/kW) is needed to make the process economically viable. We have benchmarked our analysis in Figure 4 against ammonia as a commodity (showing the global ammonia prices between 2002-2014) and as it is converted to retail electricity (prices of 0.12 GBP/kWh with a 50% ammonia-to-electricity efficiency).

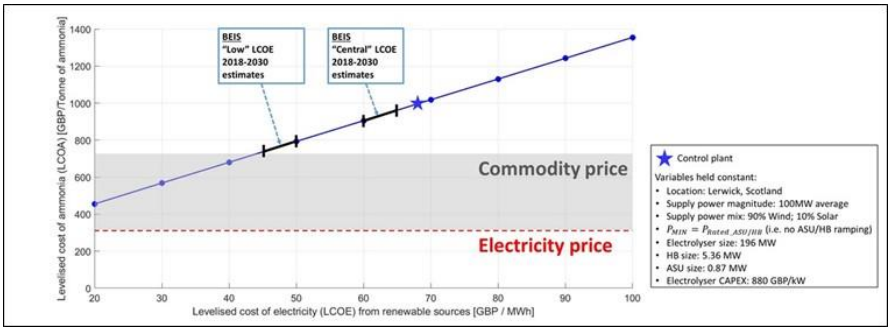


Figure 4: LCOA sensitivity to LCOE

In contrast, the process variables (shown in Figure 5) have been ignored to date and their impact had not been quantified previously. Figure 5 shows the benefits of a flexible HB process and the limitations that the process must be able to operate within, i.e. ramping faster than 5% of rated power/h, to enable these benefits to be realised. The ramping bottleneck within the process would most likely be the catalyst. These results can provide some technical requirements for future catalyst development in an electrified ammonia process.

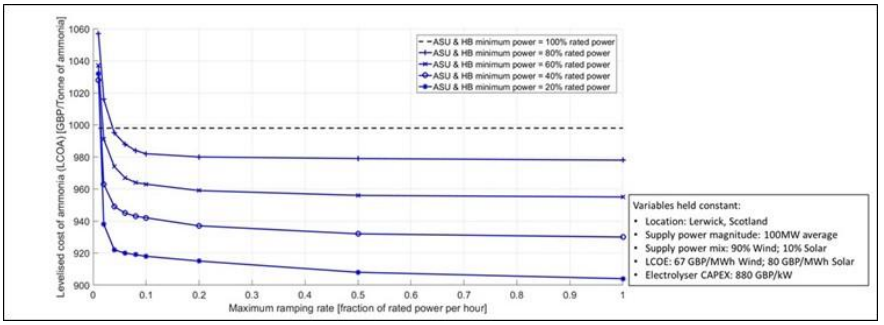


Figure 5: LCOA sensitivity to ASU & HB process minimum power and maximum ramping rate

5. Conclusions

The outline of a model implemented in MATLAB® has been presented which, for a given location, will design a 'green' ammonia plant with the optimum size (at the process level), RE mix and operation schedule. In response to a recent IEA challenge (Philibert 2017), this model makes possible detailed plant design in combination with RE supply sizing for any given location, and facilitates further optimisation (through dynamic analysis) by providing a good initial solution. The model has enabled us to identify the five most significant variables to produce 'green' ammonia, and to investigate the LCOA's sensitivity to each of them. In practice, a combination of these improvements in a suitable geographical location (i.e. with favourable supply profiles) has the potential to become the first economically viable decarbonised process substitute to conventional ammonia production.

Production cost variables

i. Levelised cost of electricity (LCOE): ±0.89 GBP/MWh results in ±10 GBP/t NH_3
ii. Electroyser CAPEX per kW rated power: ±65 GBP/kW results in ±10 GBP/t NH_3

Production process variables

iii. RE sources ratio: has a significant impact on the LCOA and optimum plant size
iv. ASU/HB process minimum power (P_{MIN}): ±12% change of rated power results in ±10 GBP/t NH_3
v. ASU/HB process maximum ramping rate (LCOA insensitive to > 4% rated power/h)

2025/2030 predictions for these variables (LCOE = 45.7 GBP/MWh (BEIS 2016), electrolyser CAPEX = 308 GBP/kW (Colella, James et al. 2014), $P_{MIN} = 20\%$ rated power and a maximum ramping rate of 6%) estimates LCOA as 588 GBP/t NH_3. This work shows how electrifying chemical production processes using intermittent RE affects process sizing and plant operation. Future work will integrate component level analysis of the synthesis loop, accounting fluctuations < 1 hour for RE and use machine learning to manage the predictability of RE.

References

Bañares-Alcántara, R., G. Dericks III, M. Fiaschetti, P. Grünewald, J. M. Lopez, A. Yang, L. Ye and S. Zhao (2015). Analysis of islanded ammonia-based energy storage systems. Oxford, UK, Univ. of Oxford**:** 158. (at www.eng.ox.ac.uk/systemseng/publications/Ammonia-based_ESS.pdf)

Beerbühl, S., M. Fröhling and F. Schultmann (2015). "Combined scheduling & capacity planning of electricity-based ammonia production to integrate renewable energies." EJOR **241**(3): 851-862.

Beerbühl, S., B. Kolbe, C. Roosen and F. Schultmann (2014). "Synthesis of ammonia from intermittently generated hydrogen." Chemie-Ingenieur-Technik **86**(5): 649-657.

BEIS (2016). BEIS Electricity Generation Costs (November 2016). EISB Dept for Business.

Colella, W. G., B. D. James, J. M. Moton, G. Saur and T. Ramsden (2014). Techno-economic analysis of PEM electrolysis for hydrogen production. Electrolytic Hydrogen Production Workshop NREL. Golden, Colorado.

Denholm, P. and M. Hand (2011). "Grid flexibility and storage required to achieve very high penetration of variable renewable electricity." Energy Policy **39**(3): 1817-1830.

Ehlers, N. (2015). The German experience. Oxford Energy Conference, Oxford.

ELEXON (2017). Demand profiling data per profile class.

Hyde, K. (2017). BIG HIT: Europe's largest integrated H_2 system. Hannover Messe. Hanover.

IEA (2016). Key world energy statistics, International Energy Agency.

IPCC (2011). Renewable energy sources and climate change mitigation: Special report of the Intergovernmental Panel on Climate Change. Cambridge University Press, (IPCC).

ISPT (2017). Power to Ammonia: Feasibility study for the value chains and business cases to produce CO2-free ammonia, Institute for Sustainable Process Technology (ISPT).

Morgan, E. R. (2013). Techno-economic feasibility study of ammonia plants powered by offshore wind. Doctor of Philosophy, University of Massachusetts Amherst.

Met Office (2006). UK Mean Wind Data, Met Office Integrated Data Archive System (MIDAS). NBAD Centre.

Philibert, C. (2017) "Producing ammonia and fertilizers: new opportunities from renewables."

Philibert, C. (2017). Renewables crossing boundaries: ammonia et al. NH3 Event, Rotterdam.

Tunå, P., C. Hulteberg and S. Ahlgren (2014). "Techno-economic assessment of nonfossil ammonia production." Environmental Progress & Sustainable Energy **33**(4): 1290-1297.

UN (2017). World fertilizer trends and outlook to 2020, UN Food and Agriculture Organization.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50303-7

Solar thermal and geothermal integration with low temperature CO_2 DENs

Raluca-Ancuta Suciu[a*], Paul Stadler[a], Luc Girardin[a], François Maréchal[a]

[a]*École Polytechnique Fédérale de Lausanne, Rue de L'industrie 17, Case Postale 440, Sion, CH-1951, Switzerland*

raluca-ancuta.suciu@epfl.ch

Abstract

Taking into account the high efficiency of solar thermal collectors and the fact that heating demands represent around 80% of the total energy needs in residential buildings, it is of interest to integrate solar thermal systems to supply HVAC services. However, both solar energy and HVAC needs have strong daily and seasonal variations, and require an efficient long term storage solution. The underground represents an efficient storage medium, which can compensate the latter mismatch; in winter, surplus heat can be extracted to satisfy the strong demand, while in summer, the rejected heat from cooling systems – in addition to the heat produced from the solar thermal collectors can be reinjected in the ground, therefore achieving long term energy storage.

Low temperature district energy networks do not only supply heating and cooling to districts, but they also have the ability to harvest heat from renewable or low temperature waste heat sources and to integrate with long term heat storage systems, such as ground storage. This work looks at the integration of low temperature CO_2 networks with solar thermal and geothermal storage to provide energy services. Solar thermal collector areas of 12.5 m^2/cap are required for typical urban centers to close yearly energy balance in the ground.

Keywords: ground long-term storage, solar thermal, low temperature district networks, geothermal probes

1. Introduction

In the 20th century geothermal energy started being harvested on a large scale for heating, industry and electricity generation. Although the worldwide use of geothermal energy amounts to 49 TWh/year of electricity and 53 TWh/year for direct use, the total potential of this resource is estimated at 2,000 TWh/year for electricity generation and 7,000 TWh/year for direct use (Groupe de travail PGG 2011). Clearly, the current use of geothermal energy represents only a very small fraction of the real potential.

One way of exploiting geothermal energy would be by extracting the heat from the ground using vertical geothermal probes (VGP) (Focaccia and Tinti 2013). VGPs are composed of one or more polyethylene U-shaped pipes, in which a heat transfer fluid flows in a circuit. The fluid heats up when it goes down the tube and transports the heat from the underground to the heating system found at the surface.

Vertical geothermal probes exploit a low temperature heat source, and are usually at a depth of no more than 300-400 m. The probes are further connected to heat pumps (HPs) which allow, with a moderate amount of electricity, to increase the temperature of

the fluid up to 35 to 45°C for heat delivery at low temperature (e.g. for space heating (SH)) or up to 60°C to produce domestic hot water (DHW) (Steiger and Kees 1981). Providing thermal services using vertical geothermal probes requires external energy (usually in the form of electricity) not just for the heat pump, but also for the pump used to circulate the heat transfer fluid inside the probe. In addition to heating, VGPs can also be used to produce cooling in summer. In this case, the heat transfer fluid evacuates the excess heat from the building in the underground, recharging the ground which was cooled down during the heating season (Wiggs 2008). Cooling services can be provided either directly or via a refrigeration cycle. In the case of direct cooling, this natural cooling technique has two advantages: it saves a considerable amount of electrical energy (i.e. used by commercial AC) and it recharges the ground with heat during summer.

However, the temperature of the ground where the probe fields are must be equilibrated over the year, since generally the heating is not equal to the cooling demand. In this work, a solar system is coupled to the network in order to close the energy balance in the ground. A typical urban center using CO_2 as the heat transfer fluid is used as case study. Celine Weber *et al* (Weber and Favrat 2010) introduced the idea of distributing CO_2 in the district energy networks at an intermediate temperature below the critical pressure of 7.4 MPa. The system has been integrated with geothermal probes for heat extraction (Suciu et al 2017), but a low temperature water source was used for heat injection. In this work, the heat injection is done using geothermal probes and the gap between the heat extraction from the ground and the heat injection into the ground is filled using heat available from solar thermal collectors. The solar thermal collector area required to close this energy balance for a typical urban center is assessed.

2. Methodology

2.1. CO_2 based district energy networks

A detailed description of the refrigerant based network using CO_2 as the heat transfer fluid is given by (Weber and Favrat 2010) and possible energy conversion technologies have been investigated by (Henchoz 2016). The decentralized energy technologies used are (i) heat pumps for SH and DHW, (ii) heat exchangers for air cooling, and (iii) vapor compression chillers for refrigeration. The centralized energy technologies used are: a central plant which balances the network (in the form of a lake or geothermal probes), photo-voltaic (PV) panels, a solid oxide fuel cell (SOFC) – solid oxide electrolysis cell (SOEC) co-generation unit, and a boiler-steam network installed at the waste incinerator. Interested readers are referred to (Suciu et al 2017) for further details. A schematic representation of the fully integrated system can be observed in Figure 1.

The system is integrated with geothermal probes for heat extraction and injection and with solar thermal collectors to close the balance on the heat extracted from the ground (i.e. additional heat injection). The problem is formulated as a mixed integer linear programming (MILP) problem and is targeting the optimal utility selection (and sizing) (Marechal et al 2003). The objective function is the minimization of the operating cost and the main constraints are the heat cascade equations and the resource balance equations.

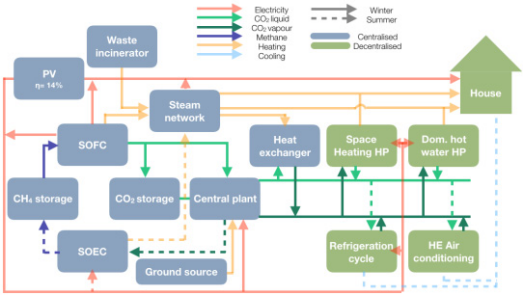


Figure 1. Schematics of CO_2 DEN

2.2. Geothermal extraction and injection

The ground storage is modeled as a fully stratified tank (Figure 2), with the temperature at the bottom fixed at T_{bottom} = 30°C and the temperature at the top fixed at the ambient temperature T_{amb}. The storage level changes by moving the thermocline up and down. The ground storage tank is modeled as a long-term storage unit:

$$SL(t+1) = SL(t) + \dot{Q}_{in} - \dot{Q}_{out} + \dot{Q}_{loss,env} \quad (1)$$

with $SL(t)$ the storage level at time t, $\dot{Q}_{in}$ the heat flow in the ground (eject heat from the network and heat from solar thermal), $\dot{Q}_{out}$ the heat flow out of the ground (heat required by the network) and $\dot{Q}_{loss,env}$ the heat losses to the environment:

$$\dot{Q}_{loss,env} = \hbar_{air} \cdot \dot{A}_{ground} \cdot (T_{bottom} - T_{amb}) \quad (2)$$

with $\hbar_{air}$ = 50 W/(m²·K) the heat transfer coefficient of air (Bergman and Incropera 2011) and A_{ground} the cross-sectional area of the ground influenced by the geothermal probe. The cross-sectional area was computed using:

$$A_{ground} = \pi \cdot \left(\frac{D_{ground}}{2}\right)^2 \quad (3)$$

with D_{ground} = 4 m (Bianchi 2006) the diameter of the ground influenced by the geothermal probe. In order to compute the electrical power required to extract heat from the ground, first the pressure drop in the pipe was computed using the friction factor f, the fluid density ρ_{H_2O}, the fluid velocity v, the probe height $\hbar_{probe}$, and its hydraulic diameter $D_{\hbar}$= 0.04 m (Groupe de travail PGG 2011):

$$dP = \frac{f \cdot \rho_{H_2O} \cdot v^2 \cdot \hbar_{probe}}{2 \cdot D_{\hbar}} \quad (4)$$

Additional details on the different parameters can be found in (Bergman and Incropera 2011). The probe height was computed assuming a geothermal gradient of 30°C/km (Amblard 2015). Using the pressure drop, the mass flow of the fluid $\dot{m}$, its density ρ_{H_2O}, and an average geothermal pump efficiency $\eta_{geot\hbar}$ = 0.8 (Self, Reddy et al 2013), the geothermal energy of extraction is calculated:

$$\dot{E}_{pumping} = \eta_{geot\hbar} \cdot \frac{dP \cdot \dot{m}}{\rho_{CO_2}} \quad (5)$$

The number of geothermal wells needed is found using:

$$N_{probes} = \frac{\max(SL)}{C_{ground} \cdot (T_{bottom} - T_{amb}) \cdot h_{probe} \cdot A_{ground}} \quad (6)$$

with C_{ground} = 2009.8 kJ/(m^3·K) (Groupe de travail PGG 2011) the volumetric capacity of the ground.

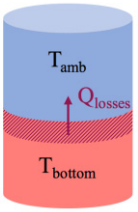


Figure 2. Schematics of stratified geothermal model

And the average temperature in the ground $T_{avg,ground}$ is computed as:

$$T_{avg,ground} = \frac{SL(t) \cdot T_{bottom} + (\max(SL) - SL(t)) \cdot T_{amb})}{\max(SL)} \quad (7)$$

2.3. Solar thermal collector

The solar thermal collector efficiency is modeled as a quadratic performance curve, depending on the external temperature T_{amb}, the incoming irradiation intensity GI $[\frac{W}{m^2}]$, a maximum efficiency η_0, and two experimental parameters a_1, a_2 (Duffie and Beckman 2013):

$$\eta = \eta_0 - a_1 \cdot \frac{T_m - T_{amb}}{GI} - a_2 \cdot \left(\frac{T_m - T_{amb}}{GI}\right)^2 \cdot GI \quad (8)$$

The heat transmission coefficient a_1 describes the thermal losses. The higher this value, the more losses the collector has. The coefficient a_2 is used to calculate the coefficient of heat transmission dependent on temperature. The average monthly solar thermal efficiency varies between 30 – 45%.

3. Results and discussion

The results are divided in three scenarios. In the first scenario, the heat from the environment is provided by a lake, in the second scenario it is provided by geothermal probes (coupled to solar thermal), and in the third scenario, besides geothermal probes and solar thermal, also the waste incineration plant is included. The number of probes/household (4 capita), the m^2/cap of PV panels and m^2/cap of solar thermal panels required for full autonomy are shown in Figure 3.

The m^2/cap of PV panels required decreases when geothermal probes are used instead of a lake since the electricity required for pumping is smaller than the electricity required in a central plant heat pump using lake water as a source. The m^2/cap of PV and solar thermal panels and the number of probes/ household (4 capita) decreases from 23.5 and 13.5 m^2/cap and 0.34 probes/household to 22.4 and 12.5 m^2/cap and 0.32 probes/household when the waste incineration plant is included, since the steam network integrated at the waste incinerator produces a part of the electricity required and of the low temperature heat which is used to vaporize CO_2 (instead of using heat from the geothermal probes).

The real potential for PV/solar thermal for a typical urban center of 11 m^2/cap (Figure 3) (Faessler et al. 2011) is not enough for full autonomy, so other sources of heat and electricity should be considered, such as wind, hydro energy or the electricity mix, and ground or waste water, respectively.

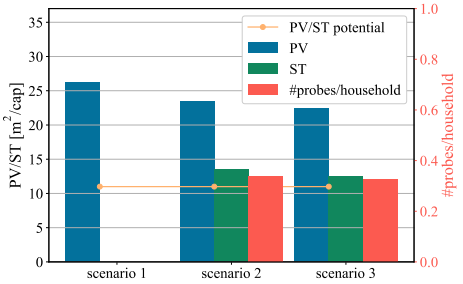


Figure 3. PV/ST area and #probes/household for a typical urban city

The heat flow profiles in and out of the ground are shown in Figure 4. One can see that heat is extracted from the ground between October and February, when the heat demand is high and injected in the ground in summer, when there is waste heat available from cooling and the solar thermal heat production is high.

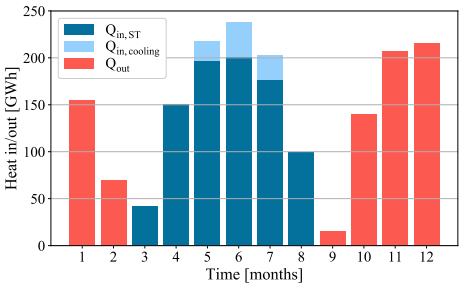


Figure 4. Heat flow in/out of the ground

The average temperature in the ground and the heating demand (space heating and domestic hot water) are shown in Figure 5. One can observe that in winter the heating demand is high, so heat is extracted from the ground and consequently the average temperature in the ground decreases, while in summer the heating demand is low, but the solar thermal heat production is high and therefore the average temperature in the ground increases.

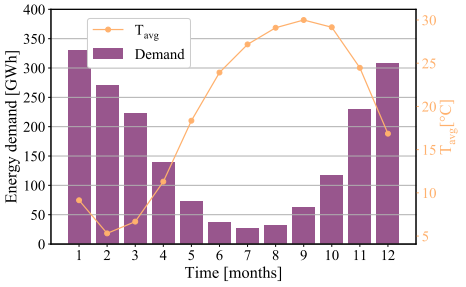


Figure 5. Average temperature in the ground and heating (SH and DHW) demand

4. Conclusion

This work studies the integration of low temperature CO_2 based district energy networks with geothermal probes as a ground energy source and solar thermal collectors as a heat source for the heat pumps and for the ground (to close the yearly balance). Three scenarios are studied depending on the source of the heat from the environment and on whether the waste incineration plant is considered or not. The PV area required for autonomy reduces when geothermal probes are used, since their electricity consumption is lower than that of a water source heat pump, and the PV and solar thermal areas and number of probes decrease when the waste incineration plant is included, since part of the electricity and of the heat required to vaporize CO_2 are produced by the steam network integrated at the waste incinerator.

The real potential for PV/solar thermal panels of 11 m^2/cap is not sufficient for autonomous urban centers, so other clean energy sources are suggested, such as wind or hydro as an electricity source and ground or waste water as a heat source.

References

Amblard, Frédéric Louis-Pierre Raphaël Marie. 2015. “Geothermal Energy Integration in Urban Systems. The Case Study of the City of Lausanne.” http://infoscience.epfl.ch/record/208829.

Bergman, Theodore L., and Frank P. Incropera. 2011. *Fundamentals of Heat and Mass Transfer*. John Wiley & Sons.

Bianchi, Mikael Andreas. 2006. “Adaptive Modellbasierte Prädiktive Regelung Einer Kleinwärmepumpenanlage.” ETH Zurich. https://doi.org/10.3929/ethz-a-005302150.

Duffie, John A., and William A. Beckman. 2013. *Solar Engineering of Thermal Processes: Duffie/Solar Engineering 4e*. https://doi.org/10.1002/9781118671603.

Faessler, Jérôme, Anthony Haroutunian, Pierre Hollmuller, and Bernard Marie Lachal. 2011. “Etude d'opportunité pour le projet GLU/GLA, https://archive-ouverte.unige.ch/unige:23583.

Focaccia, S., and F. Tinti. 2013. “An Innovative Borehole Heat Exchanger Configuration with Improved Heat Transfer.” *Geothermics* 48 (October):93–100. https://doi.org/10.1016/j.geothermics.2013.06.003.

Groupe de travail PGG. 2011. “EVALUATION DU POTENTIEL GEOTHERMIQUE DU CANTON DE GENEVE - PGG.” Vol1 : Rapport final, Vol.2 : Annexes, GADZ 5753/1, Genève.

Henchoz, Samuel. 2016. “Potential of Refrigerant Based District Heating and Cooling Networks.” https://doi.org/10.5075/epfl-thesis-6935, urn:nbn:ch:bel-epfl-thesis6935-2.

Marechal, François, and Boris Kalitventzeff. 2003. “Targeting the Integration of Multi-Period Utility Systems for Site Scale Process Integration.” *Applied Thermal Engineering* 23 (14):1763–1784.

Self, Stuart J., Bale V. Reddy, and Marc A. Rosen. 2013. “Geothermal Heat Pump Systems: Status Review and Comparison with Other Heating Options.” *Applied Energy* 101:341–348.

Steiger, D.W., and E.J. Kees. 1981. *Geothermal Heating System and Method of Installing the Same*. Google Patents. https://www.google.com/patents/US4286651.

Suciu, Raluca-Ancuta, Luc Girardin, and François Maréchal. 2017. “Energy Integration of CO2 Networks and Power to Gas for Emerging Energy Autonomous Cities in Europe.” *Proceedings of ECOS 2017*. https://infoscience.epfl.ch/record/232012?ln=en.

Weber, Celine, and Daniel Favrat. 2010. “Conventional and Advanced CO 2 Based District Energy Systems.” *Energy* 35 (12):5070–5081.

Wiggs, B. 2008. *Advanced Direct Exchange Geothermal Heating/Cooling System Design*. Google Patents. https://www.google.com/patents/US20080016894.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50304-9

Monte Carlo Simulation for Optimal Solar Cell Configuration

Aisha Alobaid[a], Raymond Adomaitis[a]

[a]*Department of Chemical and Biomolecular Engineering, Institute for Systems Research, University of Maryland, College Park, MD, 20742, USA*
aisha88@terpmail.umd.edu

Abstract

The objective of this work is to develop computational tools to identify optimal configurations when integrating a large number of solar cells under shady or faulty conditions. The photovoltaic module that will be considered in this analysis consists of 96 cells connected in series, where an equal number of solar cells, called a string, is connected to a bypass diode. Both event-driven Monte Carlo simulation and Multi-Attribute Utility Theory are used to evaluate the effect of the number of bypass diodes on the photovoltaic module maximum power point, and to find the optimal configuration needed to minimize performance losses under uncertain shading conditions. The optimal cell configuration is established when the total value function of the module is maximized. Ideally, it was found that the optimal situation corresponds to having one bypass diode connected to each cell in the module; however, by factoring in diode cost, it was found that one bypass diode wired to each group of four solar cells was economically more practical.

Keywords: Solar energy, Modeling, Monte Carlo simulation.

1. Introduction

Current trends in electrical power generation and energy storage research highlight the promise of micro- and nano-scale devices (Fan et al., 2015 and Krogstrup et al., 2013). Performance issues arise when integrating these devices, as design and performance mismatch between devices can cause deviation from the expected performance in a highly nonlinear manner. The tera- and giga-scale levels of integration necessary for producing and storing macroscopic levels of energy from nano-scale devices necessitates device integration strategies that are resilient to the inevitable faults and variability that will arise in the individual devices. Thus, new computational tools are needed to support the design, integration, and (advanced) manufacturing of these energy systems. As a first step towards developing simulation-based design tools for the integration of nano-scale energy devices, we will examine the problem of determining the optimal strategy to integrate a large number of photovoltaic (PV) solar cells under shady or faulty conditions as a proxy for our primary simulation objective.

When one cell in an array of PV cells is shaded, its voltage polarity may reverse resulting in power consumption instead of production, leading to hotspot formation, reduced overall power, and even PV module damage. Bypass diodes can be used to route current around shaded or faulty cells to partially recover the lost performance. In this paper, we focus on modeling tools useful for optimizing the number and placement of these diodes within a PV cell array.

Rao et al. (2014) investigated the effect of different interconnection schemes on the output of a PV array under partial shading conditions; the potential benefits of bypass diodes, however, was not included in their analysis. Vemuru et al. (2012) examine the effect of random shading of a PV array on the output power. They concluded that the power of the randomly shaded array is reduced significantly when no bypass diodes are used, as opposed to having a fixed number of bypass diodes in the array for the same shading conditions. Silvestre et al. (2009) used the Pspice simulation environment to study the effect of bypass diode configuration on a PV module. They presented a finite number of simulation cases in which selected cells were shaded, and demonstrated how the configuration of a fixed number of bypass diodes affected the results.

In this work, we introduce the stochastic effects and uncertainty generated by shading, which simulate real-life operating conditions. We use event-driven Monte Carlo simulations, in which each event corresponds to a separate shading case. An advantage of our choice of MC simulation is that it can be extended to any shading pattern caused by trees, falling tree leaves, standing birds, close-by buildings, and even passing clouds, once the statistical distribution of the shading effect is known or can be approximated.

2. Model and Simulation

2.1. Governing Equation

The PV module considered in this analysis consists of 96 PV cells connected in series, where strings (subsets of the full set) of solar cells are wired in parallel to a bypass diode, as shown in Figure (1). The extended diode equation, Eq.(1), is used to model the current-versus-voltage characteristics of each PV cell (Silvestre et al., 2009). The last term in the equation describes the diode avalanche breakdown at high negative voltages of the cell (Ben Or and Appelbaum, 2013).

$$I = XI_{ph} - I_o\left[exp\left(\frac{V+IR_s}{mV_{th}}\right)-1\right]-\left(\frac{V+IR_s}{R_{sh}}\right)\left[1+a\left(1-\frac{V+IR_s}{V_b}\right)^{-\beta}\right] \quad (1)$$

The diode equation parameters and their values used in this analysis are: photocurrent I_{ph}=2.443 A, dark saturation current I_o=2.935×10^{-12} A, series resistance R_s= 14.49×10^{-3} Ω, shunt resistance R_{sh}=3872 Ω, diode ideality factor m=1, thermal voltage V_{th}= k_BT/q = 2.61×10^{-2} V, breakdown voltage V_b= -25 V (Silvestre et al., 2009), and bishop's term parameters a=2×10^{-3} and β=4 (Silvestre et al., 2009). These parameters are intended to approximate a SANYO HIP-205BA3 PV module. In this simulation, the shaded cells are assumed to be 50 % shaded (X=0.5).

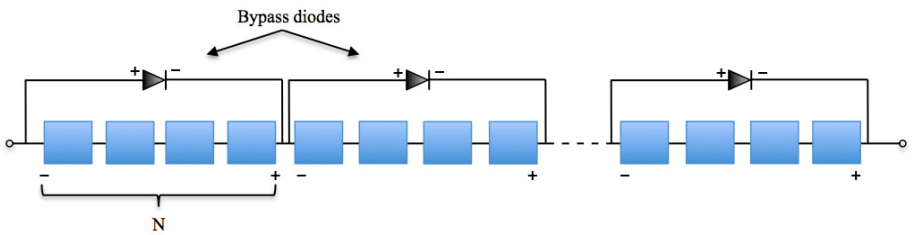


Figure (1): Schematic diagram showing the system's configuration.

2.2. Calculating the current of a string with shaded cells

To calculate the current I of a string of N solar cells wired in series, we define C_t as the number of shaded solar cells in the string, V_1 the voltage across an unshaded cell, and V is the total voltage of the string to find the voltage across a shaded cell as

$$V_{sh} = \frac{V - (N - C_t)V_1}{C_t} \tag{2}$$

Substituting V_1 and V_{sh} into Eq.(1), we obtain two diode equations: one for unshaded cells, and the other for shaded cells. Because the currents of PV cells connected in series are identical, we can set both equations equal to each other to find one function of I, V, and V_1. This equation can be solved numerically to find current values I at different string voltages V. However, for every V value we assume, we have two unknowns, I and V_1, and hence, a two-dimensional iterative loop is used. For a known value of V, first we assume I = 0, and using a Newton procedure, we find the corresponding value of V_1. Then the inner loop will re-calculate the value of I as a function of the previous V_1 and I values. The next iteration would use the most recent V_1 and I values to solve for V_1 again, and then I is re-examined in the inner loop. The stopping criterion corresponds to reaching converged V_1 and I values. This procedure is carried out for each value in the vector V, resulting in I-V data for the shaded string.

2.3. Monte Carlo Simulation

For the MC simulation, both the number of partially shaded solar cells and their position are considered random variables. The number of shaded cells is assumed to follow a uniform distribution ranging from 1 to 5 shaded cells, and the position of the shaded cells also follows a uniform distribution ranging from cell 1 to cell 96. The position of the shaded cell is important since it determines the number of shaded cells in each string, and the number of strings with shaded cells.

We developed a MATLAB program to calculate the output maximum power point for each run of the MC simulation. At the start of the program, we have a vector of values for the voltage ranging from 0 to the open circuit voltage of the string. The main idea is to determine at each voltage point whether the bypass diode for each string would be activated or not under the shading conditions randomly provided. We assigned 12 cases representing the number of bypass diodes (1, 2, 3, 4, 6, 8, 12, 16, 24, 32, 48, 96), and for every case, we run the MC simulation for 1,000 shading events, each with different random number of shaded cells and random positions, to obtain the module maximum power point P_{max}, its standard deviation, and standard error.

For each value of the voltage, we calculate its corresponding current following the model presented in sections 2.1 and 2.2. The string with the highest current will force the shaded strings to pass the same current. As a result, the bypass diode will be activated if the shaded strings resulted in a negative voltage.

When recalculating the voltage of the shaded strings, we might encounter high values of negative voltage (close to cell breakdown voltage), and the Newton procedure converges slowly or sometimes even diverges (Quaschning and Hanitsch, 1996). Hence,

in some cases, the bisection method is used to estimate the initial guess for the Newton procedure. A summary of the simulation algorithm is shown in the SysML activity diagram in Figure (2).

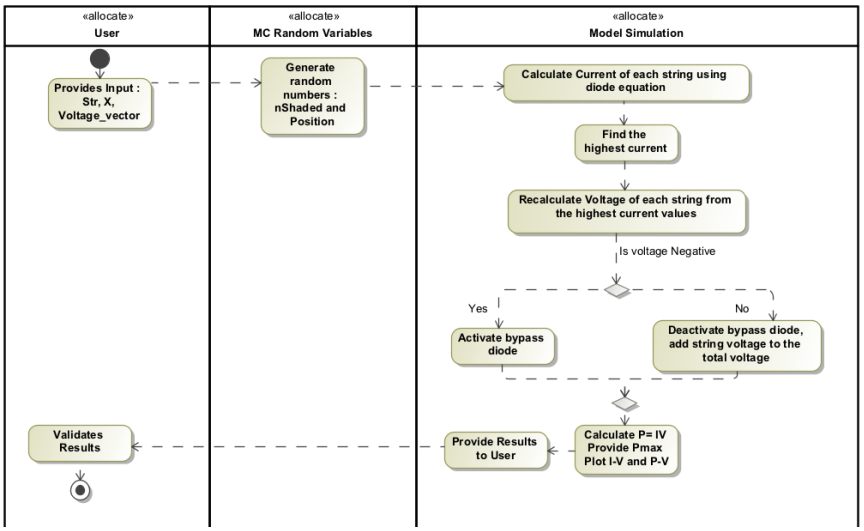


Figure (2): Simulation Algorithm Activity Diagram. Note that "Str" is the number of strings.

2.4. Multi-Attribute Utility Theory (MAUT)

Because the MC simulation results clearly demonstrated the effectiveness of using the highest number of bypass diodes (Fig. 4), we next considered the cost of adding the bypass diodes to make the final design decision. Hence, MAUT was used, with a linear additive utility function, and a weight (w) of 0.6 for cost value function, and 0.4 (1-w) for performance value function. The cost of each diode was set to be \$3, an approximate cost of Microsemi Active Bypass LX2400 (Microsemi, 2017). The cost value function V_c was normalized to be between 0 and 1, with a value of 1 corresponding to having no bypass diode, and a value of 0 corresponding to having the maximum number of bypass diodes (96×\$3=\$288). The performance value function V_p was also normalized to be between 0 and 1, with a value of 1 for the maximum power value among all bypass diodes, and 0 for the minimal power value among all bypass diodes. The total value function V(C,P) is given in Eq.(3).

$$V(C,P) = wV_c(C) + (1-w)V_p(P) \tag{3}$$

3. Analysis and Results

To examine the effect of the number of bypass diodes on the module performance, the module is simulated with only one shaded cell. As seen in Figure (3), as the number of bypass diodes increases, the output power will also increase, minimizing performance losses compared to the unshaded case. Hence, the effect of the shaded cell is minimized with increasing the number of bypass diodes.

The MC simulation results are provided in Figure (4). As the number of bypass diodes increases, the maximum power also increases. This result is expected since increasing the number of diodes decreases the number of illuminated PV cells bypassed. Even though increasing the number of bypass diodes may lead to increased leakage currents in these diodes, this effect is highly compensated by the positive benefits of adding the diodes.

For the MAUT analysis, Figure (5) shows the normalized total utility value function calculated. For the given diode cost and weights on each attribute, the optimum configuration is achieved by using 24 bypass diodes, at which P_{max} is 119.31 ± 0.20 W, and the total diode cost is $ 72.

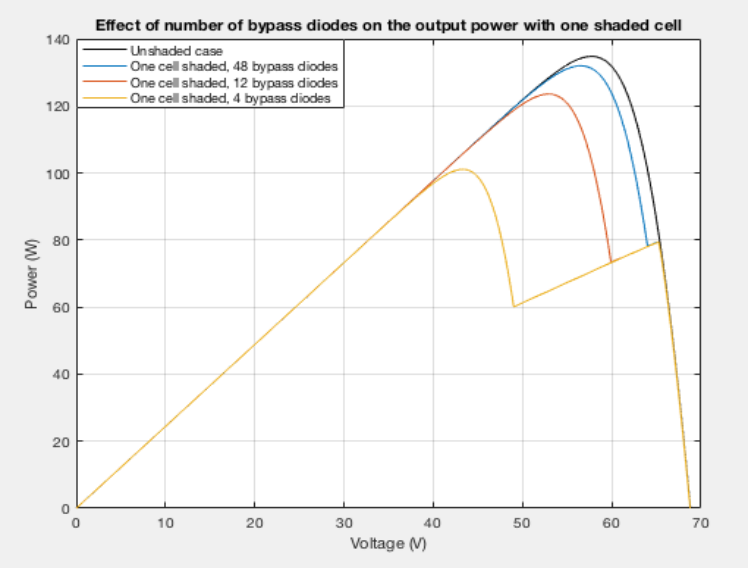


Figure (3): P-V curve for one shaded cell at different bypass diodes

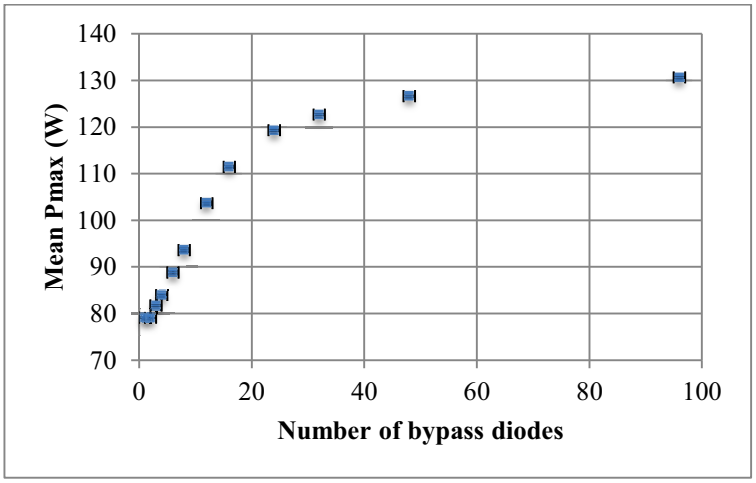


Figure (4): MC simulation results for each bypass diode case.

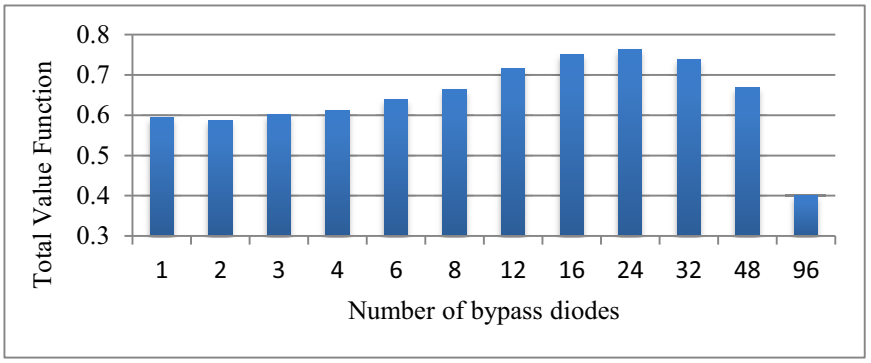


Figure (5): MAUT total value function

4. Conclusion

A computational algorithm was developed for optimizing PV cell network design. In this study, the effect of using bypass diodes in a PV module with shaded cells was examined. Ideally, it was found that the optimal situation corresponds to having one bypass diode connected to each cell in the module; however, by factoring in diode cost, it was found that one bypass diode wired to each group of four PV cells was economically more practical.

The algorithm presented in this work was able to find the optimum configuration of bypass diodes for any module with PV cells wired in series. The MC simulation allows for consideration of different shading patterns once the statistical distribution of the shading effect can be approximated. The approach developed in this paper can be extended to include randomly distributed manufacturing and other defects.

References

A. Ben Or, and J. Appelbaum, 2013, Estimation of multi-junction solar cell parameters, Progress in Photovoltaics: Research and Applications, 21, 4, 713-723.

X. Fan, C. Luo, J. Lamb, Y. Zhu, K. Xu, C. Wang, 2015, PEDOT encapsulated FeOF nanorod cathodes for high energy lithium-ion batteries, Nano Letters, 15, 7650–7656.

P. Krogstrup, H. Jorgensen, M. Heiss, O. Demichel, J. Holm, M. Aagesen, J. Nygard and A. Morral, 2013, Single-nanowire solar cells beyond the Shockley–Queisser limit, Nature Photonics, 7, 306–310.

Microsemi, 2017, Solar bypass TCO calculator spreadsheet [Microsoft Excel spreadsheet], Retrieved from https://www.microsemi.com/product-directory/diodes/622-solar-array-diode#overview.

V. Quaschning, and R. Hanitsch, 1996, Numerical simulation of current-voltage characteristics of photovoltaic systems with shaded solar cells, Solar Energy, 56, 6, 513-520.

P. Rao, G. Ilango, and C. Nagamani, 2014, Maximum power from PV arrays using a fixed configuration under different shading conditions, IEEE Journal of Photovoltaics, 4, 2, 679–686.

S. Silvestre, A. Boronat, A. Chouder, 2009, Study of bypass diodes configuration on PV modules, Applied Energy, 86, 9, 1632-1640.

S. Vemuru, P. Singh, M. Niamat, 2012, Modeling impact of bypass diodes on photovoltaic cell performance under partial shading, IEEE International Conference on Electro/Information Technology EIT.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler(Editors)Proceedings of the
13thInternational Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018,San Diego,California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50305-0

Comparative Study on Heat Transfer Efficiency of CRR Based on Entropy Analysis and Entransy Analysis

Li Xia, Yiting Zeng, Yuanli Feng, Xiaoyan Sun, Shuguang Xiang*

College of Chemical Engineering, Qingdao University of Science & Technology, Qingdao 266042, China
xsg @ qust. edu. cn

Abstract

The heat transfer efficiency of continuous reforming unit (CRR) is studied based on entropy analysis and entransy analysis. By calculating the entropy change and entransy of cold and hot streams, the transfer efficiency of entropy and entransy is obtained respectively. When the minimum temperature difference increases, the entropy transfer efficiency increases. It appears contrary to the principle of minimum entropy generation. When the minimum temperature difference increases, the entransy transfer efficiency decreases. It is proved that the greater the temperature difference, the lower the heat transfer ability. The theory of entransy can avoid the violation of the minimum principle of entropy production. Thus, compared with the entropy analysis, the entransy analysis has obvious advantages in analyzing the heat transfer efficiency of the CRR unit.

Keywords: entransy, entransy dissipation, entropy, heat transfer, the heat transfer efficiency, heat exchanger networks

1. Introduction

Many theories have been developed for the analysis of energy utilization efficiency of heat exchanger networks (HENs), such as entropy analysis and entransy analysis.

Bejan (1996) proposed the principle of minimum entropy production. However, when studying non equilibrium thermodynamics, it is found that the principle of minimum entropy production does not satisfy fourier law (Bertola and Cafaro, 2008). Guo et al. (2007) introduced the new physical quantity called entransy to describe the heat transfer capacity of objects. In recent years, some scholars have introduced the idea of entransy into HENs synthesis (Xia et al., 2016), such as heat exchanger design (Xia et al., 2017a), graphical method (Chen et al., 2012), computing the entransy transfer efficiency for the petrochemical unit (Xia et al., 2017b).

The entropy analysis and entransy analysis can describe the irreversibility of the heat transfer process. For the analysis of energy utilization efficiency of CRR unit, which is better? Then, the two methods for the HENs of the CRR unit were further compared.

2. Model of entropy analysis and entransy analysis

2.1. Entropy analysis model of HENs

The entropy analysis is based on the second law of thermodynamics. Select the N_h (hot streams) and N_c (cold streams). There is no heat loss and shaft power in the process of heat exchange. There is no fluid phase transition in the process. The heat flow rate of hot stream is CP_h and the heat flow rate of cold stream is CP_c. The temperature range from T_{in} to T_{out}, the entropy of hot and cold streams can be written as:

$$\Delta S_{h,i} = CP_{h,i} \ln \frac{T_{h,i,out}}{T_{h,i,in}} \tag{1}$$

The enthalpy change is:

$$\Delta H_{h,i} = CP_{h,i}(T_{h,i,out} - T_{h,i,in}) \tag{2}$$

$$\Delta H_h = \sum_{i=1}^{h_n} \Delta H_{h,i} \tag{3}$$

At ambient temperature T_0, the ideal work can be written as:

$$W_{id,c,i} = -CP_{c,i}(T_{c,i,out} - T_{c,i,in}) + T_0 CP_{c,i} \ln \frac{T_{c,i,out}}{T_{c,i,in}} \tag{4}$$

$$W_{id,h,i} = -CP_{h,i}(T_{h,i,out} - T_{h,i,in}) + T_0 CP_{h,i} \ln \frac{T_{h,i,out}}{T_{h,i,in}} \tag{5}$$

$$\Delta W_{id,c} = \sum_{i=1}^{c_n} \Delta W_{id,c,i} \tag{6}$$

$$\Delta W_{id,h} = \sum_{i=1}^{h_n} \Delta W_{id,h,i} \tag{7}$$

Because the efficiency of the second law of thermodynamics is

$$\eta_{II,0} = \frac{W_S}{W_{id}} = \frac{W_{id} - W_L}{W_{id}} \tag{8}$$

The ideal work done by hot stream from T_{out} to T_{in} is W_{id}. The difference between the ideal work given by the hot stream and the cold stream is W_L, which can be written as:

$$W_L = \Delta W_{id,h} - \Delta W_{id,c} \tag{9}$$

Therefore, the entropy transfer efficiency is:

$$\eta_{II,0} = \frac{W_{id,c}}{W_{id,h}} = \frac{\sum_{i=1}^{n} T_{m,h,i}(T_0 - T_{m,c,i})}{\sum_{i=1}^{n} T_{m,c,i}(T_{m,h,i} - T_0)} \tag{10}$$

$$T_{m,h,i}=\frac{T_{h,i,out}-T_{h,i,in}}{\ln\frac{T_{h,i,out}}{T_{h,i,in}}},T_{m,c,i}=\frac{T_{c,i,out}-T_{c,i,in}}{\ln\frac{T_{c,i,out}}{T_{c,i,in}}} \tag{11}$$

2.2. Entransy analysis model of HENs

The entransy analysis can be used to describe the efficiency of heat transfer (Xia et al. 2017c).According to the hypothesis of the model of a two-stream heat transfer process with entransy theory (Xia et al., 2017c), the HENs generally with constant pressure and constant volume, the constant volume specific heat of C_V and constant pressure specific heat C_P are approximately equal (Gesellschaft, 2010).

The entransy of hot and cold streams can be expressed as:

$$E_H=\sum_{i=1}^{h_n}E_{h,i},\ E_C=\sum_{i=1}^{c_n}E_{c,i},\ E_{h,i}=\frac{1}{2}CP_{h,i}(T_{h,i,in}^2-T_{h,i,\mathrm{out}}^2),\ E_{c,i}=\frac{1}{2}CP_{c,i}(T_{c,i,\mathrm{out}}^2-T_{c,i,in}^2) \tag{12}$$

The entransy dissipation can be written as:

$$\Delta E=\sum_{i=1}^{h_n}E_{h,i}-\sum_{i=1}^{c_n}E_{c,i}=\frac{1}{2}\sum_{i=1}^{h_n}CP_{h,i}(T_{h,i,in}^2-T_{h,i,out}^2)-\frac{1}{2}\sum_{i=1}^{c_n}CP_{c,i}(T_{c,i,out}^2-T_{c,i,in}^2) \tag{13}$$

Therefore, the efficiency of the entransy is:

$$\eta_0=\frac{E_C}{E_H}=\frac{\sum_{i=1}^{n}CP_{c,i}(T_{c,out,i}^2-T_{c,in,i}^2)}{\sum_{i=1}^{n}CP_{h,i}(T_{h,in,i}^2-T_{h,out,i}^2)} \tag{14}$$

Where CP denotes heat capacity flow rate, CP=mass flow m × constant pressure specific heat C_p.

3. Comparison of entropy analysis and entransy analysis for CRR unit

The pretreatment process of CRR unit is shown in figure 1. The reforming process of CRR unit is shown in figure 2.

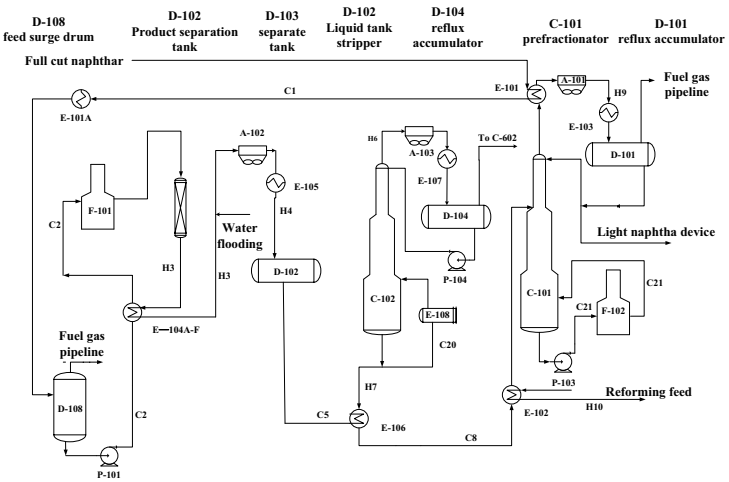


Figure 1.The pretreatment process of CRR unit

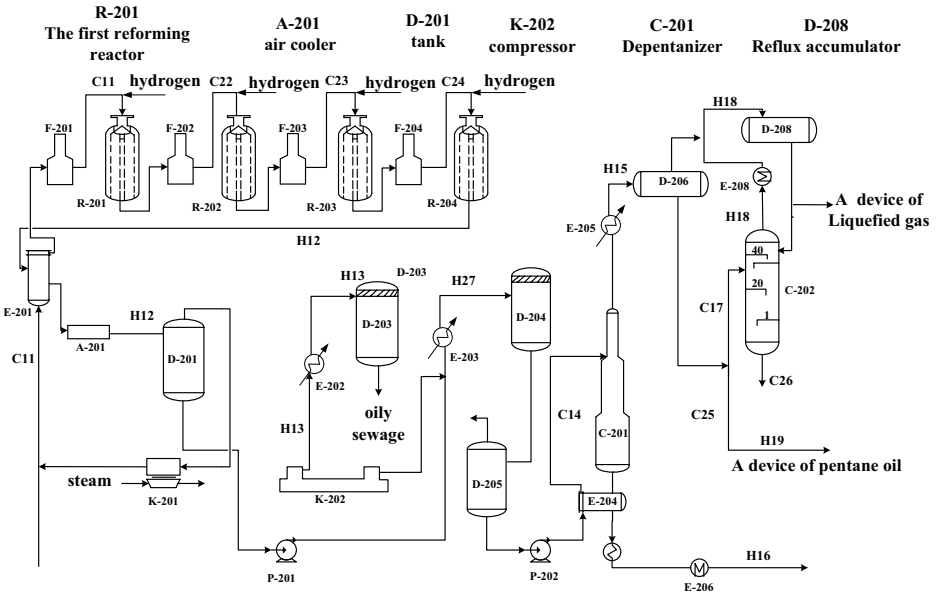


Figure 2. The reforming process of CRR unit

Extract the data of hot and cold streams, including hot and cold utilities. The extracted streams data include initial temperature, target temperature, heat capacity flow rate and heat load. The data of hot and cold streams is shown in table 1 and table 2.The environmental temperature T_0 is 298.15K, and the pressure is 0.1013MPa. The results calculated by entropy analysis and entransy analysis are shown in table 3 and table 4 respectively.

Table 1 The data of hot streams

Stream	Stream description	Supply Temperature [K]	Target Temperature [K]	CP [kW·K^{-1}]
H3A	reactor streams	278	215	53.2
H3B	reactor streams	215	144	100.5
H4	Streams mixed with water injection	109	40	61.3
H6	The stream of stripping tower	114	41	8.6
H7	The stream of stripping tower	189	112	46.3
H9	The top stream of Prefractionator	119	40	50.4
H10	The top stream of Prefractionator	182	141	58.9
H12A	Reforming product	467	132	75.8
H12B	Reforming product	132	55	136.2
H13	The stream of compressor K201 after primary compression	150	35	19.7
H15	The top stream of pentane tower	139	40	12
H16	The top stream of pentane tower	250	44	51.6
H18	The top streams of separator	80	39	6.6
H19	The top streams of separator	155	62	1.8
H27	The stream of tank	77	46	64.2

According to table 3, while the entropy transfer efficiency is 62.703%, 63.008%, 63.305%, saving utilities is 30.25 %, 27.18 %, 25.32 %. It is indicated that the lower entropy transfer efficiency it is, the more utilities you can save. The result of entropy

analysis for CRR unit is a clear violation of the minimum principle of entropy production.

According to table 4, when minimum temperature differences is 15 K, 20 K and 25 K, the entransy transfer efficiency is 75.90%, 73.97%, 72.11%, saving utilities is 25.38%, 22.62%, 19.93%. It is indicated that the higher entransy transfer efficiency it is, the more utilities you can save.

Table 2 The data of cold streams

Stream	Stream description	Supply Temperature [K]	Target Temperature [K]	CP [kW·K^{-1}]
C1	Raw oil	30	78	40.1
C2	The streams of mixture of liquid and gas at bottom of flash tank	76	271	73.6
C5	Stripper feed	40	124	42.5
C8	Prefractionator feed	108	144	67.1
C11A	Reforming feedstock	115	167	152.8
C11B	Reforming feedstock	167	517	70.5
C14	Feeding of pentane tower	46	177	51.5
C17	The separation column feed	36	75	2.7
C20	Gas stripping at the bottom of the streams	189	194	712.5
C21	Demineralized water	182	188	862.7
C22	The stream of F202	402	519	79.9
C23	The stream of F203	447	521	81.6
C24	The stream of F204	463	517	82
C25	The bottom streams	250	260	496.2
C26	The stream of separation tower	155	157	194.1

Table 3 The results of entropy analysis for CRR unit

Streams	items	ΔT_{min}=15K	ΔT_{min}=20K	ΔT_{min}=25K
Hot streams	entropy[kW·K^{-1}]	1.678×10^{2}	1.668×10^{2}	1.659×10^{2}
	ideal work[kW]	2.446×10^{4}	2.474×10^{4}	2.503×10^{4}
Cold streams	Entropy[kW·K^{-1}]	1.794×10^{2}	1.785×10^{2}	1.776×10^{2}
	ideal work[kW]	3.899×10^{4}	3.926×10^{4}	3.954×10^{4}
Hot utilities	entropy[kW·K^{-1}]	2,237	2.501	3.143
	ideal work[kW]	7.150×10^{2}	8.247×10^{2}	8.343×10^{2}
Cold utilities	entropy[kW·K^{-1}]	9.442	10.283	11.608
	ideal work[kW]	1.698×10^{3}	1.817×10^{3}	1.930×10^{3}
entropy transfer efficiency [%]		62.703	63.008	63.305
Energy saving[%]		30.25	27.18	25.32

Table 4 The results of entransy analysis for CRR unit

items	ΔT_{min}=15K	ΔT_{min}=20 K	ΔT_{min}=25 K
Hot streamsentransy[MW·K]	35.55	35.55	35.55
Cold streamsentransy [MW·K]	51.07	51.07	51.07
Hot utilitiesentransy[MW·K]	24.09	24.78	25.44
Cold utilitiesentransy[MW·K]	5.972	6.397	6.82
Entransyrecovery [MW·K]	26.98	26.29	25.63
Entransydissipation [MW·K]	2.594	2.857	3.094
Entransy transfer efficiency [%]	75.90	73.97	72.11
Entransy saving [%]	25.38	22.62	19.93

4. Conclusions

The heat transfer efficiency of the CRR unit is analyzed by entropy analysis and entransy analysis. The main conclusions are as follows:

(1) The entropy analysis and entransy analysis can describe the irreversibility of heat transfer processes. Entropy analysis pays more attention to the work of hot streams under completely reversible conditions. The entropy can not calculate the energy loss.
(2) When the minimum temperature difference is 15 K, 20 K and 25 K respectively, the entropy transfer efficiency is 62.703%, 63.008%, 63.305%. The result is a clear violation of the minimum principle of entropy production. However, the entransy transfer efficiency is 75.90%, 73.97%, 72.11%. The theory of entransy can avoid the violation of the minimum principle of entropy production.
(3) The calculation of entropy, the ideal work, the entropy transfer efficiency is related to the ambient temperature in the entropy analysis. The calculation of the entrancy and the entrancy transfer efficiency is independent of the ambient temperature in the entrancy analysis. The entrancy analysis is more convenient than entropy analysis.

Acknowledgement

Financial support for carrying out this work was provided by the National Natural Science Foundation of China (21406124).

References

A.Bejan,1996, Entropy generation minimization: the new thermodynamics of finite-size devices and finite-time processes. Journal of Applied Physics, 79, (3), 1191. https://doi.org/10.1063/1.362674

V. Bertola, E. Cafaro, 2008,A critical analysis of the minimum entropy production theorem and its application to heat and fluid flow. International Journal of Heat and Mass Transfer, 51, (7-8), 1907-1912. https://doi.org/10.1016/j.ijheatmasstransfer.2007.06.041

Q. Chen, Y.C. Xu, Z.Y.Guo, 2012, The property diagram in heat transfer and its applications,Chinese Science Bulletin,57, (35), 4646-4652. https://doi.org/10.1007/s11434-012-5476-5

Z. Y. Guo,H. Y. Zhu, X.G.Liang, 2007, Entransy-A physical quantity describing heat transfer ability,International Journal ofHeat and Mass Transfer , 50, (13-14), 2545–2556. https://doi.org/10.1016/j.ijheatmasstransfer.2006.11.034

VDI-Gesellschaft, 2010,VDI Heat Atlas, Springer, Berlin.

L.Xia,Y. l. Feng, S. G. Xiang,2016,Progress andapplication of entransy theory in energy saving of chemical processes, CIESC Journal, 67, (12), 4915– 4921. http://www.hgxb.com.cn/EN/Y2016/V67/I12/4915

L. Xia, Y. L. Feng, X. Y. Sun, S. G. Xiang, 2017a,A novel method based on entransy theoryfor setting energy targets of heat exchanger network,Chinese Journalof Chemical Engineering, 25, (8), 1037-1042. https://doi.org/10.1016/j.cjche.2017.03.015

L. Xia, Y. L. Feng, S. G. Xiang, 2017b, Comparative study on heat exchanger network adaptability based on entropy analysis and entransy analysis, Chemical Industry and Engineering Progress, 36, (10), 3657-3664. http://www.hgjz.com.cn/EN/Y2017/V36/I10/3657

L.Xia, Y.L. Feng,S.G.Xiang, 2017c, Comparative study of adaptability for S-zorb unit based on exergy analysis and entransy analysis, Chemicai Engineering Transactions, 61, (10), 1825-1830. https://dx.doi.org/10.3303/CET17613092

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50306-2

Simultaneous Removal of Organic Compounds from Wastewater Using Reverse Osmosis Process: Modelling, Simulation, and Optimisation

Mudhar A. Al-Obaidi[a,b], Chakib Kara-Zaïtri[a] and Iqbal M. Mujtaba[a*]

[a]*Chemical Engineering, School of Engineering, Faculty of Engineering and Informatics. University of Bradford. Bradford, West Yorkshire BD7 1DP, UK*
[b]*Middle Technical University, Iraq – Baghdad*
I.M.Mujtaba@bradford.ac.uk

Abstract

The modern industrialized world is generating a huge amount of wastewater containing a variety of micro-pollutants which is discharged into rivers and oceans leading to disruption of the biological ecosystem. The reverse osmosis (RO) process is one of the most promising technologies to produce high-quality recycling water at a reasonable cost. In this work we develop a new mathematical model to predict the performance of the RO process for the removal of several organic and non-organic compounds from wastewater simultaneously. The realistic operating conditions ensuring high rejection of multi-compounds is explored via the simulation of the RO process first. The model is then used for optimising the operating conditions of the process while maximising the rejection and permeate recovery.

Keywords: RO Process, Wastewater Treatment; Modelling; Simulation; Optimisation; Multi-compound rejection.

1. Introduction

The main concern of any industrial wastewater is the existence of several organic and non-organic compounds which are harmful to human beings and marine life. Despite the availability of several wastewater treatment methods such as adsorption, biological, oxidation, and electrochemical process; the Reverse Osmosis (RO) process has demonstrated its competitive performance at a reasonable cost for the removal of these compounds. Several previous studies modelled the spiral wound RO process considering the removal of a single organic contaminant from wastewater (Sundaramoorthy et al., 2011; Al-Obaidi et al., 2017). However, only a few attempted the modelling of the spiral wound RO process for the removal of several organic and non-organic compounds from waste water. Al-Bastaki (2004) developed a lumped model to study the performance of a spiral wound RO process for removing Na_2SO_4 and methyl orange dye from wastewater. Note however, the development of a distributed model for the spiral wound RO process for the removal of multi-compounds simultaneously from wastewater has not been considered yet and is the main focus of this work. Also note, in the absence of experimental data in the literature for multi-compounds removal from wastewater, the model developed in this work has been validated against experimental data available for the removal of a single compound (chlorophenol) from wastewater. The model is then used to examine the rejection of organic and non-organic compounds under various operating parameters which in turn evaluates the process performance. This is followed

by embedding the model in a multi-objective optimisation framework to simultaneously maximise the rejection and the total permeate recovery.

2. RO Process Model

In this work RO filtration system of Sundaramoorthy et al. (2011) is used to investigate the simultaneous removal of multi-compounds from wastewater. Table 1 shows the new one-dimensional model equations for the process based on solution-diffusion model and the film theory. See Al-Obaidi et al. (2017) for a similar model but for single compound rejection. Due to the multi-compounds nature of the wastewater, the contribution of all of them to the osmotic pressure was considered in the new model. The model is validated using experimental data of chlorophenol removal from wastewater (Sundaramoorthy et al., 2011) which shows a very good agreement in terms of rejection and recovery of water (Fig. 1). The process model shown in Table 1 can be written in the compact form: f(x, u, v) = 0, where x is the set of all algebraic variables, u is the set of decision variables and v denotes the constant parameters of the process. The physical properties of seawater were used to calculate the diffusivity (D_b), viscosity (μ_b, μ_p) and density (ρ_b, ρ_p) due to low concentration of wastewater. The characteristics of the membranes used are given in Table 2. Five organic compounds and three inorganic species are assumed to be in the wastewater. The solute transport parameters (B_s) of the selected compounds were gathered from the literature and given in Table 3.

3. Process simulation: Effect of operating parameters

Fig. 2 shows the variation of feed pressure, osmotic pressure, and water flux along the membrane length (x-axis). The feed pressure decreases along the x-axis due to pressure drop caused by the friction. This in turn reduces the water flux as a result of decreasing diving force. An increase of total osmotic pressure along the x-axis is noticed due to increasing accumulated concentration of the solutes at the membrane wall.

The variation of inlet feed concentration of the pollutants found in wastewater is expected. Therefore, the RO process performance is investigated here using a range (350 to 500 ppm) of each pollutant concentration carried out at fixed feed flow rate, pressure, and temperature. Fig. 3 shows the effect of increasing feed concentration of each component on the rejection and total permeate recovery. In the selected range of 350 – 500 ppm of solute feed concentration, there was no considerable effect on the rejection. However, increasing the feed concentration of all the components causes a continuous reduction in permeate recovery due to increased osmotic pressure.

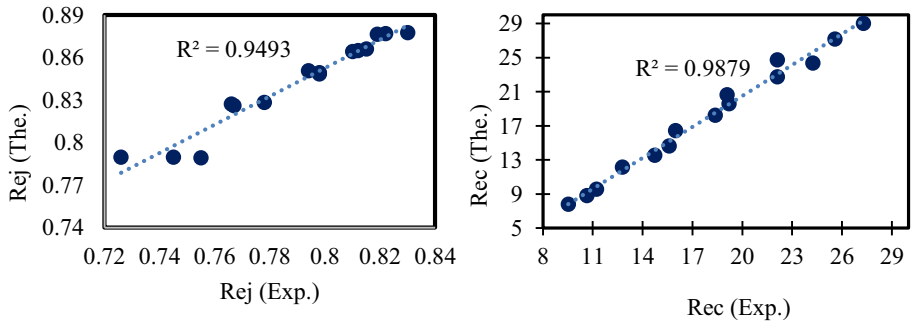


Fig. 1. The model validation results

Table 1. Model equations of a spiral wound RO for multi-compounds wastewater

Model Equations	Specifications	no.
$J_{w(x)} = A_w \left\{ \Delta P_{b(x)} - \left[\sum_i^n (R(T_b + 273.15) \left(\frac{J_{w(x)}\, C_{p(i)(x)}}{B_{s(i)}} \right) \right] \right\}$ i, n are the solute under consideration and total number of solutes	Permeate flux (m/s) at any point along the x-axis. The second term is total osmotic pressure (atm)	1
$\frac{dF_{b(x)}}{dx} = -W J_{w(x)} \quad U_{b(x)} = \frac{F_{b(x)}}{t_f W} \quad \frac{dP_{b(x)}}{dx} = -b F_{b(x)}$ $\frac{dF_{b(x)}}{dx} = -\frac{dF_{p(x)}}{dx} \quad \Delta P_{b(x)} = P_{b(x)} - P_p$	Feed flow rate (m³/s), velocity (m/s), pressure (atm) and permeate flow rate (m³/s) and operating pressure (atm)	2,3, 4,5
$\frac{C_{b(i)(x)}}{t_f W} \frac{dF_{b(x)}}{dx} + \frac{F_{b(x)}}{t_f W} \frac{dC_{b(i)(x)}}{dx} = \frac{d}{dx}\left[D_{b(i)(x)} \frac{dC_{b(i)(x)}}{dx} \right] - \frac{\left(J_{w(x)}\, C_{p(i)(x)} \right)}{t_f} + \frac{\left(J_{w(x)}\, C_{b(i)(x)} \right)}{t_f}$	Feed solute concentration (kmol/m³)	6,7
$k_{(i)(x)} = 1.177 \left(\frac{U_{b(x)}\, D_{b(i)(x)}^2}{t_f L} \right)^{0.333}$	Mass transfer coefficient (m/s) of solute (Wankat, 1990)	8
$k_{(i)(x)}(2t_f) = 246.9\, D_{b(i)(x)}\, Re_{(x)}^{0.101}\, Re_{p(x)}^{0.803}\, C_{m(i)(x)}^{0.129}$	Dimethylphenol (Srinivasan et al., 2011)	9
$k_{(i)(x)}(2t_f) = 147.4\, D_{b(i)(x)}\, Re_{(x)}^{0.13}\, Re_{p(x)}^{0.739}\, C_{m(i)(x)}^{0.135}$	Chlorophenol (Sundaramoorthy et al., 2011)	10
$C_{m(i)(x)} = \frac{C_{b(i)(x)}}{\rho_w} \qquad Re_{(x)} = \frac{2\, \rho_{b(av)(x)}\, F_{b(x)}}{W\, \mu_{b(av)(x)}}$ $Re_{p(x)} = \frac{2\, \rho_{p(av)(x)}\, t_p\, J_{w(x)}}{\mu_{p(av)(x)}}$	Dimensionless solute concentration and Reynolds number at feed and permeate channels. ρ_w: The molal density of water (55.56 kmol/m³)	11, 12, 13
$C_{p(i)(0)} = \frac{B_{s(i)}\, C_{b(i)(0)}\, e^{\frac{J_{w(0)}}{k_{(i)(0)}}}}{J_{w(0)} + B_{s(i)}\, e^{\frac{J_{w(0)}}{k_{(i)(0)}}}}$ $C_{p(i)(L)} = \frac{B_{s(i)}\, C_{b(i,L)}\, e^{\frac{J_{w(L)}}{k_{(i)(L)}}}}{J_{w(L)} + B_{s(i)}\, e^{\frac{J_{w(L)}}{k_{(i)(L)}}}} \quad C_{p(i)(av)} = \frac{C_{p(i)(0)} + C_{p(i)(L)}}{2}$	Permeate solute concentration (kmol/m³) at x=0 and x=L and average one	14, 15
$\frac{\left(C_{w(i)(x)} - C_{p(i)(av)} \right)}{\left(C_{b(i)(x)} - C_{p(i)(av)} \right)} = \exp\left(\frac{J_{w(x)}}{k_{(i)(x)}} \right)$ $J_{s(i)(x)} = B_{s(i)} \left(C_{w(i)(x)} - C_{p(i)(av)} \right)$	Wall solute concentration (kmol/m³) and solute flux (kmol/m² s)	16, 17
$Rec = \frac{F_{p(L)}}{F_{b(0)}} x\, 100 \quad Rej_{(i)} = \frac{C_{b(i)(L)} - C_{p(i)(av)}}{C_{b(i)(L)}} x100$	Recovery and rejection	18, 19

Table 2. Membrane specifications and geometry (Ion Exchange, India)

Parameter	Value
Membrane material	TFC Polyamide
Feed and permeate spacers thickness (t_f, t_p)	0.8E-3 m, 0.5E-3 m
Module length and width (L, W)	0.934 m, 8.4 m
Water Permeability Constant (A_w) and friction factor (b)	9.5188E-7 $\left(\frac{m}{atm\ s}\right)$, 8529.45$\left(\frac{atm\ s}{m^4}\right)$

Table 3. Physical and transport parameters of the eight selected organic compounds

Compound	B_s (m/s)	Membrane	Reference
Dimethylphenol	1.5876E-8	Ion Exchange, India	Srinivasan et al. (2011)
Chlorophenol	8.4680E-8	Ion Exchange, India	Sundaramoorthy et al. (2011)
Phenol	6.5367E-7	Permionics, India	Srinivasan et al. (2010)
Methyl orange dye	3.2000E-9	FilmTec SW30	Al-Bastaki (2004)
Aniline	4.1900E-6	DESAL-3B	Hidalgo et al. (2014)
Ammonium	1.1696E-7	SEPA-SSIC	Bódalo et al. (2005)
Cyanide	2.1861E-6	DESAL-3	Bódalo et al. (2004)
Sulphate	3.9869E-8	SEPA-SSIC	Bódalo et al. (2003)

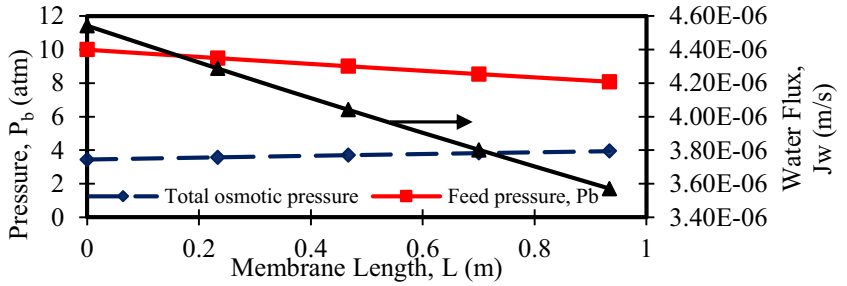


Fig. 2. The variation of operating parameters along the membrane length

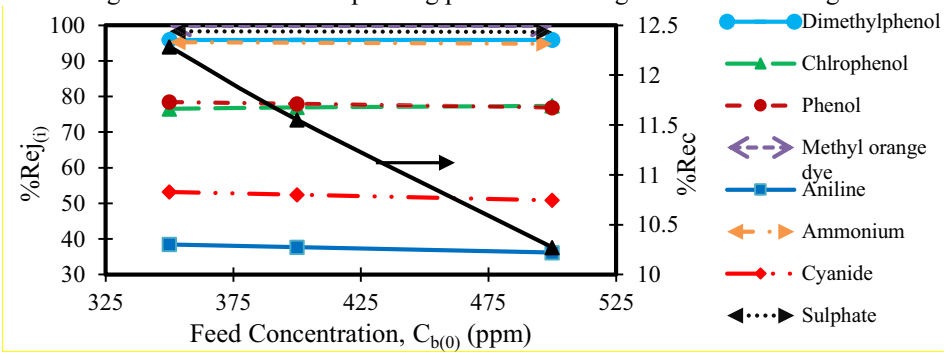


Fig. 3. Effect of compound concentration on rejection and recovery rate (operating conditions: 2.583E-4 m³/s, 10 atm, 30 °C)

Fig. 4 shows the impact of temperature on the removal of all compounds and permeate recovery. Increasing temperature causes more flexibility of membrane chains resulting in increasing convective transport by elevating the water flux. Also, it is noticed that diffusion transport increases due to increase in temperature, which is accompanied by a continuous reduction of average density and viscosity of the mixture. The effect of inlet feed pressure is shown in Fig. 5. Increasing the pressure from 10 to 20 atm causes an increase in the rejection due to increase in water flux, which dilutes the permeate. However, it seems that there is an optimum pressure which would maximise rejection of some of compounds. The impact of feed flow rate is shown in Fig. 6. The increase of feed flow rate from 2E-4 to 2.583E-4 m³/s causes a little increase in the rejection parameter but a remarkable decrease in permeate recovery. Increasing the feed flow rate causes a reduction in the osmotic pressure as a result of decreasing the membrane wall concentration. Increasing the feed flow rate from 2.583E-4 to 3E-4 m³/s causes a little

decrease in rejection and steady decrease in the permeate recovery. The reduction of permeate flux is due to reduction of wastewater residence time inside the module.

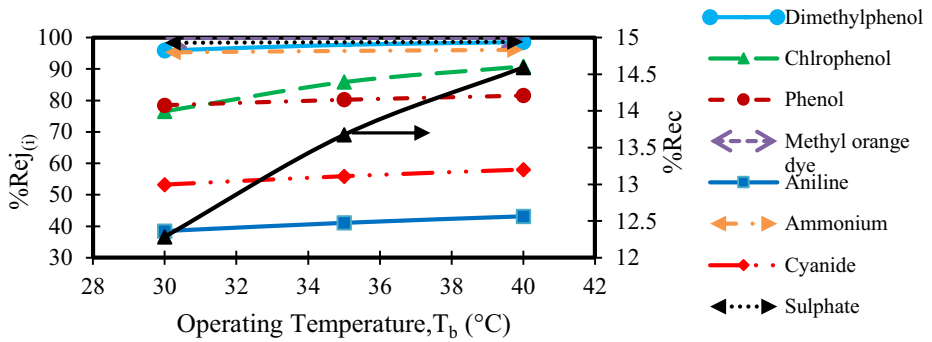


Fig. 4. Effect of operating temperature on rejection and recovery rate (operating conditions: 2.583E-4 m³/s, 10 atm, 350 ppm)

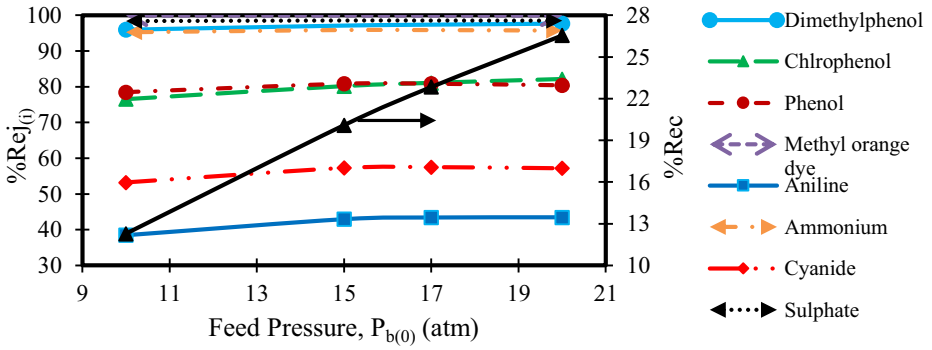


Fig. 5. Operating pressure verses rejection and recovery rate (operating conditions: 2.583E-4 m³/s, 30 °C, 350 ppm)

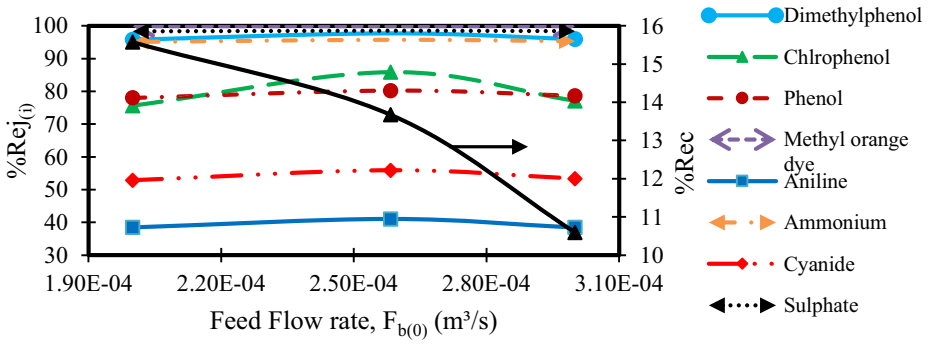


Fig. 6. Effect of operating feed flow rate on rejection and recovery rate (operating conditions: 10 atm, 30 °C, 350 ppm)

4. Process optimisation

The multi-objective function is presented below:

Max $P_{b(0)}$, $F_{b(0)}$, T_b	$Rej_{(i)}$, Rec
Subject to: Equality constraints - Process Model	f(x, u, v) = 0
Inequality constraints: Inlet pressure	$5\ atm \leq P_{b(0)} \leq 20\ atm$
Inlet feed flow rate	$1E-4\ \frac{m^3}{s} \leq F_{b(0)} \leq 1E-3\ \frac{m^3}{s}$
Operating temperature	$20\ °C \leq T_b \leq 40\ °C$

The optimal values of feed flow rate, pressure, and temperature are 7.4515E-4 m³/s, 20 atm and 40 °C respectively. The maximum permeate recovery is found to be 13.54%, while the maximum rejection of the compounds are: <Dimethylphenol, Chlorophenol, Phenol, Methyl orange dye, Aniline, Ammonium, Cyanide, Sulphate> = < 99.269, 94.922, 86.477, 99.923, 51.430, 97.253, 66.252, 99.045>. The optimisation approach leads to an increased rejection of all the selected components compared to what have been presented in Figs. 3- 6. However, the low value of permeate recovery is due to the impact of osmotic pressure of eight compounds.

5. Conclusions

A new one-dimensional model is developed for a spiral-wound RO process which is used to evaluate the impact of operating conditions on the rejection of several organic and non-organic compounds from wastewater and permeate recovery. The simulation results confirmed the importance of feed pressure and temperature to drive high performance of RO process. Finally, the multi-objective optimisation problem finds the maximum values of the rejection of all the compounds and permeate recovery.

References

N. Al-Bastaki, 2004, Removal of methyl orange dye and Na2SO4 salt from synthetic waste water using reverse osmosis, Chemical Engineering and Processing, 43, 1561-1567.

M.A. Al-Obaidi, J-P. Li, C. Kara-Zaïtri, I.M. Mujtaba, 2017, Optimisation of reverse osmosis based wastewater treatment system for the removal of chlorophenol using genetic algorithms, Chemical Engineering Journal, 316, 91-100.

A. Bódalo, J.L. Gómez, E. Gómez, G., León, M., Tejera, 2003, Reduction of sulphate content in aqueous solutions by reverse osmosis using celloulose acetate membranes, Desalination, 162, 55-60.

A. Bódalo, J.L. Gómez, E. Gómez, G. León, M. Tejera, 2004, Sulfonated polyethersulfone membrane in the desalination of aqueous solutions, Desalination, 168, 277-282.

A. Bódalo, J.L. Gómez, E. Gómez, G. León, M. Tejera, 2005, Ammonium removal from aqueous solutions by reverse osmosis using cellulose acetate membranes, Desalination, 184, 149-155.

A.M. Hidalgo, G. León, M. Gómez, M.D. Murcia, M.D. Bernal, S. Ortega, 2014, Polyamide nenofiltration membranes to remove aniline in aqueous solutions, Environmental Technology, 35(9), 1175-1181.

G. Srinivasan, S. Sundaramoorthy, D.V.R. Murthy, 2010, Spiral wound reverse osmosis membranes for the recovery of phenol compounds – Experimental and parameter estimation studies, American Journal of Engineering and Applied Science, 3(1), 31-36.

G. Srinivasan, S. Sundaramoorthy, D.V.R. Murthy, 2011, Validation of an analytical model for spiral wound reverse osmosis membrane module using experimental data on the removal of dimethylphenol, Desalination, 281, 199-208.

S. Sundaramoorthy, G. Srinivasan, D.V.R. Murthy, 2011, An analytical model for spiral wound reverse osmosis membrane modules: Part II — Experimental validation, Desalination, 277(1–3), 257-264.

P.C. Wankat, 1990, Rate-Controlled Separation. Ist. Edn., Springer, pp: 873.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50307-4

Food and bioenergy: capturing the synergies and conflicts in the design of value chains through spatio-temporal multi-objective optimisation

Sheila Samsatli*

Department of Chemical Engineering, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom
*s.m.c.samsatli@bath.ac.uk

Abstract

An MILP model was developed for biomass value chains that can maximise synergies and minimise conflicts within the food-energy-water-environment nexus. The model accounts for the spatial dependencies of the system, at a 50 km level, with a long planning horizon from 2010 to 2050 with seasonal time steps. A scenario was presented in which the model was used to examine competition between food and energy for arable land in the UK. Results suggest that the demands for food can be met along with 10% of total energy demands from bioenergy but food production must move to different locations to make way for bioenergy.

Keywords: biomass value chains; food vs. energy; land use and resource allocation; spatio-temporal supply chain optimisation; GIS analysis.

1. Introduction

Biomass can be used to produce electricity, heat and transport fuels to help meet energy and emissions targets, diversify energy supply and create growth and jobs (Demirbas, 2009). Biomass and bioenergy are intricately linked with the food-energy-water-environment nexus. Biomass production and utilisation involves chains of activities, including growing the feedstocks, transporting them to processing facilities, converting them to various products and so on. Each of these activities can have profound environmental, economic and social impacts. Effective and careful management of all activities in biomass supply chains (Samsatli and Samsatli, 2018) is crucial to ensure that the production and utilisation of biomass are sustainable, do not contribute to global warming and maintain the balance between all of the elements in the nexus. Many biomass supply chain optimisation models have been developed over the last decade (see reviews of De Meyer et al. (2014), Mafakheri and Nasiri (2014), Yue et al. (2014) and Mitchell (2000)). However, nearly all of these models did not account for the interaction of the biomass supply chains with the nexus. The aim of this work is to develop a spatio-temporal (Samsatli and Samsatli, 2015) mixed integer linear programming (MILP) model that can determine the important decisions concerning the design and operation of biomass value chains and identify sustainable chains that are synergistic with the nexus.

2. Problem statement

Many different types of biomass can be grown and converted to many different products. There are many decisions that need to be determined simultaneously, such as: what biomass types to grow, where and when to grow them; what energy services/products to generate (e.g. heat, biofuels, electricity, chemicals); what processing technologies to use, where to locate them, when to invest in them; centralised or distributed processing; whether or not to densify/pre-process biomass before transportation; how to transport

biomass and distribute products to customers; and how to manage inventory (as the production of most types of biomass is seasonal and not available all year round).

3. Mathematical model

The model considers a planning horizon from 2010 to 2050, with 10-year time steps, $d \in \mathbb{D}$, in order to model the investment in and retirement of technologies and land use changes, as well as seasonal time steps, $t \in \mathbb{T}$, in order to capture the seasonal production and storage of biomass. For the spatial representation, the UK is divided into 50 km × 50 km square cells, with each cell, $i \in \mathbb{I}$, characterised by: the area of each type of land cover available for growing crops, existing technologies, transport distances between each pair of cells and existing transport infrastructures. The total land area is divided into three categories, $c \in \mathbb{C}$, according to the existing land use: arable, grassland and forest (CEH, 2015). In Figure 1, the black circles are the total area of the three categories while the grey circles are the area of each individual category. These categories correspond closely to the different types of biomass that can be grown, e.g. wheat and other arable crops, energy crops such as Miscanthus and forestry. The yields for the different crops will depend on the category of land on which they are grown. These categories are also used to restrict crops to certain land categories, e.g. disallowing production of arable or energy crops on existing forest. Further, these land categories are required in order to track land use change and its associated emissions over the decades. The use of land can be limited according to various constraints, such as suitability for cultivation and harvesting, as well as social and environmental constraints. Specifically, constraints can be applied to exclude the following land areas in an optimisation run: elevation > 250 m; slope > 15 %; topsoil organic carbon > 30%; urban areas, roads and rivers; parks, scheduled monuments and world heritage sites; designated areas and cultural heritage areas; natural and semi-natural habitats; areas of outstanding natural beauty; and woodlands. Figure 1 shows how much area is available without any constraints and with all of the constraints applied.

3.1. Objective function

The objective function, to be minimised is given by Eq. (1).

$$Z = \sum_{d \in \mathbb{D}} \sum_{m \in \mathbb{M}} \omega_m \left(I_{dm}^{\mathrm{BP}} + I_{dm}^{\mathrm{TI}} + I_{dm}^{\mathrm{TO}} + I_{dm}^{\mathrm{Q}} + I_{dm}^{\mathrm{RP}} + I_{dm}^{\mathrm{SI}} + I_{dm}^{\mathrm{SO}} + I_{dm}^{\mathrm{RD}} - I_{dm}^{\mathrm{R}} \right) \quad (1)$$

Each term in the parentheses represents an impact in decade d on metric m. These are: biomass production, technology investment, technology operation, transportation, resource purchase, storage invesment, storage operation, resource disposal and revenue from sales of resource. The set of metrics $\mathbb{M}$ comprises costs and GHG emissions, including CO_2 and other emissions such as NO_x, SO_x and PMs.

3.2. Constraints

The resource balance, given by Eq. (2), considers various source and sink terms for each resource, $r \in \mathbb{R}$, in each cell, $i \in \mathbb{I}$. These include production of biomass, production of a resource through a conversion technology, transportation of resource into and out of each cell, compulsory demand satisfaction, purchase/import of a resource, sale of any bi/co-products and disposal of any waste resources that cannot be sold. The resource balance also tracks the inventory of stored resource, accounting for any losses in storage. The biomass produced, in oven-dry tonnes per season, is given by Eq. (3), where A_{rcid} is the area allocated to production of biomass type $r \in \mathbb{R}^{\mathrm{B}}$ on land class c in cell i during decade d and Y_{rcidt} is its yield potential in season t. M_r in Eq. (2) is the moisture content of biomass type $r \in \mathbb{R}^{\mathrm{B}}$. If wheat is grown in order to use just the grain in further processes, then an amount of straw can also be collected. The factor ϕ_{cid} defines the mass

of straw that can be collected if 1 odt of wheat grain is harvested, assuming that only half of the straw left in the field can be recovered (Eqs. (4) and (5)).

If wheat is grown, other crops must be rotated with it to maintain soil quality (Eq.(6)). The crops that can be rotated with wheat are defined by the set $\mathbb{R}^{\mathrm{WR}}$; y_r^{WR} is the number of years that wheat can be grown for each year of the rotation crop.

The total area allocated to crops in each cell, i, on each land class, $c \in \mathbb{C}$, is restricted by the maximum available area, $A_{cid}^{\max}$ (Eq.(7)).

The rate of operation of a conversion technology, $p \in \mathbb{P}$, is restricted by the total installed capacity in any cell, i, and the fraction of the year that it is available (e.g. accounting for maintenance and other downtime), a_{pd} (Eq.(8)).

The capacity of installed technologies is determined by a set of investment constraints (Eq.(9–11)). C_{pid} is the total installed capacity of technology p in cell i in decade d. It is equal to the installed capacity in the previous decade (or the initial capacity, for the first decade) plus any newly installed capacity, CI_{pid}, minus the capacity that is retired, CR_{pid}. CI_{pid} is constrained by the minimum and maximum capacity of a new technology, $C_{pd}^{\min}$ and $C_{pd}^{\max}$, multiplied by the number of new technologies installed, NI_{pid}. The retired capacity is calculated from the invested capacity and a "retirement fraction", $\psi_{pdd'}$, which defines the fraction of capacity that retires in decade d of technologies that were invested in decade d'; it also includes the retirements of existing technologies, CR_{pid}^{0}.

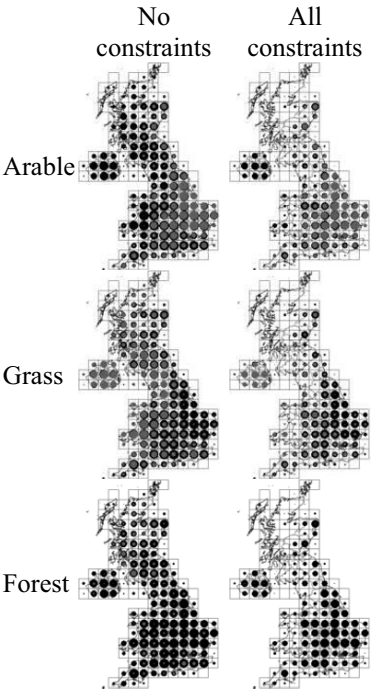


Figure 1. Available land area with and without constraints applied. Black circles: available area; grey circles: area of the land cover type.

$$\frac{B_{ridt}|_{r\in\mathbb{R}^{\mathrm{B}}} + B_{idt}^{\mathrm{straw}}|_{r=\mathrm{Straw}}}{1-M_r} + \sum_{p\in\mathbb{P}} \mathcal{P}_{pidt}\alpha_{pdr}\gamma_{pdr} + \sum_{l\in\mathbb{L}}\sum_{i'} Q_{ri'ildt}|_{r\in\mathbb{R}^{\mathrm{Q}}} - \sum_{l\in\mathbb{L}}\sum_{i'} Q_{rii'ldt}|_{r\in\mathbb{R}^{\mathrm{Q}}} - D_{ridt} + RP_{ridt} - RS_{ridt} = \left(I_{ridt} - I_{rid,t-1} + S_{ridt}^{\mathrm{loss}}\right)|_{r\in\mathbb{R}^{\mathrm{S}}} + RD_{ridt} \qquad \forall r\in\mathbb{R}, i\in\mathbb{I}, d\in\mathbb{D}, t\in\mathbb{T} \tag{2}$$

$$B_{ridt} \le \sum_{c} A_{rcid}Y_{rcidt} \qquad \forall r\in\mathbb{R}^{\mathrm{B}}, i\in\mathbb{I}, d\in\mathbb{D}, t\in\mathbb{T} \tag{3}$$

$$B_{idt}^{\mathrm{straw}} \le \sum_{c} \phi_{cid}A_{\mathrm{WG},cid}Y_{\mathrm{WG},cidt} \qquad \forall i\in\mathbb{I}, d\in\mathbb{D}, t\in\mathbb{T} \tag{4}$$

$$\phi_{cid} = 0.5\frac{Y_{\mathrm{WC},cid} - Y_{\mathrm{WG},cid}}{Y_{\mathrm{WG},cid}} \qquad \forall i\in\mathbb{I}, d\in\mathbb{D} \tag{5}$$

$$\sum_{r\in\mathbb{R}^{\mathrm{WR}}} y_r^{\mathrm{WR}}A_{rcid} = A_{\mathrm{WC},cid} + A_{\mathrm{WG},cid} \qquad \forall c\in\mathbb{C}, i\in\mathbb{I}, d\in\mathbb{D} \tag{6}$$

$$\sum_{r\in\mathbb{R}^{\mathrm{B}}} A_{rcid} \le A_{cid}^{\max} \qquad \forall c\in\mathbb{C}, i\in\mathbb{I}, d\in\mathbb{D} \tag{7}$$

$$\mathcal{P}_{pidt} \leq C_{pid} a_{pd} \qquad \forall p \in \mathbb{P}, i \in \mathbb{I}, d \in \mathbb{D}, t \in \mathbb{T} \tag{8}$$

$$C_{pid} = C_{pi,d-1} + C_{pi}^{0}|_{d=1} + CI_{pid} - CR_{pid} \qquad \forall p \in \mathbb{P}, i \in \mathbb{I}, d \in \mathbb{D} \tag{9}$$

$$C_{pd}^{\min} NI_{pid} \leq CI_{pid} \leq C_{pd}^{\max} NI_{pid} \qquad \forall p \in \mathbb{P}, i \in \mathbb{I}, d \in \mathbb{D} \tag{10}$$

$$CR_{pid} = CR_{pid}^{0} + \sum_{d'} CI_{pid'} \psi_{pdd'} \qquad \forall p \in \mathbb{P}, i \in \mathbb{I}, d \in \mathbb{D} \tag{11}$$

4. Case study

Scenario: bioenergy can satisfy up to 10% of the total UK energy demand, with a trajectory from 2010 to 2050 given in Figure 2 (Samsatli et al., 2015). Only arable crops are considered: wheat, oilseed rape and sugar beet, which are the major food crops produced in the UK. Wheat has to be rotated with either oilseed rape or sugar beet: 2 years of wheat for every year of rotation crop. In order to capture the competition for arable land between food and energy, the model will determine the allocation of arable land for production of crops for either food or bioenergy. All of the land constraints discussed in Section 3 are applied. Food, energy services (i.e. heat, electricity and transport fuels) and by-products can be sold at market price to maximise the net present value (NPV) of the overall system. Figure 3 presents the value web superstructure diagram for wheat, which shows the wide range of products that can be produced from this resource: different types of transport fuel, electricity and heat as well as various by-products (e.g. DDGS from ethanol and butanol production, which can be sold as animal feed). Similar value web diagrams exist for oilseed rape and sugar beet but are not shown due to limited space. The model will choose what crops to grow and what products and by-products to generate but all demands for food (equivalent to current production) must be met. The model will determine the best locations for food and energy production. The aim is to determine: the design and operation of the value chains to maximise the NPV; how much land area is required and where to plant the crops; how much of the energy service demands can be met; in what form the energy will be; etc.

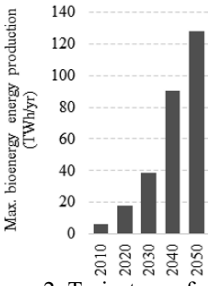


Figure 2. Trajectory of max bioenergy production (Samsatli et al., 2015).

5. Results and discussion

Figure 4 shows the wheat production from 2010 to 2050. Over the decades, additional wheat is grown for bioenergy and this has displaced some of the food production, which has moved to other locations. As the circles for bioenergy appear and grow over the decades, some of the circles for food disappear (e.g. in the central part of the country) while the other circles for food increase in size. In all decades, the same amount of wheat is produced for food but in different locations – this suggests that bioenergy will have local and regional impacts on food production. Bioenergy production starts in the eastern part of the country in 2020 and then in the later decades expands towards the northern and southern western parts of the UK and then grows westward. The spatial distribution of oilseed rape follows that of wheat as it is the main crop rotated with wheat. Large amounts of the oilseed rape are being used for energy, since its demands for food are much lower than those for wheat. There is also a small amount of sugar beet produced – mostly for food. Figure 5(a) shows the optimal biomass value chains in 2010. Oilseed rape is used for food as well as heat and electricity by converting it first to rapeseed oil, which is then used in internal combustion (IC) engine combined heat and power (CHP) plants. Additional value is obtained from the rapeseed meal (by-product from oil extraction), which is sold for animal feed.

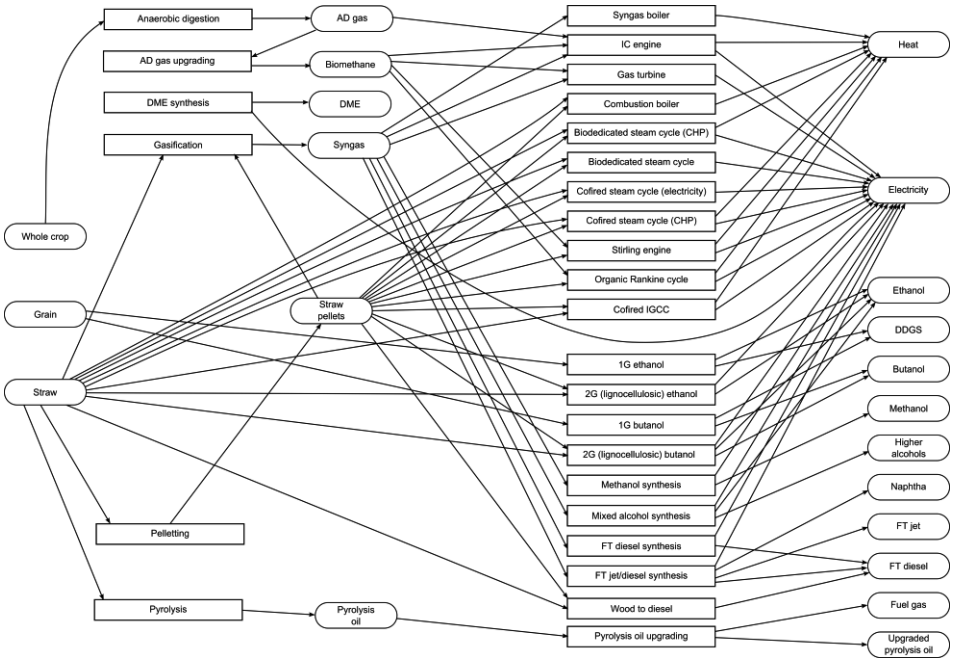


Figure 3. Value web network superstructure for wheat.

The map in Figure 6(a) shows the location of the oil extraction and the IC engine CHP plants. Both technologies are in the same location so only oil extraction can be seen on the map. The blue arrow represents the transport of the rapeseed oil. In this decade, all of the wheat grain produced is used to satisfy its food demands. Some of the straw from wheat production is sold as animal bedding and the remainder is used in CHPs for heat and electricity. In the 2050s, the limit on the amount of bioenergy that can be produced is much larger (see Figure 2) and the amount of wheat grain produced for bioenergy is almost the same as the amount grown for food (see Figure 4). Figure 5(b) presents the optimal value chains in 2050, where large amounts of wheat grain are used to produce butanol. The by-product DDGS is sold as animal feed. Figure 6(b) shows the location of the butanol plants: wheat is harvested in different locations and transported to the butanol plants. Wheat straw is used in a combination of biomass and syngas boilers to produce heat, the syngas being produced by the gasification of straw. Some straw is sold as animal bedding but because of the large amount of wheat grain produced in this decade, there is an excess of straw beyond its market demand that cannot be used for heat, the demands for which are also being met by the rapeseed oil. Sugar beet is only grown to meet its demand for food. The structure of the chain for oilseed rape did not change from 2010 but the amount of energy produced from this chain increased significantly. Overall, the optimal value chains result in an NPV of £96bn with a GHG emissions saving of 729 million tonnes over the 40 year planning horizon.

6. Conclusions

A multi-objective, spatio-temporal MILP model was developed to optimise the strategic design and tactical operation of biomass value chains. The model was used to examine food vs. energy scenarios in the UK and the results indicate that the UK has sufficient arable land to satisfy demands for main food crops (wheat, sugar beet and oilseed rape) and the target of meeting 10% of the energy demands through biomass. However, food

production is moved to other locations in order to give way for energy production, suggesting local and regional impacts of bioenergy on food production. By-products (wheat straw, DDGS, rapeseed meal, sugar beet pulp) can have a significant contribution to the profitability and emissions reduction of the whole value chain. Overall, exploiting the synergies between food and energy production can result in profitable value chains with negative GHG emissions.

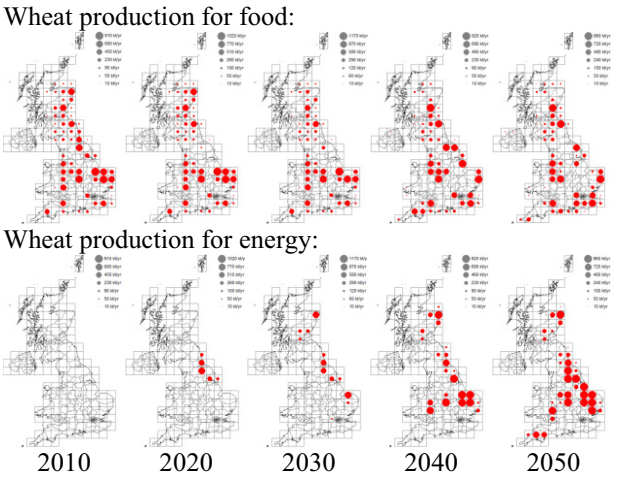


Figure 4. Wheat production from 2010 to 2050. Food production has been moved to give way for bioenergy.

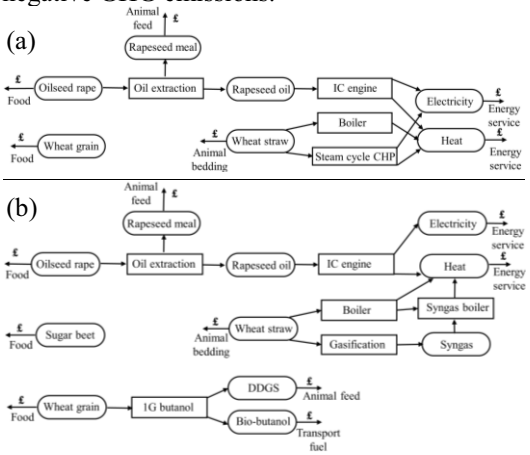


Figure 5. Optimal biomass value chains in (a) 2010 and (b) 2050

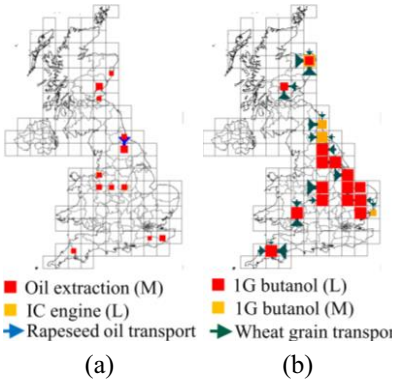


Figure 6. Location of technologies in (a) 2010 and (b) 2050.

7. Acknowledgements

Dr Samsatli would like to thank the RCUK/EPSRC and the Newton Fund for partial funding of her research through the BEFEW project (Grant No. EP/P018165/1).

References

Centre for Ecology and Hydrology (CEH), Land Cover Map 2015.
De Meyer A, Cattrysse D, Rasinmäki J, Van Orshoven J., 2014, Renewable and Sustainable Energy Reviews, 31, 657-70.
Demirbas, A., 2009, Energy Conversion and Management, 50, 2239-2249.
Mafakheri F., Nasiri F., 2014, Energy Policy, 67, 116-26.
Mitchell, C.P., 2000, Biomass and Bioenergy, 18, 265-278.
Samsatli, S., Samsatli, N.J., 2018, Applied Energy, DOI: 10.1016/j.apenergy.2017.09.055.
Samsatli, S., Samsatli, N.J., 2015, Computers and Chemical Engineering, 80, 155-176.
Samsatli, S., Samsatli, N.J., N. Shah, 2015, Applied Energy, 147, 131-160.
Yue, D., You, F., Snyder, S.W., 2014, Computers and Chemical Engineering, 66, 36-56.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50308-6

Dewatering Screw Pressing: Model Development and Sensitivity Analysis for Process Understanding

Franco Mangone[a*], Adrián Ferrari[a], Soledad Gutiérrez[a]

[a]*Chemical and Process Systems Engineering Group – Chemical Engineering Institute - Engineering School – Universidad de la República, Julio Herrera y Reissig 565, PC 11300, Montevideo, Uruguay*

Abstract

Dewatering screw presses (DSP) are a type of dewatering unit for sludge, which operate by compressing the material and letting water permeate to the outside. This is desirable to reduce downstream transportation costs, since sludge is often an unwanted product and needs to be disposed of. No simple mathematical models for DSP were found in open literature, and so one is proposed and analyzed in this work. Sensitivity and identifiability analysis are carried out in order to select the best parameter set for future parameter estimation.

Keywords: Dewatering screw press, sensitivity, identifiability.

1. Introduction

Sludges often result from separation processes, and they broadly represent a problem for industries because of the need to handle them. This of course has a cost, whether if the sludge is transported to a final disposal site, or if it is subjected to any kind of further treatment. Due to its high moisture content (> 90%), removing water from sludge can be advantageous in order to reduce downstream transportation costs.

To accomplish this goal, several methods and equipment exist, for instance the centrifuge decanter, dewatering tubes, diverse filter-type units, etc. Dewatering screw presses (DSP) are one of the options available for dewatering, and consist of a drum to which sludge is fed, and that also allows water to squeeze to the outside. The material is transported and compressed by an internal screw conveyor, providing the necessary potential for dehydration. Process understanding and modelling will be the main focus of this work.

2. Sludge behavior

The sludge dewatering process is more complex than an ideal sedimentation and it involves the need of additional pressure in order to dewater the material further once a certain threshold concentration is reached, called the *gel point* (ϕ_g).

This phenomenon is described by Buscall and White (1987) in a theory called "Compressive Rheology", which explains that the two parameters that describe sludge dewatering are its resistance to compression, and the permeability to water of the material. Intuitively, the former should increase and the latter should decrease as the solids concentration (ϕ) rises, and the authors describe numerical values that account for both phenomena and should follow those trends, which are the compressive yield stress, $P_y(\phi)$, and the hindered settling factor, $r(\phi)$. Both will be incorporated in is contribution.

3. Model description

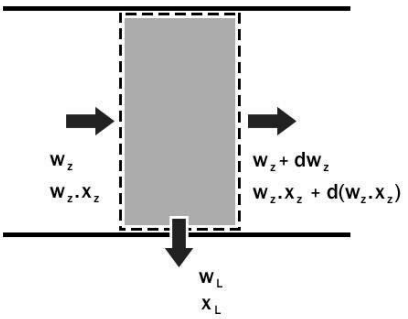


Figure 1. Selected control volume for mass balance

To start developing a mathematical description of the equipment, some assumptions need to be made. Initially, it is assumed that the composition of the sludge across a section of the unit will be uniform because of how the fluid is being pushed forward by the screw conveyor. This equates to saying that it resembles a plug flow behaviour, and in this sense Figure 1 shows the corresponding mass balance, where w_z is the total cake mass flow (kg/s), x_z is the mass solids fraction (dimensionless), and w_L and x_L represent the clarified liquid's total flow and solids content respectively.

3.1. Constitutive equation

With a plug flow structure assumed, the balances in a differential control volume can be analysed. The process is similar to a filtration, because of how water escapes through a semi-permeable drum wall that retains the solids inside of the unit, which is depicted in Figure 1. However, as a filtration-based model is not described in the original theory of Buscall and White (1987), a different approach is presented here by considering the work of Ruth et al. (1933) on filtration processes, who developed Equation (1):

$$\frac{dV}{A.dt} = \frac{\Delta P}{\mu.(R_m + R_c)} \tag{1}$$

Where dV/dt represents the permeate volumetric flow (m^3/s), A the filtration area (m^2), ΔP the differential pressure applied to the system (Pa), μ the water viscosity (Pa.s), and finally R_m and R_c being membrane (the equipment wall in this case) and cake resistance (m^{-1}) respectively. The original equation describes a non-steady state phenomenon and can be adapted to our model by substituting the dV/dt term with suitable process variables, namely w_L. At first sight, this equation does not incorporate the parameters described by Buscall and White (1987), but since the properties of the sludge are not dependent on the selected dewatering method, then the parameters in different models for cake formation should be related (De Kretser et al. 2003).

First, the pressure drop used in the filtration model can be represented by an applied pressure (P_T, the total pressure applied by the screw) minus the atmospheric pressure, since the permeate escapes to the outside of the drum. When the gel point is reached, extra pressure will be needed to continue dehydrating the formed cake, and, as a consequence, the original ΔP will be reduced by a certain amount. This is where it is possible to incorporate the P_y variable.

On the other hand, the parameter that accounts for permeability is clearly R_c, so it should be somehow related to the hindered settling factor. Matteson and Orr (1987) demonstrated that R_c can be defined as a function of a specific cake resistance α, in accordance with Equation (2):

$$\alpha(\phi) = \frac{R(\phi)}{(1-\phi).\mu} \tag{2}$$

Where $R(\phi)$ is the hindered settling factor multiplied by a set of constants, and ϕ is the volumetric solids fraction (dimensionless). At this point it might be helpful to clarify that ϕ and x_z which are both a function of z (position along the unit), are a measure of concentration and can be mutually related.

3.2. Parameters

All that remains is to define the functional expression of $P_y(\phi)$ and $R_c(\phi)$ in order to complete the constitutive equation for the model. Many authors have carried out studies to achieve this purpose, for instance Channel and Zukoski (1997) who state that P_y can be expressed as Equation (3):

$$P_y(\phi) = P_{y,0}.\left[\left(\frac{\phi}{\phi_g}\right)^n - 1\right] \tag{3}$$

Where $P_{y,0}$ is a pre-exponential factor which has the same units as P_y, expressed in Pa, and n is an exponential factor that accounts for the "microstructure" of the sludge. Note that ϕ_g is introduced as a fitting factor, since P_y should be zero for concentrations lower than or equal to the gel point.

For cake resistance, by finding a function that describes the hindered settling factor it is possible to propose a function for R_c by using Equation (2). Landman and White (1992) proposed and verified a functional description of $r(\phi)$, and when combined with (2) it is possible to obtain Equation (4):

$$R_c(\phi) = R_{c,0}.\frac{1}{\left(1-\frac{\phi}{\phi_{cp}}\right)^m} \tag{4}$$

Where $R_{c,0}$ is simply the set of constants that results from combining Equation (2) with the results found by Landman and White (1992), and ϕ_{cp} is a "random close-packing limit" that expresses a maximum possible concentration for the sludge, which has an empirical value of 0.64 (Berryman, 1983).

3.3. Mass balance

By performing a global and solids mass balance on the selected control volume, and adequately substituting the result of the former in the constitutive equation, the two model equations that allow for numerical simulation of the system can be obtained:

$$\frac{dw_z}{dz} = -\frac{\Delta P.\rho.\pi.D}{\mu.(R_m + R_c)} \tag{5}$$

$$\frac{dx_z}{dz} = \frac{dw_z}{dz} \cdot \frac{(x_L - x_z)}{w_z} \quad (6)$$

It is assume that x_L is this fraction of x_z, since no information was found regarding this variable. A value of 5% will be used, which is close to standard for this kind of equipment, and therefore we assign this percentage to a variable *C* so that $x_L = C.x_z$. The functional relationship between these two variables might be more complex, and that can be the focus of further model improvements if this proves to be influential.

4. Simulation, analysis, and discussion

All simulations were performed in the software Matlab R2010a, using the parameter values shown in Table 1.

The model should show an asymptotic behaviour for w_z and x_z. The cake flow rate should decrease until reaching a minimum value, which means no more dewatering occurs, and the opposite should happen for the solids fraction. Results are shown in Figure 2, supporting these conceptualizations.

4.1. Sensitivity analysis (SA)

Due to the lack of experimental data, sensitivity and identifiability analysis will be carried out in order to correctly select the parameter subsets for further estimation. Local (differential) sensitivity is defined as the derivative of a target output with respect to a model input or parameter. The values are then normalized in order to compare dimensionless sensitivities obtained for different outputs and from different inputs/parameters (Sin et al., 2010).

Table 1. Parameter values used for model simulation

Parameter	Value	Units
ϕ_g	0.07	-
m	5	-
n	5	-
C	0.05	-
P_T	9000	Pa
$P_{y,0}$	1000	Pa
R_m	$1.0x10^{11}$	m^{-1}
$R_{c,0}$	$1.0x10^{11}$	m^{-1}

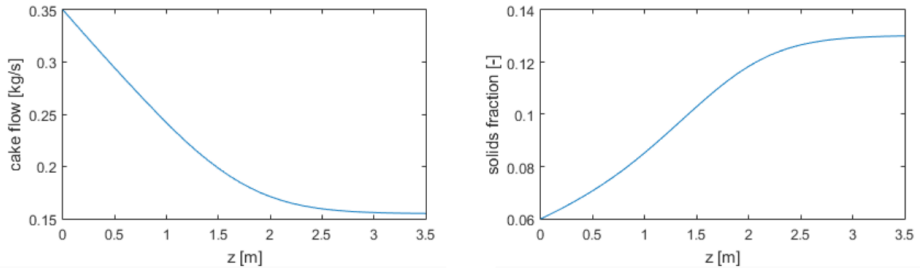


Figure 2. Cake flow (left) and solids fraction (right) simulation results for the developed model

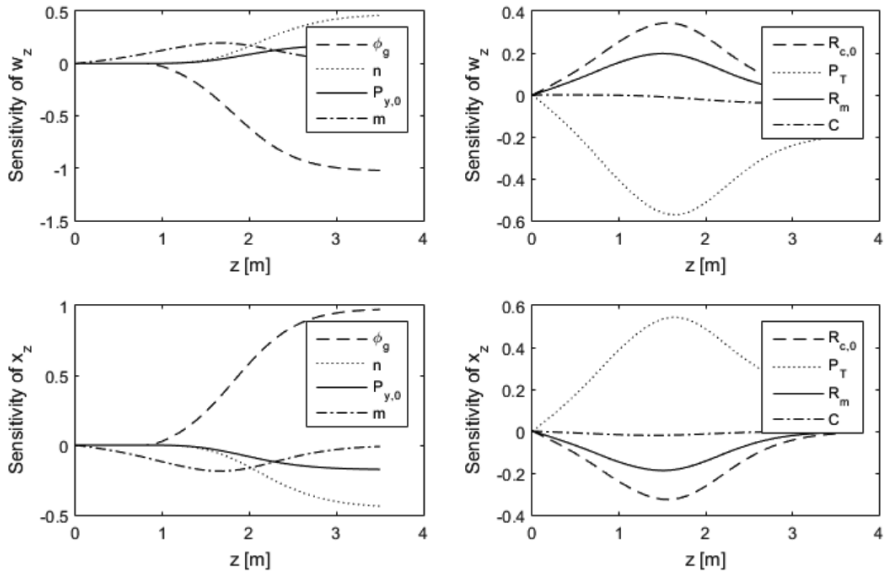


Figure 3. Dimensionless sensitivities to every parameter for each process output

In this case, the target outputs will be cake flow rate and solids content. To calculate the sensitivities, a positive and negative deviation of 0.1% around the nominal condition applied for all parameters (Sin et al., 2010). SA results are shown in Figure 3. Low sensitivity to C can be observed, meaning that this parameter will not be influential on the model. In addition, the results provide evidence about some degree of over-parameterization in the system, indicated by the similarity in some sensitivity trends, which represent correlation between parameters. These could represent difficulties in obtaining unique estimates for the parameters (Sin et al., 2010), hence identifiability analysis becomes necessary to identify an adequate subset for parameter estimation.

4.2. Identifiability analysis (IA)

This implementation was based in the collinearity index calculation in accordance to Sin et al., (2010) using the local SA results. Parameter subsets are analyzed and their linear correlation is evaluated through a measure called collinearity index (γ). A high value of γ means more correlation between the considered parameters. The analysis is used to find the largest size identifiable subset assuming a threshold γ value for the identifiability. Table 2 summarizes the obtained results for IA.

Results indicate that a maximum of two parameters are identifiable assuming a threshold of 5 for the collinearity index. In order to select an adequate subset for further analysis, the one with the lowest index can be chosen. The ranking of identifiable subsets is resumed in Table 2. The minimum collinearity index corresponds to the subset (n, R_m), meaning that after parameter estimation a minimal correlation between them would be expected, obtaining the best possible unique estimators. Nevertheless, it is important to consider the physical meaning of the parameters. Only two of them are DSP specific, P_T and R_m, while the rest can be proven to be material properties related (De Kretser et al., 2003). This means that they can also be determined via other experiments on the sludge itself, and not using data from DSP operation since they are

not strictly process related. Consequently, (P_T, R_m) could be selected for parameter estimation since it is also an identifiable subset.

Table 2. List of largest size identifiable subsets of model parameters

Parameter subset	Collinearity index	Parameter subset	Collinearity index
(ϕ_g, m)	1.0432	($P_{y,0}$, m)	0.9627
(ϕ_g, $R_{c,0}$)	0.9819	($P_{y,0}$, $R_{c,0}$)	0.9142
(ϕ_g, P_T)	1.2591	($P_{y,0}$, P_T)	1.1444
(ϕ_g, R_m)	0.9527	($P_{y,0}$, R_m)	0.8909
(n, m)	0.9155	(m, P_T)	4.7968
(n, $R_{c,0}$)	0.8754	($R_{c,0}$, P_T)	3.9556
(n, P_T)	1.0802	(P_T, R_m)	3.4781
(n, R_m)	**0.8560**		

5. Conclusion

A mathematical model for DSP, including the sludge behavior, was developed and implemented in Matlab. It was shown that the model follows the expected tendencies. SA results indicate the presence of some degree of over-parameterization, and thus suggest a potential for model reduction, which can be a future challenge. IA results indicated that a maximum of two parameters are identifiable. DSP specific parameters, such as P_T and R_m, can be selected to be obtained by parameter estimation from process data, while the sludge properties could be determined via other non-process related experiments. Important factors such as cake thickness variation throughout the length of the tube and dynamic conditions are not considered in this study and should be the focus of further work in order to develop a model that more accurately describes the process.

References

J. G. Berryman, 1983, “Random close packing of hard spheres and disks”, Physical Review A, 27, 1054-1061

R. Brun, P. Reichert, and H. R. Künsch, 2001, “Practical identifiability analysis of large environmental simulation models”, Water Resources Research, 37, 4, 1015-1030

R. Buscall and L. R. White, 1987, “The Consolidation of Concentrated Suspensions”, J. Chem. Soc. Faraday Trans. I, 3, 873-891

G. M. Channell and C. F. Zukoski, 1997, “Shear and Compressive Rheology of Aggregated Alumina Suspensions”, AIChE, 43, 7, 1700-1708

R. G. De Kretser, D. V. Boger, and P. J. Scales, 2003, “Compressive Rheology: An Overview”, *Rheology Reviews*, 125-165

K. A. Landman and L. R. White, 1992, “Determination of the Hindered Settling Factor for Flocculated Suspensions”, AIChE, 38, 2, 184-192

M. J. Matteson and C. E. Orr, 1987, “Filtration: Principles and Practices”

B. F. Ruth, G. H. Montillon, and R. E. Montonna, 1933, “Studies in Filtration II. Fundamental Axiom of Constant-Pressure Filtration”, Ind. Eng. Chem, 25, 153-161

G. A. Sin, S. Meyer, and K. V. Garnaey, 2010, “Assessing reliability of cellulose hydrolysis models to support biofuel process design – Identifiability and uncertainty analysis”, Computers & Chemical Engineering, 34, 1385-1392

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50309-8

A hierarchical Food-Energy-Water Nexus (FEW-N) decision-making approach for Land Use Optimization

Styliani Avraamidou[a,b,c], Burcu Beykal[a,b], Ioannis P.E. Pistikopoulos[a,b], Efstratios N. Pistikopoulos[a,b*]

[a]*Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station TX 77843, USA*
[b]*Texas A&M Energy Institute, Texas A&M University, College Station TX 77843, USA*
[c]*Department of Chemical Engineering, Centre of Process Systems Engineering (CPSE), Imperial College London, London SW7 2AZ, UK*
stratos@tamu.edu

Abstract

The land use allocation problem is an important issue for a sustainable development. Land use optimization can have a profound influence on the provisions of interconnected elements that strongly rely on the same land resources, such as food, energy, and water. However, a major challenge in land use optimization arises from the multiple stakeholders and their differing, and often conflicting, objectives. Industries, agricultural producers and developers are mainly concerned with profits and costs, while government agents are concerned with a host of economic, environmental and sustainability factors. In this work, we developed a hierarchical FEW-N approach to tackle the problem of land use optimization and facilitate decision making to decrease the competition for resources and significantly contribute to the sustainable development of the land. We formulate the problem as a Stackelberg duopoly game, a sequential game with two players – a leader and a follower (Stackelberg, 2011). The government agents are treated as the leader (with the objective to minimize the competition between the FEW-N), and the agricultural producers and land developers as the followers (with the objective to maximize their profit). This formulation results into a bi-level mixed-integer programming problem that is solved using a novel bi-level optimization algorithm through ARGONAUT. ARGONAUT is a hybrid optimization framework which is tailored to solve high- dimensional constrained grey-box optimization problems via connecting surrogate model identification and deterministic global optimization. Results show that our data-driven approach allows us to provide feasible solutions to complex bi-level problems, which are essentially very difficult to solve deterministically.

Keywords: Food-Energy-Water Nexus, Bi-level Optimization, Land Use Optimization, Data-Driven Optimization

1. Introduction

A combination of rapid population growth, urbanisation, and economic development has spurred an overwhelming increase in world demands for food, energy, and water. With the availability of those goods being constantly decreasing, it is essential, now more than ever, to tackle effectively this issue.

This phenomenon can be linked to an economic theory called "The tragedy of the commons" (Lloyd, 1833), that refers to a situation within a shared resource system where individual users acting independently according to their own self-interest behave contrary to the common good of all users by depleting or spoiling that resource through their collective action (Hardin, 1968). Water, energy and food can be regarded as common resources, that can be overused and under-maintained. To add to the complexity, water, energy and food are interconnected resources forming what we referred to as Food-Energy-Water Nexus. For example, to produce food we need energy in the form of fertilizers or electricity, and water for irrigation. To produce clean water, we need energy for cleaning, and to produce energy we need water for cooling, and food if we were to use biofuels.

Even though we know that a holistic approach is required to tackle the problem of FEW increasing demand, the objectives of individual companies or organizations are usually based on short term profit.

To tackle this issue, Ostrom, E., 1990 discussed that enabling governments at multiple scales to interact with community organizations so that we have a complex nested system can give solutions to 'the tragedy of commons' problem. In other words, governments can shape strategies that lead organizations into actions that would avoid or limit stresses on the FEW Nexus. This can be done through rules and regulations, or subsidies that drive the objective and actions of the individual organizations towards the common good. This leads to a hierarchical system, with the government deciding (at the highest decision level) the rules, regulations and subsidies to motivate organizations towards deciding on their actions (at a lower decision level) to achieve the government's objective, i.e. best sustainable long-term FEW-N solution.

In this work, we attempt to model and solve hierarchical systems as the one described above. We are focusing on the case of land use allocation which is an important issue for sustainable development and its optimization can have a profound influence on the provisions of interconnected elements that strongly rely on the same land resources, such as food, energy, and water. In Section 2 the Land Use allocation case study is explained and the hierarchical model is developed and presented. In Section 3 a novel data-driven algorithm for the solution of the complex hierarchical problem is presented, while the solution for land use problem is presented in Section 4. Section 5 concludes this paper.

2. Land Use Allocation Problem

The chosen case study is a land use allocation problem that involves a fictional piece of land owned by an organization. The organization will invest in this land, and may choose different types of developments, which can involve a mixture of both agricultural and energy production land processes, consisting of solar energy, wind energy, fruit production, vegetable production, and livestock grazing. The interest of the land developer is to maximize its profit, while the interest of the government that rules this piece of land is to minimize the stresses on the FEW-N. This leads to a hierarchical optimization problem.

We are formulating this problem as a Stackelberg duopoly game (Stackelberg, 2011), a sequential game with two players – a leader and a follower. The government agents are

treated as the leader, and the land developer as the follower. The formulation of this problem as a bi-level problem is presented in Eq. (1).

$$\begin{aligned} &\min \ \ FEW\ Nexus\ Stress \\ &s.t. \ \ Goverment\ Budget \\ &\quad \max \ \ Developer's\ Profit \\ &\quad s.t. \ \ Land\ Properties\ , Developers\ Budget, Land\ Process\ Models \end{aligned} \tag{1}$$

2.1. Government's Optimization Problem – Leader's Problem

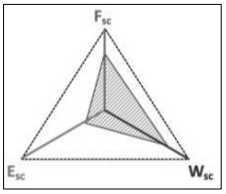


Figure 1 FEW metric

As mentioned before the government's move is only constrained by its budget. The objective of the government is to limit the stresses between food energy and water. While the importance of the FEW-N has been widely accepted, a quantitative index assessing the integrated FEW-N performance is rather lacking (Garcia et al. 2016). In this work, we have defined a FEW-N metric that aggregates FEW elements into a single 'geometric' metric, that can be used by the government as an objective.

The metric system involves scaling the quantitative value for each FEW element, to values between 0 and 1. A value of 0 indicates the worst possible scenario for a given element while a value of 1 implies the converse. The bounds on each element were determined by minimizing and maximizing the Land use allocation problem with each element as an objective. The three scaled elements are then used to create the geometric shape illustrated in Fig. 1. The area of the triangle formed corresponds to the FEW metric, the bigger the area the better the FEW Nexus solution.

$$\begin{aligned} FEW_{metric} = \Bigg[&\frac{Etotal - Emin}{Emax - Emin} \times \left(1 - \frac{Wtotal - Wmin}{Wmax - Wmin}\right) + \frac{Etotal - Emin}{Emax - Emin} \\ &\times \frac{Ftotal - Fmin}{Fmax - Fmin} + \left(1 - \frac{Wtotal - Wmin}{Wmax - Wmin}\right) \times \frac{Ftotal - Fmin}{Fmax - Fmin} \Bigg] \frac{\sin 120}{2} \end{aligned} \tag{2}$$

where Etotal, Ftotal are the total energy and food produced by the system correspondingly and Wtotal is the total water consumed by the system. Emin and Fmin are the minimum amount of energy and food that can be produced, and Wmin is the minimum amount of water that can be consumed. The same follows for Emax, Fmax, and Wmin.

For this case study, the government is using subsidies for each FEW sector. Therefore, the optimization problem of the government is as follows:

$$\begin{aligned} &\max_{sub} \ \ FEW_{metric} \\ &s.t. \quad sub_X \leq bg \quad \forall x = E, F, W \end{aligned} \tag{3}$$

where $sub_F, sub_E,$ and sub_W are the amount of dollars the government is subsidizing food, energy and water respectively.

2.2. Land Developer's Optimization Problem – Follower's Problem

As mentioned before the objective of the land developer is to maximize its profit. To achieve this, the land developer must take into account the characteristics of i) each land process, ii) the land itself, and iii) the subsidies given by the government. The specific case study for this work, assumes a fairly small piece of land in a climate similar to that of Texas, split into 8 evenly sized square plots, each with different properties. These properties, shown below in Table 1, define the limitations on which land processes can occur on each plot, and thus portray the specific situation of the assumed land developer.

Also, seasonal differences in climate can greatly impact the characteristics of the land as a whole, and change the appeal of various processes. These seasonal differences contribute to changes in a) energy profit based on consumers' demands, b) cost of transporting water due to availability, c) efficiency of the different systems in different weathers, and d) minimum water requirements for crops (due to rainfall), which are all modeled in the process. Therefore, the model includes parameters for each season that reflect the variations in energy and crop production efficiency, demand for energy, water cost, and water availability during different periods of the year.

For agricultural processes, the amount of water and energy used for the agricultural processes determines their yield (Table 2). The amount of Energy produced by solar and wind are only varied by land properties and seasonal changes. When both land properties and seasonal changes are at their optimal the amount of energy produced by one land piece is 50kW and the amount of energy produced by wind turbines is 1000kW.The profit gained from agricultural processes is proportional to the yield, whereas the profit from the energy processes is proportional to the kW produced.

The model also allows wind energy and agricultural processes to be developed on the same piece of land.

The model was developed in GAMS and it consisted of 1,723 equations, 216 integer variables and 1,107 continuous variables. The detailed model can be provided upon request as it was excluded from the document due to page restrictions.

Table 1 Land Properties

	Land Section Number							
	1	2	3	4	5	6	7	8
Good soil	√	√	√	√	×	×	√	√
Adequate sun	×	×	√	√	√	√	√	√
Adequate wind	√	√	√	√	√	√	×	×
Water available	×	×	×	×	√	√	×	√

Table 2 Yield of agricultural processes

Agricultural land use type	Yield (tonnes)
Fruits	$10E + 10^{-4}W$
Vegetables	$10E + 15^{-4}W$
Livestock	$E + 40^{-4}W$

3. Solution Method: ARGONAUT

A novel data-driven solution algorithm was developed for the solution of the large scale bi-level mixed-integer programming problem. The main idea behind the algorithm is to approximate the bi-level problem into a single level problem by collecting data from the optimality of the lower level problem. The main steps of the proposed algorithm are illustrated in Fig. 2.

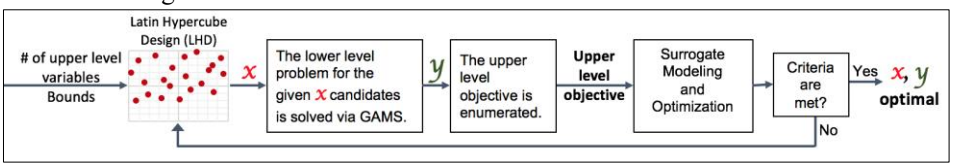


Figure 2 Illustration of the main steps of the Data-Driven Bi-level algorithm

3.1. Data-Driven Solution Method

The proposed algorithm is tested with different constrained data-driven optimization strategies, and AlgoRithms for Global Optimization of coNstrAined grey-box compUTational problems (ARGONAUT) (Boukouvala et al., Boukouvala and Floudas, 2017) was chosen for the solution of the hierarchical FEW Nexus problem. The selection of solver was based on its ability to perform constrained optimization on black/grey-box problems with a global optimization strategy. The ARGONAUT algorithm performs global optimization on general constrained grey-box problems by providing accurate surrogate formulations to each unknown equation while incorporating adaptive parallel sampling, bounds tightening and variable selection within the algorithm for improved accuracy in obtaining the global solution.

3.2. Benchmarking

This algorithm was tested by solving the challenging set of bi-level test problems from Mitsos et al. 2017. The developed algorithm was able to converge to the global solution for 72% of the test problems, and was able to find feasible local solution for the rest of problems.

4. Results and Discussion

Using ARGONAUT as the data-driven solution method the results in Table 3 were attained. It is observed that different government strategies are able to give the same upper level objective with the same land development solution (Fig. 3). One can choose the strategy that poses the lowest expenditure for the government, i.e. Run 10. It is worth noting that without any governmental subsidies the FEW metric would have been 20.4% less than the optimal calculated with the subsidies, and with solar energy being developed in sections 7 and 8 instead of fruit production.

1. Wind & Fruits	2. Wind & Fruits	3. Wind & Fruits	4. Wind & Fruits
5. Wind	6. Wind	7. Fruits	8. Fruits

Figure 3 Land allocation solution

Table 3 ARGONAUT solutions

Run #	FEW metric	# of Samples	Subsidies (k$)					Total Subsidies(k$)
			Solar	Wind	Fruits	Veg.	Liv.	
1	1.2299	147	0	19,796.2	50,000	0	0	418,777.2
2	1.2299	140	0	50,000.0	50,000	0	0	600,000.0
3	1.2299	234	0	20,363.5	49,811	0	0	421,047.0
4	1.2299	103	0	49,366.4	35,193.9	0	0	507,361.8
5	1.2299	161	0	0	50,000	0	0	300,000.0
6	1.2299	112	0	50,000	50,000	0	0	600,000.0
7	1.2299	288	0	50,000	21,847.1	0	0	431,082.6
8	1.2299	115	0	4,2368.1	50,000	0	0	554,208.0
9	1.2299	255	0	993.7	50,000	0	0	305,962.2
10	**1.2299**	**267**	**0**	**8,218.2**	**25,533.9**	**0**	**0**	**202,512.6**

5. Conclusions

It is evident that the developed model and approach constitutes a greatly effective means of optimizing land use through a Food-Energy-Water Nexus approach. The flexibility, robustness and scalability of the model facilitates its application to a real-life case study, with the various parameters and variables being redefined to fit the desired situation, and the solutions made to fit a client's demands. This strong adaptability to real world situations can thus help create more sustainable methods of agriculture and energy production in a world where demands for resources are constantly increasing, whether this be on a local, regional or national scale.

6. Acknowledgement

We are grateful to NIEHS Superfund Research Program (P42 ES027704), Texas A&M Superfund Research Center, Texas A&M University and Texas A&M Energy Institute for the financial support.

References

Boukouvala, F., Floudas, C.A., 2017. ARGONAUT: AlgoRithms for Global Optimization of coNstrAined grey-box compUTational problems. Optim. Lett. 11, 895-913.

Boukouvala, F., Hasan, M.M.F., Floudas, C.A., 2017. Global optimization of general constrained grey-box models: new method and its application to constrained PDEs for pressure swing adsorption. J. Glob. Optim. 67, 3-42.

Garcia, D.J., You, F., 2016 The water-energy-food nexus and process systems engineering: A new focus, Computers and Chemical Engineering, 91, 49-67.

Hardin, G, 1968. "The Tragedy of the Commons". Science. 162 (3859): 1243–1248.

Lloyd, W.F., 1833. Two lectures on the checks to population. England: Oxford University.

Mitsos, A., Barton, I.P., 2017. A test set for bilevel programs.

Ostrom, E., 1990. Governing the Commons: The Evolution of Institutions for Collective Action. Cambridge University Press.

Stackelberg, V.H., 2011. Market structure and equilibrium: 1st Edition Translation into English, Bazin, Urch & Hill, Springer, 1-134.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50310-4

Optimization of Membrane Processes at the Water-Energy Nexus

Mingheng Li[a, *], Yoshihito Adachi[a], Cuong Ngo[a], Eduardo Guillen[a]

[a]*Department of Chemical and Materials Engineering California State Polytechnic University Pomona, 3801 W. Temple Ave, Pomona, CA 91768, United States*
minghengli@cpp.edu

Abstract

Water and energy are two of the most pressing problems faced by the world today. Osmotic membrane processes such as Reverse Osmosis (RO) and Pressure Retarded Osmosis (PRO) can help alleviate both of these challenges. This work focuses on systematic optimization of a two-stage RO-PRO process for energy-efficient seawater desalination which integrates RO for desalination and PRO for power generation into a single process. RO-PRO is of great significance at the water-energy nexus given that water production, pump energy consumption, and hydraulic/osmotic energy recovery are essentially intertwined. The optimization problem is formulated to minimize the Specific Energy Consumption (SEC) subject to specified total membrane area and total water recovery. By solving the optimization model, the best design (e.g., allocation of membrane area among different units) and operating conditions (e.g. applied pressure in each membrane unit) are simultaneously determined. It is shown that the proposed two-stage RO-PRO is energy-efficient than single- and two-stage ROs as well as one-stage RO-PRO.

Keywords: Optimization, Water-Energy Nexus, Reverse Osmosis, Pressure Retarded Osmosis

1. Introduction

Energy and water are the top two grand challenges faced by the world (Smalley, 2003). The most modern industrial water desalination technology employs reverse osmosis (RO), which consumes roughly 0.5 kWh/m^3 and 3.5 kWh/m^3 electricity for groundwater and seawater sources, respectively. In seawater RO (SWRO), pump energy consumption accounts for a major portion of the total cost, which motivates research and development in energy-efficient RO (Zhu et al., 2009). For example, employing energy recovery devices (ERDs) has drastically reduced the pump energy consumption by recovering hydraulic potential energy in the RO brine (Stover, 2007).

Recently, Pressure Retarded Osmosis (PRO) has emerged as an osmotically driven membrane process that can generate power from a salinity gradient (Achilli, et al. 2009). The RO brine, due to its high osmotic pressure, is naturally suitable for this application if low-salinity impaired water is also available. The PRO and SWRO can be integrated into a RO-PRO process in which the power produced in the PRO may partially subsidize the power required by the SWRO. The research efforts are particularly encouraged by recent development of high-performance membranes with desired structural, mechanical, and permeative properties suitable for PRO applications (Han et al., 2017).

A RO-PRO membrane process is an interesting problem at the energy-water nexus (Gabriel et al. 2016, Garcia et al. 2016, Li et al. 2018). In the RO-PRO, water production, pump energy consumption, and hydraulic/osmotic energy recovery are essentially intertwined. Process systems engineering plays a very important role in the system-level understanding and optimization of such a process, which may guide experimental investigations undertaken by many research groups across the world.

2. Optimization Model

The schematics of two membrane process - RO for water desalination and PRO for power generation, are shown in Figure 1. The feed water in PRO is usually from low-salinity impaired wastewater resources and the draw solution could be seawater or brine from a SWRO process.

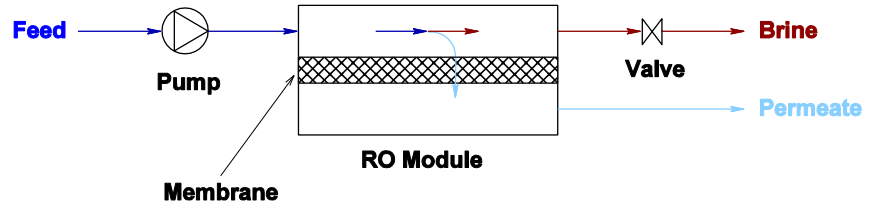


Fig. 1 Schematics of (a) RO for seawater desalination and (b) PRO for power generation.

A detailed description of transport phenomena in commercial spiral wound membrane units may require Computational Fluid Dynamic simulations (Li et al., 2016). If ignoring the effects of pressure drop and concentration polarization, both membrane processes shown in Fig. 1 can be described by the model shown below (Li, 2010; Li, 2015):

$$dQ = dA \cdot L_p \cdot (Q_0\pi_0/Q - P) \quad (1)$$

where Q, π and P are the flow rate, osmotic pressure, and hydraulic pressure of the solution with the higher salinity. A and L_p are membrane area and hydraulic permeability, respectively. Subscript 0 represents membrane inlet. An analytical solution to Eq. (1) can be derived as follows (Li, 2010; Li, 2015):

$$\gamma = \alpha\left[(1-q) + \alpha \ln \frac{\alpha - 1}{\alpha - q}\right] \quad (2)$$

where $\alpha = \pi_0/P$, $q = Q_1/Q_0$, and $\gamma = AL_p\pi_0/Q_0$. Subscript 1 represents membrane outlet. Eq. (2) clearly reveals the nonlinear coupling amongst the design parameter γ, the operating parameter α, and the performance parameter q. In RO, the fractional recovery Y = 1-q. Both α and q are less than 1 in RO and greater than 1 in PRO.

Concentration polarization is an important phenomenon that reduces water flux across the membrane. External concentration polarization may be supressed by optimizing flow conditions and spacer designs. However, internal concentration polarization, which occurs only in the PRO but not in the RO, cannot be easily mitigated. To account for this detrimental effect, an efficiency factor η_{PRO} is introduced so that the actual volume gain ratio in PRO *q-1* is only a fraction of the value predicted by Eq. (2).

In the RO community, Specific Energy Consumption (SEC) is a very important term describing pump power consumption per unit flow of the permeate product. If normalized by the feed osmotic pressure, it becomes a dimensionless quantity, or NSEC. Similarly, NSEP denotes the power produced by the PRO per unit flow rate normalized by the draw solution osmotic pressure. The author's group have demonstrated that employing multi-stage designs and increasing the total dimensionless membrane area γ_{total} ($\gamma_{total} = A_{total}L_p\pi_0/Q_0$) reduce NSEC in RO and enhance NSEP in PRO, as shown in Fig. 2.

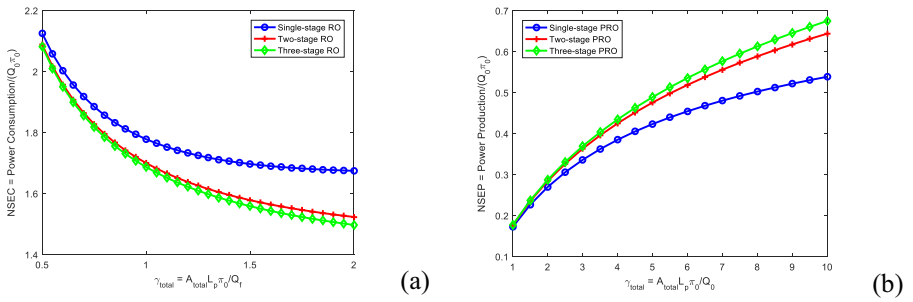


Fig. 2 Effects of total dimensionless membrane area γ_{total} and number of stages on (a) NSEC in RO at 40% recovery (Li, 2011) and (b) NSEP in PRO (Li, 2018).

Motivated by this, a two-stage RO-PRO process is proposed in this work for potential energy-efficient seawater desalination. As shown in Fig. 3, seawater is pressurized by several ERDs in addition to a feed pump before entering the first stage of RO. The concentrate from the first stage of the RO is pressurized further via a booster pump and sent to the second stage for additional water recovery. The osmotic and hydraulic energies in the RO brine are then recovered by the PRO and ERD systems downstream. The first two ERDs are necessary in order for the hydraulic osmotic pressure ratio at the inlet of each PRO to reach its optimal value which is unknown priori (Li, 2015; Li, 2017).

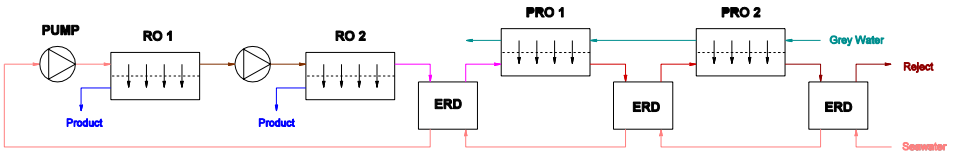


Fig. 3 Schematic of a hybrid two-stage RO-PRO process for energy-efficient seawater desalination.

The RO is a mature technology while the PRO is still under development. In this work it is assumed that both RO and PRO membranes have comparable permeative properties and costs per area. The optimization model is formulated to minimize the NSEC by simultaneously adjusting design parameters γ_j and operating parameters α_j while maintaining the total water recovery Y_{total} and total dimensionless area γ_{total}:

$$
\begin{aligned}
\min_{\alpha_j,\gamma_j,q_j} \eta_{pump} NSEC &= \frac{\frac{Y_1}{\alpha_1} + \frac{1-\eta_{ERD}(1-Y_2)}{\alpha_2} - \frac{\eta_{ERD}(q_3-1)}{\alpha_3} - \frac{\eta_{ERD}(q_4-1)}{\alpha_4}}{Y_{total}} \\
s.t. & \\
0 &= \gamma_j - \alpha_j \left[Y_j + \alpha_j \ln \frac{1-\alpha_j}{1-Y_j-\alpha_j} \right], j = 1,2 \\
0 &= \gamma_j - \alpha_j \left[1 - q_j^{ideal} + \alpha_j \ln \frac{\alpha_j - 1}{\alpha_j - q_j^{ideal}} \right], j = 3,4 \\
0 &= (q_j - 1) - \eta_{PRO}(q_j^{ideal} - 1), j = 3,4 \\
0 &= \gamma_{total} - \left[\gamma_1 + \gamma_2(1-Y_1)^2 + \gamma_3(1-Y_1)^2(1-Y_2)^2 + \gamma_4(1-Y_1)^2(1-Y_2)^2 q_3^2 \right] \\
0 &= 1 - (1-Y_1)(1-Y_2) - Y_{total} \\
0 &\leq \gamma_j, j = 1,\ldots,4 \\
0 &\leq \alpha_j, j = 1,2 \\
0 &\leq 1 - \alpha_j, j = 1,2 \\
0 &\leq \alpha_j - 1, j = 3,4 \\
0 &\leq \alpha_1 - \alpha_2(1-Y_1) \\
0 &\leq \alpha_4 q_3 - \alpha_3 \\
0 &\leq Y_j, j = 1,2 \\
0 &\leq 1 - Y_j, j = 1,2 \\
0 &\leq q_j - 1, j = 3,4
\end{aligned}
\tag{4}
$$

where η_{pump} is the pump efficiency, η_{ERD} is the ERD efficiency, subscripts 1, 2 represent RO stages, and 3,4 PRO stages. The optimization model is solved using constrained nonlinear optimization function `fmincon` in Matlab.

3. Results and Discussion

The optimization is carried out at different values of dimensionless membrane area γ_{total} (0.4-2), overall water recovery (30% and 50%) and PRO efficiency (70% and 100%). The optimization results of two-stage RO-PRO (RO2PRO2) are compared with those in one-stage RO (RO1), two-stage RO (RO2) as well as one-stage RO-PRO.

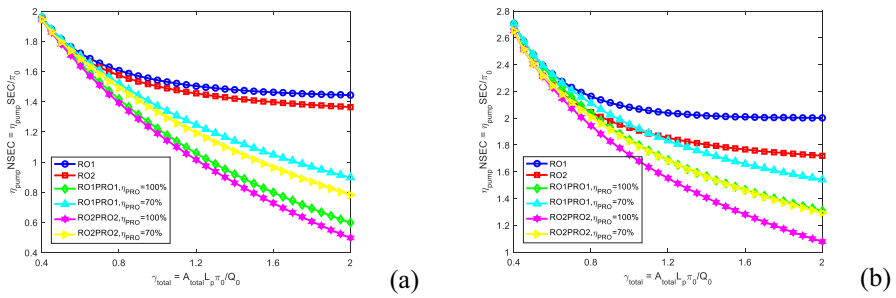
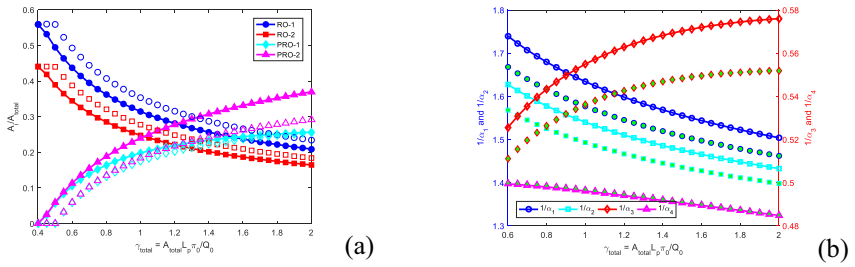


Fig. 4 Optimal NSEC in different membrane configurations under different η_{PRO} (70% and 100%) and different water recoveries Y_{total} (a) 30% and (b) 50%.

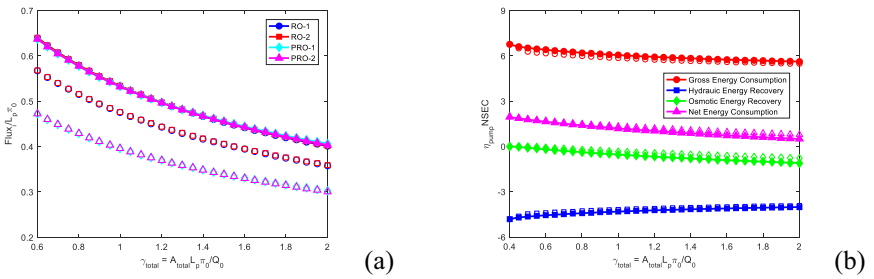
It is seen from Fig. 4 that the two-stage RO-PRO is the most energy-efficient configuration at fixed values of Y_{total} and γ_{total}. Moreover, the RO-PRO configurations keep the downhill trend of NSEC as γ_{total} increases. This implies that additional investment in membrane area and/or advances in membrane technologies may lead to a substantial reduction in SEC. It is worth noting that a representative value for γ_{total} in industrial SWRO plants is about 0.8 (Li, 2013). Mitigating the detrimental effects in PRO is very beneficial in the RO-PRO configurations, especially when water recovery is low.

Optimal allocation of membrane area and dimensionless applied pressure among all four membrane units are shown in Fig. 5. If γ_{total} is small, the PRO units are eliminated by the optimization model, and the two-stage RO-PRO collapses to a two-stage RO. When γ_{total} gets larger, the fraction of membrane area in the PRO units for energy recovery increases while that in the RO units for water recovery reduces. This is because $1/\alpha_1$ and $1/\alpha_2$ approach values closer to 1, or the RO moves towards its thermodynamic limit, and its energy efficiency approaches a plateau (as shown by the red and blue curves in Fig. 4). Therefore, the benefit of energy saving in RO resulting from additional membrane area is not as much as that of energy recovery in PRO from the same additional membrane area. Fig. 5 also implies that if the detrimental effects in PRO are small, membrane area allocation shifts from RO to PRO. Moreover, it allows a larger hydraulic pressure ($1/\alpha_3$) to be applied in the first PRO stage, enhancing osmotic energy recovery (Li, 2017).



(a) (b)

Fig. 5 Optimal (a) membrane area and (a) pressure among all membrane units in two-stage RO-PRO under different η_{PRO} (100%, solid symbols and 70%, hollow symbols) at 30% water recovery.



(a) (b)

Fig. 6 Optimal (a) dimensionless flux and (b) dimensionless energy consumption (positive) and recovery (negative) in two-stage RO-PRO under different η_{PRO} (100%, solid symbols and 70%, hollow symbols) at 30% water recovery.

The optimal flux in each membrane unit at a 30% water recovery is shown in Fig. 6(a). If the detrimental effects in PRO can be completely eliminated, it appears that all membrane units should have about the same flux to minimize the SEC. The presence of detrimental effects in PRO leads to a reduction of water flux in not only PRO, but RO as well.

The objective function $\eta_{pump}NSEC$ in Eq. (4) is in fact the sum of several terms including the gross energy consumption (or $(Y_1/\alpha_1 + 1/\alpha_2)/Y_{total}$), hydraulic energy recovery, (or $-\eta_{ERD}(1 - Y_2)/\alpha_2/Y_{total}$), and osmotic energy recovery (or $-\eta_{ERD}[(q_3 - 1)/\alpha_3 + (q_4 - 1)/\alpha_4]/Y_{total}$). The relative magnitude of these terms and their sum are shown in Fig. 6(b). It appears that the osmotic energy recovery by PRO is by no means close to the hydraulic energy recovery if γ_{total} is no greater than 2. Because of competition between the ROs and the PROs, the osmotic energy recovery can only be partially done. Considering the presence of detrimental effects in PRO, its value may be even smaller.

4. Conclusion

RO-PRO designs continue the downhill trend in the NSEC vs γ curve, so that advances in membrane technology, and/or additional investment in membrane, will bring about a further substantial reduction in SEC in SWRO. Two-stage RO-PRO offers more savings than single-stage RO-PRO, especially at higher recoveries. However, the osmotic energy recovery is much lower than the hydraulic energy recovery in RO-PRO processes if $\gamma \leq 2$. Therefore, even if the RO-PRO technology is commercialized, the SWRO industry may still not observe an energy saving on the same magnitude of the one brought about ERDs, unless a breakthrough in membrane technology allows a large γ in plant design.

References

A. Achilli, T. Y. Cath, A. E. Childress, 2009, Power generation with pressure retarded osmosis: An experimental and theoretical investigation. J. Membr. Sci., 343, 42-52.

G. Han, Z. L. Cheng, T. S. Chung, 2017, Thin-film composite (TFC) hollow fiber membrane with double-polyamide active layers for internal concentration polarization and fouling mitigation in the osmotic process, J. Membr. Sci., 523, 497-504.

K. J. Gabriel, M. M., El-Halwagi, P. Linke, 2016, Optimization across the water-energy nexus for integrating heat, power, and water for industrial processes, coupled with hybrid thermal-membrane desalination. Ind. Eng. Chem. Res., 55, 3442-3466.

D. J. Garcia, F. You, 2016, The water-energy-food nexus and process systems engineering: a new focus, Comput. Chem. Eng., 91, 49-67.

M. Li, 2011, Reducing specific energy consumption in reverse osmosis (RO) water desalination: An analysis from first principles, Desalination, 276, 128-135.

M. Li, 2013, A unified model-based analysis and optimization of specific energy consumption in BWRO and SWRO, Ind. Eng. Chem. Res., 52, 17241-17248.

M. Li, 2017, Reducing energy consumption of seawater desalination: Staged RO or RO-PRO? Desalination, 422, 124-133.

M. Li, 2018, Systematic analysis and optimization of power generation in pressure retarded osmosis: Effect of multi-stage design, AIChE J., 64, 144-152.

M. Li; T. Bui, S. Chao, 2016, Three-dimensional CFD analysis of hydrodynamics and concentration polarization in an industrial RO feed channel, Desalination, 397, 194-204.

M. Li, 2015, Analysis and optimization of Pressure Retarded Osmosis for power generation, AIChE J., 61, 1233–1241.

Z. Li, A. Siddiqi, L. D. Anadon, V. Narayanamurti, 2018, Towards sustainability in water-energy nexus: Ocean energy for seawater desalination. Renew. Sust. Energ. Rev., 82, 3833-3847.

R. E. Smalley, 2003, Top Ten Problems of Humanity for Next 50 Years, Energy & NanoTechnology Conference, Rice University.

R. L. Stover, 2007, Seawater reverse osmosis with isobaric energy recovery devices, Desalination, 203, 168-175.

A. Zhu, P. D. Christofides, Y. Cohen, 2009. Effect of thermodynamic restriction on energy cost optimization of RO membrane water desalination. Ind. Eng. Chem. Res. 48, 6010-6021.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50311-6

Energy Network Optimization in an Oil Refinery

Elif Mete[a], Metin Turkay[b*]

[a]*TUPRAS R&D Center, Turkish Petroleum Refineries, Korfez, Kocaeli, 41780 Turkey*
[b]*Koc University, Rumelifeneri Yolu, Sariyer, Istanbul, 34450 Turkey*
mturkay@ku.edu.tr

Abstract

Management of energy is a critical factor in refinery operations, having a significant impact on the production costs. The effective management of the energy system in the refinery can improve the economic performance significantly. The energy demand in refineries changes continuously in the presence of changing crude oil properties, operation conditions of process units, and cost of fuels. We present a decision support system to manage the complex energy network of the refinery by determining the optimum operational combinations of the equipment for minimizing the energy costs. In addition to operational optimization of the energy network, we also examine the impact of policy decisions through scenario analysis. We show that around 3.5 % cost reduction is possible without capital investment.

Keywords: energy network optimization, utility systems, mixed-integer programming

1. Introduction

The production of refined petroleum products is highly energy intensive. Although refinery processes involve both energy generating and consuming operations, extra energy is required to keep operational stability. Most of the refineries operate its own utility system to satisfy the energy demand. Refinery energy production expenses comprise a significant portion of the refinery budget. Effective energy management is essential to improve the effectiveness while minimizing the operational costs in a very competitive environment. However, because of the complexity of energy network, the decision-making process is a very problematic task. In addition, the energy demand for refining operations changes time due to varying crude oil properties and operating conditions. Optimization in the energy network system minimizes the operation cost and allows to get the most value from the existing system.

Important researches and contributions have been addressed for the problem of utility system optimization. One of the first important studies in this field has been published by Papoulias and Grossmann (1983). A MILP model was proposed in their study to the synthesis of utility systems with fixed steam and electricity demands. Bruno et al. (1998) states a MINLP approach for structural and parameter optimization in utility systems. The optimization algorithm decides the configurations and loads of the equipment but also the operating conditions like enthalpies and efficiencies.

In the study of Iyer and Grossmann (1997), a MILP mathematical model is developed to determine the optimum operational combinations of the equipment at minimum cost of operation. The objective function takes into account changeover costs for units. In solution, a shortest path algorithm is preferred instead of branch and bound to shorten the computation time. The study of Velasco-Garcia et al. (2011) accounts the cost of

shut-downs and start-ups of utility operating units, as well. A SMILP model approach proposed for varying utility demand of process.

Zhang et al. (2008) presented an overall optimization model for process and utility operation. In the traditional approach for oil refinery-wide production planning, utility system and process system are optimized, separately. LP models of the process units are developed first to obtain a planning strategy. Production planning model determines crude oil selection, product and allocation of utilities. Secondly, the utility system is optimized to meet the energy demean at minimum operational cost. The production planning model is much more complicated than the utility optimization model. By separating these two models, the interrelationship between them is simplified and the global optimum result may not be achieved. Therefore, the utility model and process model are composed together to increase the optimization model efficiency.

Micheletto et al. (2008) developed a MILP model for utility plant optimization of RACAP Refinery in Sao Paulo, Brazil. The objective function is minimization of operation cost of the overall plant. The mathematical model can be simplified from MINLP to MILP, because each header temperature and pressure are taken as constant. They also. stated that MILP modelling gives very accurate optimization results.

Turkay et al. (2004) illustrated the strategic benefits of collaboration among different cogeneration systems in a single period Setting. This approach was extended to cover the investment decisions to facilitate collaboration and two-way exchange with the energy grid. (Soylu et al., 2006). Ozturk and Turkay (2016) showed that incorporation of new technologies to reduce the Green House Gas emission can be analysed to provide insight into long-term energy system planning using multi-objective mixed-integer optimization.

In this paper, power generation equipment in the refinery is analyzed thermodynamically to determine the variables that have an impact on the equipment efficiency. A mathematical programming model is developed to represent the energy demand of the refinery with the minimum production cost of energy. The model is formulated as an MILP problem and includes all operational constraints, mass and energy balances, demand satisfaction constraints, efficiency models and operational status of the equipment. All energy consuming/producing equipment in the refinery are defined in the model. The interactions between refinery units and the utility plants are taken into consideration. In addition to optimization of the operations in the energy network we examine two scenarios where the energy policy changes are examined in detail. The approach that is presented in this paper had been tested for one year in online and real-time decision-making purposes.

2. The energy network model

Although typical refineries have a single utility plant, the refineries with high Nelson Complexity Index (Nelson, 1976) usually has more than one utility plant due to high production capacity and investments to upgrade these refineries that span several decades. Tupras Izmit Refinery has a Nelson Complexity Index of 14.5 with two interconnected refining facilities and three utility plants. The first refining facility (F1) contains two utility plants (PA and PB) whereas the second refining facility (F2) contains one utility plant (PC). PA has four boilers to produce very high-pressure steam, four steam turbines to generate electricity, a gas turbine and heat recovery units. PB has four boilers to produce high-pressure steam that is used in back pressure turbines to

drive pumps and compressors in the refinery. PC contains two sets of gas turbines and heat recovery units to generate ultra-high-pressure steam that is used to generate electricity in turbines. In addition, deaerators and preheaters are also present in PC that also produce low-pressure steam to be used in pumps and compressors. The excess steam produced in F2 can be used in F1 and it is possible to transfer electricity between two facilities. It is also possible to exchange electricity (both ways) with the national grid.

All of the major units in these three utility plants are modelled using the conservation of energy, conservation of mass and real-time data from the process monitoring system. We present the model for the gas turbines in this section. A gas turbine is a type of combustion engine that is used to transform fuel energy to mechanical energy. This mechanical energy drives a generator to produce electricity. There are 3 main parts in a gas turbine: the compressor where the pressure of the incoming air is increased, the combustor that burns the fuel and produce high-pressure and high-velocity gas, the turbine where the high-pressure gas is expanded to produce work which is used to drive the generator. In the compression process, the compression ratio is calculated using,

$$r_p = Pr_2/Pr_1 \tag{1}$$

where r_p is the compression ratio, Pr_1 and Pr_2 are the inlet and outlet pressures respectively. The isentropic efficiency of the compressor (nc) is estimated with:

$$n_c = (T_{2s} - T_1)/(T_2 - T_1) \tag{2}$$

where T_{2s} is the isentropic compressor outlet temperature and T_1 and T_2 are compressor inlet and outlet temperatures respectively. Then, the work of compressor (W_c) can be calculated from:

$$W_c = C_{pa}T_1\left(r_p^{\gamma_a-1/\gamma_a} - 1\right)/n_m\, n_c \tag{3}$$

where C_{pa} is the specific heat of air, γ_a is a constant (=1.4), n_m and n_c are mechanical and isentropic efficiency of the compressor respectively. The energy balance in the combustion chamber can be expressed as,

$$m_a C_{pa} T_2 + m_f HV_f + m_f C_{pf} T_f = (m_a + m_f)TIT \tag{4}$$

where m_a and m_f are the mass flow of air and fuel, C_{pf} is the specific heat of the fuel, HV_f is the heating value of the fuel, T_f is the temperature of the fuel, TIT is the turbine inlet temperature. Then, the shaft work in the turbine section (W_t) can be expressed as,

$$W_t = C_{pf}\, TIT\, n_t \left(1 - \frac{1}{r_p^{\gamma_g-1/\gamma_g}}\right)\Big/ n_m \tag{5}$$

where n_t is the turbine efficiency and γ_g is a constant with a value of 1.33. The net work in the gas turbine system is the difference between W_t and W_c. In addition to the fundamental equations above, the efficiency of the gas turbines are directly affected by different factors including air temperature at the compressor inlet, air humidity, fuel

characteristics, steam and water injection during combustion. The process data were used to perform a regression analysis for power production in the gas turbines. The agreement between the process data and the regression model is very high as shown in Fig.1.

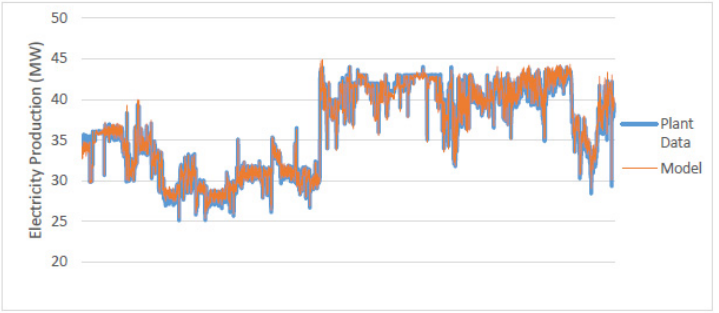


Fig.1 Comparison between the regression model and the plant data.

After the mathematical models for every major equipment are constructed, the optimization model for the entire energy network is assembled. The objective function is defined as the minimization of the operating cost that is subject to all material and energy conservation laws, equipment characteristics and efficiencies and electricity trade (buy/sell) options and emission limits and penalties. The binary variables are used to model the on/off status of the major equipment during the operations.

3. Results

The optimization problem is routinely solved to support online real-time decision-making regarding the operation of the energy network of the refinery. Typical MILP model for this purpose includes 28,104 variables (190 are 0-1 variables) and 10,656 constraints. The optimization is based on hourly operation plans.

3.1. Base Case

In order to assess the validity of the optimization results and analyse the benefits from the optimization approach by satisfying the electricity and steam demand in the two facilities of the refinery. The optimization of the energy network favors the most efficient equipment and allocates higher loads to them and shuts down some of the boilers. The optimal operation decreases the fuel gas consumption and the natural gas consumption increases. In order to eliminate the problem of drastic changes in the operation of the energy network, the furnace operations were not changed so the fuel consumption at each furnace remains unchanged. The green house gas emission levels are also calculated considering the operation before and after the optimization. By optimizing the energy network, the total energy of consumed fuel in the utility plants is decreased from 427.53 Gcal/hr to 418.97 Gcal/hr. This decrease also decreases CO_2 emissions from 101.1 ton/hr to 99.1 ton/hr while SO_2 emission is kept at 0. The total cost of operations includes demineralized water production cost and fuel cost consumed in the energy production. The overall savings for the base case are calculated as 2.51 %.

In addition to this regular decision support in the operations, the system is also used for scenario analysis to support policy decisions.

3.2. Scenario 1: Electricity buy/sell

Most of the refineries prefer to generate the electricity in their own utility plants for the continuity of their operations. The impact of electricity exchange with the national grid is analyzed in this scenario. The electricity buy/sell prices were obtained from the national grid web site and the option to meet the electricity demand of the refinery fully or partially from the national grid is considered as well as selling electricity if the refinery demand is satisfied and generating additional the electricity to sell to the national grid is economically advantageous. When selling electricity is economically favorable, the steam demand in the refinery stays the same and all of this demand is satisfied, electricity generation in gas turbines increase to 2.5 MW, the fuel gas consumption is reduced by 14.73 % and the natural gas consumption increases by 3.5 %. The total energy of consumed fuel in the utility plants is reduced from 427.53 Gcal/hr to 418.97 Gcal/hr and the CO_2 emissions are reduced from 101.1 ton/hr to 99.8 ton/hr. The total operating cost of the energy production is decreased by 2.75 %.

A sensitivity analysis study is carried out to examine the effect of the electricity price on the electricity trade with the maximum electricity trade limits for buying and selling are determined as 15 MW. Figure 2 shows the optimum amount of electricity purchased or sold at the corresponding electricity price. It is more profitable to buy electricity from the national grid for the prices under the approx. 105 TL/MW.

Fig.2 Sensitivity of electricity trade with respect to electricity price.

3.3. Scenario 2: Rolling drives

In this scenario, the rolling equipment are involved in the optimization algorithm; sparing the rolling drives allows multiple operational configurations of the units. For the least cost operation, it is important to choose the optimal set of drives to operate. Switching sparing drives for optimization changes the steam and electricity demand of the refinery. In addition, the electricity trade is not limited to 15 MW as in the previous scenario. The main difference in this case is the switching of 9 rolling drives that powered by steam to electricity. Thus, the electricity demand of the refinery increases by 0.49 MW, while HP steam demand decrease by 44.88 ton/hr. The total fuel gas consumption in the utility plants is decreased, while total natural gas consumption is

increased. The CO_2 emission level is reduced from 101.1 ton/hr to 98.9 ton/hr. The overall saving in the total operating cost is 3.451 %.

4. Conclusions

In this paper, a mathematical model for optimization the operations of the complex energy network in the refinery is presented. The problem is formulated as a mixed-integer linear program with the objective function of minimizing the total energy production cost. The mathematical model includes all operational constraints, mass and energy balances, demand satisfaction constraints, efficiency models and operational modes of the equipment in the energy network. The developed approach is analyzed and tested on real plant data of the Tupras Izmit Refinery. The online utility demand of the refinery is determined by receiving live plant data for steam, fuel, condensate, BFW, and electric system metering devices via plant information system. The problem is modeled in GAMS and solved by CPLEX solver for online real-time decision support.

In scenario analysis, two optimization cases are conducted to analyze the policy decisions in the operation of the refinery. The main objective of these optimization studies is to validate the extent of the developed model for policy analysis and to evaluate the economic impact of the optimization. In addition, a detailed sensitivity analysis is carried out to examine the effect of the electricity price on the electricity trade. Up to 3.5 % cost reduction is achieved comperes to base cases.

References

J.C. Bruno, F. Fernandez, F. Castells, and I.E. Grossmann, 1998, A rigorous MINLP model for the optimal synthesis and operation of utility plants. Chemical Engineering Research and Design, 76(3) , 246-258.

R.R. Iyer and I.E. Grossmann, 1997, Optimal multiperiod operational planning for utility systems. Computers and Chemical Engineering, 21 (8), 787–800.

S.R. Micheletto, 2008, Operational optimization of the utility system of an oil refinery. Computers and Chemical Engineering, 32(1-2), 170-185.

W.L. Nelson, 1976, Guild to refinery operating cost (process costimating). 3rd ed., Petroleum Publishing Co, Tulsa, OK.

A. Ozturk, and M. Turkay, 2016, Bicriteria Optimization Approach to Analyze Incorporation of Biofuel and Carbon Capture Technologies. AIChE J.,62(9), 3473-3483.

S.R. Papoulias and I.E. Grossmann, 1983, A structural optimization approach in process synthesis-i. utility systems. Computers and Chemical Engineering, 7(6), 695-706.

A. Soylu, C. Oruc, M. Turkay, K. Fujita, and T. Asakura, 2006, Synergy Analysis of Collaborative Supply Chain Management in Energy Systems using Multi-Period MILP. European Journal of Operational Research, 174(1), 387-403.

M. Turkay, C. Oruc, K. Fujita, and T. Asakura, 2004, Multi-Company Collaborative Supply Chain Management with Economical and Environmental Considerations. Comput. Chem. Eng., 28(6-7), 985-992.

P. Velasco-Garcia, P.S. Varbanov, H Arellano-Garcia and G. Wozny, 2011, Utility systems operation: Optimisation-based decision making. Applied Thermal Eng., 31(16) 3196-3205.

B.J. Zhang and B. Hua, 2007, Effective MILP model for oil refinery-wide production planning and better energy utilization. Journal of Cleaner Production, 15(5), 439-448.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50312-8

Optimal Supply Chain for Biofuel Production under the Water-Energy-Food Nexus Framework

Dulce C. López-Diaz[a], Luis F. Lira-Barragán[a], J. Betzabe González-Campos[a], Medardo Serna-González[a], Mahmoud M. El-Halwagi[b], Jose M. Ponce-Ortega[a*]

[a]*Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Michoacan, 58060, México*
[b]*Chemical Engineering Department, Texas A&M University, College Station, TX, 77843, USA*
jmponce@umich.mx

Abstract

This work proposes a mathematical optimization model for the design of a biorefining system through an efficient supply chain considering economic (minimization of the total annual cost) and environmental objectives (minimization of the overall CO_2 emissions), through the water-energy-food nexus. The proposed model is a mixed-integer linear programming (MILP) problem, which includes the material flow analysis (MFA) technique in order to simulate the changes that take place in a watershed system impacted by the implementation of the biorefining system and determine the water consumption caused mainly by the production of the feedstock and for the operation of the biorefineries. The methodology is configured to account for the uncertainty inherent to prices of feedstocks and products, biofuel demands, biomass requirements, weather conditions, grain production, etc. Finally, a case study located in Mexico is solved for a set of scenarios with the purpose of illustrating the capabilities of the approach. The results show the trade-offs among the considered objectives impacted by the uncertainty in the biorefinery system.

Keywords: Biorefinery, optimization, supply chain, material flow analysis, uncertainty.

1. Introduction

Nowadays, biofuels have been recognized as potential options to partially replace fossil fuels due to the decrease in the production projections for fossil fuels, and the stricter environmental legislations for the use of them, which have produced huge environmental impacts during decades. In spite of the economic advantages that fossil fuels have with respect to renewable alternatives, biofuels offer several environmental advantages including the reduction of the carbon footprint on a life cycle basis and the enhancement of renewability and sustainability. A biorefinery is an industrial processing plant where value-added chemicals and/or biofuels are produced using different renewable raw materials (grains and starch crops, agricultural residues, food waste, forestry materials, etc.). The design of a biorefinery is based on the optimal use of natural resources. For example, Khalidi et al. (2016) defined the pathways for producing biofuels. Because the biorefinery design is a complex task, which interacts with many other sectors, the entire associated supply chain must be simultaneously considered. Cucek et al. (2014) proposed a method for synthesizing a biorefinery supply chain through a mixed-integer linear programming model. Murillo-Alvarado et al. (2013) proposed a methodology dealing with the selection of feedstocks, products,

processing routes and processes for the separation of products for a biorefinery. Recently, López-Díaz et al. (2017) developed a mathematical model to incorporate a supply chain for the production of biofuels with a material flow analysis (MFA) approach with the purpose of designing an integrated biorefinery system with the environment to implement a study for the water-biofuel nexus. It should be noted that none of the previously reported methods have considered the relationship between the water and food nexus in the design of biofuels supply chains. Therefore, this work presents a superstructure to account for the interaction of the water-food nexus in the design of a biofuel supply chain, which incorporates the MFA technique to track the impact of the wastewater streams on the environment and to quantify the overall water requirements of the project). The proposed approach considers simultaneously the optimal location of a set of biorefineries, the type and amounts of feedstocks and biofuels (attempting to satisfy a given demand), the overall CO_2 emissions (including the CO_2 fixed by the crops during their growth period), the use of resources (water and land requirements), the routes of transportation and distribution for raw materials and biofuels, and the types of processing technologies. The optimization approach is driven by economic (maximization of the total annual cost) and environmental (minimization of the overall greenhouse gas emissions) goals. Furthermore, the optimization model accounts for the interactions by the sale of grains associated with the feedstock production.

2. Model formulation

A superstructure was proposed to consider the relationships between the agricultural products and the production of biofuels as well as the implication with food production in the cultivating sites. Additionally, the scheme is able to determine the impact in the watershed in terms of the consumption and discharges of effluents. One of the most important contributions is the consideration of the involved uncertainty in the model, because a large number of parameters are uncertain including the water availability, weather conditions, raw material prices, and biofuel prices. The proposed optimization model is described as follows:

2.1 Watershed model

2.1.1 Watershed at initial conditions before the implementation of the biorefining system. In this balance, the flowrate in reach *r* includes discharges and extraction to the watershed from the agricultural, biorefinery and urban activities surrounded the watershed as well as the weather condition (precipitation and evaporation).

$$Q_{r,s}^{Out} = Q_r^{In} + P_{r,s} + D_{r,s} + H_{r,s} + \sum_{t=1}^{N_{(t)}} FT_{r,t,s} + \sum_m w_{m,r,s}^{Discharge-feedstocks} + \sum_k w_{k,r',s}^{Discharge-biorefinery} - \sum_m w_{m,r,s}^{Used-feedstocks} - \sum_k w_{k,r',s}^{Used-biorefinery} - L_{r,s} - U_{r,s}, \qquad \forall r, \forall s \tag{1}$$

2.1.2 Overall water balance for each tributary of the watershed. The tributaries are little rivers that flow into the watershed, the balance considers the interaction in the reaches directly.

$$FT_{r,t,s} = S_{r,t}^{Untreated} + S_{r,t}^{Treated} + I_{r,t,s} + P_{r,t,s} + D_{r,t,s} - L_{r,t,s} - U_{r,t,s}, \qquad \forall r, \forall t, \forall s \tag{2}$$

2.2 Modelling of the biorefining system

In the design of the corresponding supply chain are estimated the capacities of process operation, biomass production, cultivating area, demand of water, logistic of distribution, etc.

2.2.1 Production at the cultivating fields.

$$f_{m,r,s}^{Prod-feedstock} = \alpha_{m,r,s}^{fedstock}\left[a_{m,r}^{exist} + a_{m,r}^{new}\right], \qquad \forall m, \forall r, \forall s \tag{3}$$

2.2.2 Activation of binary variables for biorefineries.

$$F_{m,r',s} \le \mathrm{F}_{\mathrm{m,r'}}^{\max}\, y_{r',s}^{Biorefinery}, \qquad \forall r', \forall s \tag{4}$$

2.2.3 Production of biofuels.

$$P_{k,m,r',s}^{Bio} = \alpha_{k,m,s}^{Bioref} F_{m,r',s}, \quad \forall k, \forall m, \forall r', \forall s \tag{5}$$

2.2.4 Products obtained in markets.

$$\sum_{r'} P_{k,r',j,s}^{Prod-market} \le P_{k,j,s}^{Prod-market-Max}, \qquad \forall k, \forall j, \forall s, \forall r' \tag{6}$$

2.3 Gains and costs

In the economic objective, the capital and operation costs are considered in the total annual cost for biorefineries and the agriculture activity.

2.3.1 Gains obtained by the selling of products.

$$G_s^{Products} = \sum_k \sum_{r'} UC_{k,s}^{Product} P_{k,r',s} \tag{7}$$

2.3.2 Gains obtained by selling grains.

$$G_s^{grains} = \sum_m \sum_r UC_{m,s}^{Grains} f_{m,r,s}^{\mathrm{Prod-grains}}, \ \forall s \tag{8}$$

2.3.3 Capital cots for biorefineries.

$$CapCost_s^{Biorefineries} = k_F\left[\sum_m \sum_{r'} FC_{m,r'}^{Biorefinery} y_{r',\mathrm{s}}^{Biorefinery} + \sum_m \sum_{r'} VC_{m,r'}^{Biorefinery}\left(F_{m,r',\mathrm{s}}\right)^{\sigma_{m,r'}^{Biorefinery}}\right], \ \forall s \tag{9}$$

2.3.4 Operating costs for biorefineries.

$$OpCost_s^{Biorefineries} = \sum_m \sum_{r'} UC_{m,r'}^{Biorefinery} F_{m,r',\mathrm{s}}, \ \forall s \tag{10}$$

2.4 Objective functions

The first objective is economic that is defined as the net gains in the supply chain, then the environmental objective functions represent the total used water and emissions.

2.4.1 Total annual profit

$$Profit_s = G_s^{Products} + G_s^{Grains} - CapCost_s^{Biorefineries} - OpCost_s^{Biorefineries} - OpCost_s^{Feedstocks} - OpCost_s^{Water-biorefineries} - OpCost_s^{Water-feedstocks} - TCost_s^{Feedstocks} - TCost_s^{Grains} - TCost_s^{Products} \quad (11)$$

2.4.2 Total water requirement

$$Water_s^{required} = \mathrm{Q}_{s\ \mathrm{final}}^{\mathrm{Initial}} - Q_{s\ final}, \quad \forall s$$

2.4.3 CO_2 emissions

$$CO2_s^{total} = CO2_s^{prod} P_s^{biototal} - CO2_s^{abs} \mathrm{A}_s^{total}, \quad \forall s \quad (12)$$

3. Results and discussion

The proposed model was coded in the General Algebraic Modelling System (GAMS) as a mixed-integer lineal programing problem. A case study was solved to evaluate the capacities of the proposed model considering the installation of a biorefinery system in the center region of México. The Balsas river is considered as watershed to study the behavior of the water resources, which was divided in 23 fixed reaches defined accounting for studies by the National Water Council (CONAGUA, 2016). The total biofuel demands are presented as requirements by ten markets (important cities along the watershed), where the ten percent of the total energy demand is satisfied by biofuels. A biorefinery can offer a wide platform of products, but for this case only was considering the bioethanol and biodiesel productions. For this system, the agricultural information such as soil conditions, potential cultivation area, crop yields, demand of water by crops, etc., are provided by SAGARPA (2016). The considered biorefineries are configured to use second generation biomass in the production of biofuels in specific agricultural residues of corn, wheat, sorghum and sugar cane to produce bioethanol, and jatropha and oil palm to produce biodiesel. The case study works with the next distribution of prices: $US 0.45/L and US$0.66/L for bioethanol and biodiesel respectively, for agriculture products as grains of corn wheat, sorghum we have $US 140/t, $US 142/t, $US 138/t, $US 30/t, and for sugar cane, jatropha, and oil palm $US 4000/t and $US 304/t provided by World Bank database. In this case, some parameters for products and feedstocks, weather conditions like precipitation, evaporation and filtration as well as some agricultural parameters are modeled under uncertainty. In Excel®, n-scenarios are generated through the Latin hypercube sampling method under certain probability of distribution using a normal distribution of 90% and 95%. Then, fifty scenarios were generated using Monte Carlo simulation to determine the optimal design of the supply chain necessary for defining the location and operation of the biorefinery system.

The problem was analyzed considering four situations; pessimist and optimistic conditions for the profit and for the consumption of water in the project. Then, the multi-objective optimization problem was solve using the solver CPLEX with 309,650 continuous variables, 23 binary variables, and 84,947 constraints. The CPU time depends on each scenario, where the corresponding solution for the analyzed specific scenario is presented in Table 1. The time horizon is a year with an investment time of ten years. With the purpose of making a wide analysis, it is selected 3 specific scenarios: the first has 20% of probability, the second has 70% of occurrence and finally the scenario 3 has 94% of occurrence. The model also was analyzed for economic and environmental expected and pessimistic conditions.

Table 1. Case study solutions.

Case	CPU TIME (s)	OPTIMAL SOLUTION
Maximizing Mprofit	2426	US$6.79x10^8/y
Maximizing Wprofit	90.4	US$6.44 x10^8/y
Minimizing Mwater	244.7	0.639 m^3
Minimizing Wwater	27.05	0.016 m^3

Figure 1 shows the results for the scenarios maximizing the expected profit and the worst or pessimistic profit against the demand of water under these conditions. In this sense, Figure 2 presents the configuration for the solution at scenario with 94% of probability of occurrence (point 3). The expected profit is \$US 7.58 x10^8 /y, and the required fresh water is 3.24 x10^7 m^3/y, where there are required 7 biorefineries to satisfy the biofuel demands in the markets. The total biofuel production is 6.71 x10^7 L/y of bioethanol and 8.66 x10^7 L/y of biodiesel. Comparing with the worst or pessimist case, the profit was obtained with a value of US\$ 7.32 x10^8 /y and the fresh water consumption increases to 3.25 x10^7 m^3/y. This scenario requires a new cultivating area of 19 377 ha to satisfy the biomass demands. The decision about how much of specific crop can be required as biomass depends of the land characteristics, cost of farm and harvesting, as well as the transportation logistic. The supply of the markets is satisfied considering the interaction between the agricultural, food and biofuel to define an efficient supply chain and optimal location of a biorefinery system.

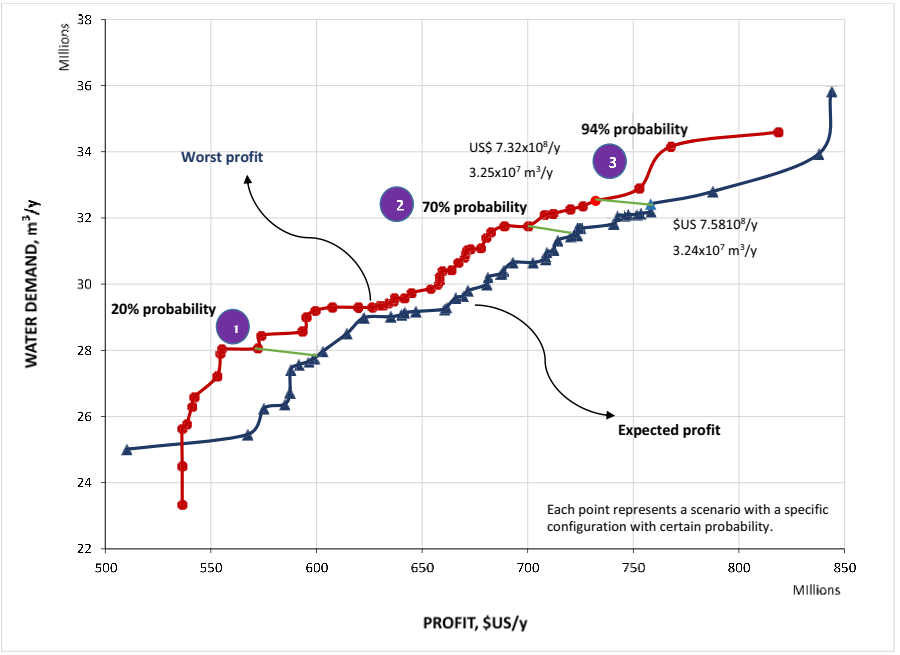


Figure 1. Pareto curve for the profit vs water demand.

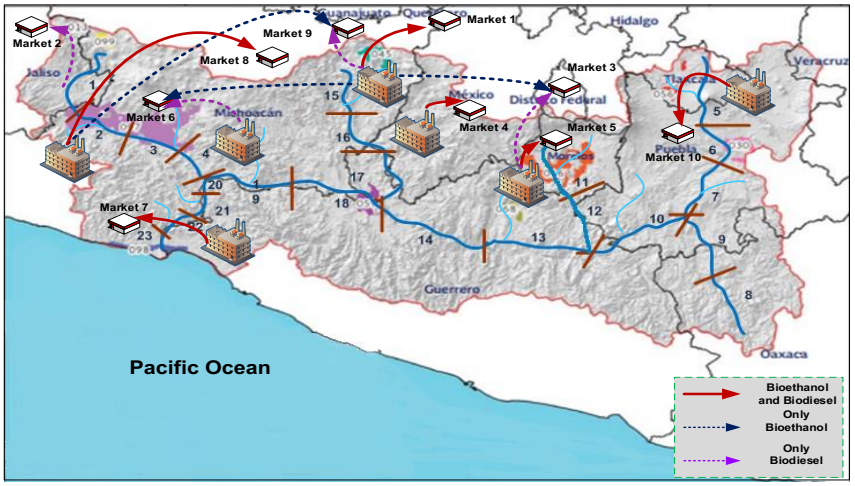


Figure 2. Optimal configuration for the biorefinery system in scenario 3 for expected profit.

4. Conclusions

The proposed optimization approach determines the location, operation and logistic of distribution for biomass and biofuels, as well as new cultivation areas, fresh water demand, the impact on the watershed by design and efficient supply chain for the biorefinery system considering simultaneously economic and environmental objectives. The biofuels have proved to be a potential energy alternative but the disadvantage is found in the nexus that this sector has with the food sector. The cultivation of new land to cultivate biomass can be compromised by food production and in some cases the impact on the watershed for extraction and discharges by agricultural and industrial uses is greater that the benefit of greenhouse emission. This work is a powerful tool for decision makers to have a more complete analysis for the production of biofuels.

References

CONAGUA, 2016, Mexican National Water Commission. Water Statistics 2016, https://www.gob.mx/cms/uploads/attachment/file/115145/SGP-2-14Web.compressed.pdf, Accessed March 2017.

L. Cucek, M. Martín, I.E. Grossmann, Z. Kravanja, 2014, Multi-period synthesis of optimally integrated biomass and bioenergy supply network, Comput Chem Eng, 66, 57–70.

O. Khalidi, S. Guleria, M.A.G. Koffas, 2016, Chapter 4-Pathway and strain design for biofuels production, Biotechnology for biofuel production and optimization, Elsevier B.V., New York, 97–116.

D.C. López-Díaz, L.F. Lira-Barragán, E. Rubio-Castro, J.M. Ponce-Ortega, M.M. El-Halwagi MM, 2017, Optimal location of biorefineries considering sustainable integration with the environment, Renew Energy, 100(1), 65–77.

P.E. Murillo-Alvarado, J.M. Ponce-Ortega, M. Serna-González, A.J. Castro Montoya, M.M. El-Halwagi, 2013, Optimization of pathways for biorefineries involving the selection of feedstocks, products, and processing steps, Ind Eng Chem Res, 52(14), 5177–5190.

SAGARPA-SIAP, 2016, Mexican System of Information about Agriculture and Fishing. Advance of planting and harvesting for Mexico. Mexico City, Mexico, www.siap.gob.mx/index.php?op tioncom_wrapper&;viewwrapper&Itemid347, Accessed March 2016.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50313-X

Systematic framework development for the construction of surrogate models for wastewater treatment plants

Resul Al, Chitta Ranjan Behera, Alexandr Zubov, Gürkan Sin*

Process and Systems Engineering Center (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, Søltofts Plads 227, 2800 Kgs. Lyngby, Denmark.
gsi@kt.dtu.dk

Abstract

Surrogate modeling (also referred to as metamodeling) has attracted increased attention from researchers in various fields of engineering due to its use in computationally expensive engineering tasks such as Monte Carlo based global sensitivity analysis and process design optimization. However, the applications of surrogate models in the field of wastewater treatment modeling have not been extensively explored in the literature. In this work we present a systematic methodology for construction of powerful surrogate models to be used for the global sensitivity analysis of Benchmark Simulation Model 1 (BSM1) plant. A quasi-random design of experiments technique, Sobol sampling, is employed to generate an experimental design, which is further used to build surrogate models. A class of advanced metamodeling algorithms such as sparse polynomial chaos expansion (PCE) using least angle regression, Kriging interpolation, polynomial chaos Kriging (PCK), radial basis function (RBF) interpolation, multivariate adaptive regression splines (MARS), and a multilayer perceptron type feedforward neural network (ANN) are applied. The generalization error of the developed models has been estimated using holdout cross-validation with coefficient of determination (R^2) and root mean squared error (RMSE) being used for evaluation of model predictive accuracy and model selection, respectively. The framework was further investigated for its suitability in a Monte Carlo based global sensitivity analysis using Sobol' method. The results obtained suggest that by following the framework, ANN and Kriging type surrogate models can effectively be constructed and used to estimate Sobol' sensitivity indices of WWTP design parameters.

Keywords: surrogate modelling, wastewater treatment modelling, design of experiments, metamodeling, global sensitivity analysis.

1. Introduction

Wastewater treatment plant (WWTP) design is motivated towards meeting effluent quality requirements, which are regulated by governmental and intergovernmental bodies. Among the most important effluent discharge limits are chemical oxygen demand (COD), total nitrogen (TN), total suspended solids (TSS), and ammonium (SNH) concentrations. In the WWTP community, significant effort has been invested into developing activated sludge process mechanistic models and benchmarking of wastewater treatment plant control strategies, which produced widely used Activated Sludge Models (ASMs) and Benchmark Simulation Models (BSMs) (Gernaey *et al.*,

2014). Regarding the use of surrogate models in the community, Raduly *et al*. (2007) developed artificial neural networks for rapid performance evaluation of WWTPs. Moral *et al*. (2008) developed an ANN for ASM. However, none of them focused on performance comparison of different type of surrogate models. In this work we present a new systematic framework for efficient construction of plant-wide surrogate models using advanced surrogate modeling and data-driven algorithms including Kriging, polynomial chaos expansions, and neural networks from a process design point of view. In what follows, we describe our modeling approach of WWT process, and the surrogate model construction framework with an end use application of a global sensitivity analysis.

2. Surrogate Modeling

MARS is a powerful nonparametric regression algorithm using two-sided truncated basis functions to explain the relationships between the predictor and response variables. More details about this algorithm can be found in Friedman (1991). Originally developed for scattered multivariate data interpolation, RBF is an exact interpolation algorithm (model predictions goes through the training data points) using various basis functions such as biharmonic, multiquadric, inverse multiquadric, thin plate spline, and Gaussian. Jekabsons implemented MARS (2016) and RBF (2009) in MATLAB environment, which were used for this study. Marelli and Sudret (2016) developed UQLab framework for uncertainty quantification, which also implemented surrogate modeling algorithms including Kriging and polynomial chaos expansions using least angle regression, and polynomial chaos Kriging. More explanations about how these algorithms work can be found in the user manuals of the framework. As these algorithms allow defining many configurable model parameters, we employ a grid search approach to find the optimum set of parameters, which performs a combinatorial optimization by screening over each different combinations and finding the model with minimum root mean square error on the test sample. The same approach is also used for finding optimum MARS and RBF models. A multilayer perceptron type feedforward neural network comprising of one hidden layer (with different number of neurons) and one output layer was developed using the Neural Network Toolbox of MATLAB (Mathworks Inc.). The grid search is also used to find the optimum number of neurons in the hidden layer, transfer functions (i.e. logarithmic, tangential, and linear sigmoid functions), and training algorithms (i.e. Bayesian regularization backpropagation, Levenberg-Marquardt backpropagation, and scaled conjugate gradient backpropagation).

3. Methodology

The BSM1 model has a relatively simple plant layout consisting of a five-compartment activated sludge reactor (two anoxic tanks followed by three aerobic tanks) and a secondary settler. From a process design point of view, we create an initial input space with five design-related parameters taken from Sin *et al.* (2009). These include hydraulic residence time in the anoxic tank (HRT_{an}) (varied between 1 and 7 hours), volume ratio (VR) of anoxic tank to aerobic tank (varied between 0.2 and 0.8), ratio of external recycle flowrate to influent flowrate (RR_e) (varied between 0.4 and 1.4), ratio of internal recycle flowrate to influent flowrate (RR_i) (varied between 2.8 and 4.2), and wastage flowrate (Q_w) (varied between 300 and 450 m^3/day). To reach a steady state,

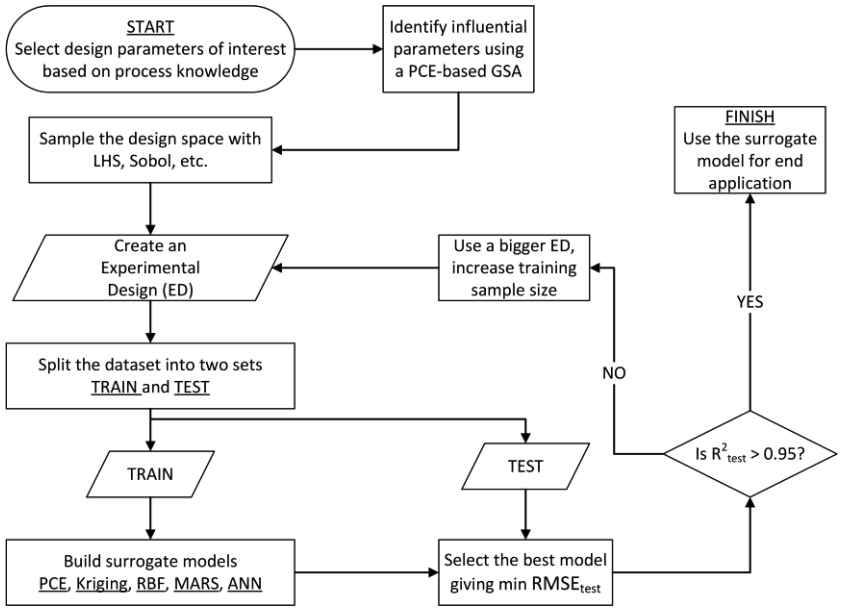


Figure 1: Systematic framework for developing surrogate models to be used in WWTP modeling

the BSM1 model is simulated for 150 days using the dry weather influent file. Model outputs of interest included effluent quality indices , such as COD, TSS, TN, and SNH.

Figure 1 shows the methodological framework used to build the surrogate models. After deciding on the design parameters, in order not to include an uninfluential parameter in the design space, the framework proposes to perform a PCE-based global sensitivity analysis, following from Sudret (2008) whose ground-breaking work showed the effectiveness of PCE models in the global sensitivity analysis. This step can be achieved with a relatively small experimental design (ED) size of 200. With the new design parameters sampled between their ranges of variation using the quasi-random Sobol sampling algorithm, the framework creates a larger ED (size of which is 1000 in this case) and splits it into training set size of 500 and a global test set size of 500. The global test set was never shown to any of the algorithms during training and just used for estimating the generalization error of the models (holdout validation). By doing so, one can make sure to avoid overfitting, a challenge faced when models perform well on training sample but not on the test sample. For cases where such large test sample sizes are unaffordable, 5-fold or 10-fold cross-validation techniques can also be used. The framework further splits the training sample into smaller sub-training samples and builds surrogate models. The grid search uses unused training sample to optimize models from each algorithm whereas the global test sample provides performance score comparison of different algorithms. The computational cost of building these models depends on the training sample size and the complexity of the original model as well as the adopted computing strategy. In order to test the usability of the final developed models, a variance based global sensitivity analysis (Sobol method), which produces first and total order Sobol' indices, is performed. The theory behind evaluation of these indices can be found in Saltelli (2008). Basically, first and total indices respectively indicate individual and total effects of inputs on the variance of the model outputs.

4. Results and Discussions

As stated in the methodology, we first performed a PCE-based global sensitivity analysis, results of which are shown in Figure 2. Since they directly relate to the effective size of aerobic and anoxic tanks, HRT_{an} and VR are found to be the most influential parameters in terms of their effects on the model outputs, whereas the internal recycle ratio played no significant role on any of the outputs. Hence this was excluded from further analysis. These results also agree well with those obtained in Sin *et al.* (2011).

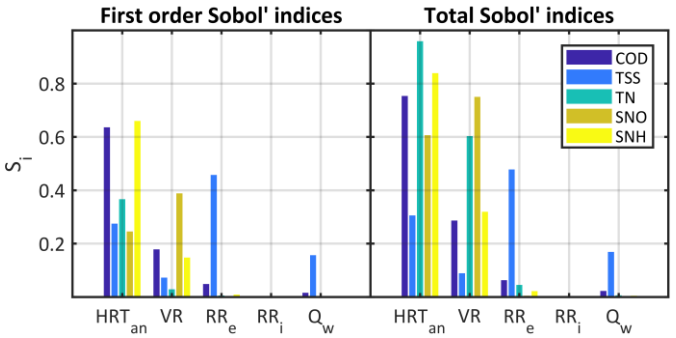


Figure 2: PCE-based global sensitivity analysis results for models outputs of interest.s.

Table 1 shows performance test scores of the developed surrogate models for the model output of COD. Scores are tabulated for training sample sizes of 100, 200 and 400 and the test sample was kept the same size of 500. Kriging and ANN type models outperform the other type of models. Table 2 shows performance scores obtained for the model output of SNH, which exhibits strong nonlinearities with the model inputs. Kriging and ANN models show good results even for the training sample size of 100. All the models show improvement as the training sample size increases.

Table 1: Performance scores of surrogate models developed for COD model output

Model type	$N_{training} = 100$		$N_{training} = 200$		$N_{training} = 400$	
	R^2_{test}	$RMSE_{test}$	R^2_{test}	$RMSE_{test}$	R^2_{test}	$RMSE_{test}$
PCE	0.88	1.33	0.89	1.19	0.95	0.78
Kriging	0.98	0.40	0.99	0.28	0.99	0.21
PC-Kriging	0.86	1.54	0.94	0.84	0.97	0.45
MARS	0.88	0.96	0.95	0.62	0.98	0.35
RBF	0.73	1.78	0.76	1.41	0.80	1.25
ANN	0.98	0.45	0.99	0.31	0.99	0.29

Table 2: Performance scores of surrogate models developed for SNH model output

Model type	$N_{training} = 100$		$N_{training} = 200$		$N_{training} = 400$	
	R^2_{test}	$RMSE_{test}$	R^2_{test}	$RMSE_{test}$	R^2_{test}	$RMSE_{test}$
PCE	0.62	9.04	0.84	5.47	0.84	5.55
Kriging	0.95	3.04	0.97	2.53	0.99	1.59
PC-Kriging	0.95	3.13	0.96	2.88	0.98	1.73
MARS	0.89	4.64	0.92	3.96	0.93	3.73
RBF	0.54	9.31	0.56	9.12	0.57	8.96
ANN	0.98	2.10	0.98	1.76	0.99	1.57

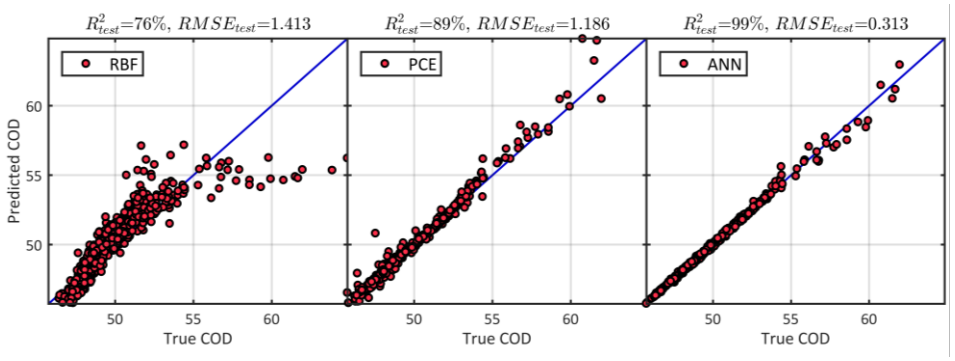


Figure 3: Predictions of RBF, PCE, and ANN type surrogate models trained on a sample size of 200 are compared for the original model response of COD.

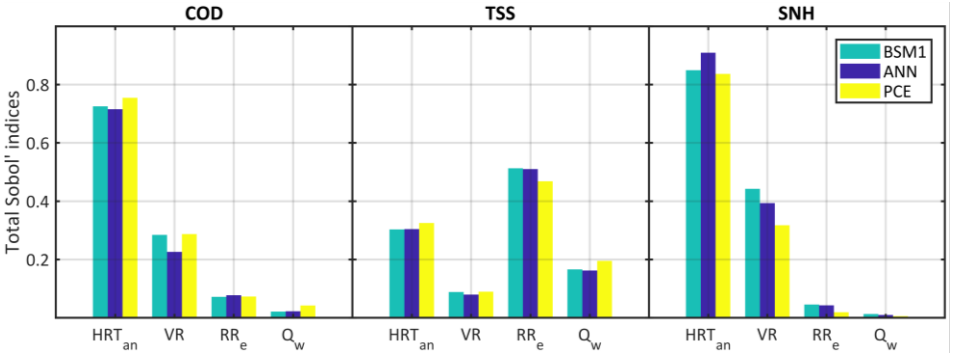


Figure 4: Total Sobol' sensitivity indices are calculated for the model outputs of COD, TSS, and SNH using both the original model (BSM1) and the surrogate models of type ANN and PCE.

Figure 3 shows a comparison of predictions of RBF, PCE, and ANN models for the model output of COD. RBF overpredicts lower responses and underpredicts higher responses. Figure 4 shows total Sobol' sensitivity indices for COD, TSS, and SHN outputs obtained using the original model (BSM1) and two surrogate models (ANN and PCE). A sample size of 2000 is used with Sobol' sampling strategy. This results in 12000 runs of the original model and the ANN model, which was parallelized using 20 MATLAB workers (CPU threads). Advantageously, the PCE model does not require any new model runs after its creation to compute these sensitivity indices. Table 3 shows the computation time comparison of these simulations and the speedup gain obtained due to the use of surrogate models. Although the time required to perform a single original model run was no more than ten seconds and the power of parallel computing was used, performing a global sensitivity analysis with the original model costed several hours. Use of surrogate models instead resulted in significantly lower computation time.

Table 3: Computation time of global sensitivity analysis are compared for different models.

Model used	Computing power used	Computation time (sec)	Speedup
BSM1	20 MATLAB workers in parallel	15890	1x
ANN	20 MATLAB workers in parallel	94	169x
PCE	1 MATLAB engine in series	7	2270x

5. Conclusions

In this work, we have presented a new systematic framework, which can be used to build plant-wide surrogate models for wastewater treatment plants using advanced surrogate modelling algorithms. The list of algorithms tested included artificial neural networks, polynomial chaos expansions, Kriging, among others. For predicting effluent COD and ammonium, Kriging and neural networks type surrogate models are found to be the most effective. A training sample size 100 was found to be enough to build models with test R^2 values above 0.95. Also important is the results obtained from the use of surrogate models in the case of a global sensitivity analysis, which revealed very close results compared to the ones obtained using the original model, yet with much faster computation times.

Acknowledgements

The authors acknowledge funding from the EU Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no.675251 as well as the funding from Water Joint Programming Initiative under the PIONEER_STP project.

References

Friedman J.H., 1991, Multivariate Adaptive Regression Splines (with discussion), The Annals of Statistics, Vol. 19, No. 1

Gernaey K.V., Jeppsson U., Vanrolleghem P.A., Copp J.B (Eds.), 2014, Benchmarking of control strategies for wastewater treatment plants, IWA Scientific and Technical Report No. 23, IWA Publishing, London, UK

Jekabsons, G., 2009, Radial Basis Function (RBF) interpolation, available at http://www.cs.rtu.lv/jekabsons/regression.html

Jekabsons, G., 2016, ARESLab: Adaptive Regression Splines toolbox for Matlab/Octave, 2016, available at http://www.cs.rtu.lv/jekabsons/

Marelli, S., & Sudret, B., 2016, UQLab: The Framework for Uncertainty Quantification, *Proc. 2nd Int. Conf. Vulnerability Risk Anal. Manage.*, Liverpool, U.K., 2014, pp. 2554–2563.

Moral, H., Aksoy, A., & Gokcay, C. F., 2008, Modeling of the activated sludge process by using artificial neural networks with automated architecture screening. *Computers & Chemical Engineering*, *32*(10), 2471–2478. https://doi.org/10.1016/j.compchemeng.2008.01.008

Ráduly, B., Gernaey, K. V., Capodaglio, A. G., Mikkelsen, P. S., & Henze, M., 2007, Artificial neural networks for rapid WWTP performance evaluation: Methodology and case study. *Environmental Modelling and Software*, 22(8), 1208–1216.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008, Global Sensitivity Analysis: The Primer. Wiley.

Sin, G., Gernaey, K. V., Neumann, M. B., van Loosdrecht, M. C., & Gujer, W., 2011, Global sensitivity analysis in wastewater treatment plant model applications: prioritizing sources of uncertainty. Water research, 45(2), 639-651.

Sin, G., Gernaey, K. V., Neumann, M. B., van Loosdrecht, M. C., & Gujer, W., 2009, Uncertainty analysis in WWTP model applications: a critical discussion using an example from design. Water Research, 43(11), 2894-2906.

Sudret, B., 2008, Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering and System Safety*, *93*(7), 964–979. https://doi.org/10.1016/j.ress.2007.04.002

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50314-1

Land Availability, Utilization, and Intensification for a Solar Powered Economy

Yiru Li[a], Caleb K. Miskin[a], Rakesh Agrawal[a*]

[a]*Davidson School of Chemical Engineering, Purdue University, West Lafayette, IN, 47906, USA*
agrawalr@purdue.edu

Abstract

Solar energy, though promising as the energy source for a fossil fuel-deprived future, is a dilute resource and harvesting it requires vast tracts of land. In this study, we develop an extensive process system model for land requirement analysis in each of the 48 contiguous states of the United States for a solar powered economy to address the likely land competition. Land requirement analysis in this study takes into account several issues that are usually ignored. Efficiencies of major energy conversion steps from primary energy to end use are accounted and intermittent solar availability, actual solar farm output and land availability are all considered. In addition, we prefer local photons for local use for consideration of minimizing transmission loss and energy security. Under this preferred scenario, our land requirement analysis shows that 16 of the 48 contiguous states have insufficient available land and that the land competition for energy and food will be intense. Thus, in a solar economy, land use intensification will be required to avoid conflict between our competing land use needs.

Keywords: solar, PV, land, requirement

1. Introduction

Solar energy is our most ubiquitous and promising renewable energy resource that has the potential to meet our future energy demands due to its sustainable availability and abundance (Lewis and Nocera, 2006). Replacing fossil resources with solar power provides an opportunity to eliminate much of the world's pollution, greenhouse gas emissions, and energy security challenges. However, solar energy is a dilute source with power density several orders of magnitude lower than fossil fuels (Smil, 2015). Harnessing solar energy therefore requires vast tracts of land. A number of studies have envisioned a fossil fuel-deprived future powered by solar energy. In most of these previous studies, land availability is not a constraint to the feasibility of a solar powered future. For example, Turner (1999) has estimated that no more than 1 % of the land in the United States, if used for photovoltaics (PV), could provide all the energy needed for the entire country. While these researchers are optimistic about the total solar economy, several recent land requirement analyses have raised the concern in land availability in some densely populated countries. MacKay (2013) has suggested that transition to a total solar economy in Britain would require land area similar to the area of the entire country and that several other high-density, high-consuming countries, such as Japan and Germany, are facing the same challenge. A recent land requirement analysis by Capellán-Pérez et al. (2017) over 40 countries around the world has also pointed out that it could be physically infeasible for many densely populated countries to domestically harness renewable energy to maintain the current energy consumption.

According to these recent land requirement estimations, countries with relatively low population density, such as the US, have sufficient land for solar farms to meet all energy demand.

However, the US is large country where coastal areas are densely packed while other states have relatively less population and energy demand. We would like to see if solar photons could locally fulfil the energy demand. In this work, an extensive systems model for detailed land requirement analysis is developed for each of contiguous 48 states of the US for a full solar economy. In our system model, electricity demand in a solar powered future is first estimated according to current end use energy consumption and energy conversion efficiencies. To take into account the intermittency of solar availability, storage capacity of solar production should be included and various storage methods are deployed. We use the actual PV production data reported (NREL, 2013) to calculate the land area requirement. Furthermore, available land for PV installation is limited; only part of the miscellaneous and urban land could be directly utilized for PV. Considering all the issues stated above, we find out that currently available land is insufficient for many states in the US, which suggests that land competition will be intense in a full solar economy. Novel concepts will be needed to meet food and energy needs in a "Full Earth" maintained by primarily solar energy.

2. System model for land requirement and analysis for a full solar economy

2.1. Methodology for estimating electricity demand in a solar powered future

For the purpose of this study, PV panels are the only devices harnessing solar energy and converting it to electricity supplied to end uses. Therefore, most end use systems are to be electrified for the transition and we need to estimate the total electricity demand (E_{total}^{future}) for electrified end use systems through detailed modeling of each end-use sector (Figure 1). Currently, fossil fuels are either supplied directly to end use sectors or converted to electricity before consumed. In our model, we treat power generation as a separate sector and only direct primary energy consumption is considered for the residential, commercial, industrial and transportation sectors. The systems that might not be feasibly electrified would be powered by synthetic fuels, which are originated from biomass under our assumption. In addition, fossil fuels consumed as feedstocks are also replaced by bio-fuels. The bio-fuel production is also treated as a stand-alone sector.

For the electricity generation, current electricity generated from fossil fuels (EIA, 2016[a]) will be replaced by electricity from solar energy (Eq. (1)).

$$E_{electricity}^{future} = \text{Actural electricity generated from fossil fuels} \quad (1)$$

For other end-use sectors, the basis for estimating the electricity demand is to assure that the end-use energy consumptions are the same for current and future scenarios under different energy conversion efficiencies. The calculation paradigm is expressed by Eq. (2)

$$E_i^{future}\eta_i^{future}\eta_{transmission}^{future} = \sum_j E_{i,j}^{current}\eta_{i,j}^{current} \quad (2)$$

where the subscript *i* refers to the end-use, *j* refers to the fuel and $E_{i,j}^{current}$ refers to the current primary energy consumption from fossil fuel *j* for use *i*. The efficiencies for various end uses and fuels are listed in Table 1.

Table 1: Efficiencies for various end uses and fuels

Sector	End-use (i)	Fuel (j)	$\eta_{i,j}^{current}$	$\eta_{i,j}^{future}$
Residential	Residential	Natural gas, petroleum, coal	65 %	95 %
Commercial	Commercial	Natural gas, petroleum, coal	65 %	95 %
Transportation	Light duty vehicle	Motor gasoline, LPG	14.8 %	75 %
Transportation	Truck	Diesel	20.2 %	75 %
Industrial	Industrial heat	Natural gas	80 %	95 %
Industrial	Industrial heat	Petroleum	80 %	95 %
Industrial	In-house electricity	Natural gas	46 %	-
Industrial	In-house electricity	Petroleum	35 %	-

The primary energy consumption data for each of the end-uses can be acquired from EIA (EIA, 2016[b]) and according to the efficiencies listed in Table 1 we can estimate the electricity demand for a future solar economy. Note that the electricity transmission loss is taken to be 4.7 % and therefore a $\eta_{transmission}^{future}$ value of 95.3 % is accounted (EIA, 2016[a]).

Now what has been left out is the means of transportation that could not feasibly be electrified and the industrial feedstocks. The fractions of these fossil fuels (e.g. 25.2 % of the natural gas going to industrial sector is for feedstock) are shown in Figure 1. The conversion process from biomass to bio-fuels requires hydrogen to get rid of the oxygen in biomass and the hydrogen is assumed to come from water electrolysis. The electricity required for water electrolysis ($E_{bio-refineing}^{future}$) is added to the total solar electricity demand. Therefore, the total electricity demand is calculated by Eq. (3)

$$E_{total}^{future} = E_{electricity}^{future} + E_{bio-refineing}^{future} + \sum_{i,j} E_{i,j}^{future} \quad (3)$$

where *i, j* are listed in Table 1.

2.2. Intermittency of solar energy, PV output and land availability

One of the essential challenges of solar energy utilization is the intermittency of solar availability. In the US, the average solar availability is only 4.8 hours per day, which means that all the solar electricity demand must be produced in 1/5 time of a day (Lewis and Nocera, 2006). The electricity produced in the solar available time period needs to be stored. Currently, the battery storage efficiency varies in the range of 75 - 94 % (Gençer and Agrawal, 2016) and carbon storage cycle has an efficiency of 55 – 58 % (Al-Musleh et al., 2014). We adopt a mixed energy storage method and the overall system storage efficiency is taken to be 65 %. Considering the intermittency and the storage efficiency, the corrected total electricity for a solar economy can be calculated by Eq. (4) where $f_{solar\ available}$ refers to the fraction of solar available time in a 24-hour day, which is 20 % here.

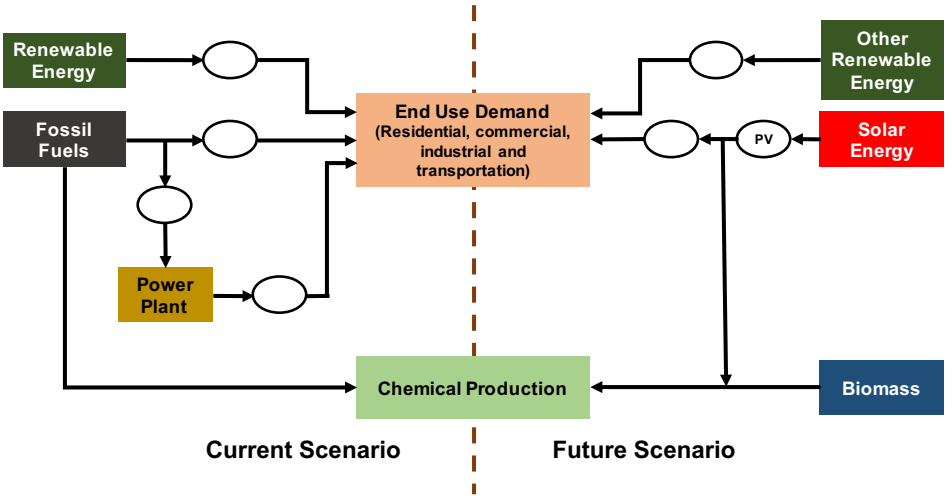


Figure 1: The energy supply infrastructure of current and future scenarios. The current end use demand will be maintained by solar energy in the future. Ovals in the figure indicate the energy conversion efficiency.

$$E_{corrected}^{future} = E_{total}^{future} f_{solar\ available} + E_{total}^{future}(1 - f_{solar\ available})/\eta_{storage} \quad (4)$$

A reasonable PV output number is crucial for our land requirement analysis. Although the maximum solar cell efficiency reported to date can be up to 46 % (NREL, 2017), the actual PV power output from PV parks is generally much lower. Instead of estimating PV power output by looking at the solar flux and PV efficiency as in most of the previous studies, we use actual production data of existing PV parks. An average power production of ~7 W/m^2 is reported for PV installations that are greater than 1 MW in size (Ong, 2013) and we take this number for our land requirement analysis. In addition, we average the total electricity demand over the total time of a year to obtain the total power output requirement (P^{future}) for a solar economy (Eq. (5)).

$$P^{future} = \frac{E_{corrected}^{future}}{Total\ time\ of\ a\ year} \quad (5)$$

Then we can calculate the land requirement for each of the contiguous states as we know the total power output requirement and power output density for PV parks ((Eq. (6)).

$$Land\ area = \frac{P^{future}}{power\ output\ density} \quad (6)$$

3. Miscellaneous and urban land are insufficient

Land area available for PV installation is only limited to certain types of land. Agriculture land (croplands, grassland pastures and ranges) cannot be used for PV installation as food production will be harmed by PV panels. Forest-use land and special-use land (national and state parks, etc.) are also unsuitable for PV for the consideration ecosystem protection. Agriculture land, forest-use land and special-use land account for 93.2 % of the total land of the US. Only part of urban-use land can directly be used for rooftop PV and other miscellaneous land (marshes, deserts and

other barren land generally of low value for agricultural purposes) is currently available for PV parks.

If we compare the land area requirement estimated in Section 2 and the area of miscellaneous and urban land for each of the contiguous 48 states in the US (Figure 2), we can see that there are 16 states that need more than 50 % of the miscellaneous and urban area. For Wyoming, West Virginia and Texas, even more than 100 % of the miscellaneous and urban land are required for PV. What we have to note is that 50 % is a rather high utilization ratio, especially for urban area. Existing studies have shown that only a small percentage of urbanized area could be covered by solar panels, generally no more than 10 %, with acceptable solar conversion efficiency (Capellán-Pérez et al., 2017). Therefore, for the highly urbanized coastal areas, such as California and New Jersey, the land challenge could be more severe. Clearly, a significantly large number of states will not able to meet their need locally and will have to import energy from other parts of the country. Alternatively, options to use other land areas such as agricultural land should be explored. This could be challenging as current practice of using photovoltaics on farmland is known to hurt food production or alter crops that can be grown (Armstrong et al., 2016).

Moreover, our energy demand estimation is based on the energy consumption data for the year of 2014. With population growth and economic development, the energy demand could be higher in the future, which would require more land dedicated for power production. In addition, population growth would require more land for human activities. Therefore, the land competition will more challenging in a full earth. To resolve this land issue for the realization of solar economy, additional land apart from urban and miscellaneous land will be locally required.

4. Conclusions

To transition to a full solar economy, an extensive systems model is developed accounting for harnessing solar energy as electricity, its storage, transportation and use. The model is then used for detailed land requirement analysis in each of the contiguous 48 states of the Unites States to examine the possible land usage for energy. Results show that for a future solar economy, 16 states are likely to require more than 50 % of the miscellaneous and urban land for solar power generation. Our results are in sharp contrast to the common understanding that only small land areas will be needed in a future solar economy to satisfy entire energy need. These results show a need for innovation in land use intensification including the possibility of using farmland while causing no reduction in food production.

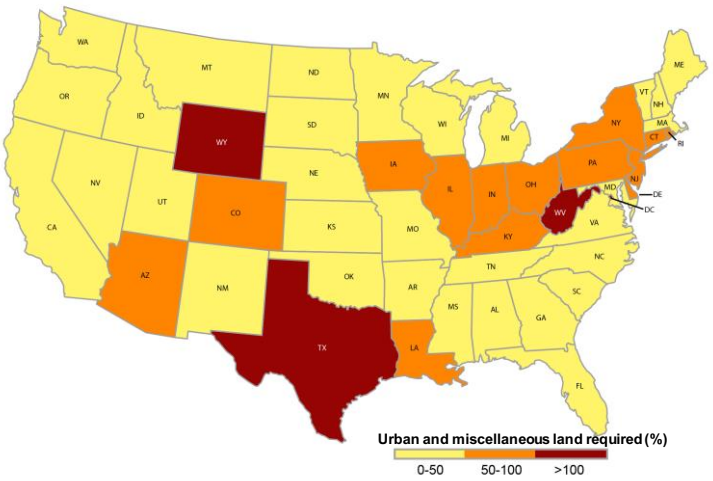


Figure 2: The percentage of miscellaneous and urban land required for PV

References

Al-Musleh, E.I., Mallapragada, D.S. and Agrawal, R., 2014. Continuous power supply from a baseload renewable power plant. Applied Energy, 122, pp.83-93.

Armstrong, A., Ostle, N.J. and Whitaker, J., 2016. Solar park microclimate and vegetation management effects on grassland carbon cycling. Environmental Research Letters, 11(7), p.074016.

Capellán-Pérez, I., de Castro, C. and Arto, I., 2017. Assessing vulnerabilities and limits in the transition to renewable energies: Land requirements under 100% solar energy scenarios. Renewable and Sustainable Energy Reviews, 77, pp.760-782.

EIA, 2016. Net Generation by State by Type of Producer by Energy Source.

EIA, 2016, State Energy Data System

EIA, 2017. State Electricity Profiles

Gençer, E. and Agrawal, R., 2016. A commentary on the US policies for efficient large scale renewable energy storage systems: Focus on carbon storage cycles. Energy Policy, 88, pp.477-484.

Lewis, N.S. and Nocera, D.G., 2006. Powering the planet: Chemical challenges in solar energy utilization. Proceedings of the National Academy of Sciences, 103(43), pp.15729-15735.

MacKay, D.J., 2013. Solar energy in the context of energy use, energy transportation and energy storage. Phil. Trans. R. Soc. A, 371(1996), p.20110431.

NREL, 2014. Best Research-Cell Efficiencies. National Renewable Energy Laboratory, CO.

Ong, S., Campbell, C., Denholm, P., Margolis, R. and Heath, G., 2013. Land-use requirements for solar power plants in the United States (No. NREL/TP-6A20-56290). National Renewable Energy Laboratory (NREL), Golden, CO..

Smil, V., 2015. Power density: a key to understanding energy sources and uses. MIT Press.

Turner, J.A., 1999. A realizable renewable energy future. Science, 285(5428), pp.687-689.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50315-3

Modeling and Design of a Solar-Assisted High-Performance Direct Contact Membrane Distillation System

Mahdi Mohammadi Ghaleni[a], Abdullah Al Balushi[a], Mona Bavarian[a], and Siamak Nejati[a*]

[a]*University of Nebraska-Lincoln, Department of Chemical and Biomolecular Engineering, 820 N 16th Street, Lincoln, NE 68588*
snejati2@unl.edu

Abstract

Hollow fiber membrane (HFM) modules have recently drawn considerable attention in membrane distillation (MD) because of their high packing density compared to other type of modules. In this work, a mathematical model is developed to study the impact of design and process parameters on both permeate flux and energy efficiency of hollow fiber membrane modules. The membrane module is integrated with a heat exchanger to utilize the heat of waste brine . Also, a solar collector is integrated in the system to increase the temperature of feed stream before entering the module. The simulation data were in good agreement with the experimental results. We found that the close-packed configuration provides the maximum permeate flux and energy efficiency for the module when the inter-spacing parameter, the ratio of the fiber radius to the center-to-center distance between neighboring fibers, is properly adjusted. The simulation results suggested that the permeate flux for this system can be enhanced as high as 18% by choosing the close-packed configuration and adjusting the inter-spacing parameter such that neighboring fibers are not in contact with each other.

Keywords: membrane distillation; hollow fiber membrane; solar energy; water desalination, modeling and simulation.

1. Introduction

Water scarcity is a severe universal problem, demanding new methods for desalination of saline resources such as seawater and brackish water [1–4]. Membrane distillation (MD) is an emerging technology for producing freshwater. In MD, a hydrophobic membrane acts as a physical barrier between a cold (distillate) and a warm stream (saline water). The membrane is typically made of a hydrophobic material allowing the water vapor to pass through but hinders the liquid flow. The driving force for vapor permeation is the difference in vapor pressure across the membrane induced by the temperature gradient.
Although MD is a promising technology, it has not been commercially developed for water desalination mainly because of the lack of high-performance materials and modules specifically designed for MD. Among the different membrane modules, hollow fiber membrane modules offer the best performance because of their high packing density, high mechanical strength, and small footprints [5, 6].
In the present work, a mathematical model is developed to simulate the membrane distillation process, operated in direct contact mode, in a hollow fiber membrane module. The effect of different packing configurations and design parameters on the module

performance were investigated. The membrane module is integrated with a shell and tube heat exchanger to utilize the heat of brine in the system. Also, a solar collector is used to provide the thermal energy needed for heating of the feed stream.

2. Mathematical Model

We built a three-dimensional (3D) mathematical model to simulate a coupled system of momentum, heat, and mass transfer in a hollow fiber membrane module using COMOSL Multiphysics software.

2.1. Membrane Distillation Model

In the DCMD process, two water streams including the saline and distillate streams pass through the lumen and in the shell side of the module. The hydrodynamic behavior of the system can be modeled by solving Navier-Stokes equations for an incompressible fluid [6,7]. Figure 1 shows the schematic of a single hollow fiber membrane in its control volume, the location of feed (in the shell) and distillate (in the lumen) streams, as well as a diagram of a laboratory-scale solar-assisted membrane distillation process.

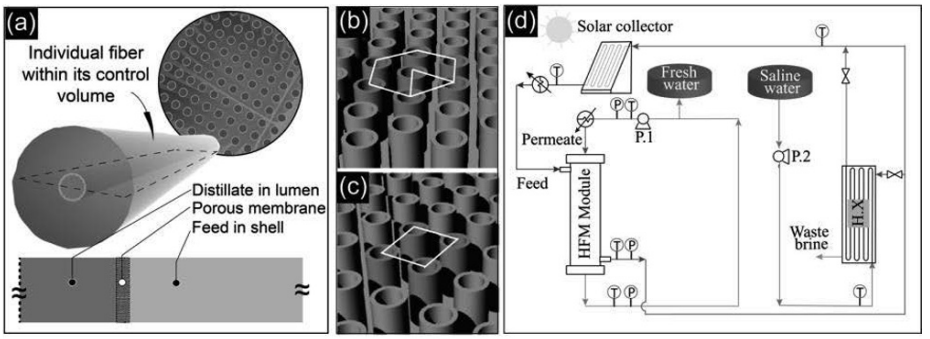


Figure1. The schematic of the hollow fiber membrane module configuration and a lab-scale membrane distillation setup. (a) an individual fiber in its control volume (the cross-section representation is not to scale); (b) hollow fiber membranes in the close-packed configuration; (c) hollow fiber membranes in the square configuration; (d) the schematic of a lab-scale solar-assisted membrane distillation process.

2.1.1. Energy Transfer

The energy balance equation for the feed and distillate streams includes both convection and conduction terms [8]. The heat is transferred between feed and distillate streams by conduction through the membrane.The energy transfer equations were written for the three subsystems of the saline stream, membrane, as well as distillate stream.

2.1.2. Mass transfer

The vapor transfer in the hollow fiber membrane can be described using two primary diffusional terms, Knudsen and molecular diffusion, whose significance can be determined using the membrane pore size and mean free path of molecules [8,9].

The mass transfer equation in the hollow fiber membrane in the radial direction can be defined as [8]:

$$J = -\frac{D_{eff}}{RT_m}\Delta p_w \quad (1)$$

where, Δp_w is the gradient of the partial pressure of water vapor across the membrane, determined using Antoine's equations [6,10].

The effective diffusion coefficient can be defined as follow [9,11]:

$$D_{eff} = \frac{\varepsilon}{\tau}\left(\frac{1-\frac{\bar{p}_{w,pore}}{P_t}}{D_{w-a}} + \frac{1}{D_{kn}}\right)^{-1} \tag{2}$$

where, ε and τ are porosity and tortuosity of the hollow-fiber membrane; D_{w-a} and D_{kn} are molecular and Knudsen diffusion coefficients, respectively [8,12]:

$$D_{w-a}\left[m^2 s^{-1}\right] = \frac{1.895\times10^{-5}\ T_m^{2.072}}{P_t} \qquad 273\ [K] < T_m < 373\ [K] \tag{3}$$

$$D_{kn} = \frac{4d_p}{3}\sqrt{\frac{R_g T_m}{2\pi M_w}} \tag{4}$$

$\bar{p}_{w,pore}$ in equation (2) is the average partial pressure of water in the pore. It can be determined using Antoine's equations at the average temperature of the membrane (T_m)– the average temperature of feed and distillate streams at the membrane interface.

2.2. Heat Exchanger and Solar Heater

A shell-tube heat exchanger can be utilized to preheat the feed entering the system using the heat of waste brine. The effective heat transfer area of the heat exchanger was assumed to be equal to 1.22 m^2. The overall heat transfer coefficient, U, was calculated to be equal to 1500 W/m^2K [13]. The feed recirculating in the system is mixed with a constant amount of preheated fresh feed (to control the salinity in the feed line) and flows through a solar collector before entering the hollow fiber membrane module. The heat transfer rate, Qc, gained by the feed in the solar collector is given by:

$$Qc = \dot{m}_f C_{p,f}\left(T_{co} - T_{ci}\right) \tag{5}$$

where, $\dot{m}_f$ and $c_{p,f}$ are the mass flow and specific heat of the feed, respectively; T_{co} and T_{ci} are the outlet and inlet temperatures of the feed from and to the collector, respectively.

The instantaneous efficiency of the solar collector, η, is defined by the following equation:

$$\eta = \frac{Qc}{A_c G_T} \tag{6}$$

where, A_c is the collector area, and G_T is the averaged solar irradiance on a tilted surface. The efficiency of solar collector is related to the optical properties of the collector, as well as the ambient conditions during the operation. The following correlation is used to find the efficiency of solar collector [14]:

$$\eta = \eta_0 - a\frac{T_{cm} - T_a}{G_T} - b\left(\frac{T_{cm} - T_a}{G_T}\right)^2 \tag{7}$$

where, η_o, a, and b are constants that can be found experimentally. T_{cm} and T_a are the mean temperature in the collector and the ambient temperature, respectively. The parameters η_o, a, and b are provided by the manufacturers of the collectors. In this study, the parameters of "Apricus ETC-30" collector that has a gross area of 4.4 m^2 ($\eta_o = 0.46$, $a = 0.802$, and $b = 0.005$) were used [14]. Accordingly, the efficiency of the solar collector was calculated to be 0.4299 for an averaged solar irradiance, G_T, of 800 W/m^2 and ambient temperature of 25°C.

3. Results and Discussions

3.1. *Effect of packing configuration*

Two different packing configurations—close-packed and square— were considered to investigate the effect of packing on the performance of hollow fiber membrane modules. Figure 2.a shows the normalized permeate flux for the two packing arrangements. The results indicate that the permeate flux for the close-packed configuration is on average 18% higher than that of square configuration.. Figure 2.b also displays the temperature difference between membrane surface and bulk flow at the feed side for both close-packed and square configurations. This temperature difference can be translated into the temperature polarization parameter (θ) — an important parameter in MD showing the temperature gradient existing between bulk flow and membrane-flow interface. The feed-side temperature polarization of the two different configurations was calculated using the following equation [6]:

$$\theta = \frac{T_{mf} - T_{md}}{T_f - T_d} \tag{8}$$

Where T_f and T_d are the bulk temperature of feed and distillate streams, respectively. Also, T_{mf} and T_{md} are the temperature of the membrane at the feed and distillate sides, respectively.

The calculations showed that the temperature polarization of both configurations remains almost constant along the length of the membrane. However, the temperature polarization issue is less for the close-packed configuration; the value of θ for close-packed and square configurations (module length=20 cm) was 0.90 and 0.85, respectively, where the value of unity is an ideal case. For this reason, the close-packed configuration was selected to further study the effect of system parameters on the permeate flux and thermal efficiency.

To determine the effect of the packing design on the performance of close-packed configuration, a dimensionless parameter (R/a), the ratio of outer diameter to the center-to-center distance of hollow fiber membranes, is defined. The R/a parameter is equal to 0.5 when the fibers are adjacent to each other. In this case, significant active surface areas in the module will be unavailable. Additionally, this condition results in a significant pressure drop and flow misdistribution due to spatial constriction. The simulations were performed for different R/a parameters. Figure 2.c shows that, for all fiber diameters tested, the increase in the R/a does not affect the permeate flux, because the process is mass transfer limited at R/a value of less than 0.45. The permeate flux drops when R/a reaches the values higher than 0.45. This observation can be attributed to the overlapping thermal boundary layers and that the increase in the transmembrane thermal conduction, pushing the process towards a heat transfer limited regime at R/a=0.5.

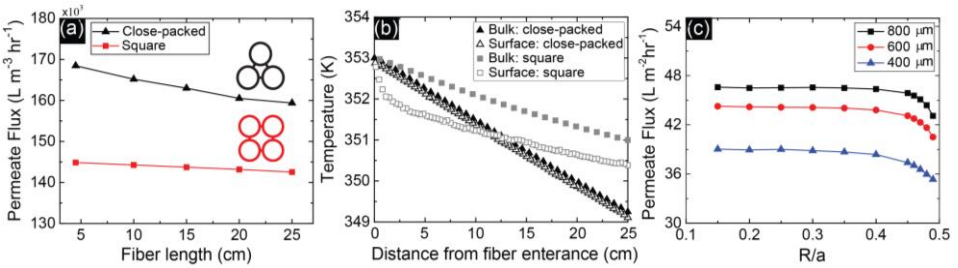


Figure 2. (a) Effect of packing configuration on the permeate flux; (b) membrane surface and bulk temperature of fiber with 25 cm length in close-packed and square configurations. (c) The effect of fiber-to-fiber spacing on the normalized permeate flux for membranes with various diameters. Here, R is the outer radius of the membrane; a is the center-to-center distance of neighboring fibers. The feed and distillate inlet temperatures are 17 °C and 80 °C, respectively. The fiber porosity and tortuosity are set at 74% and 2.5, respectively. For (a) and (b), the membrane outer and inner radii were kept constant at 400 μm and 300 μm, respectively.

4. Comparison of experimental and computational results

Feed temperature is one of the most critical parameters in the MD process because it influences water partial pressure gradient existing across the membrane. Figure 3.a shows the effect of feed temperature on the permeate flux for two different hollow fiber membranes. The results indicate that the permeate flux increases with feed temperature. The modeling results were in good agreements with the experimental data when the tortuosity parameter was set to be equal to 2.5 and 8. The effect of tortuosity and feed temperature on the permeate flux of the system is also shown in Figure. 3b. The permeate flux increases with the increase in the feed temperature and the consequent increase in the temperature gradient across the membrane.

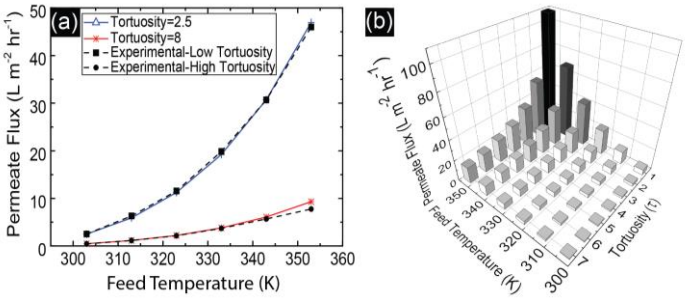


Figure 3. (a) Comparison between experimental results for the permeate flux of two hollow fiber membranes with low and high tortuosity and the prediction of the model for the membranes in a close-packed configuration[15]; (b) effect of feed temperature and tortuosity on the permeate flux.

5. Concluding Remarks

We developed a multiphysics model to investigate the impact of design and process parameters on both permeate flux and energy efficiency of hollow fiber membrane modules used in a direct contact membrane distillation (DCMD) process. The simulation

results showed that the permeate flux for this system can be enhanced as high as 18% by choosing and fine-tuning the packing configuration of the hollow fiber membranes. It was found that the close-packed configuration provides the maximum permeate flux and energy efficiency for the module, if the ratio of the fiber radius to the center-to-center distance between neighboring fibers, is properly set.

References

[1] M. Elimelech, W.A. Phillip, The Future of Seawater Desalination: Energy, Technology, and the Environment, Science (80-.). 333 (2011) 712–717. doi:DOI: 10.1126/science.1200488.

[2] C. Boo, M. Elimelech, Thermal desalination membranes: Carbon nanotubes keep up the heat, Nat Nano. 12 (2017) 501–503. http://dx.doi.org/10.1038/nnano.2017.114.

[3] S. Lin, N.Y. Yip, M. Elimelech, Direct contact membrane distillation with heat recovery: Thermodynamic insights from module scale modeling, J. Memb. Sci. 453 (2014) 498–515. doi:10.1016/j.memsci.2013.11.016.

[4] M.A. Shannon, P.W. Bohn, M. Elimelech, J.G. Georgiadis, B.J. Mariñas, A.M. Mayes, Science and technology for water purification in the coming decades, Nature. 452 (2008) 301–310. doi:10.1038/nature06599.

[5] D.M. Warsinger, J. Swaminathan, E. Guillen-Burrieza, H.A. Arafat, J.H. Lienhard V, Scaling and fouling in membrane distillation for desalination applications: A review, (2014). doi:10.1016/j.desal.2014.06.031.

[6] M.S. Khayet, T. Matsuura, Membrane distillation : principles and applications, Elsevier, 2011.

[7] M. Mohammadi-Ghaleni, S. Nejati, Modeling of hollow fiber membrane module-performance in direct contact membrane distillation, in: Sep. Div. 2016 - Core Program. Area 2016 AIChE Annu. Meet., 2016: p. 414.

[8] L. Li, K.K. Sirkar, Influence of microporous membrane properties on the desalination performance in direct contact membrane distillation, J. Memb. Sci. 513 (2016) 280–293. doi:10.1016/j.memsci.2016.04.015.

[9] J. Lee, T. Laoui, R. Karnik, Nanofluidic transport governed by the liquid/vapour interface., Nat. Nanotechnol. 9 (2014) 317–23. doi:10.1038/nnano.2014.28.

[10] C.L. Yaws, The Yaws Handbook of Vapor Pressure : Antoine coefficients, Elsevier, 2005.

[11] J. Lee, R. Karnik, Desalination of water by vapor-phase transport through hydrophobic nanopores, J. Appl. Phys. 108 (2010). doi:10.1063/1.3419751.

[12] W. Kast, C.R. Hohenthanner, Mass transfer within the gas-phase of porous media, Int. J. Heat Mass Transf. 43 (2000) 807–823. doi:10.1016/S0017-9310(99)00158-1.

[13] T.L. Bergman, F.P. Incropera, Fundamentals of heat and mass transfer., Wiley, 2011.

[14] J.A. Duffie, W.A. Beckman, Solar engineering of thermal processes, Wiley, 2013.

[15] K.Y. Wang, T.S. Chung, M. Gryta, Hydrophobic PVDF hollow fiber membranes with narrow pore size distribution and ultra-thin skin for the fresh water production through membrane distillation, Chem. Eng. Sci. 63 (2008) 2587–2594. doi:10.1016/j.ces.2008.02.020.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50316-5

Heat transfer enhancement for the coil zone of closed wet cooling towers through field synergy analysis

Xiaocui Xie, Chang He, Bingjian Zhang, Qinglin Chen*

School of Chemical Engineering and Technology / Guangdong Engineering Center for Petrochemical Energy Conservation, Sun Yat-Sen University, Guangzhou 510275, China
chqlin@mail.sysu.edu.cn

Abstract

This work proposes one kind of closed wet cooling tower (CWCT) equipped with longitudinal finned tubes, and develops a two-dimensional numerical model using FLUENT. The Navier-Stokes and *k-ε* equations are applied to calculate the vapor flow and Reynolds stress. Important operational/structural parameters such as the inlet air angle and velocity are analyzed by the multi-fields synergy theory, in which the increment in synergy degree between the velocity field, temperature field and humidity field can enhance the heat and mass transfers. The results show that the increase in fin number, inlet air velocity and decrease in fin height are conducive to the distribution uniformity of vortex and air turbulence in CWCT, which improves the field synergy degree between the velocity, temperature and humidity. With the alteration of air inlet angle, the field synergy degree is optimal at the angle of 45°, in which the heat transfer coefficient and mass transfer coefficient are about 1.15 times and 1.5 times of those of 30°, respectively. Moreover, the field synergy principle revealed that near the fin where the thermal boundary layer is thin, the α of the fin tube cases were about 1.2~1.5 times.

Keywords: closed wet cooling tower; CFD simulation; heat transfer enhancement; multi-field synergy

1. Introduction

Closed wet cooling tower (CWCT) is one type of cooling system that widely used in chemical engineering, buildings, metallurgy enterprises and so on (Xie et al, 2017). With the largely increasing consumption of energy and water in the process industry, improving the thermal performance and water-saving of CWCTs has received substantial concerns. Up to now, various studies have been conducted to investigate the heat and mass transfer process in CWCTs. These investigations mainly focused on exploring the optimal design and operating conditions (Xia et al, 2011) of CWCTs equipped with different tubes(M. Sarker et al,, 2009), such as plain tubes, finned tubes and oval tubes, et al. For example, Hasan and Sirén (2005) compared the oval tube with a plain tube in the CWCTs by experiment researches, and they concluded that the oval tubes were better than the plain tubes by 1.93~1.96 times in terms of the combined thermal-hydraulic performance index. Compared to time-consume by experimental measurements, numerical simulation by CFD has become a helpful tool in thermal hydraulic design and analysis of cooling towers. Hasan and Gan (2002) found that CFD results agree well with the analytical results when the air flow is simulated with an air

supply from the bottom of the tower. Jiang et al. (2013) further revealed that the heat and mass transfer driving forces of a cross-flow CWCT were fairly uniform by establishing a two-dimensional CFD model. The heat transfer enhancement in CWCTs mainly focuses on optimizing operating conditions and novel designed CWCTs with special tubes, e.g. oval tubes. The heat transfer enhancement of the CWCTs installed with longitudinal fin tubes is rarely reported.

Guo *et al.* (1998, 2005) proposed field synergy theory to analyze the convection heat transfer by relating the velocity and temperature gradient fields. Field synergy theory provides theoretical basis for the development of enhanced heat transfer technology. According to field synergy theory, the intensity of convective heat transfer can be changed with the alteration of synergy degree of velocity field and heat flow in the CWCT. Thus, it is important to find a relation between heat transfer enhancement and multiple fields (velocity field, humidity field, temperature filed) for a better application of field synergy principle in CWCTs.

The objective of this study is to investigate the heat and mass transfer performances of CWCTs equipped with longitudinal fin tubes using the field synergy theory. The operating conditions of CWCTs are discussed and analyzed in detail, including the air flow angle and inlet air velocity. On this basis, an optimization model is built for the process of heat and mass transfers. The influence of process parameters (velocities of airflow, humidity, and heat exchange surface characteristics) on the heat and mass transfer performances can be thoroughly analyzed.

2. Theory and model

2.1 field synergy principle

Field synergy theory (Guo et al, 1998) is used to analyse the thermal performance in CWCTs from a local point of view. For a two-dimensional steady simulation, the energy equation can be written as follows:

$$\rho c_p (u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y}) = \frac{\partial}{\partial y}(k\frac{\partial T}{\partial y}) \tag{1}$$

The heat source term of the conduction equation is expressed as follows:

$$-q = \frac{\partial}{\partial y}(k\infty\frac{\partial T}{\partial y}) \tag{2}$$

This analogy between convection and conduction heat transfer implies that tube wall heat flux can be obtained by the integration of the convective term in the thermal boundary layer. Integrating Eq. (1) over the thermal boundary layer and writing the convective part in the scalar product form, is given by:

$$\vec{u}\cdot\nabla\vec{T} = \left|\vec{u}\right|\left|\nabla\vec{T}\right|\cos\theta \tag{3}$$

where θ is the synergy angle. It is observed that one of the ways to enhance the heat transfer is to increase the scalar product of the velocity and the temperature gradient, either by reducing the synergy angle θ.

2.2 Governing equations

The species transport models are adopted to simulate the concentration of different species such as humidity in the CWCT, and the general format of the convection-diffusion equation to solve the mass fraction of different species is provided as follows:

$$\frac{\partial}{\partial t}\left(\rho Y_i\right)+\nabla\cdot\left(\rho\vec{v}Y_i\right)=-\nabla\cdot\vec{J}_i+S_i \tag{4}$$

This work employs DPM to simulate the spray water by tracking a large number of droplets, whose momentum governing equation follows the form of Newton's Second Law:

$$\frac{d\vec{u}_p}{dt}=F_D\left(\vec{u}-\vec{u}_p\right)+\frac{\vec{g}\left(\rho_p-\rho\right)}{\rho_p}+\vec{F} \tag{5}$$

2.3 Performance parameters

The performance parameters depend on the geometry and flow conditions in CWCTs. the heat and mass transfer of CWCTs can be described by the heat transfer coefficient α and the mass transfer coefficient β, respectively.

The heat transfer coefficient is expressed as:

$$\alpha=\frac{m_{\mathrm{a}}c_{\mathrm{pa}}}{A}\ln\left(\frac{t_{\mathrm{w}}-t_{\mathrm{a,in}}}{t_{\mathrm{w}}-t_{\mathrm{a.out}}}\right) \tag{6}$$

where m_{a} is the mass flow rate of the dry air, and t refers to the temperature .

The mass transfer coefficient is given by:

$$\beta=\frac{m_{\mathrm{a}}}{A}\ln\left(\frac{i_{\mathrm{w}}^{'}-i_{\mathrm{a,in}}}{i_{\mathrm{w}}^{'}-i_{\mathrm{a.out}}}\right) \tag{7}$$

where i_{a} refers to the enthalpy of the air, the subscripts w refers to the process water, and A is the heat or mass transfer area, respectively. The thermal performances of CWCTs equipped with plain and longitudinal fin tubes are investigated and compared. Specific parameters are shown in Figure 1.

3. Solving model

The FLUENT software was applied for the present numerical simulations, and the governing equations were solved by the finite volume method using the SIMPLE algorithm, and the convective terms in governing equations were discretized by the QUICK scheme with second-order precision. In addition, the convergence criterions, namely the normalized residuals are less than 10^{-4} for the momentum equations and 10^{-6} for the energy equation. The quantity and quality of the meshes are of great important to the accuracy and time of the simulations. In order to generate the most appropriate grid and capture the exact heat and mass transfer, the structured meshes were employed in the computational domain, shown in Figure 2.

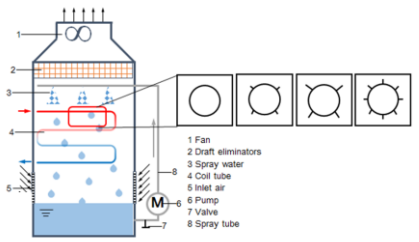


Figure 1 Schematic of the closed wet cooling tower

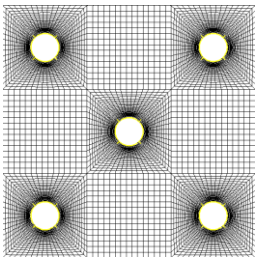


Figure 2 Mesh generation

4. Results and discussion

4.1. Influence of inlet air angle

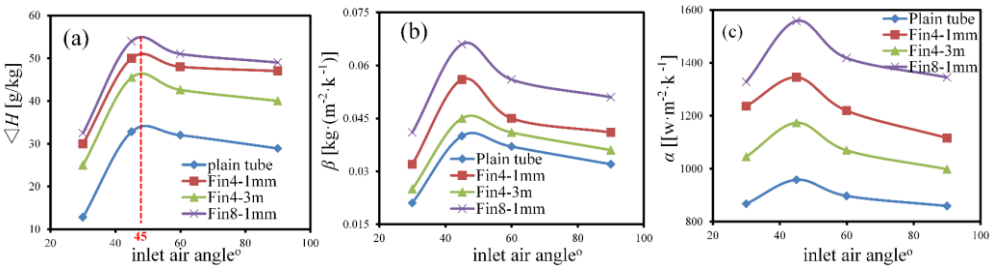


Figure 3 Influences of inlet air angle on (a) humidity difference (ΔH), (b) air-water mass transfer coefficient β, and (c) heat transfer coefficient α

Through numerical simulation, the influences of the inlet air angle on the heat and mass transfer coefficients (Eqs. 10~11) have been investigated under the operating conditions as follows: inlet air temperature 293 K, inlet air velocity 1 m/s, process water temperature 323 K. As depicted in Figure 3 (a~c), the humidity difference(ΔH), mass transfer coefficient(β) and heat transfer coefficient (α) show increasing trends with the inlet air angle ranging from 30° to 45°. However, when the inlet air angle exceeds 45 °, all these performances decrease with the increase in the inlet air angle, indicating that 45 ° is potentially a critical value with respect to the inlet air angle. Moreover, the α and β for the fin tubes are higher than that of plain tubes and these differences become more significant under different fin numbers. The the fin tube case is around 75% higher than that of the plain tube case at the the inlet air angle of 45 °. This is mainly due to the fact that the synergy angle gradually decreases with the inlet air angle ranging from 30° to 45°. This phenomenon can be further explained by the synergy effect of multiple fields in Figure 4 and 5, in which we present the coupling of temperature, humidity and velocity field at the the inlet air angle of 45 °. As shown, the temperature and humidity distribution of the fin tube between the wall surface and bulk airflow can change more rapidly in comparison with those of other tubes. The field synergy degree is optimal at the angle of 45°, in which the heat transfer coefficient and mass transfer coefficient are about 1.15 ~1.5 times than other angles.

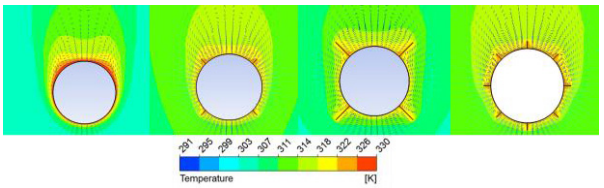


Figure 4 Coupling of temperature field and velocity vector field

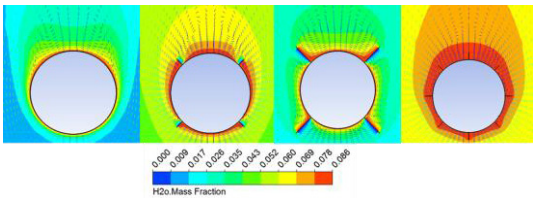


Figure 5 Coupling of humidity field and velocity vector field

4.2. Influence of inlet air velocity

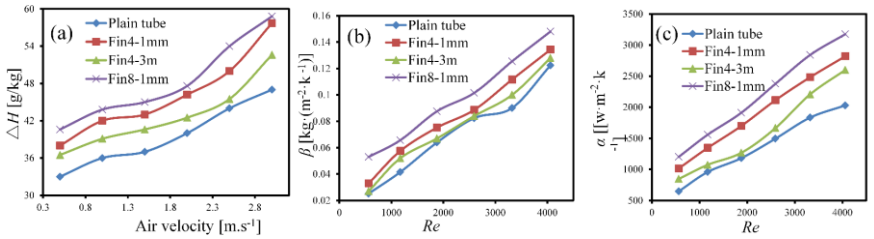


Figure 6 (a~c) shows the humidity difference($\triangle H$), mass transfer coefficient(β) and heat transfer coefficient (α) of the four tube cases with respect to a variable inlet air velocity.

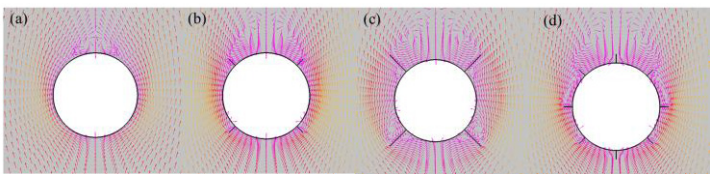


Figure 7 velocity vector field: (a) plain tube, (b) fin4-1mm, (c) fin4-3mm, (d) fin8-1mm

One of the objectives of this study is to explore the influence of inlet air velocity on the heat transfer enhancement of CWCTs. The higher air flow velocity can enhance the heat transfer rate from the spray water to air in the process of the countercurrent contacting of vapor-liquid phase. This enhancement helps to decrease the temperature of the outlet process water to the wet-bulb temperature of the inlet air. Thus, with the rise in air flow velocity, as shown in Figure 6(a~c), the α of fin and plain tube cases increases around 62% and 68%, respectively. Moreover, the α of the fin tube case is about 1.2~1.5 times that of the plain tube case, this is because the fins on the outer surface which leads to a reduction in the synergy angle compared to the plain tube, can effectively disturb and accelerate the movement of falling film. Herein, that is, the Reynolds number and heat transfer coefficient are enhanced, especially in the case of high speed airflow. This phenomenon can be further explained by the synergy effect of velocity fields shown in Figure 7, in which the air velocity of the fin tubes close to the wall surface have the

stronger turbulence intensity among these tubes cases investigated. Overall, we can conclude that the increase in inlet air velocity has a positive effect on the heat and mass transfer in CWCTs.

5. Conclusions

A numerical study had been carried out using plain tubes and fin tubes to enhance the thermal performance of the CWCT. The influences of structure and operating parameters were discussed and the main conclusions have been achieved: The effect of operational/structural parameters including the inlet air angle and velocity were systematically analyzed with the field synergy principle. The α of the fin tube cases was the highest when the air air angle at a critical value 45^o. The field synergy principle revealed that near the fin where the thermal boundary layer is thin, the α of the fin tube cases were about 1.2~1.5 times that of the plain tube case, and the increases of β of the fin tube cases were around 150%.

Acknowledgement

Financial supports from the National Natural Science Foundation of China (No. U1462113, 21606261) and the Science and Technology Planning Project of Guangzhou city (No. 201704030136) is gratefully acknowledge

References

Xie, X., Zhang,Y,et al., 2017, Bench-scale Experimental Study on the Heat Transfer Intensification of a Closed Wet Cooling Tower using Aluminum Oxide Nanofluids. Industrial & Engineering Chemistry Research, 56, 6022-6034.

Xia, Z., Chen, C., Wang R., 2011, Numerical simulation of a closed wet cooling tower with novel design, International Journal of Heat and Mass Transfer, 54, 2367-2374.

Sarker M., Shim, G. et al, 2009, Enhancement of cooling capacity in a hybrid closed circuit cooling tower, Applied Thermal Engineering, 29, 3328-3333.

Hasan A., 2005, Thermal-hydraulic performance of oval tubes in a cross-flow of air, Heat and mass transfer, 41, 724-733.

Zhu, D.S., Zheng W.Y., et al, 2013, Computational Analysis of Closed Wet Cooling Towers, Numerical Heat Transfer, Part A: Applications, 63, 396-409.

Hasan A., Gan G., 2002, Simplification of analytical models and incorporation with CFD for the performance predication of closed−wet cooling towers. International journal of energy research, 26, 1161-1174.

Jiang J.-J., Liu X.-H., Jiang Y., 2013, Experimental and numerical analysis of a cross-flow closed wet cooling tower, applied thermal engineering, 61, 678-689.

Guo Z.Y., Li, D., Wang, B., 1998, A novel concept for convective heat transfer enhancement, International Journal of Heat and Mass Transfer, 41, 2221–2225.

Guo, Z.Y., Tao,W.Q., Shah, R., 2005,The field synergy principle and its applications in enhancing single phase convective heat transfer, International Journal of Heat and Mass Transfer, 48, 1797–1807.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50317-7

Nutraceuticals Production Under a Water-Food-Energy-Waste Integration Concept

Carla V. García Prieto, Fernando D. Ramos, Vanina Estrada, M. Soledad Díaz*

Planta Piloto de Ingeniería Química (PLAPIQUI), Departament de Ingeniería Química, Universidad Nacional del Sur (UNS)-CONICET, Bahía Blanca (8000), Argentina
sdiaz@plapiqui.edu.ar

Abstract

In this work, we address the optimization of an integrated microalgae-based biorefinery through the formulation of a mixed integer nonlinear programming model (MINLP) for the production of astaxanthin and co-products (renewable fuels, bioplastics and green chemical products) reducing the dependency of external energy sources. The proposed model is implemented in GAMS to maximize net present value. Numerical results show the economic feasibility of microalgae-based nutraceuticals production in an integrated biorefinery under the water-food-energy-waste nexus concept (NPV: 336 MM$). Gross profit of the optimal integrated biorefinery is $8,434 per kg of astaxanthin, approximately 30% higher than the obtained when astaxanthin is produced as a single product, with no considerations on water-food-energy and waste nexus.

Keywords: biorefinery, microalgae, nutraceuticals, optimization.

1. Introduction

In the last years, the water-food-energy nexus concept has been promoted as a tool for achieving sustainable development accounting the relationship among these three resources (García & You, 2017). From a nexus perspective, integrated biorefineries, mainly the ones that involve second and third generation feedstocks (Martín & Grossmann, 2017), can help to solve some water-food-energy issues. For instance, developments on algae cultivations as a biorefinery feedstock have promising prospects to address these concerns (Martinez-Hernandez & Samsatli, 2017). Furthermore, many recent emerging waste management technologies lead to incorporate the management of waste streams to the nexus concept, contributing in this way to the global challenges context presented by water-food-energy nexus.

The worldwide nutraceutical market has been increased due to current population and the health trends (Sabol et al., 2017). Astaxanthin global market is estimated at 280 t in 2014 and is further projected to reach 670 t by 2020. Nutraceuticals are food or part of food playing an important role in normal physiological functions that maintain healthy human beings. Powerful antioxidants with application in nutraceuticals, pharmaceuticals, cosmetics and food industries, such as the natural ketocarotenoid astaxanthin, can be obtained under the integrated biorefinery concept.

In this sense, we address the optimization of an integrated microalgae-based biorefinery for the production of astaxanthin including the possibility of co-produce renewable fuels, bioplastics and green chemical products. We also incorporate a potential anaerobic digestion process in order to reduce waste streams and mitigate the dependency of energy sources by means of a combined heat and power cycle.

2. Process Description

The proposed superstructure, shown in Figure 1, involves the production of natural astaxanthin from the microalgae *Haematococcus pluvialis* as well as different route alternatives in order to make it sustainable, under the water-food-energy-waste concept. Main processes are described below.

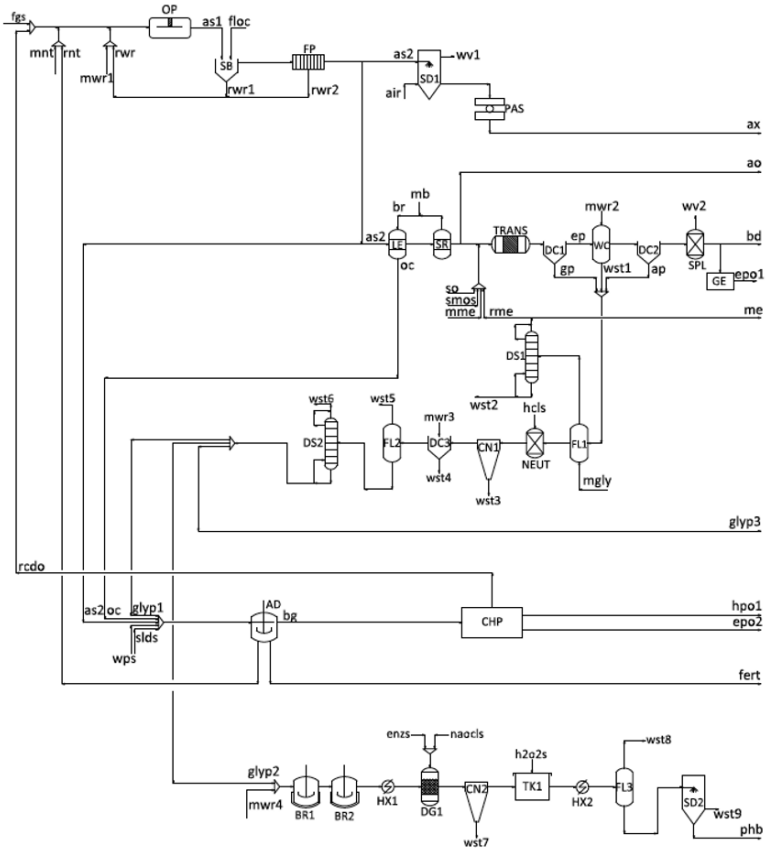


Figure 1. Integrated microalgae-based biorefinery superstructure

2.1. Astaxanthin production process

In this work the cultivation of *Haematococcus pluvialis* in open ponds (OP) is taken into account due to their capability of astaxanthin accumulation, under certain culture conditions. Carbon dioxide can be fed from neighbouring industries flue gases and from process streams, reducing GHG emissions. Nutrients, mainly phosphorous and nitrogen, are incorporated to the cultivation system from residual process streams and/or from external sources. After harvesting (SB) and dewatering (FP) the algae biomass is dried in a spray dryer (SD1) and then cracked by a bed airflow pulveriser (PAS) to obtain the final product.

2.2. Energy, water and waste management

The integrated biorefinery is proposed to be thermal energy self-sufficient and at least 20% of the electrical energy consumption is imposed to be from renewable sources,

aligned with state policies in Argentina. In order to reduce the dependency of external energy sources, biodiesel and its subsequent transformation to electric energy (GE) is proposed from microalgae and/or soybean oil. A combined heat and power cycle (CHP) to transform biogas generated by the anaerobic digestion of substrates is considered in the superstructure in order to satisfy energy constraints. Potential substrates include microalgae biomass, residual microalgae biomass from lipid extraction (oil cake), sludge from waste water treatment plant and waste paper. Water from residual streams in harvesting and dewatering steps as well as a nutrient-rich stream from the anaerobic digestor (AD), are fed back into the microalgae cultivation system, thus reducing external nutrients and water requirements. The potential inclusion of an anaerobic digestor can help into the reduction of waste streams not only from the process, but also from external sources (sludge and waste paper).

2.3. Transportation fuel and packaging

The proposed model considers bulk and packed product transportation from the facility location to a distribution center. Biodiesel can be used as transportation fuel and the requirement amount is considered as a lower bound on biodiesel production. Biodiesel is produced from soybean and/or microalgae oil transesterification (TRANS), with glycerol as a by-product. Glycerol is mostly used for food, cosmetics and pharmaceutical products, but can also be used as raw material for higher added value products as poly(hydroxybutyrate) (PHB). PHB is a biopolymer produced by microorganisms, with similar properties to fossil fuel based polymers. In this work, PHB production must be, at least, the required for main product packaging. Crude glycerol is also included as potential raw material.

2.4. Product portfolio

Under the concept of integrated biorefinery, it is well known that revenues can be increased by producing and selling one or more value-added products that enlarge the existing product portfolio. The development of product portfolio leads to several benefits, such as adjustment of supply to the market, using process flexibility. Potential co-products include: algae oil, biodiesel, purified glycerol, methanol, fertilizers from the anaerobic digestion process and PHB.

3. Mathematical Modelling

The proposed superstructure is formulated as an MINLP problem and implemented in GAMS (Brooke, 2013). The model includes mass balance constraints, energy and correlation constraints, detailed equipment design and cost capital correlations, integer constraints and economic objective function.

3.1. Mass balance constraints

Mass balances are formulated for each unit in the superstructure shown in Figure 1. For reactive units (θ') mass balances are formulated as follows:

$$f_{r,j}^{\theta'} = \sum_{k \in K} f_{\theta',j}^{k} + \sum_{h \in H} \xi_{j,s_h}^{h} \cdot C_h \cdot \sum_{k \in K} f_{\theta',s_h}^{k} \quad \forall\, j \in J \tag{1}$$

where $f_{r,j}^{\theta'}$ is the mass flowrate of component j from reactive unit θ' to outlet stream r, in kg j/day. $f_{\theta',j}^{k}$ is the mass flowrate of component j from inlet stream k to reactive unit θ',

in kg j/day. ξ_{j,s_h}^{h} is the mass coefficient between components j and s_h for the reaction h, in kg$j$/kg s_h. s_h is the limiting reactant for reaction h. C_h is the limiting reactant conversion for reaction h. f_{θ',s_h}^{k} is the mass flowrate of component s_h from inlet stream k to reactive unit θ', in kg s_h/ day.

3.2. Energy balance constraints

The energy consumed by each process unit is represented by Eq. (2), as follows:

$$EC_{\vartheta} = ECR_{\vartheta} \cdot m_{\vartheta} \tag{2}$$

where EC_{ϑ} is the energy consumption of unit ϑ (it includes θ and θ' units), in kWh/day. ECR_{ϑ} is the energy consumption ratio of unit ϑ, in kWh/kg. m_{ϑ} is the mass flowrate relative to unit ϑ, in kg/day.

3.3. Cost capital constraints and equipment design correlations and

Capital cost for equipment related to the anaerobic digestion section, microalgae cultivation step, combined heat and power cycle and astaxanthin production process has been calculated using the six-tenths rule. Cost equations involved in the remaining section of the integrated biorefinery and detailed equipment design correlations for each unit have been formulated following the methodology presented by García Prieto et al. (2017).

3.4. Integer constraints

Potential units in the proposed superstructure are associated to binary variables y_i. Big M constraints are formulated in order to establish the existence (y_i=1) or non-existence (y_i=0) of the unit.

3.5. Inequality constraints

Upper bounds of the main variables involved in the process are shown in Table 1.

Table 1. Major model constraints

Variable		Upper bound
CN_{op}	Maximum Carbon: Nitrogen ratio	25
$-CN_{op}$	Minimum Carbon: Nitrogen ratio	-20
wsp_{AD}	Available waste paper	38,500 t/y
sld_{AD}	Available sludge	12,700 t/y
NER_{EE}	Minimum electric energy self-supplied	5
NER_{HE}	Thermal energy self-sufficiency	1
PHB_{prod}	Industrial scale of a PHB plant	10,000 t/y
$-PHB_{prod}$	Minimum requirement for packaging	-2,520 t/y
BD_{prod}	Industrial scale of a biodiesel biorefinery	47,500 t/y
$-BD_{prod}$	Minimum requirement for transport	-198.25 t/y

Particularly, net energy ratio (NER) constraint is given for ensure a minimum of self-supplied energy. For electric and heat NER calculations (NER_{EE} and NER_{HE}), the following equations are implemented in the model.

$$NER_{EE} = \frac{EE_{C_{bd}} + EE_{C_{ast}} + EE_{C_{pha}} + EE_{C_{AD}}}{EE_{P_{bg}} + EE_{P_{bd}}} \quad (3)$$

$$NER_{HE} = \frac{HE_{C_{bd}} + HE_{C_{ast}} + HE_{C_{pha}} + HE_{C_{AD}}}{HE_{P_{bg}}} \quad (4)$$

where, $EE_{C_{bd}}$, $EE_{C_{ast}}$, $EE_{C_{pha}}$, $EE_{C_{AD}}$, $HE_{C_{bd}}$, $HE_{C_{ast}}$, $HE_{C_{pha}}$, $HE_{C_{AD}}$, are the electric and heat energy consumed in biodiesel, astaxanthin, PHA and anaerobic digestion process, respectively. $EE_{P_{bg}}$ and $HE_{P_{bg}}$ are the electric and heat energy obtained by biogas generation and $EE_{P_{bd}}$ is the electric energy given by biodiesel production.

3.6. Objective function

Net present value (NPV) is the objective function given by Eq. (5). It includes total capital investment (I), which involves fixed capital and working capital (10% of fixed capital). Furthermore, the objective function considers revenues from sales (Rev), manufacturing cost (C_{manuf}), raw materials cost (C_{rawmat}) and utilities cost (C_{utilit}), in \$/y. In order to take into account time value of money, each cash flow is discounted back to its present value (a^{-1}) considering a project lifetime of 15 years and a discount rate of 10% (Ramos et al., 2017).

$$NPV = -I + a^{-1}.\left(Rev - C_{manuf} - C_{rawmat} - C_{utilit}\right) \quad (5)$$

4. Numerical Results

The MINLP model has 601 continuous variables and 7 binary variables, with 961 constraints. It was solved with DICOPT (CONOPT and CPLEX) in 65 s. The optimal configuration and the numerical results for 12 t/y of astaxanthin production, corresponding to NPV maximization, is shown in Figure 2. The NPV and gross profit of the integrated biorefinery is 336 MM\$ and 8,434 \$/kg astaxanthin, respectively. $\boldsymbol{NER_{EE}}$ and $\boldsymbol{NER_{HE}}$ are 2.134 and 1, respectively, for the optimal configuration. If astaxanthin is produce as a sole product and no renewable energy and materials are considered, NPV for the production of the same amount of astaxanthin is 377 MM\$ and a gross profit of 6,419 \$/kg astaxanthin.

5. Conclusions

In this work, a detailed MINLP model for an integrated microalgae-based biorefinery is formulated to determine the optimal process configuration, based on NPV maximization. Optimization results show the economic feasibility of astaxanthin production in an integrated biorefinery under the water-food-energy-waste nexus concept, with a gross profit of 8,434 \$/$kg_{astaxanthin}$, 30 % higher than the obtained when astaxanthin is produced as a single product. The nutraceuticals and co-products sales (renewable fuels, bioplastics and green chemical products) allow obtaining a NPV of nearly \$ 336 MM. Even that this NPV value is lower than the one where astaxanthin is a sole product (377 MM\$), this could be reverted with a higher project lifetime. Our future work is to develop a bicriteria

optimization problem, including not only the economic benefits but also de environmental impact.

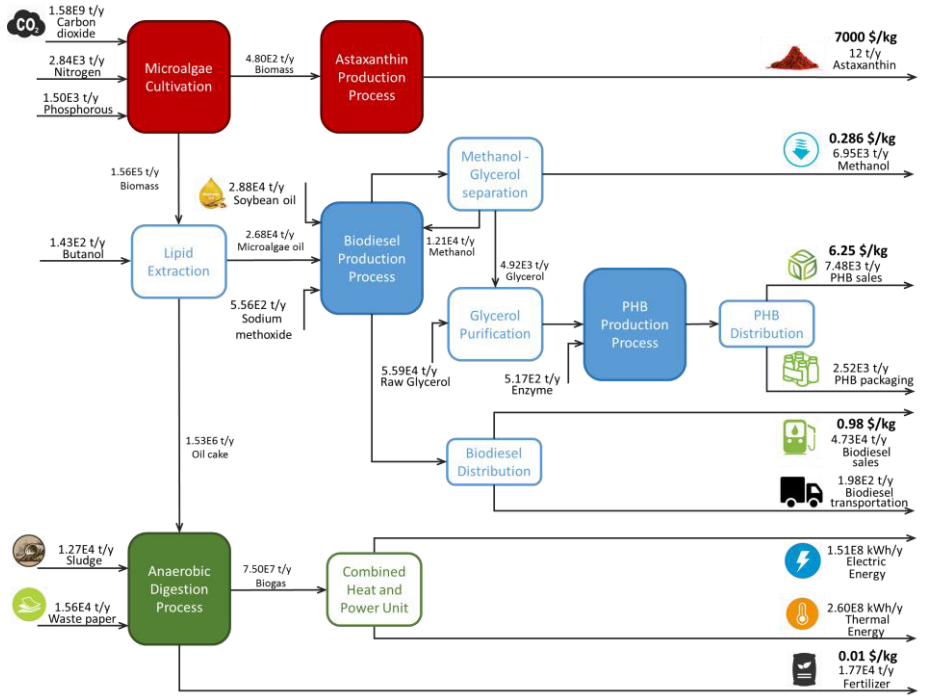


Figure 2. Optimal configuration and numerical results

References

A. Brooke, D. Kendrick, A. Meeraus, R. Raman, 2013, GAMS-A User's Guide, GAMS Development Corporation: Washington.

D. García, F. You, 2017, Systems engineering opportunities for agricultural and organic waste management in the food-water-energy nexus, Current Opinion in Chemical Engineering, 18, 23-31.

C. V. García Prieto, F. D. Ramos, V. Estrada, M. A. Villar, S. Diaz. 2017, Optimizationof an integrated algae-based biorefinery for the production of biodiesel, astaxanthin and PHB, Energy, 139, 1159-1172.

E. Martinez-Hernandez, S. Samsatli, 2017, Biorefineries and the food, energy, water nexus – towards a whole systems approach to design and planning, Current Opinion in Chemical Engineering, 18, 16-22.

M. Martín, I. E. Grossmann, 2017, Optimal integration of a self sustained algae based facility with solar and/or wind energy, Journal of Cleaner Production, 145, 336-347.

F. D. Ramos, M. A. Villar, M. S. Diaz, 2017, Optimal Design of Poly(3-hydroxybutyrate) Production using alternative Carbon Sources, Computer Aided Chemical Engineering, 40, 877-882.

A. M. Sabol, M.-O. Bertran, J. P. Raftery, J. M. Woodley, R. Gani, M. N. Karim, 2017, Separation and recovery of intracellular beta-carotene using a process synthesis framework, Computer Aided Chemical Engineering, 40, 2851-2856.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50318-9

Land use modeling and optimization based on food-energy-water nexus: a case study on crop-livestock systems

Yaling Nie[a,b,c,d], Styliani Avraamidou[c,d], Jie Li[e], Xin Xiao[a,b*], Efstratios N. Pistikopoulos[c,d*]

[a]*Division of Environment Technology and Engineering, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, China*
[b]*University of Chinese Academy of Sciences, Beijing 100049, China*
[c] *Department of Chemical Engineering, Texas A&M University, College Station, TX 77843 USA*
[d] *Texas A&M Energy Institute, Texas A&M University, College Station, TX 77843 USA*
[e] *School of Chemical Engineering and Analytical Science, The University of Manchester, Manchester, M13 9PL UK*
xxiao@ipe.ac.cn, stratos@tamu.edu

Abstract

Efficient land use in agricultural systems is a complicated decision-making problem with resource competitions and conflicting objectives. Systematic thinking based on food-energy-water (FEW) nexus is a necessity for modeling and optimization of the systems. However, challenges arise in making decisions while encountering conflicting objectives, limited data and coupling components. To address these challenges, we developed a global optimization-based land allocation framework, which provides an adaptive data-driven modeling method based on limited realistic data to predict yields for production components, a FEW index to help solve the multi-objective optimization problem and carry out assessments. Computational results indicate that the framework can provide valuable production models and a comprehensive FEW index to select strategies for optimal land allocation and limit stresses in the FEW nexus.

Keywords: land use, nexus, data-driven modeling, global optimization

1. Introduction

Agricultural land use optimization is a complicated process as it consists of allocating not only various species to different areas but also activities and resources based on different strategies of land allocation (Bergstrom et al., 2013). It often involves policy-makers, stakeholders, spatial factors, constraints, dynamic conditions and multiple components, which need systematic thinking for modeling and optimization. To achieve a better understanding of the relationships among nexus elements and make optimal decisions, all these challenges have to be addressed. (Miralles-Wilhelm et al., 2016).

Complications arise form conflicting interests over different stakeholders, such as profit, food demand, environmental penalty, and use efficiency of resources. Data limitation is another complicating factor. Many studies focus on modelling the production components as complicated model sets that take into consideration many factors such as climate change and biological properties (Memmah et al., 2015). However, modeling

with increasing considerations need more data and parameters, which is often unrealistic for application purposes or optimization (Jones et al., 2017). Since agricultural systems with multiple components need a model family to represent production processes, the FEW nexus interactions among them also need to be quantified to facilitate nexus wide decision making.

These challenges call for systematic modelling and optimization methods that can make nexus wide decisions for the systems. In this work, we develop a global optimization-based land allocation framework, which includes an adaptive modelling method based on limited realistic data for predicting production components, and a series of FEW index for achieving multi-objective solutions and quantitative assessments.

2. The Modeling and Optimization Framework

Figure 1 briefly illustrates the framework for a land use allocation problem in agricultural systems. The tasks of the problem are to decrease the stresses between land resources (food-energy-water) and maximize the benefits based on land conditions and balance the interests from different stakeholders. The resulting multi-objective optimization problem can be solved by a three-step method.

Step 1: Data collection and pre-processing. Data in the system includes input-output data from production units, economic data from social surveys, and other conditions, such as climate and soil data. The data pre-processing includes outlier detection, data normalization, data grouping based on FEW flow for production units.

Step 2: Adaptive Data-driven modelling. As the input-output data have been processed, the most important step is to simulate the output variables response to changes in the input variables for production components. These models are constructed by using a data-driven modelling method. Take crop models as examples, a number of simulated experiments are defined based on local conditions (soils, climate, irrigation, crop properties and sowing dates). By parametrically modifying the amount of input variables X, such as water, fertilizer and seeds applied to a crop in each simulated experiment, simulators can estimate its yield response $F^{sim}(X)$. The resulting input-output data points $s \in S^{modeling}$ can be fitted to yield prediction models M.

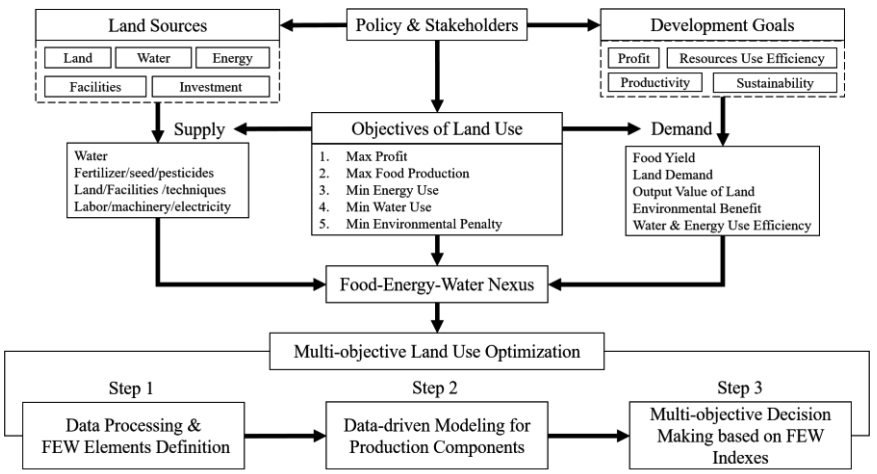


Figure 1. Framework of land use optimization.

For model M with certain size K, parameters α_k are estimated by minimizing the sum of absolute model error $e_n(X_n)$ over training data points $n \in \mathrm{S}^{\mathrm{train}}$ (Cozad et al., 2014). The optimization problem formulations are as follows (Eq. (1)):

$$
\begin{aligned}
&min \sum_{n \in \mathrm{S}^{\mathrm{train}}} |e_n(X_n)| \\
&\text{s.t.} \quad F^{sim}(X) = \sum_{k=1}^{K} \alpha_k f^k(X), k = 1, 2, \ldots, K \\
&F_n^{obs}(\mathrm{X_n}) = F_n^{sim}(X_n) + \epsilon_n(X_n), n = 1, 2, \ldots, N_1 \\
&\mathrm{ER}(X_m) = \frac{1}{N_2} \sum_{m \in \mathrm{S}^{test}} |F_m^{sim}(X_m) - F_m^{obs}(X_m)| / F_m^{obs}(X_m), m = 1, 2, \ldots, N_2 \\
&e_n(X_n) \geq \epsilon_n(X_n), \ \mathrm{e_n(X_n)} \geq -\epsilon_n(\mathrm{X_n}) \\
&\sum_k y_k = \mathrm{K}, \ \mathrm{y_k} \in \{0,1\}, -\mathrm{M}y_k \leq \alpha_k \leq M\mathrm{y}_k
\end{aligned}
\tag{1}
$$

The surrogate function $F^{sim}(X)$ is a linear combination of K nonlinear basis functions. $F^{obs}(X)$ is the observed output data from simulating experiments. By solving this MIP problem with increasing size K until the test error ratio $\mathrm{ER}(X_m)$ meet the tolerance, candidate size K of model M can be obtained. This modelling procedure is repeated for I times by using 10-fold cross validation methods. Candidate K with maximal frequencies is selected as the optimal size of model M, then the parameter estimation of model M with size K can be solved by using the whole modeling data $S^{modeling}$.

Though accurate surrogate model M with small size K can be achieved by data-driven modelling methods, available realistic data for modelling are sparse. As the optimal surrogate model $\mathrm{M^K}$ based on simulating experiments is generated, an adaptive strategy based on limited realistic data is designed to decrease the true error of surrogate models. Key ideas of the strategy include (a) calculation of the similarity scores $Sim(s, \mathrm{S^{real}})$ between real data $\mathrm{S^{real}}$ and modelling data points $s \in S^{modeling}$, modeling data points with lowest similarity scores are defined as bad samples $\mathrm{S^{bad}}$; (b) renewal of the modeling data $S^{modeling}$ by adding real data points $\mathrm{S^{real}}$ and deleting bad samples $\mathrm{S^{bad}}$, then the surrogate model M is rebuilt by previous modelling methods. An independent real data set $\mathrm{S^{validate}}$ is used to estimate true model errors during the adaptive modelling.

Step 3: Multi-objective Optimization based on integrated FEW index. Since a family of unit models are developed to predict yields by proposed modeling methods, connections among them are quantified by using FEW flows. The land use problem is defined as a multi-objective MINLP problem by considering maximum profit (P), maximum food yield (F), minimum energy use (E), minimum water use (W) and minimum environmental penalty (En). These objectives are calculated based on resource use, cost and yield output from different pieces of land. $\mathrm{P_{sc}}$, $\mathrm{F_{sc}}$, $\mathrm{E_{sc}}$, $\mathrm{W_{sc}}$ and $\mathrm{En_{sc}}$ are indexes for these objectives, respectively. To comprehensively assess multiple solutions and achieve a balanced final solution, a FEW-Nexus index GA is developed to solve the multi-objective optimization problem by maximizing its value (Eq. (2)).

$$d = \frac{d-d_{\min}}{d_{\max}-d_{\min}}, d = \mathrm{P_{sc}},\ \mathrm{F_{sc}}\ ;\ d' = 1 - \frac{d'-d'_{\min}}{d'_{\max}-d'_{\min}},\ d' = \mathrm{E_{sc}}, \mathrm{W_{sc}},\ \mathrm{En_{sc}}$$
$$\mathrm{GA} = \frac{1}{2}(\mathrm{P_{sc}F_{sc}} + \mathrm{F_{sc}E_{sc}} + \mathrm{E_{sc}W_{sc}} + \mathrm{W_{sc}E_{sc}} + \mathrm{E_{sc}P_{sc}})\sin(72°) \quad (2)$$

3. Case Study: Land Allocation for A Crop-livestock System

The methodology for the framework is illustrated through a case study for a crop-livestock system in Yucheng Experimental Station belonging to Chinese Academy of Sciences, China. Figure 2 presents the superstructure for the system. The design supplies two-by-two land pieces for crops and livestock production units, respectively. The production units are two crops (wheat, corn) and two livestock (two kinds of layer chicken). It also includes multiple decision options for water supply, energy use, food production, and treatments of waste and byproducts. Previous modelling method is used to build the prediction models for crops, CPLEX solver is used to solve the MIP problem. Empirical models are used as prediction models for livestock. The resulting land use problem of this system is a MINLP problem, which is solved using solver ANTIGONE.

4. Results and Discussion

4.1. Component models and good-of-fit performance

Models for production components in the system are reported by Table 1. The good-of-fit performance for crop models by using proposed methods is illustrated in Figure 3. Strategy 1 is the common method by adding new realistic samples without adjusting bad samples, strategy 2 is the proposed strategy by considering bad samples. The results show strategy 2 has better and more robust performance especially when increasing adaptive times of realistic samples. The yield response functions of livestock include production ratio functions $Y_e(t)$ and growth curve models $Y_w(t)$, which are generated based on empirical functions (Savegnago et al., 2012).

Table 1. Models for production components in the crop-livestock system.

Component	Production Models
Crop 1	$\mathrm{F(x)} = -12150.4\log(0.001x_1) + 1197.3(1 - \frac{1}{0.01x_2+1}) - 2.03 \cdot 10^{-4}x_2x_3 + 13767.2\sqrt{0.001x_1}$
Crop 2	$\mathrm{F(x)} = -73497.3(0.01x_1)^2 + 0.01036x_2x_3 + 10235\sqrt{0.001x_1}$
Livestock 1	$Y_e(t) = 0.945(1 - e^{-0.968t+3.796})e^{-0.006t}$ $Y_w(t) = 1.771e^{-3.158e^{-0.132t}}$
Livestock 2	$Y_e(t) = 0.933(1 - e^{-0.469t+0.0005})e^{-0.017t}$ $Y_w(t) = 1.964e^{-3.777e^{-0.15t}}$

4.2. Optimal solutions for multiple objectives

The results of optimization for the five objectives and the integrated objective based on index GA are reported in Table 2. Optimal solutions for different objectives provide multiple decisions for land use, food yield, energy and water use, and choices of technique treatments in the system. The radar maps of proposed FEW indexes in Figure 4 show that, the solutions by maximizing index GA can give more balanced designs for decision making rather than the other five solutions based on single objectives.

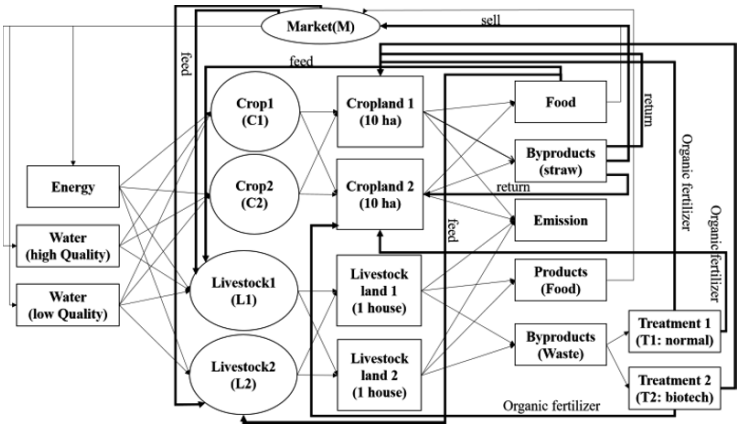


Figure 2. A superstructure design for the crop-livestock land use system.

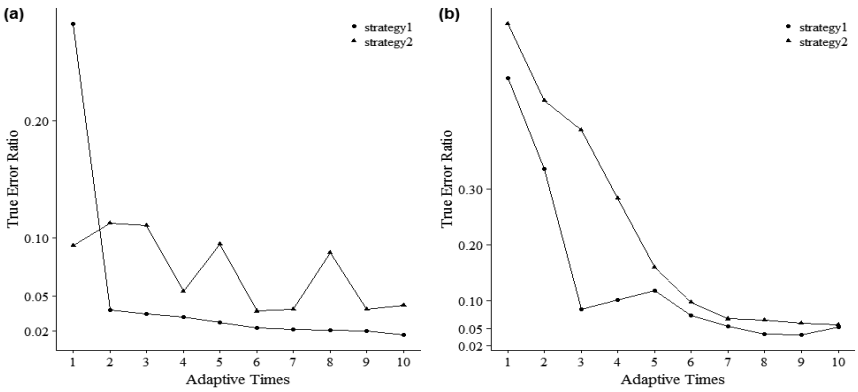


Figure 3. Good-of-fit performance of adaptive modelling methods.

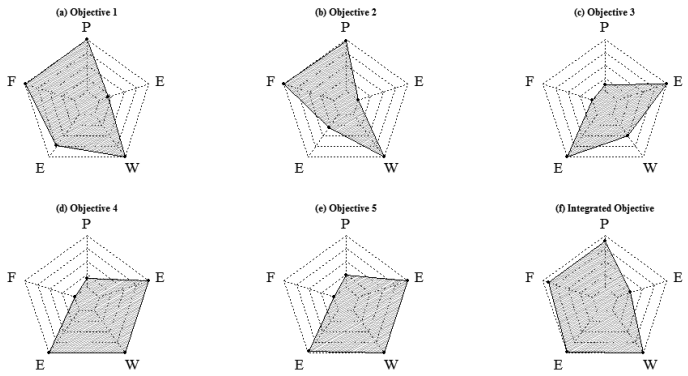


Figure 4. Evaluation of solutions for multi-objective problems based on FEW indexes.

Table 2. Optimal designs of land and resources for multiple objectives.

Optimal design	Max P	Max F	Min E	Min W	Min En	Max GA
Profit (yuan)	871,567	823,715	-220,218	-145,218	-83,423	743,005
Food (MJ)	501.27	502.44	10.9	10.9	10.92	448.96
Energy (MJ)	3,166,367	8,441,562	19,9387	19,942.2	474,840	242,484.9
Water (t)	220.5	220.5	50,014	14	14	196
En (yuan)	141,232	171,456	3,112.91	3,112.91	3,110.76	125,995
Land use (C1)	1	2	0	0	1	0
Land use (C2)	1	0	2	2	0	2
Land use (L1)	2	2	2	2	2	2
Land use (L2)	0	0	0	0	0	0
Straw return	YES	YES	NO	NO	NO	NO
Waste disposal	T1	T2	T1	T1	T1	T1(50%)
Crop water use	high	high	high	high	low	low(50%)
Feed water use	low	high	high	high	high	low
Feed source	C(22.9%)	M(100%)	C(100%)	C(100%)	M(100%)	C(100%)
Layer time (w)	63	63	4	4	4	56

5. Conclusions

Agricultural land use decision-making problems call for a systematic framework considering conflicted objectives, limited data and coupling components. This work takes a crop-livestock system as an instance, and illustrates a global optimization-based land allocation framework, including methods to achieve multi-objective land use solutions based on a FEW index, and develops an adaptive data-driven modelling method to predict yields for production components based on limited realistic data.

6. Acknowledgement

The authors gratefully acknowledge financial support from STS project of Chinese Academy of Science [KFJ-EW-STS-054-3] and China Scholarship Council.

References

Bergstrom, J.C., Goetz, S.J. and Shortle, J.S., 2013. Land use problems and conflicts: Causes, consequences and solutions. Routledge.

Cozad, A., Sahinidis, N.V. and Miller, D.C., 2014. Learning surrogate models for simulation‐based optimization. AIChE Journal, 60(6), pp.2211-2227.

Jones, J.W., Antle, J.M., Basso, B., Boote, K.J., Conant, R.T., Foster, I., Godfray, H.C.J., Herrero, M., Howitt, R.E., Janssen, S. and Keating, B.A., 2016. Brief history of agricultural systems modeling. Agricultural Systems.

Miralles-Wilhelm, F., 2016. Development and application of integrative modeling tools in support of food-energy-water nexus planning—a research agenda. Journal of Environmental Studies and Sciences, 6(1), pp.3-10.

Memmah, M.M., Lescourret, F., Yao, X. and Lavigne, C., 2015. Metaheuristics for agricultural land use optimization. A review. Agronomy for sustainable development, 35(3), pp.975-998.

Savegnago, R.P., Cruz, V.A.R., Ramos, S.B., Caetano, S.L., Schmidt, G.S., Ledur, M.C., El Faro, L. and Munari, D.P., 2012. Egg production curve fitting using nonlinear models for selected and nonselected lines of White Leghorn hens. Poultry science, 91(11), pp.2977-2987.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50319-0

Optimal Design of Transport and Reaction Pattern in Premixed Methane-air Micro-combustor

Xuepu Cao[a], Shengkun Jia[a], Yiqing Luo[a], Xigang Yuan[a,b*], Kuo-Tsong Yu[a,b]

[a] *School of Chemical Engineering and Technology, Tianjin University, 300072, Tianjin, China*
b State Key Laboratory of Chemical Engineering, Tianjin University, 300072, Tianjin, China
yuanxg@tju.edu.cn

Abstract

The importance of developing micro-combustor with high efficient is apparent, due to the demand for miniature power source with high energy density. A two-dimension model of micro-combustor for a premixed methane/air mixtures combustion process, under the condition of stable burning, was constructed. The availability and the entropy generation of the combustion were analyzed. The minimum of entropy generation was set as the objective function, subject to a constraint of constant viscous dissipation. Through variation calculation, an optimal velocity field and associate body force field were obtained. Such a solution provided a thermodynamic limit to the enhancement of the combustion process. Inspired by the flow field obtained by the optimization, an improved internal structure of the micro-combustor that gave a higher efficiency was suggested and analyzed. The results obtained can provide a guideline for the design of micro-combustor with high burning efficiency.

Keywords: Micro-combustors; combustion efficiency; entropy generation minimization; optimal design

1. Introduction

Micro-combustors first suggested by Epstein and Senturia (1997) are kinds of miniature reactors, which have attracted considerable attentions. This is due to the increasing demand for portable power source of high energy density, such as miniaturized heat engines, gas turbines, air vehicle, thruster, thermo-photovoltaic etc. The importance of developing micro-combustors with higher efficiency is apparent facing to the growing concern of energy, economy and environment. The efficiency of the combustion process depends on the fluid flow pattern that is coupled with the heat and mass transfers and the reaction inside a micro-combustion chamber. Hüseyin Yapıcı et al. (2005) investigated the entropy generation rates of a methane-air burner for different equivalence ratios and swirl numbers. M. Bidi et al. (2010) optimized the burner by entropy generation minimization method, and improved the efficiency by changing the flame location. Uttam Rana et al. (2014) identified the major sources of exergy destruction and calculated the second law efficiency in different conditions. H. R. Arjmandi et al. (2015) investigated the effects of different parameters on the entropy generation and proposed the optimal value of each parameter based on EGM method. Most of these previous works on increasing the combustion efficiency, however, rely on the analysis-improvement scheme, in which the improvement of the structural or

operational parameters is based on the feedback from the performance tests by means of either experiment or numerical simulation.

For any operating condition, there should be an optimal fluid flow pattern that gives the highest combustion efficiency in terms of the reaction and heat or mass transfer inside a micro-combustion chamber. In the present work, basic ideas of Process Systems Engineering were used to develop a methodology for the optimal design of the flow pattern to give maximum combustion efficiency in a micro-combustion chamber. The combustion process was regarded as a system, and its behaviours can be manipulated by free variables so that a criterion that has relevance to the combustion efficiency can be minimized or maximized. In our work we took the entropy generation rate, which reflects the irreversibility of the combustion process as the objective function to be minimized and the mass and heat conservation equations as the constraints. We regard entropy generations due to burning reaction and heat transfer as the major sources of the irreversibility in the combustion system, and so the sum of these two entropy generations was taken as the criterion. The optimization problem was solved through variation calculation for a velocity field and associate body force field that give the combustion process the maximum thermodynamic efficiency. The proposed method was illustrated by an example, and optimal structure designs approaching to the thermodynamic limit of the highest efficient micro-combustor were discussed.

2. Optimization method

2.1. Numerical model

Figure 1: Schematic diagram of the micro-combustor

The schematic diagram of the problem is shown in figure.1. The width of the micro-combustor model was selected 2 mm, the length was 4 mm. For combustion stability (Norton and Vlachos, 2003, Norton and Vlachos, 2004), the boundary conditions are set as follows: the inlet velocity is 0.3 m/s; the temperature of the mixture at the inlet is 500 K, and the equivalence ratio, φ, is 0.9. The wall of the combustor is assumed adiabatic.

The governing equations include the continuity, momentum, energy conservation and mass conservation equations for a steady state gas flow.

The momentum equation is given by Eq. (1), where $\boldsymbol{F}$ is the body force which can be interpreted as the source of the enhancement of chemical reaction process.

$$(\rho\mathbf{U})\bullet\nabla\mathbf{U} = -\nabla P + \mu\nabla^2\mathbf{U} + \mathbf{F} \tag{1}$$

2.2. Availability analysis

The availability, or exergy, the maximum theoretical work obtained when the system interact with the environment, can be an evaluation criteria to measure the energy efficiency of a process.

The second law efficiency as the evaluation criteria of the process is defined as:

$$\eta = \frac{(\sum A)_{in} - I}{(\sum A)_{in}} \tag{2}$$

where I is the thermodynamic irreversibility, and is given by

$$I = T_{ref} \iiint \sigma \tag{3}$$

In the process of combustion of methane, the irreversibility is mainly due to the heat transfer and chemical reaction (Arjmandi and Amani, 2015, Bidi et al., 2010, Rana et al., 2014). The entropy generation rate can be approximately expressed as the sum of contributions of these two effects.

$$\sigma \approx \sigma_{cond} + \sigma_{ch} \tag{4}$$

$$\sigma_{cond} = \frac{\lambda}{T^2}(\nabla T)^2 \tag{5}$$

$$\sigma_{ch} = -\frac{1}{T}\omega_{ov}\Delta_r G_m \tag{6}$$

2.3. Optimal calculation

Based on the availability analysis in section 2.2, the second law efficiency can then be increased by minimizing the irreversibility of the process. So the integral of the total entropy generation is employed as the objective function in the optimization problem. And this objective function is given by

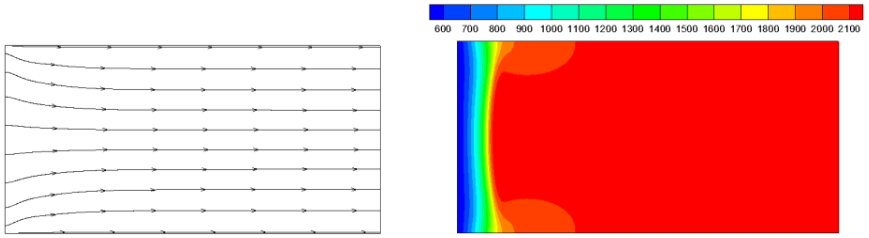
$$\min J^* = \iiint_{\Omega}\left[-\frac{1}{T}\omega_{ov}\Delta_r G_m + \frac{\lambda}{T^2}(\nabla T)^2\right] d\Omega \tag{7}$$

The combustion process must satisfy the continuity, the momentum, energy conservation and the mass conservation equations. By restricting the total viscous dissipation to a constant, the problem becomes the minimization of a functional with constraints. To solve the problem, a Lagrange function, which incorporates the constraints with the objective function via Lagrange multipliers, can be formulated. The minimum of the functional can then be calculated by using the Euler-Lagrange equations. Using the obtained velocity $\boldsymbol{U}$, the external body force $\boldsymbol{F}$ which can be known as the driving force to the flow field giving the minimum entropy generation rate, can be computed.

3. Results and discussion

To illustrate the proposed method above, a combustion process in the chamber shown in figure 1 was solved. The package FLUENT 14.5™ was used in the solution for the optimization problem. The UDS (user defined scalar) module was employed to solve the equation set calculated by variation method of the Lagrange function. The UDF (user defined function) module in the package was used to evaluate the body force $\boldsymbol{F}$ through the momentum Eq. (1).

For comparison, the combustion reaction process without optimization was firstly simulated under the condition of $\boldsymbol{F}$ = 0. The results of the streamline and the temperature distribution without optimization are shown in figure 2.



(a) Streamlines (b) Temperature distribution

Figure 2: Streamlines and Temperature distribution without optimization

And then, through solving the optimization problem of formulation (7) in section 2.3 under conditions of different values of viscose dissipation, entropy generations were obtained as shown in figure 3. Figure 3 shows that the minimum entropy generations due to the chemical reaction and heat transfer decreased with the increase of the viscous dissipation, and so did the total entropy generation.

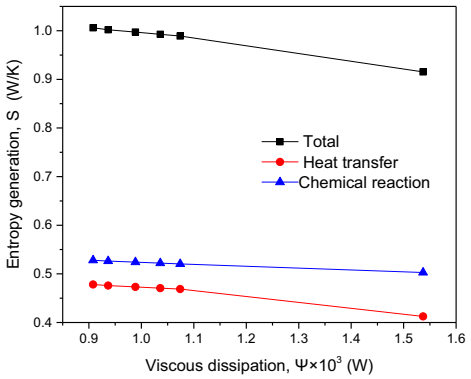
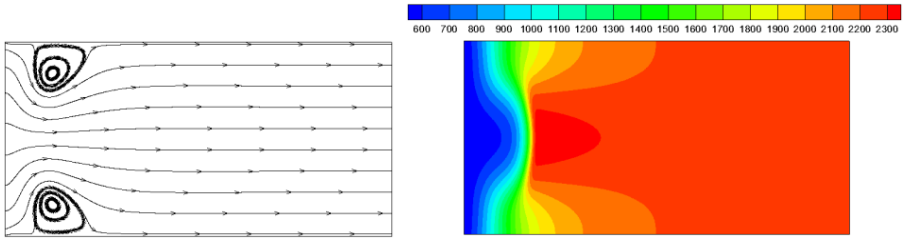


Figure 3: Variations of the entropy generation due to the chemical reaction and heat transfer, and the total entropy generation, with respect to the viscous dissipation



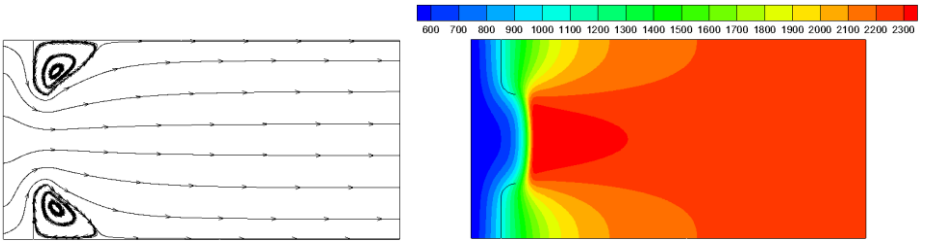
(a) Streamlines (b) Temperature distribution

Figure 4: Streamlines and Temperature distribution with optimization

The streamlines and the temperature distribution corresponding to the last point in figure 3 are shown in the figure 4. Compared with figure 2(a), it can be found that the optimization calculation led to the appearance of two axisymmetric eddies near the inlet of the combustor as shown in figure 4(a). The optimized flow field improved the mixing effects of the reactants and products and reduced the temperature gradient. As a result, the entropy generation of heat transfer decreases. Figure 4(b), compared with figure 2(b), shows the reaction (burning) region was extends towards the outlet of the chamber,

indicating that the reaction rate was more smoothed out with less availability losses of both the chemical reaction and the heat transfer. As a result, the outlet temperature of the combustion with the optimized flow field is higher than that in normal one, as shown in figure 5.

The total entropy generation reduced through optimization calculation and the second law efficiency calculated by Eq. (2) is increased from 71.87 % to 74.40 %, at a little more cost of viscous dissipation.



(a) Streamlines (b) Temperature distribution

Figure 5: Streamlines and temperature distribution in micro-combustor with baffles

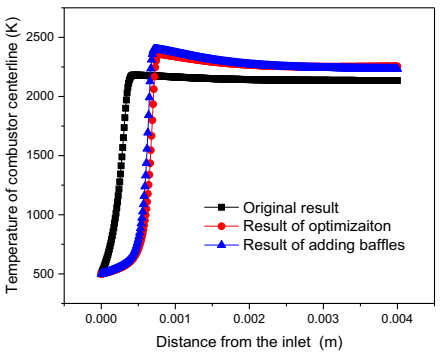


Figure 6: Centerline temperature profile comparison with the results of optimization, adding baffles and original results

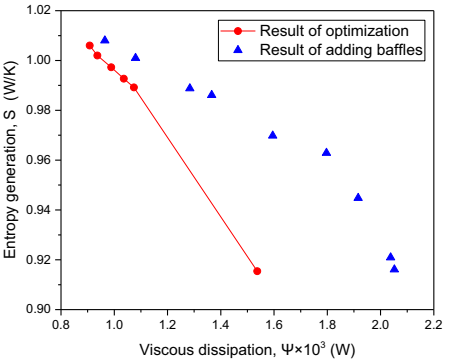


Figure 7: Entropy generation comparison after optimization and adding baffles

According to the Eq. (1), the optimized flow field can be known being generated by the body force **F**. In reality, however, it is difficult or even impossible to implement a force field to generate the optimized flow field as shown in figure 4(a). However, based on the optimization results in figure 4, we can design the internal structures of the micro-combustor so that the flow field can approach the optimal one. For example, inspired by figure 4, two axisymmetric baffles can be added into the micro-combustor near the inlet to generate similar eddies as shown in figure 4(a). The result close to the optimum one was obtained by changing the position, shapes and sizes of the baffles, as shown in figure 6. The streamlines and temperature distribution in figure 6 are shown similar to those in figure 4. Figure 5 shows that the combustion temperature for the improved combustor gives temperature lower than that for the optimal one, but much higher than the normal one.

Particularly, the value of the total entropy generations at the certain viscous dissipation in figure 3 can be regarded as the thermodynamic limit to the combustion enhancement by modifying the inner structure of the combustor. Figure 7 shows the comparison of total entropy generation after optimization and adding baffles. It can be seen that the availability destruction for adding baffles is never lower than that in the corresponding optimized one, which is the minimum of entropy generation of the process under the certain value of viscous dissipation.

4. Conclusions

Availability of the combustion process was used to define as the criterion for optimal design of a micro-combustor. The minimum of total entropy generation of the combustion process was set as the objective function, under fixed viscous dissipation. The calculus of variation was use to solve the optimization problem. The optimization results can be known as the thermodynamic limit for the enhancement of combustion process optimization. The thermodynamic limit can be used as the target of combustor structure design, although it is theoretical and difficult to reach in practice. Optimization design of the micro-combustor structures, according to the optimal flow field obtained, was obtained. The combustors with the modified structure had high second law efficiency, but lower than that by optimization at the same viscous dissipation.

In this paper, we proposed a method to the lowest entropy generation values at different values of viscous dissipation for the combustion process, which can provide a guideline for the micro-combustor design. The optimization design method of the micro-combustor can be extended to other types of combustor. Through the optimization method presented in this paper, one can retain more available energy in the combustion process, i.e., improve the grade of the energy, which is very meaningful in the age of resource and energy scarcity.

Acknowledgements

Support by National Natural Science Foundation of China (Grant 91434204) is acknowledged.

References

Arjmandi, H. R. & Amani, E. (2015) A numerical investigation of the entropy generation in and thermodynamic optimization of a combustion chamber. *Energy,* **81,** 706-718.

Bidi, M., Nobari, M. R. H. & Avval, M. S. (2010) A numerical evaluation of combustion in porous media by EGM (Entropy Generation Minimization). *Energy,* **35** (8)**,** 3483-3500.

Epstein, A. H. & Senturia, S. D. (1997) Macro Power from Micro Machinery. *Science,* **276** (5316)**,** 1211-1211.

Norton, D. G. & Vlachos, D. G. (2003) Combustion characteristics and flame stability at the microscale: a CFD study of premixed methane/air mixtures. *Chemical Engineering Science,* **58** (21)**,** 4871-4882.

Norton, D. G. & Vlachos, D. G. (2004) A CFD study of propane/air microflame stability. *Combustion and Flame,* **138** (1-2)**,** 97-107.

Rana, U., Chakraborty, S. & Som, S. K. (2014) Thermodynamics of premixed combustion in a heat recirculating micro combustor. *Energy,* **68** (8)**,** 510-518.

Yapıcı, H., Kayataş, N., Albayrak, B. & Baştürk, G. (2005) Numerical calculation of local entropy generation in a methane–air burner. *Energy Conversion and Management,* **46** (11-12)**,** 1885-1919.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018,San Diego,California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50320-7

Development of a model to identify combined use in residential water end use events

Eduardo S. Soares[a*], Karla P. Oliveira-Esquerre[a], Adelmo M. de Aguiar[a], Gabriella L. P. Botelho[a], Asher Kiperstok[b].

[a]*Federal University of Bahia (UFBA), Growing with Applied Modeling and Multivariate Analysis (GAMMA), Industrial Engineering Program (PEI), Rua Prof. Aristides Novis, 02, Federação, 40210-630, Salvador, Bahia, Brazil.*
[b]*Clean Technology Network of Bahia (TECLIM), Industrial Engineering Program (PEI), Rua Prof. Aristides Novis, 02, Federação, 40210-630, Salvador, Bahia, Brazil.*
soares.eduardo.sampaio@gmail.com

Abstract

The determination of water end use patterns by specific in-home equipment (appliance), the behaviour inherent to the users of the equipment, as well as the daily flow rate of water consumption, is important for the monitoring of water use, allowing the understanding of the system and the development of intelligent strategies for the management of this natural resource and its demand. Thus, the objective is to develop a model that identifies when an event is simple or combined (that is, simultaneous events, an event that includes several individual overlapping events), based on data obtained from low-income families in Brazil. An integrated approach was used combining water meters, home appliance audits and a self-reported household water use diary. Random Forest-based algorithms have been developed to aggregate flow tracking and to identify when an event is individual or combined. With the developed methods, it is easy to classify simple uses, but the combined uses still need more attention, being a future challenge the identification of its critical variables and improvement of classification.

Keywords: Water end-use event; Residential water flow trace disaggregation; Machine Learning. Pattern recognition; Random Forest.

1. Introduction

The world population has taken a major leap in numerical and industrial growth rates since the middle of the last century. In many cases, as in third world countries, socioeconomic development, as well as the supply chain, does not accompany population growth, with great disparity. Brazil, for example, is the 9th world economy (IPRI, 2017) and is home to about one fifth of the world's water reserves. However, although approximately 82.5% of the population has access to water, less than half of the population has access to sewage (48.6%), and the loss of about 37% in water supply (The World Bank, 2016). São Paulo, the largest and most well-developed city in the country, is experiencing a water crisis of the worst that has already been faced and there are no concrete plans to face the disaster that the crisis has caused (Carmo et al., 2016). Thus, the debate and the commitment to the development of strategies of water resources management are of great importance. Understanding how water use works in residential environments generates clear benefits for public water utilities, end users and for decision making regarding resource management strategies by allowing predictive

management of water supply during in periods of water scarcity or not, identification of water use peaks for water demand and cost management and identification of water use habits or water leaks at the facility, are some of these benefits (Carboni et al., 2016; Gurung et al., 2015.; McKenna et al., 2012; Morrison et al. 2014).

In order to provide the best knowledge of the residential consumption profile and help the understanding of the systems for the development of better public strategies, several approaches are used for the disaggregation of consumption and its subsequent characterization. The measurement can be based on different types of meters, such as those based on water flow, pressure evaluation, acoustic evaluation and multi-modal (article-sensors) and, as the measurement forms vary, the classifiers also do. For this work, we will focus on the approach of methods based on water flow, since they have a better adaptation to the residences in which the study was conducted. For this approach, the main commercial classifiers are *TraceWizard* (Aquacraft, 2016) and *Identflow* (WRc Plc, 2016), which have satisfactory identification of singular uses (based on Decision Trees algorithms), but still require great interaction with the user. In addition, they have low quality with respect to the analysis of multiple uses, which show the cases of greater difficulty. This problem has a good approach by Nguyen et al. (2013) which makes use of Hidden Markov Model (HMM) and Dynamic Time Warping to make the problem solvable. It is noteworthy that both HMM and DTW alone do not achieve such positive results, requiring a joint approach. Thus, we use this work to propose a new method based on Random Forest to identify the simple and multiple uses, which presents good quantitative and qualitative results.

2. Methodology

2.1. Obtaining Data

Data collection was conducted in the Plataforma and Chapada do Rio Vermelho neighbourhood, in the city of Salvador, state of Bahia, Brazil. This is the largest state of the Northeast Region of Brazil, with an estimated population of 15,344,447 inhabitants. Its population density is 24.82 inhab / km² and the average water consumption per capta is 110.57 L / inhab.day (IBGE, 2017). The state capital, Salvador, consumes comparatively most of the water, characterizing an uneven distribution. According to the Brazilian Institute of Geography and Statistics (IBGE), in the last census conducted (2010), the city of Salvador had a population density of 3,859.44 inhabitants / km² with an area of 692,818 km².

During the studies, ten families were monitored with the implanted system. This system was composed of a LAO®, DN 20 and metrological class B multi-jet water meter installed at the reservoir outlet, with a sensor to 0.1 L of water used. The system was then configured to collect data every 10 seconds, characterizing the period of each pulse. This gives us total of 2713 days of observations (plus users) between 2011 and 2014, which represents about 1.6 million pulses to be worked. The final data generated by the system were analyzed by *TraceWizard* software version 4.1 and after analyzed by trained technical analysts. Before starting the analysis using the software, it was necessary to define the flow chart corresponding to each device in the residence. Subsequently, they were entirely revised based on the usage log and, in the end, the

complete data was organized for each pulse, with the complete date (including hour, minute, second), the recorded flow, the utensil related to that use (including Toilet, Shower, Water Filter, Bathroom faucet, Washing Machine, Kitchen faucet, Clotheswasher faucet), family, ID of the house. It is worth mentioning that *TraceWizard* worked only as a process accelerator, with only 48.86% of the original data being preserved.

2.2. Data analysis and aggregation

All data were processed and aggregated, defined as single or multiple uses. The total duration of use was estimated as well as its volume and the appliances involved. It was organized as follows: contend fixture type, date of event, start time, stop time, duration, volume, peak flow rate, modal flow rate - as *TraceWizard* (DeOreo, 2011), mean flows, mode, number of repetitions of the mode, median, volume and the probability of the use to be simple (in each hour), besides the peak referring to the flows. All manipulations and analyzes were performed with the aid of the programming language R.

2.3. Random Forest

The Random Forest algorithm used was proposed by Breiman (2001). This algorithm is characterized with a supervised learning algorithm from a set of Decision Trees. These Decision Trees are trained by the bagging method, based on the idea that the combination of learning models improves the result. This differs from the standard trees because each node is divided using the best combination of all the variables. As the decision trees grow, an extra randomness is brought in for the random forest model and it searches for the best feature within a random subset of features. It does not make the algorithm more complicated to use since it requires only few hyperparameters and these generally produce a good classification results. Other important aspects to be considered are overfitting and the sensitive analysis of the variables. Because there are enough trees in the forest, the classifier is not expected to overfit the model. And by providing information of how much each feature is related to the model result (dependent variable), it is possible to select which features to use in the model, since features with low importance can end up, often, interfering negatively to the result.

For this research, the models were constructed using the R programming language and RandomForest package (Liaw and Wiener, 2002). A ratio of 80/20 was applied for the division of the training and test data set, the sets being sampled respecting the proportion of the categories contained in the database.

3. Discussion and Results

3.1. Model Accuracy

The Random Forest model was created with the variables: number of repetition of the mode, the mode, mean flow, peak flow rates, median, probability of the use in a given hour to be simple, volume of a use and the household, as with 500 trees. The obtained results show a precision, recall and F1 score of 0.935, 0.975 and 0.954 for the train set and 0.936, 0.975 and 0.955 for the test set. Precision indicates the number (or

percentage) of simple/combined uses that are actually simple/combined. Recall shows the number of accurate answer we labeled as simple/combined. The F1 score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. The results have good sensitivity, but the model is able to predict much better simple uses than combined ones.

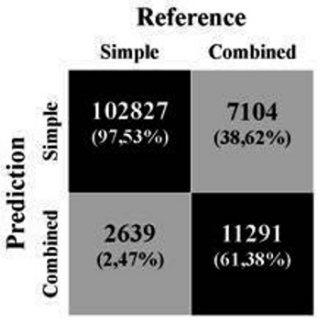


Figure 1. Confusion Matrix.

The major difficulty between differentiating single uses and combined uses is that the way in which the equipment works depends greatly on the behaviour of the user. There is no regular behaviour determined by mechanical parts, thus, the user can use the flow rate as desired. So often, a high flow from the shower can be confused with a high flow rate from the use of the sink and toilet, with similar durations, for e.g. In Figure 1, the confusion matrix can be evaluated.

3.2. Variable Importance

For the model created, for the prediction of simple uses, there are no outstanding variables. The most important variable is the median, but it does not have much difference in relation to the others. With respect to the combined uses, the most important variable is the mean flow, with great emphasis, as can be seen in Figure 2, where VM1 is the number of repetition of the mode; VM2 is the mode; VM3 is the mean flow; VM4 is the peak flow rate; VM5 is the median; VM6 is the probability of the use in a given hour to be simple; VM7 is the volume of a use and; VM8 is the household. The definition of the mean flow for the classification of combined uses as the most important variable comes along with the mode and the median, characteristics that say about the distribution of the flows in the time space of the use. Simple uses, in general, have more regular profiles, so the variables have similar amounts. The combined uses have distortions generated by the overlapping of the flows, which directly impact on the central tendency of the uses.

3.3. Model Threshold

As can be seen in Figure 3 and Figure 4, the model performs the correct prediction most of the time. The scale of the votes represents the degree of certainty that is in each vote, being 1.0 a vote of absolute certainty of the model. Thus, when the model has great certainty about the vote, it misses very little. The errors only surpass the hits when the vote is certain of 0.5, but for simple uses only.

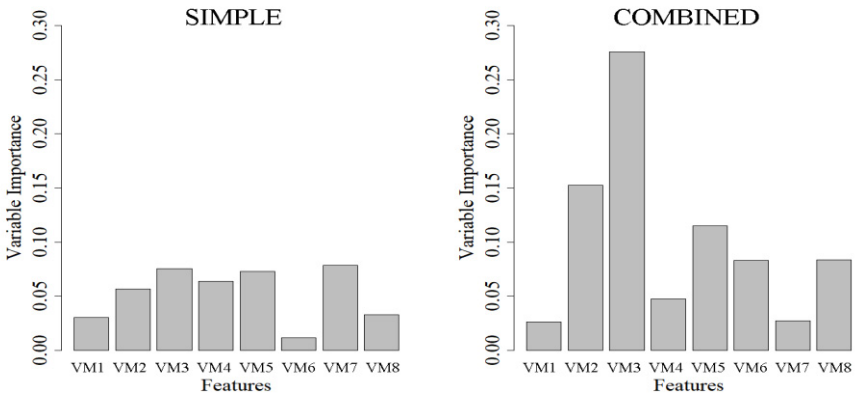


Figure 2. Variable Importance.

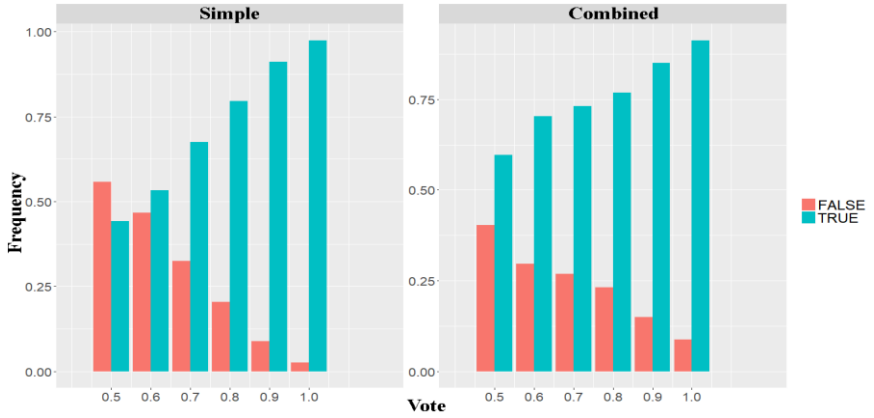


Figure 3. Model Threshold (Frequency vs. Vote).

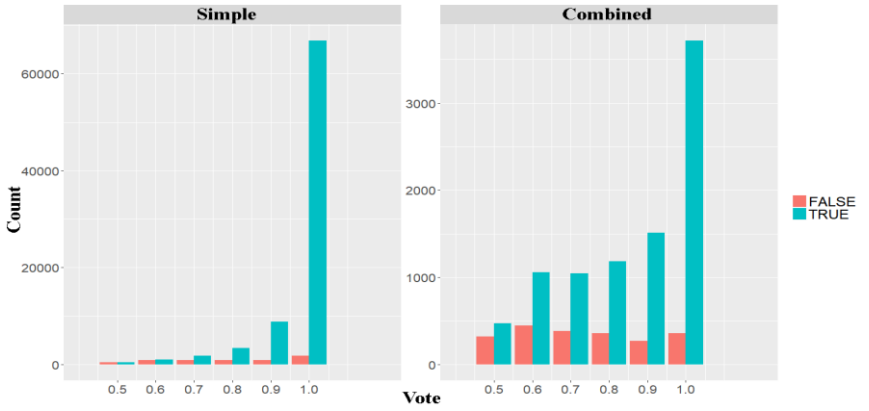


Figure 4. Model Threshold (Count vs. Vote).

4. Conclusion

The classification of uses by Random Forest is promising, with satisfactory results when compared with others available in the literature. In previous work of the group, we have already been able to classify and predict the category of simple uses. This work serves as a complement to the identification of combined uses, which is a major challenge for most approaches to the problem. The model has good sensitivity, but the accuracy can still be better. Thus, it is necessary to evaluate what could be new variables that add more information to the model. For future work, it is expected to be able to describe the profile of the use and to use a more integrated approach, without the need of specific algorithms for each step (identification of simple or combined uses and later classification). Other approaches in addition to machine learning algorithms are also evaluated for future work.

References

Breiman, 2001. Random forests. Machine Learning, 45(1), 5-32, 2001.

Carmo, R. (Org) et al. Demographic dynamics and environmental change: disasters and water (Dinâmica demográfica e mudança ambiental: desastres e água, in Portyguese only] Campinas, SP: Núcleo de Estudos de População "Elza Berquó" / Unicamp, 2016.

Carboni, D., Gluhak, A., McCann, J., & Beach, T. Contextualising water use in residential settings: A survey of non-intrusive techniques and approaches. Sensors, 16(5), 738, 2016. https://doi.org/10.3390/s16050738.

DeOreo, W. B. California Single Family Water Efficiency Study, Aquacraft, www.aquacraft.com, 2011.

Gurung, T. R.; Stewart, R. A.; Beal, C. D.; Sharma, A. K. Smart meter enabled water end-use demand data: platform for the enhanced infrastructure planning of contemporary urban water supply networks. Journal of Cleaner Production, Volume 87, Pages 642-654, 2015.ISSN 0959-6526. https://doi.org/10.1016/j.jclepro.2014.09.054.

IBGE - Brazilian Institute of Geography and Statistics: Cidades 2017 <http://cidades.ibge.gov.br/> (accessed March 19 2018).

IPRI - Institute of International Relations Research (Brazil): The 15 biggest economies of the World (As 15 maiores economias do mundo, in Portuguese only]. <http://www.funag.gov.br/ipri/index.php/o-ipri/47-estatisticas/94-as-15-maiores-economias-do-mundo-em-pib-e-pib-ppp> (accessed March 19 2018).

Liaw A.; Wiener, M. Classification and Regression by Random Forest. R News 2(3), 18-22, 2001.

McKenna, E.; Richardson, I.; Thomson, M. Smart meter data: Balancing consumer privacy concerns with legitimate applications. Energy Policy, 41, 807–814, 2012. DOI: 10.1016/j.enpol.2011.11.049.

Morrison, J., Friedler, E. A critical review of methods used to obtain flow patterns and volumes of individual domestic water using appliances. Urban Water Journal, 12(4), 328–343, 2014. https://doi.org/10.1080/1573062x.2014.900090.

Nguyen, K. A.; Zhang, H.; Stewart, R. A. Development of an intelligent model to categorise residential water end use events. Journal of Hydro-environment Research, 7(3), 182-201, 2013. ISSN 1570-6443.https://doi.org/10.1016/j.jher.2013.02.004.

The World Bank (USA) (Org.): Brazil may be the owner of 20% of the world's water supply but it is still very thirsty. 2016. <http://www.worldbank.org/en/news/feature/2016/07/27/how-brazil-managing-water-resources-new-report-scd> (accessed March 19 2018).

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50321-9

Simultaneous Energy and Water Optimization in Shale Exploration

Doris Oke[a], Thokozani Majozi[a]*, Rajib Mukherjee[b], Debalina Sengupta[b] and Mahmoud El-Halwagi[b]

[a]*School of Chemical and Metallurgical Engineering, University of the Witwatersrand, 1 Jan Smuts Avenue, Braamfontein, Johannesburg, 2000, South Africa*
[b]*Chemical Engineering Department,Texas A&M University,College Station, TX 77843-3122,USA*
Thokozani.majozi@wits.ac.za

Abstract

This work presents a mathematical model that simultaneously optimizes water and energy usage in hydraulic fracturing using continuous time formulation. This is achieved through purification of the highly contaminated water using membrane technology which is energy intensive. The membrane regenerator considered in this work is membrane distillation (MD). Detailed design model for this technology is developed and incorporated within the water network superstructure in order to allow for simultaneous optimisation of water, operation, capital cost as well as energy used. The study also looked at the feasibility of utilising the co–produced gas that is traditionally flared (in the case of oil producing shale) as energy source for the membrane regenerator. The resultant model is applied in a case study to demonstrate its applicability. The application of the model results in 22.4 % reduction in freshwater consumption and 23.2 % savings in the total cost of freshwater. The thermal energy consumption is in the order of 244×10^3 kJ/m^3 of water.

Keywords: Hydraulic fracturing, water and energy, membrane technology, MD, flaring

1. Introduction

The "shale revolution" has stimulated tremendous production of oil and natural gas globally. From 2007 to 2015, shale gas production in the US has grown from 2 to 15 trillion cubic feet per year with estimates of continued growth to support massive monetization projects (Zhang and El-Halwagi, 2017). However, the process by which it is carried out known as hydraulic fracturing is associated with several environmental challenges, i.e. excessive amount of water usage, generation of significant volume of wastewater and widespread flaring of co-produced gas.

Water management decisions within shale gas production fall into two primary categories, i.e. water utilization within hydraulic fracturing operations and disposal of wastewater from drilling and production (Gay et al., 2012). Shale gas production is a highly water-intensive process with a typical well requiring around 5 million gallons of water to drill and fracture. Water contamination is another aspect of shale gas production that has generated significant resistance to current shale production processes. Thus, water consumption in shale gas production has become a major concern making water resource management an important operational and environmental issue (Hasaneen and El-Halwagi, 2017). The efficient use of water resources involves the treatment of

wastewater for reuse or recycling which can be achieved using water purification units often classified as membrane and non-membrane processes. The use of membrane system for water treatment is an energy intensive process, therefore minimising energy within the water network is also of great importance.

A number of studies have been carried out on the optimisation model for water management in shale gas production. Yang et al. developed mixed integer linear programming (Yang et al., 2014a) and mixed integer nonlinear programming (Yang et al., 2015) investment and scheduling model for optimal water management in shale gas which dealt with short and longer term operation respectively. Gao and You (2015) addressed a similar problem but assumed a fixed schedule for the wellpad fracturing. Bartholomew and Mauter (2016), also developed a multiobjective mixed integer linear programming for assessing the trade-offs between financial cost and human health and environment costs. However, most of these studies have adopted the discrete time scheduling formulation for the wellpad fracturing which does not guarantee model accuracy and leads to suboptimal solution due to the discrete approximation of the time horizon. In addition, most of the research done in this area has represented the wastewater treatment unit as a "black box "which does not give the true cost representation of the project or uses "short cut" regenerator model (Yang et al., 2014b) due to the complexity of the regenerator design.

Furthermore, studies available in literature for the utilisation of the co-produced gas that is flared in the case of oil producing well is either focused on the onsite atmospheric water harvesting (Wikramanayake and Bahadur, 2016) using the captured gas or channelling the gas that would have been flared as a source of heat (Glazer et al., 2014) for heat based regenerators. The work by Glazer et al. (2014) is done based on analytical framework and not in the context of mathematical optimization.

This paper addresses the synthesis and optimisation of an integrated water and membrane network that simultaneously optimizes water and energy consumption in hydraulic fracturing using continuous time formulation. The membrane technology considered is membrane distillation. A detailed design of this regenerator is incorporated to determine the optimal operating conditions for efficient energy use.

2. Problem statement

The problem statement in this work can be stated as follows. Given are: (i) number of freshwater sources, (ii) set of wellpads S with known flow rates and known maximum allowable contaminant concentrations, (iii) total number of frac stages for each wellpad, (iv) earliest fracturing date for each wellpad, (v) set of wastewater injection wells, (vi) volume of water required per stage, (vii) the minimum and maximum number of stages that can be fractured per day (viii) time horizon of interest, (ix) regenerator with variable removal ratio and (xi) a gas storage facility.

Using the given data, it is required to determine (i) the fracturing schedule of the wellpads that optimizes water and energy usage, (ii) optimum design of MD regenerator such as number of membrane modules and energy consumption and (iii) the feasibility of using captured flared gas as energy source for the regenerator.

3. Model formulation

Based on the problem statement, the superstructure in Figure 1 is developed.

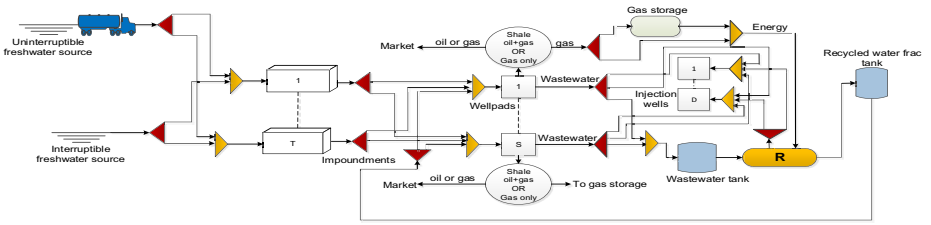


Figure 1. Superstructure representation of the water network.

The objective of the problem is defined in Eq. (1) as:

$$\max\ profit = R - \left(c^{fw} + c^{ww,trt} + c^{ww,st} + c^{dis}\right) \tag{1}$$

The objective function is subject to the following constraints:

Eq. (2) is the allocation constraint which specifies that each wellpad *s* has to be fractured exactly once at time point *n.* A time point is a precise moment within a given horizon when an event occurs.

$$\sum_{n} wx_{s,n} = 1 \qquad \forall s \in S \tag{2}$$

Eq. (3) states that the time at which time point *n* occurs must correspond with the availability time of wellpad *s.*

$$Tt_{n} \geq \sum_{s}\left(st_{s} wx_{s,n} - H\left(1 - wx_{s,n}\right)\right) \qquad \forall n \in N \tag{3}$$

Water requirement by each wellpad is supplied with freshwater and/or recycled water from the frac tank, which is obtained by Eq. (4).

$$tw_{s,n} = fw_{s,n} + rw_{s,n} \qquad \forall s \in S, n \in N \tag{4}$$

Eq. (5) states that the flowback water generated after wellpad fracturing can be sent to regenerator R for treatment or injection well.

$$f_{s,n}^{bw} = f_{s,n}^{trtb} + f_{s,n}^{dis} \qquad \forall s \in S, n \in N \tag{5}$$

The heat required by the feed into the regenerator R is given by Eq. (6)

$$Hrf = QfCp\left(Tbf - Tsf\right) \tag{6}$$

The regeneration network involves both capital and operational costs. These functions are included in the overall objective function of the water network as $c^{ww,trt}$, such that the energy consumption and subsequent cost of regeneration are optimized in conjunction with water consumption.

The annualized fixed and operating cost of the MD network excluding heating is given by Eq. (7)

$$AC = \left(58.5A_{m} + 1115Qf\right) + \left(1411 + 43\left(1 - LR\right) + 1613\left(1 + V\right)\right)Qf \tag{7}$$

The annual heating cost is given by Eq. (8)

$$AHC = AOT\left(khHrf\right) \tag{8}$$

Additional constraints include water and contaminant balance around the storage vessel and the regenerator, sequencing constraint for the water and regenerator network, and the design constraints for the regenerator.

4. Case study

To illustrate the application of the proposed model, an example taken from Yang et al. (2014) is considered as illustrated in Table 1. Three scenarios are considered to ensure the detailed analysis of the model. Scenario 1 is the base case which is the water integration without regeneration. Scenario 2 is the case where black box model is used; i.e. water minimisation only. Scenario 3 considers water integration involving detailed regenerator where water and energy are optimized simultaneously.

Table 1. Wellpad Data

Wellpads	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14
Earliest frac day	1	1	1	1	1	39	1	273	273	273	396	379	379	1
# of stages	57	61	54	55	64	26	97	88	86	76	63	100	100	87

The solution comparison between the three scenarios is given in Table 2 and the fracturing schedule for each scenario is presented in Figures 2-4. The results show that the model proposed to make use of freshwater from interruptible sources thereby reducing the cost and environmental issues that is associated with trucking. Figures 3 and 4 show that the tightness in fracturing schedule of each group of wellpads improves the recycling efficiency of flowback water. Due to efficient reuse of flowback water, a saving of 183,534 m^3 of freshwater is achieved out of the total volume of 818,000 m^3 required for the 14 wellpads. The saving is found to be 21.2 % higher than the study in literature (Yang et al., 2014a) using discrete time formulation. In scenario 2, 96.7 % of the flowback water is sent to regenerator *(R)* and the remaining 3.3 % is sent to the injection well to be disposed while in scenario 3, 99.4 % of the flowback water is sent to regenerator (*R)* while the remaining 0.6 % is disposed.

In order to calculate the cost and energy associated with the wastewater regeneration, a cost analysis based on the black box model and detailed model was performed. The results obtained are given in Table 3. The results show that the deviation of the cost function from the actual cost of regeneration in scenario 2 is 12.7 %. The result obtained in scenario 3 show that the optimized cost of regeneration is 6.6 % higher than the cost of MD standalone model. This is due to the fact that optimizing the MD-feed temperature results in a decrease in the water flux thereby increasing the membrane area required which in turn leads to an increase in the fixed cost of the membrane. The simultaneous optimization of both energy and water within the water network results in 12.7 % reduction in the amount of energy required by the regenerator based on the throughput per day. The amount of energy required is reduced from 699×10^6 kJ to 610×10^6 kJ. The thermal energy consumption is in the order of 244×10^3 kJ/m^3 of water, which is found to be lesser than the range of thermal energy consumption values reported for membrane distillation in literature. The range of thermal energy required by membrane distillation is between

120 and 1700 kWh/m^3 equivalent to between 432×10^3 kJ/m^3 and 6.12×10^6 kJ/m^3 (Suarez and Urtubia, 2016). Using gas that would otherwise be flared as the source of heat for the regenerator rendered the heating cost in the objective function to become zero.

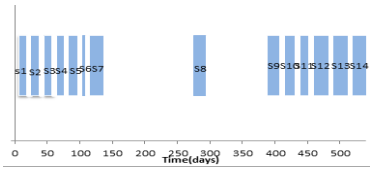


Figure 2. Fracturing schedule (base case)

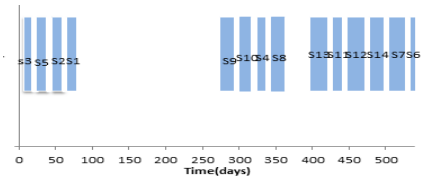


Figure 3. Fracturing schedule (scenario 2)

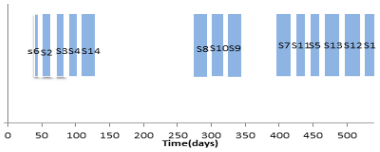


Figure 4. Fracturing schedule (scenario 3)

Table 2. Solution comparison

	Scenario 1	Scenario 2	Scenario 3
Freshwater pumped (1,000 m^3)	818.8	640.3	635.3
Freshwater trucked (1,000 m^3)	0	0	0
Regenerated water (1,000 m^3)	0	178.5	183.5
Freshwater save (%)	0	21.8	22.4
Freshwater trucking cost ($ 1,000)	0	0	0
Freshwater pumping cost ($ 1,000)	13,043	10,019	10,012
Revenue ($ 1,000)	--	261,240	261,240
Profit ($ 1000)	--	235,860	237,340

Table 3. Summary of regeneration costs, values are given in $1000/annum

	Estimated cost (Scenario 2)	Standalone cost	Water and energy optimization (scenario 3)
MD	11,307	9,875	10,575

5. Conclusions

This work explores simultaneous water and energy optimization in shale play using continuous time formulation with the incorporation of detailed MD model within the water network. The goal is to balance the tradeoff between water acquisition from interruptible and uninterruptible water sources. The efficient recycling of wastewater leads to 22.4 % reduction in the amount freshwater required. The approach indicates that optimizing energy and water simultaneously results in significant reduction in the amount of thermal energy required for regeneration. Based on the data gathered from literature, the amount of gas that is flared in most of the shale play is sufficient to provide the energy needed for regeneration.

Nomenclatures

R revenue from gas production	c^{fw} freshwater cost
AC annual fixed and operating cost	AHC annual heating cost
$c^{ww,trt}$ wastewater treatment cost	A_m required membrane area
$c^{ww,st}$ cost of wastewater storage	kh cost of heating
c^{dis} cost of disposal	$f_{s,n}^{bw}$ flowback water at time point *n*
$tw_{s,n}$ total water needed for fracturing wellpad *s* at time point *n*	Qf flowrate into regenerator *R*
	Cp specific heat of the feed into *R*
$fw_{s,n}$ freshwater required at time point *n*	Tbf temperature of the feed in the bulk of *R*
Hrf heat required by the regenerator feed	Tsf supply temperature of the feed into *R*
LR Liquid recovery for the regenerator	Tt_n time corresponding to time point *n*
H time horizon of interest	
v ratio of recycled reject to raw feed	$f_{s,n}^{trtb}$ amount of water sent to regenerator *R* at time point *n*
$rw_{s,n}$ recycled water needed at time point *n*	
st_s Availability date of wellpad *s*	$f_{s,n}^{dis}$ amount of water sent to disposal at time point n
$wx_{s,n}$ binary variable for the assignment of task *s* at time point *n*	

References

T.V. Bartholomew, M.S. Mauter, 2016, Multiobjective optimization model for minimizing cost and environmental impact in shale gas water and wastewater management, Journal of ACS Sustainable Chemistry and Engineering 4, 3728-3735.

J. Gao, F. You, 2015, Optimal design and operations of supply chain networks for water management in shale gas production: MILFP model and algorithm for the water-enegy nexus, Journal of American Institute of Chemical Engineers 61,6, 1184-1208.

M.O. Gay, S. Fletcher, N. Meyer, S. Gross, 2012. Water management in shale gas plays. Technical report. IHS Water White Paper, 1-10.

Y.R. Glazer, J.B. Kjellsson, K.T. Sanders, M.E. Webber, 2014, Potential for using energy from flared gas for on-site hydraulic fracturing wastewater treatment in Texas, Environmental Science and Technology Letters 1,300-304.

R. Hasaneen, M.M. El-Halwagi, 2017. Integrated Process and Microeconomic Analyses to Enable Effective Environmental Policy for Shale Gas in the United States, Journal of Clean Technologies and Environmental Policy 19,6, 1775-1789.

F. Suarez, R. Urtubia, 2016. Tackling the water-energy nexus: an assessment of membrane distillation driven by salt-gradient solar ponds, Journal of Clean Technology Environmental Policy, 1-16

E. Wikramanayake, V. Bahadur, 2016, Flared natural gas-based onsite atmospheric water harvesting (AWH) for oilfield operations, Environment Research Letters 11, 1748-9326

L. Yang, I.E. Grossmann, M.S. Mauter, R.M. Dilmore, 2015, Investment optimization model for freshwater acquisition and wastewater handling in shale gas production, Journal of American Institute of Chemical Engineers 61,6, 1770-1782.

L. Yang, R Salcedo-Diaz, I.E. Grossmann, 2014b. Water network optimization with wastewater regeneration models, Journal of Industrial Engineering Chemistry Research 53,17680–17695.

L. Yang, I.E. Grossmann, J. Manno, 2014a, Optimization models for shale gas water management, Journal of American Institute of Chemical Engineers 60,10, 3490-3501.

C. Zhang, M.M. El-Halwagi, 2017. Estimate the Capital Cost of Shale-Gas Monetization Projects, Chemical Engineering Progress 113,12, 28-32.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50322-0

Modelling a water-air conditioning system of a large commercial building for energy consumption evaluation

Rodrigo Luiz Schil Pellegrini[a], Natália Magalhães de Araújo[a], Míriam Tvrzská de Gouvêa[a*]

[a]*Mechanical Engineering Deparment, School of Engineering, Universidade Presbiteriana Mackenzie, Rua da Consolação 896, São Paulo, ZIP 01302-907, Brazil*
miriamtg_br@yahoo.com, miriam.gouvea@mackenzie.br

Abstract

In this paper, a mathematical model used to represent the thermal behavior of an air conditioning water-air system applied to a large commercial building is proposed to represent its operation. Thermodynamic properties of refrigerant are rigorously modelled as well as subcooling and superheating of refrigerant inside the chiller. A simplified first principle model is presented for the description of the several conditioned rooms. The proposed model is used to evaluate the dynamic behavior of the system when subjected to external ambient air and thermal load disturbances. System response to step changes in the main manipulated variables are also shown.

Keywords: Air conditioning, vapour compression system, chiller, air-water systems, process simulation.

1. Introduction

In commercial buildings, energy consumption by heating, ventilation, and air conditioning (HVAC) can achieve 50% of the total energy consumed and it is reported that air conditioning may be responsible for 10 to 20% of the national energy demand in developed countries (Vakiloroaya et al., 2014). Therefore, being able to better control or optimize the operation of air conditioning systems is an up-to-date challenge.

Air-water HVAC systems are widely used in large commercial applications because they can ensure good temperature and air quality control. Figure 1 shows one possible implementation of an air-water HVAC, which is considered in this paper. Cool air inflated in the conditioned zones is refrigerated by cold water in fan-coils. The water into the fan-coils comes from chillers. Each chiller mainly consists of a vapor compression system, possessing a compressor, a condenser, an evaporator and an expansion valve. In order to control the temperature in the conditioned zones, the flowrate of the air into the zones is maintained fixed and a water bypass is introduced in each fan-coil, as is shown in figure 1. The bypass fraction is manipulated in order to control the temperature in the zone. Although water condensers are preferably used in large applications, air condensers can also be used, as is the case here. As for the compressor and expansion valve, in this paper, a positive displacement compressor and an electronic expansion valve (EEV) are used.

In order to test new control or optimization strategies a suitable model of the process is needed. Afram and Janabi-Sharifi (2014) reviewed modelling methods for HVAC

systems. Models that are based on physics and system performance data have appropriate generalization capabilities and accuracy and are more largely used. However, several models used in the literature for testing more advanced control strategies make use of significant simplifications. For example, Ma et al. (2009) developed a model to test several control strategies. Instead of calculating the compressor work using a rigorous thermodynamic model, they evaluated the compressor work from the work obtained from a compression cycle where the vapour phase at the exit of the compressor is saturated. In this paper, we will see that this approach is not adequate for the HVAC considered here.

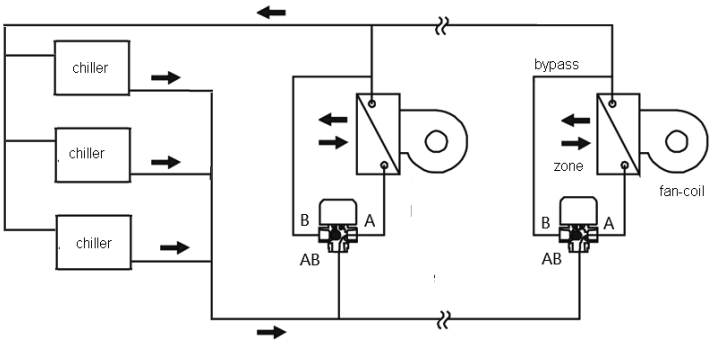


Figure 1: HVAC system considered in the present work.

2. Building description and mathematical modelling

Nomenclature					
A_{eev}	EEV opening	$\dot{V}$	volumetric flowrate	e	entry
A_g	geometric throat area of EEV	VU	specific volume	ev	evaporator
c_p	heat capacity at constant pressure	W_{cp}	shat work of compressor	ext	external air
c_v	heat capacity at constant volume			fan	*fan-coil*
f_{bpr}	bypass fraction	γ	heat capacity ratio	fc	from chiller
h	enthalpy	η_i	isentropic efficiency	in	inflated air
m	mass	η_v	volumetric efficiency	lw	liquid water
$\dot{m}$	mass flowrate	ω	shaft speed	ret	return air
P	pressure			rf	refrigerant
q	heat rate	Subscript		s	exit
T	temperature	c	chiller	tc	to chiller
ua	humidity ratio	cd	condenser	vw	vapour water
UR	relative humidity	cp	compressor	w	water
		da	dry air	z	zone

The building considered in this paper is situated in the city of Belo Horizonte in Brazil and has an area served by the HVAC system of 19,193.40 m^2, which is divided into 73 zones served by 73 fan-coils. The nominal capacity of each of the 3 chillers needed to refrigerate the building at operating point is 1064,4 kW. The refrigerant is R-410A. In the design, ambient dried and bulb air temperatures were considered to be, respectively, 32 ºC and 24 ºC. Ambient conditions of all zones are to be controlled at 24ºC and a relative humidity of 50%. Evaporator´s and condenser´s saturation temperatures are, respectively, -5ºC and 50ºC at nominal point. The design superheating is of 10ºC and subcooling (ΔT_{sub}) was designed to be 5ºC.

Air and liquid water properties are modelled as in ASHRAE (2017). Specific volume of R-410A is modelled by Martin-Hou equation of state (DuPont, 2004). Enthalpy, internal energy and entropy of superheated vapour at low pressure conditions were modelled considering an ideal gas behaviour. The heat capacity of the ideal gas was considered as in Dupont (2004). All other properties were obtained from the data presented in Çengel and Boles (2011).

The focus of the modelling presented here is on the thermal behaviour of the HVAC system. Friction losses in piping and heat exchangers are therefore neglected. The energy balances of each zone and of its corresponding fan-coil, as well as of each evaporator and condenser are modelled assuming a mixed chamber as usually done in HVAC modelling (Afram and Janabi-Sharifi, 2014). Eq.(1) to Eq.(4) exemplify the procedure. Note that although the equations simplify the physical reality, they can adequately predict subcooling and superheating of the refrigerant in the chillers. Also, the effect of external air state on system performance is explicitly taken into account. Water flowing through the evaporator of each chiller has constant flowrate. Heat rate for each heat exchange device is related to the logarithmic mean temperature. In this paper, every global heat exchange coefficient is assumed constant. The idea is that those values can be obtained on-line from data reconciliation (Özyurt and Pike, 2004). Thermal load to the zone may be varied accordingly to the design thermal profile or to simulate abnormal ambient conditions.

$$\frac{dT_{ret}}{dt} = \frac{\dot{m}_{air}\left(cp_{da} + ua \cdot cp_{vw}\right)\left(T_{in} - T_{ret}\right) + q_z}{m_z\left(cv_{da} + ua \cdot cp_{vw}\right)} \tag{1}$$

$$\frac{dT_{fan}}{dt} = \frac{\dot{m}_{fan} cp_{lw}\left(T_{fc} - T_{fan}\right) + q_{fan}}{m_{fan,w} cp_{lw}} \tag{2}$$

$$\frac{du_{cd,air}}{dt} = \frac{1}{m_{cd,air}}\left(\frac{\dot{V}_{ext}}{VU_{ext}}\left(h_{ext} - h_{cd,air}\right) + q_{cd}\right) \tag{3}$$

$$\frac{du_{cp,e}}{dt} = \frac{\dot{m}_{rf}\left(h_{cd,s} - h_{cp,e}\right) + q_{ev}}{m_{ev,rf}} \tag{4}$$

All water streams from all fan-coils are mixed into one stream that is divided into 3 streams, each entering one chiller. Similarly, all water streams from the 3 chillers are mixed and then split into the 73 streams that will feed the fan-coils. In this work, we use the model proposed by Li (2013) to simulate the EEV. Under chocked condition, Eq. (5) can be used to model the mass flowrate of R-410A. The parameters a_1 to a_4 can be retrieved from Li (2013), ΔT_{sub} is the subcooling at the condenser and the critical temperature T_c is 345.28K (Dupont, 2004).

$$\dot{m}_{rf} = 0.676123\left(-a_1 + a_2 \cdot A_{eev} - a_3 \cdot A_{eev}{}^2 + a_4 \cdot A_{eev} \cdot \Delta T_{sub} T_c^{-1}\right) \cdot A_g \sqrt{\rho_{cd,s} \cdot P_{cp,s}} \tag{5}$$

In this paper, the positive-displacement compressor is modelled at steady-state, assuming that its dynamic is very fast. The volumetric flowrate is modelled as usually done in the literature by Eq. (6) and the volumetric efficiency is modelled by Eq. (7). The shaft work is given by Eq. (8). Roskosch et al. (2017) suggest a dependency of the

isentropic efficiency on the pressure ratio of the compressor. This is the approach considered here in Eq. (9).

$$\dot{V} = V_{cp} \cdot \omega \cdot \eta_v \tag{6}$$

$$\eta_v = 0{,}91 - 0{,}011279\left(P_{cp,s} \cdot P_{cp,e}^{-1}\right)^{\gamma} - P_{cp,s} \cdot P_{cp,e}^{-1} \tag{7}$$

$$W_{cp} = \dot{m}_{rf}\left(h_{cp}^{*} - h_{cp,e}\right)\eta_i^{-1} \tag{8}$$

$$\eta_i = 0{,}0046362 P_{cp,2}^{3} \cdot P_{cp,e}^{-3} - 0.068534 P_{cp,2}^{2} \cdot P_{cp,e}^{-2} + 0.34692 P_{cp,s} \cdot P_{cp,e}^{-1} + 0.11229 \tag{9}$$

3. Results

The equations were solved with the aid of Matlab®. The differential equations were solved using Runge-Kutta 4th order method with a maximum integration step of 1s. The non-linear set of equations were solved through Matlab®´s SQP algorithm. For the lack of space, we will only present open-loop simulation results around the nominal operating point with constant chiller´s condenser pressure.

Figure 2 presents some system responses to step changes in the main manipulated variables. Both EEV opening and external air flowrate were perturbed by +/- 5% of their nominal point value. Bypass fraction was changed from 10% at normal operating point to 15% and 5% in all zones. Figure 2 includes simulations obtained by solving the model presented here and the model to evaluate compressor work by Ma et al. (2009). For the latter, the coefficients for the quadratic empirical model relating the real compressor shaft work with the compressor work obtained from what the authors called the fictitious vapour compressor cycle, were obtained by regression to six different operating points. We can see that some responses exhibit a higher order dynamic behaviour with a very fast response. Suction temperatures varied over 10% of their nominal value, somewhat affecting the compressor shaft work. Suction temperature must be controlled at least within a specified range so that no liquid enters the compressor and compressor blades be not damaged. From figure 2c we can see that bypass fraction does not significantly affect compressor shaft work, but as can be seen from figure 2d, external flowrate will somewhat affect it. When analyzing the results with the model of Ma et al. (2009), acceptable responses are obtained only for perturbation in EEV opening. Wrong directions are obtained by increasing bypass fraction or air flowrate. This probably occurred because their model does not take into account variations in the thermal load of the condenser. The difference in CPU time using the current model and the model by Ma et al. (2009) is very small (only 0,6 s for a time horizon of 600 s in an Intel T5500 Core 2 Duo 1.66GHz processor).

Figure 3a shows system response to simultaneous perturbations in the bypass ratio and in the thermal load (-5%) for all zones. It can be seen that the temperature of the zones exhibits a non-linear behaviour. The zone temperature can be controlled at the desired value. Noteworthy is also the large variation in time response of zone temperature dependent on the water bypass ratio value. In figure 3b the thermal load was perturbed in only four zones, respectively by -5%, + 3%, -10% and +3%. The temperatures of these zones were able to be controlled by manipulating the bypass ratio. The increase of 3% in the thermal load saturated the manipulated variable. We can also see that although the thermal load of the 5th zone was not varied, there was the need of manipulating the bypass ratio in order to control the temperature. All temperature were controlled within a range of +/- 0,1ºC.

Figure 4 shows how the external air state may affect both the temperatures of the zones and the compressor shaft work. Relative humidity was considered 50% and the air dry bulb temperature was varied to 31ºC and to 33ºC. Similar types of dynamic responses as in the previous cases are observed. It is noteworthy to say that the shaft work does not significantly vary. Variations were below 1% for the relatively small perturbations performed.

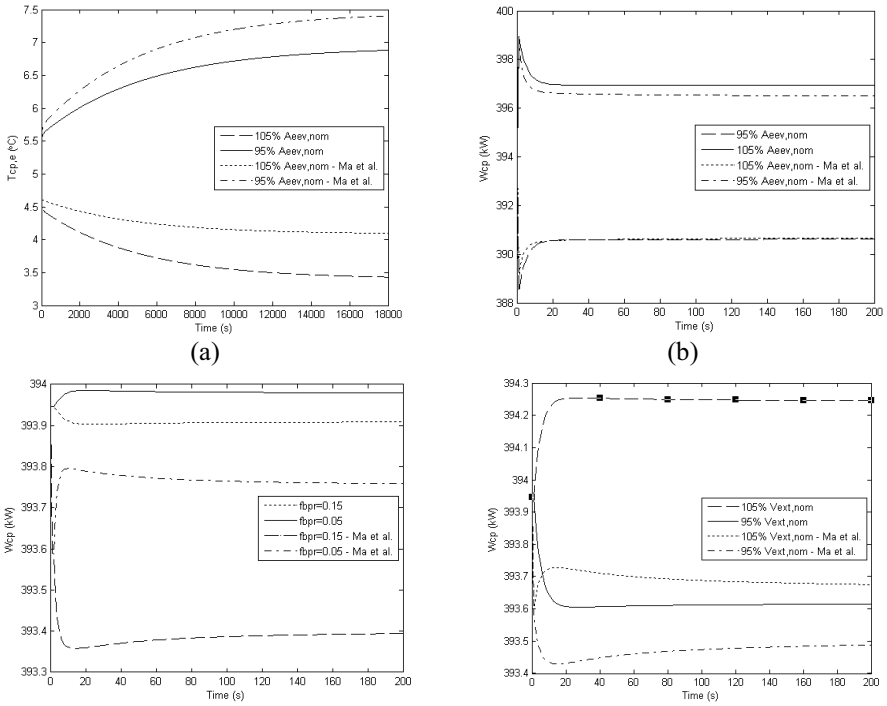


Figure 2: Responses to step changes in manipulated variables in open-loop: (a) EEV effect on compressor suction temperature (b) EEV effect on compressor shaft power (c) Changes in water bypass fraction effect (d) External air flowrate into condenser effect

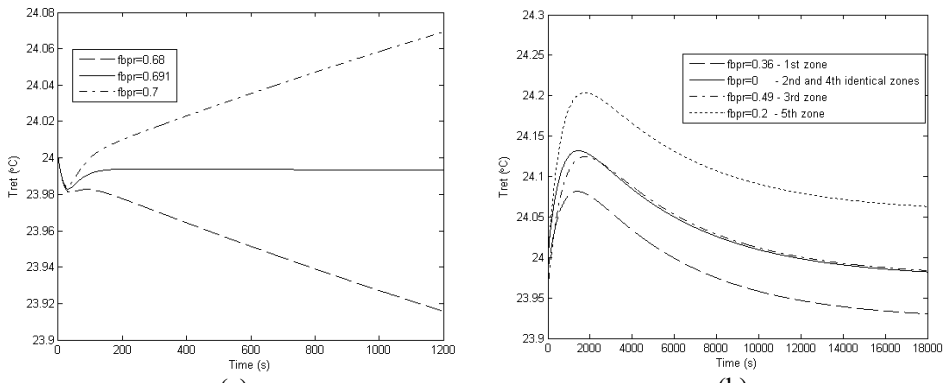


Figure 3: Response to perturbation in thermal load and bypass ratio

The responses show that the system exhibits some non-linear response and that there are different dynamic behaviours with too different time constants. There are also degrees of freedom to control zone temperatures and at the same time minimize compressor shaft work while maintaining important process variables within required ranges. So, it seems that system performance would benefit from advanced control strategies and that the model presented here can be used to test more advanced control strategies.

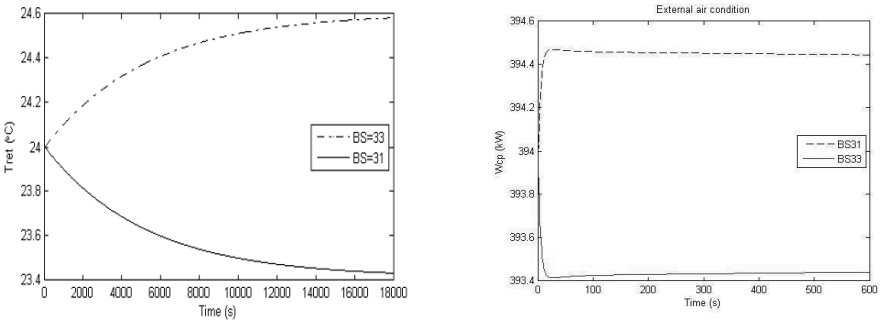


Figure 4: Effect of external air temperature

4. Conclusions

In this paper a model of an air-water HVAC system operating with an air condenser and using water bypass to control the temperature of the conditioned zones was presented. The modelling approach is based on rigorous thermodynamic modelling and explicitly relates the effect of disturbances like external ambient condition and thermal load to the conditioned zones. The thermal dynamic behaviour of the HVAC was presented and the dynamic response of the system could be analysed. The system exhibited a non-linear behaviour on some variables, large variations in time constants and different types of dynamic responses. The model seems adequate to test advanced control strategies.

References

A. Afram, F. Janabi-Sharifi, Review of modeling methods for HVAC systems, Applied Thermal Engineering, v. 67, n. 1-2, p. 507-519, 2014.

ASHRAE.; Handbook of Fundamentals. ASHRAE, Inc., 2017.

Y.A. Çengel, M.A. Boles Thermodynamics: na engineering approach, 7th edition. McGraw-Hill, 2011.

Dupont. Thermodynamic Properties of DuPont Suva 410A Refrigerant. Dupont Technical Information, 2004.

W. Li, Simplified modeling analysis of mass flow characteristics in electronic expansion valve. Appl. Thermal Engineering, v. 53, p. 8-12, 2013

Z. Ma, S. Wang, F. Xiao, Online performance evaluation of alternative control strategies for building cooling water systems prior to in situ implementation. Applied Energy, v. 86, p. 712-721, 2009

D.B. Özyurt, R.W. Pike, Theory and practice of simultaneous data reconciliation and gross error detection for chemical processes, Comp. and Chem. Engineering, v. 28, p. 381-402, 2004

D. Roskosch, V. Venzik, B. Atakan, Thermodynamic model for reciprocating compressors with the focus on fluid dependent efficiencies, Int. J. of Refrigeration., v. 84, p. 104-116, 2017

V. Vakiloroaya, B. Samali, A. Fakhar, K. Pishghadam, A review of different strategies for HVAC energy saving. Energy Conversion and Management, v. 77, p. 738-754, 2014.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50323-2

Economic Optimization of Integrated Nutrient and Energy Recovery Treatment Trains Using a New Model Library

Céline Vaneeckhaute[a,b*], Enrico Ulisse Remigi[c], Evangelina Belia[d], Erik Meers[e], Filip M.G. Tack[e], Peter A. Vanrolleghem[f]

[a] *BioEngine, Université Laval, 1065 avenue de la Médecine, Québec, QC, G1V 0A6, Canada*
[b] *CentrEau, Université Laval, 1065 avenue de la Médecine, Québec, QC, G1V 0A6, Canada*
[c] *MIKE by DHI Software for water environments, Guldensporenpark 104, 9820 Merelbeke, Belgium*
[d] *Primodal Inc., 145 Rue Aberdeen, Québec, QC, G1R 2C9, Canada*
[e] *EcoChem, Ghent University, Coupure Links 653, 9000 Ghent, Belgium*
[f] *modelEAU, Université Laval, 1065 avenue de la Médecine, Québec, QC, G1V 0A6, Canada*
celine.vaneeckhaute@gch.ulaval.ca

Abstract

In order to hasten the implementation of optimal, cost-effective, and sustainable treatment trains for resource recovery, a nutrient recovery model (NRM) library has been developed and validated at steady state. The reported research aims to use the NRM library to establish the operational settings of a sustainable and cost-effective treatment scenario with maximal resource recovery and minimal energy and chemical requirements. Under the optimized conditions and assumptions made, potential financial benefits for a large-scale anaerobic digestion and nutrient recovery project were estimated at 2.8-6.5 USD m^{-3} manure based on net variable cost calculations, or an average of ± 2 USD m^{-3} y^{-1}, equivalent with 40 USD t^{-1} total solids y^{-1}, over 20 years in the best case when also taking into account capital costs. Hence, it is likely that in practice a full-scale ZeroCostWRRF (water resource recovery facility at zero cost) can be constructed.

Keywords: anaerobic digestion; circular economy; mathematical modelling; process optimization; resource recovery

1. Introduction

To hasten the implementation and integration of sustainable nutrient recovery strategies and to adequately put together an optimal treatment train of unit processes for resource recovery, a generic nutrient recovery model (NRM) library has recently been developed and validated at steady state (Vaneeckhaute et al., 2018). The proposed models are dynamic mathematical models, based on detailed solution speciation and reaction kinetics. Key unit process models were developed for anaerobic digestion (NRM-AD), phosphorus precipitation/crystallization (NRM-Prec), nitrogen stripping (NRM-Strip) and absorption using an acidic air scrubber (NRM-Scrub). In view of simulating

complete treatment trains for nutrient and energy recovery, also ancillary unit process models for solid-liquid separation (NRM-Settle), chemical dosing (NRM-Chem) and a heating unit (NRM-Heat) were built. To facilitate numerical solution, a highly efficient interface between the geochemical modelling software PHREEQC and the water quality modelling software WEST (DHI) was established and verified. Global sensitivity analyses (GSA) were performed in order to define the most important factors impacting a wide range of 25 performance indicators of a nutrient and energy recovery treatment train, such as methane and biogas production, digestate composition and pH, ammonium sulfate recovery, struvite production, product particle size and density, and air and chemical (acid, base) requirements (Vaneeckhaute et al., 2015).

Model simulation outputs were very sensitive to input waste stream characteristics through their direct effect on pH, which is adequately determined by means of the chemical speciation calculation integrated in the process models. Moreover, important generic insights in the interactions between process inputs and outputs were obtained through GSA (Vaneeckhaute et al., 2015). Based on the results, it was possible to define an optimal sequence of unit processes in a treatment train for energy and nutrient recovery aiming at the production of high-quality fertilizers at minimal cost (Figure 1).

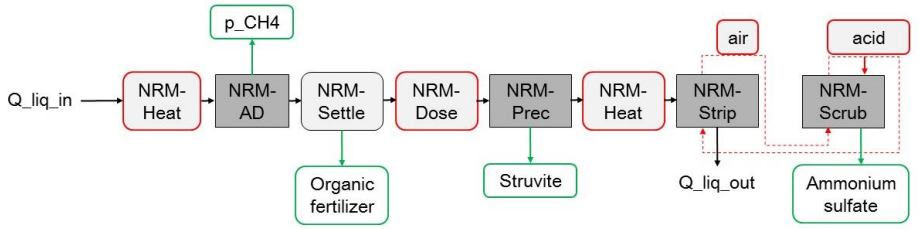


Figure 1 Optimal treatment train configuration targeting bio-based struvite and ammonium sulfate fertilizer production; Red = consumable (= cost); Green = recovered resource (= revenue); AD = anaerobic digestion; Dose = chemical dosing; Heat = heat exchanger; Prec = precipitation/crystallization; p = partial pressure in the biogas; Q_liq = liquid flow rate; Scrub = scrubber; Strip = stripper.

This paper aims to present the use of the NRM library to establish the operational settings of a sustainable and cost-effective treatment scenario with maximal resource recovery and minimal energy and chemical requirements. To this end, an economic analysis was programmed in the process model library, and the operational settings of the above treatment train (Figure 1) were optimized for pig manure as a case study.

2. Methodology

First, realistic design parameters for the unit processes in the nutrient recovery treatment train (Figure 1) were obtained by distributing a technical questionnaire to key technology suppliers in the field. A cost estimate for a design flow of 2,000 m^3 d^{-1} as input to the anaerobic digester was requested using input ranges for nitrogen (N), phosphorus (P), chemical oxygen demand (COD), volatile suspended solids (VSS), total solids (TS), and alkalinity from Cesur and Albertson (2005). The resulting digestate composition (Cesur and Albertson, 2005) was used as input to the nutrient recovery units. Based on the data obtained from the budget proposals, the operational envelope for optimization was compiled. It includes: i) the operational temperature, liquid flow rate, and amount of base/alkalinity dosing for the anaerobic digester, ii) the fraction of

non-settleable precipitates and particulate COD for the phase separation unit, iii) the amount of base dosing, the concentration of seed material in the input flow, and precipitate extraction rate for the precipitation unit, iv) the operational temperature and gas flow rate for the stripping unit, and v) the acid dose and liquid recycle flow rate for the scrubbing unit. The initial values for the optimization experiment were set at the design values provided by the technology providers. The lower and upper limits were set at the values for the unit process GSAs defined in Vaneeckhaute (2015).

The key performance indicators evaluated in the optimization experiment were:

i. **Net costs = chemical costs + energy costs – revenues** (objective = minimize), where:
   a. energy cost items are related to raising the liquid temperature for anaerobic digestion and stripping (with potential for heat exchange, see below), as well as to air pumping for stripping;
   b. chemical cost items refer to the addition of alkalinity or base to the digester, of acid for N absorption in the scrubber, and of base for pH-increase prior to precipitation and stripping;
   c. revenues are related to CH_4 production (energy recovery was assumed, see below), the marketing of mineral fertilizer N, P, and potassium (K), and the potential marketing of organic fertilizer.

ii. **Resource recovery** (objective = maximize), which includes:
   a. methane recovery in NRM-AD;
   b. mineral N, P, and K recovery in NRM-Prec;
   c. mineral N and sulfur (S) recovery in NRM-Strip/NRM-Scrub;
   d. organic (+ N/P/K) fertilizer recovery (settled solids) in NRM-Settle.

iii. **Use of consumables** (objective = minimize), involving:
   a. net thermal energy use = heat required for stripping + heat required for digestion – heat recovered from CH_4 production – potential heat recovered in heat exchangers (see below);
   b. net electricity use = blower energy (air) – electricity recovered from CH_4 production;
   c. chemical use = acid use + base/alkalinity use.

Biogas CH_4 was assumed to be valorized as energy in a combined heat and power generation (CHP) unit, with a conversion efficiency of 40 % as heat, 38 % as electricity, and assuming 22 % heat losses. In terms of heat requirements, both a worst and best-case scenario was considered. In the best case, 10 % heat losses in the digester and 50 % internal heat recovery in the stripping system were assumed as indicated by technology providers. In the worst case, the heat requirements in the digester were 1.9 times higher than the theoretical heat required to heat the input flow. In that case, no internal heat recovery in the stripping system was considered.

To perform the calculations, the GN_Direct algorithm, i.e. DIviding RECTangles algorithm for global optimization (Gablonsky and Kelley, 2001), available from the NLOpt solver package in WEST, was used with a tolerance of 10^{-8} and a maximum of 10,000 evaluations. The latter was identified as the most convenient optimization algorithm/strategy for this type of complex problems following virtual testing and evaluation of various global optimization methods. As compared to other methods included in WEST, it allowed to reduce computation times significantly while ensuring

convergence to the global optimum. The operational envelope of the treatment train was optimized, after which a detailed economic analysis for the treatment train with optimized operational settings was performed, including all operational costs, labor, material and maintenance costs, revenues from CO_2-emission reduction credits, and capital costs.

3. Results and Discussion

The optimized values of the operational factors considered in the optimization experiment are compiled in Table 1. Key performance indicators that were calculated using the optimized factors are also provided. An overview of the annual treatment train operational costs and revenues under the optimized conditions, as well as the capital costs for each unit process in the treatment train is presented in Table 2. The estimation is based on an operational basis of 8,000 hours per year. All costs are expressed in USD.

As depreciation costs and loan service costs vary depending on when and where the money is borrowed, stakeholders are interested in the **yearly net cash flows** determined by the variable costs and revenues. Based on the optimized values obtained and all assumptions made in this case study, the yearly net variable cost balance can be positive. Financial benefits could even be obtained, estimated at about 2.8-6.5 \$ m^{-3} manure y^{-1} (55-130 \$ t^{-1} TS^{-1} y^{-1}) for the large-scale project and associated assumptions in this case. Hence, in terms of net variable cash flows, it is likely that in practice a ZeroCostWRRF (water resource recovery facility at zero cost) could be achieved. As one could be critical on the optimized (low) digester temperature and hydraulic residence time (HRT) obtained in this study (Table 1), the economic analysis was also performed for a digester operated at a temperature of 50 °C with a HRT of 15 d. The financial benefits in this scenario amounted to 2-6 \$ m^{-3} manure y^{-1}, which is competitive with the above optimal scenario. Hence, if a high-temperature treatment is required for end-product pasteurisation, the latter scenario may be targeted, though it is less sustainable in terms of consumables (heat and chemical use). At a HRT of 30 days, the financial benefits amounted to about 3 \$ m^{-3} y^{-1} in the best case, but a loss of 1.5 \$ m^{-3} y^{-1} was obtained in the worst case. The most important factor impacting the operational cost balance, next to the HRT, is the potential for heat recovery. Hence, process and design engineers should focus on the optimization of heat balances in the configuration of future integrated nutrient and energy recovery facilities.

Furthermore, when considering capital costs, stakeholders may be interested in the **net present value** (NPV), which is the sum of the present values of incoming and outgoing cash flows over a period of time, including the investment cost at time 0. Assuming an average discount rate of 6 % and a depreciation period of 20 years for all unit processes, except for the stripping unit, for which a depreciation period of eight years was assumed as advised by technology providers, the nutrient recovery project presented above would have a positive NPV in year 7 of operation in the best case. This value is at the lower end of the range of payback times for existing anaerobic digestion plants without a nutrient recovery treatment train in the US, i.e. 6.9-8.9 years based on a survey of 24 plants (Vik, 2003). The NPV after 20 years amounted to about 3.5 M \$, resulting in average net financial benefits of ± 2 \$ m^{-3} manure y^{-1} (40 \$ t^{-1} TS y^{-1}) over 20 years.

Table 1 Value of the optimized factors in the treatment train optimization experiment and of the resulting performance indicators; Underlined values impact costs, while italicized values impact revenues; COD = chemical oxygen demand; f_ns_P = fraction of non-settleable precipitates; f_ns_X = fraction of non-settleable biological particulate solids.

Unit process	OPTIMIZATION Optimized factor	Value	PERFORMANCE Indicator	Value
Anaerobic digester	Temperature (°C)	28	Heat input (best/worst case; MWh_{th} d^{-1})	24-41
	Flow rate (m^3 d^{-1})	2,700	HRT (d)	15
	Ca-dose (kg d^{-1})	0	*COD degradation (%)*	55
			VSS degradation (%)	45
			CH_4 production (m^3 m^{-3} manure)	5.8-7.4
			Heat recovery (MWh_{th} d^{-1})	72
			Electricity recovery (MWh_{el} d^{-1})	68
Phase separation	f_ns_P	0.25	*Organic fertilizer production (ton X_COD d^{-1})*	15
	f_ns_X	0.05		
Precipitation unit	$Mg(OH)_2$ dose (ton d^{-1})	1.5	*Mineral fertilizer production (ton P d^{-1})*	1.5
	Seeding Kstruvite (g m^{-3})	3.1	*P recovery (%)*	99
	Seeding Struvite (g m^{-3})	3.1		
	Precipitate flow rate (m^3 d^{-1})	150		
Stripper	Temperature (°C)	55	Heat input (best/worst case; MWh_{th} d^{-1})	42-85
	Gas flow rate (Mm^3 d^{-1})	1.5	Electricity input (MWh_{el} d^{-1})	2.9
	Gas pressure (atm)	4		
Scrubber	Acid flow rate (m^3 d^{-1})	17.5	*Mineral fertilizer production (ton N d^{-1})*	5.0
	Liquid recycle rate (m^3 d^{-1})	2.5	*NH_4-N recovery (%)*	84

Table 2 Costs and revenues (k$ y^{-1}) for the optimized nutrient recovery treatment train; $ = USD; AD = anaerobic digester; CHP = conventional heat and power; CAPEX = capital expenditures; OPEX = operational expenditures.

COSTS (k$ y^{-1})	FIXED COSTS	VARIABLE COSTS					REVENUES RESOURCE RECOVERY		
		OPEX							
	CAPEX	Heat (best)	Heat (worst)	Electricity	Chemicals	Maintenance, material & labor	Biogas + fertilizer best	Biogas + fertilizer worst	CO_2 credits
AD + CHP	22,500	694	1,198	621	-	977	3,547	3,547	1,334
Phase separation	1,250	-	-	2.5	to be evaluated	226	1,741	0	-
Precipitation	4,750	-	-	6.3	102	48	1,468	1,468	-
Strip/Scrub	680	1,034	2,069	74	913	6.8	2,365	2,365	-
Other	2,000	-	-	-	-	-	-	-	-
Rounded total	**31,000**	**1,750**	**3,250**	**700**	**1,000**	**1,250**	**9,100**	**7,400**	**1,350**

The **internal rate of return** (IRR), i.e. the discount rate that makes the NPV equal to zero, after 20 years in this case was 18 %, which approximates the estimated best-case IRR (including subsidies) after 20 years for an operational full-scale WRRF in the Netherlands, i.e. 19-21 % (Gebrezgabher et al., 2010). In the worst-case scenario, the IRR after 20 years was only 5 %. Hence, based on the analysis (worst vs. best case), it can be stated that the feasibility of implementing a resource recovery project will highly depend on the heat recovery potential, the marketing potential of the fertilizers, as well as the subsidies obtained. For instance, when accounting for an income of 40 $ t^{-1} net saved CO_2-equivalents (= current global market price of carbon; LLC, 2012) instead of the conservative US carbon prices (15 $ t^{-1} CO_2-equivalents), the IRR would be around 26 % and 14 % in the best and worst case, respectively, resulting in a revenue of 1.3-3.4 $ m^{-3} manure y^{-1} (25-70 $ t^{-1} TS y^{-1}) averaged over 20 years.

4. Conclusions

The potential of the NRM library for optimization of the operational settings of a selected nutrient and energy recovery treatment train was presented by means of a case study for pig manure. An economic analysis indicated that in the best-case scenario a ZeroCostWRRF can be constructed. The NPV after 20 years amounted to about 3.5 M

USD, resulting in average net financial benefits of ± 2 USD m^{-3} manure y^{-1} or 40 USD t^{-1} total solids y^{-1} over 20 years. The IRR after 20 years was 18 %. Results indicate that subsidies, fertilizer marketing potential and heat balances are key factors determining the feasibility of resource recovery projects. Hence, process and design engineers should focus on the optimization of heat balances in the configuration of future integrated nutrient and energy recovery facilities. Fertilizer regulations and subsidies should be adjusted accordingly.

References

D. Cesur, M.L. Albertson, 2005. Modification of anaerobic digestion model no. 1 for accumulation and biomass recycling, Proceedings of the AGU Hydrology Days, Colorado State University, Fort Collins, Colorado, USA.

J.M. Gablonsky, C.T. Kelley, 2001. A locally-biased form of the DIRECT algorithm. Journal of Global Optimization, 21, 27-37.

S.A. Gebrezgabher, M.P.M. Meuwissen, B.A.M. Prins, A.G.J.M. Oude Lansink, 2010. Economic analysis of anaerobic digestion - A case of Green power biogas plant in The Netherlands, NJAS – Wageningen Journal of Life Science, 57(2), 109-115.

LLC, 2012. An overview of methane digesters: Costs, benefits and the potential for revenue in a carbon economy, the Clark Group, LLC, Washington, DC, USA.

C. Vaneeckhaute, F.H.A. Claeys, F.M.G. Tack, E. Meers, E. Belia, P.A. Vanrolleghem, 2018. Development, implementation, and validation of a generic nutrient recovery model (NRM) library, Environmental Modelling and Software, 99, 170-209.

C. Vaneeckhaute, F. Claeys, E. Belia, F.M.G. Tack, E. Meers, P.A. Vanrolleghem, 2015. Nutrient recovery treatment train set-up using a new model library and global sensitivity analyses, Proceedings of the 9th IWA Symposium on Systems Analysis and Integrated Assessment (WATERMATEX2015), International Water Association (IWA), Gold Coast, Australia.

C. Vaneeckhaute, 2015. Nutrient recovery from bio-digestion waste: From field experimentation to model-based optimization, Joint Ph.D. Dissertation, Ghent University, Ghent, Belgium, and Université Laval, Québec, QC, Canada.

T.E. Vik, 2003. Anaerobic digester methane to energy: A statewide assessment, DEE McMahon Associates Inc., Neenah, Wisconsin, USA.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50324-4

Operation of Intermittent Water Distribution Systems: An Experimental Study

Saravanan Chinnusamy[a], Prasanna Mohandoss[a], Varghese Kurian[a], Sridharakumar Narasimhan[a]* and Shankar Narasimhan[a]

[a]*Department of Chemical Engineering; Indian Institute of Technology Madras; Chennai, India 600036*
sridharkrn@iitm.ac.in

Abstract

With the demand for water increasing rapidly, optimal operation of Water Distribution Networks (WDNs) is necessary to provide consumers with the maximum amount of water possible in an equitable manner. This paper presents the outcomes of an experimental investigation of supply policies implementable on rural WDNs. Tests conducted on a fully automated lab scale network, configured to represent rural WDNs, shows significant variations in supply time with the operational policy followed. Further, a systematic set of experiments are carried out to determine the flow rates in different network configurations and this data is later used to formulate a linear programming problem that identifies an optimal operational policy for the system.

Keywords: Water distribution networks, intermittent water supply, scheduling, linear programming.

1. Introduction

The global demand for water is increasing rapidly and many Water Distribution Networks (WDNs) are unable to cope with increased demands (Vairavamoorthy et al., 2008). Network operators in developing countries, often resort to demand management through intermittent supply as a temporary solution (Vairavamoorthy et al., 2008). In such intermittent systems, it is common for the supply to be deficient as well as inequitable (Bhave and Gupta, 2006). Even in other networks that normally provide adequate supply, infrastructure becomes a bottleneck for a short time while the system recovers from failures (Chandapillai et al., 2012). These scenarios necessitate an optimal operation of WDNs so that the available water is distributed equitably, in the minimum time possible.

Operation of WDNs has been actively studied in the past few decades and in most of these studies, the objective has been the reduction of operational expenditure (Sankar et al., 2015; Mulholland et al., 2013; Shi and You, 2016). Maintaining the quality of water has also been the objective for few studies (Nichita and Oprea, 2007). As the techniques proposed in the above-mentioned works were mostly intended for urban water networks, availability of sufficient water has been an assumption that is common to all of them. Even though millions of people still receive water for only few hours in a day, the available literature on the operation of WDNs in water deficient scenarios is limited. The method for quantifying the fairness in distributing available water put forth by Solgi et al. (2014) and the model predictive control for water deficient networks proposed by Sankar et al. (2015) are noteworthy efforts in this regard.

In this work, we propose a technique for operating a class of networks with intermittent supply, commonly referred to as Regional Rural Water Supply Systems (RRWSSs). In these systems, one or two sources provide the water required for a group of beneficiary villages with the help of pumps or by gravity. The supply is intermittent and during the supply hours and the objective for the operational policy is to deliver as much water as possible during the day equitably. The control elements available for this are the valves and pumps in the system. Though there have been few recent works that deals with similar networks (Amrutur et al., 2016), the authors have assumed the system to be equipped with continuous control valves. As a feedback loop with flow measurements is necessary for using control valves, they are difficult to implement for rural WDNs and therefore, in this work we limit our scope to networks with only ON/OFF valves. The other important aspect that distinguishes this paper is that it presents the results of an experimental study where experimental data is used instead of a calibrated model. The configuration of the network, experiments carried out on it and results obtained are given in the following sections.

2. Problem description

A typical rural water supply scheme serves several communities or villages in a region. Here, water is stored in a main balancing reservoir and supplied under gravity to village or local tanks using a piped network. The inlet to each tank is fitted with a valve that is either fully open or fully closed and hence behaves like an ON/OFF valve. The basic operation problem of such a WDN is to supply the desired amount of water to each community or village. There are several reasons for poor performance of the network which can be attributed to issues related to design and operation, e.g., mismatch between actual and forecast demand, poor design practices, aged pipes. Even in well designed, relatively new networks, villages at higher elevations or located far away from the source receive lower supplies than those situated near the source or in low-lying areas. The end result is poor performance as defined in terms of limited availability of water, reduced hours of operation, uncertainty and inequity in supply.

The problem we address is the maximal distribution of water to the beneficiary villages maintaining equity. A simple policy is to fill each tank one at a time. An alternative policy is to open all valves simultaneously and close the valves leading to a particular tank when the demand is met. However, these heuristics may not be feasible or guarantee equitable supply. We propose to solve this problem by formulating an optimization problem (P) where the objective is to meet the cumulative demands while minimizing the time required by manipulating the ON/OFF valves in the inlet of the respective tanks. The decision variables are the time span between event points in the day (τ) and the state of the ON/OFF valves (x) during this time. Additionally, the variables Q and h indicate the flow and head in pipes and junctions respectively. Sets E_{pipe}, E_{valve} N_T, N_S and N_I represent the pipes, valves, beneficiary villages (demand points), source nodes and intermediate nodes in the network. I_{max} denotes the number of intervals the available time is divided into. Other parameters used here are the length of pipes (z), diameters (d), Hazen William's coefficients (ϕ) and demand for water at the villages (D). In this Mixed Integer Non-linear Program (MINLP), constraint equations (2) & (3) model the flow across pipes and valves respectively, (4) specifies the head at the source and villages, (5) represents mass balance at intermediate nodes and (6) ensures the demand satisfied at all downstream villages.

$$(\mathrm{P}) \min_{\tau,x,Q,h} \sum_i \tau_i \tag{1}$$

$$\text{s.t.} \quad h_{i,l} - h_{i,m} = sgn(Q_{i,(l,m)}) \frac{10.67\ |Q_{\{i,(l,m)\}}|^{1.85} z_{(l,m)}}{\phi_{(l,m)}^{1.85} d_{(l,m)}^{4.87}} \qquad (l,m) \in E_{pipe},\ 1 \le i \le I_{max} \tag{2}$$

$$\left(h_{i,l} - h_{i,m}\right) x_{i,(l,m)} = 0 \qquad (l,m) \in E_{valve}, 1 \le i \le I_{max} \tag{3}$$

$$h_{i,j} = h_{i,j0} \qquad j \in N_T \cup N_S, 1 \le i \le I_{max} \tag{4}$$

$$\sum_{(l,m)\in E} Q_{i,(l,m)} = 0 \qquad m \in N_I, 1 \le i \le I_{max}, E \in E_{pipe} \cup E_{valve} \tag{5}$$

$$\sum_i Q_{i,(l,m)} \tau_i = D_m \qquad j \in N_T \tag{6}$$

Upon solving the problem P, if the minimum time is less than the maximum time of operation (e.g., 24 hours if daily demand is specified), the solution is feasible. However, if the minimum required time is greater than the available time, the solution is infeasible. In such a case, we reduce the supply time for every community by the same factor and such a policy would distribute the maximum amount of water equitably.

The above problem is a computationally challenging MINLP and requires a well calibrated hydraulic model to be solved. We address both these issues and present an efficient technique for solving the same. To this end, we propose to decouple the optimization from the hydraulic modeling completely. Each of the valve configurations results in a unique set of flow rates in the network. Given n tanks in the network, each inlet valve can be either in the ON or OFF state and hence the number of states is $N=2^n$. Hence, the network state space is finite and the flow rates in the pipes and the tank inlets are uniquely determined by the state of the valves. The problem would determine the time for which the system had to be operated in each of the above-mentioned valve configurations to meet the demand requirements. The objective is to minimize the time required for supplying the water. The decision variables t_p are the time intervals for which valve configuration p had to be active. $q_{j,p}$ denotes the flow rate into tank j in valve configuration p.

$$(P1) \min_{t\ge 0} \sum_{p=1}^{N} t_p \qquad s.t \quad \sum_{p=1}^{N} q_{j,p}\, t_p = D_j\ , 1 \le j \le N_T \tag{7}$$

The flow rates q_{jp} can be determined using a hydraulic solver. However, if the number of tanks is small, the flow rates can be determined experimentally by recording flow rates into the tank in all the N valve configurations. Thus, the decomposition approach allows us to solve a Linear Program ($P1$) to be solved instead of a challenging MINLP and obviating the need for a well calibrated hydraulic model.

3. Experimental setup

This study was conducted on a lab scale WDN indicating a RRWSS with a single source supplying water to five downstream villages. A schematic of the setup is given in Fig.1. It consisted of one Over Head Tank (OHT) representing the source and five Small Tanks (STs - numbered T1 to T5) representing the storage at the villages. The OHT was made of PVC with 75 cm height and 45 cm diameter and had a capacity of 100 litres. It

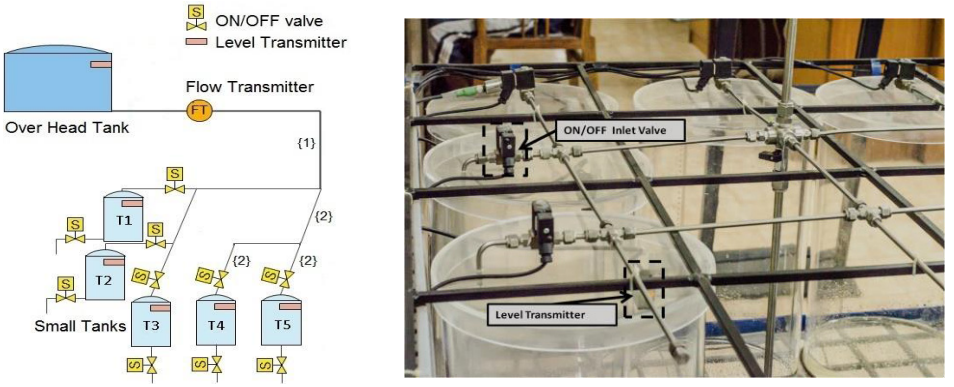


Figure 1 a) Schematic of network b) Lab scale water distribution network

was placed at an elevation of 4 m from the ground. The STs were made of acrylic glass with 50 cm height, 30 cm diameter and 25 litres capacity. The main pipe denoted as 1 in Fig.1(a) had 300 cm length and 1.27 cm diameter and smaller pipes denoted as 2 were of 30 cm length and 0.32 cm diameter. All pipes were made of stainless steel and the complete setup was supported on a 4 × 2.5 × 2 m steel structure.

The out flow of the OHT was controlled by a flow control valve connected to its outlet. The flow into the STs were controlled by ON/OFF valves (Burkert 6011, solenoid valve) placed at their inlet and out flow from them were controlled by ON/OFF valves connected to their outlet. All tanks were equipped with ultrasonic level transmitter (Baumer U500) to measure the water level online. Additionally, there was an ultrasonic flow transmitter (Burkert 8081) connected downstream of the control valve to measure the flow inline. All the instruments were interfaced to a computer through National Instruments DAQ card and LabVIEW was used to program and control the devices. The water level in the OHT was maintained at a constant level of 42.5±2.5 cm.

Here, the problem of supplying water to beneficiary villages translates into the problem of supplying water to the STs from the OHT. For this, the water requirements at the STs were decided as given in Table 1. The problem at hand was to satisfy the water requirement of the STs in the minimum time possible. As in the RRWSS discussed by Bhave and Gupta (2006), the control elements available for our experiments were the ON/OFF valves at the inlet of the STs. The outlet of these tanks was kept closed and therefore, the amount of water supplied could easily be inferred from the level measurements. Levels were measured at a rate of 5 Hz and filtered within LabVIEW. The control valve immediate downstream of the OHT was always kept fully open.

4. Operational policies

Two heuristic supply policies were tested for the case. In the first case (HEURISTIC-I), the inlet valves to all the five STs were opened simultaneously. The supply to each tank was stopped as when its respective demand was satisfied. This mode of operation required 2284 s to meet the requirements as shown in Figure 2. A colored cell indicates that the valve is open and white cell indicates that the valve is closed. In the next policy tested (HEURISTIC-II), the inlet to only one tank was opened initially. Once the

requirement for the tank was satisfied, it was closed and the next was opened. Continuing likewise for all tanks, it took 3712 s to provide the water. The actual amount of water supplied following each of the above-mentioned control policies is given in Table 1.

The optimal operation here is the one which meets the requirements of the STs in the minimum time. The setup involved five tanks with ON/OFF valves connected at their inlet, there were a total of 2^5 valve configurations possible. Therefore, 32 different experiments were conducted to determine the flow rates into each tank at different valve configurations. The optimal supply scheme (OPT) obtained by solving $P1$ required 2197 s to supply the water. The Gantt chart depicting this is shown in Fig. 3. The actual amounts of water supplied (last column of Table 1) are almost equal to the demands. The slight deviations between them is due to flow measurement errors and marginal variation in the head within operating range set for OHT. While the heuristic supply schemes have only one downstream tank receiving supply for all or some of its operational time, the optimal scheme provides supply to at least two tanks always. In this way, it makes a better use of the pipes in the network and delivers water in the minimum time.

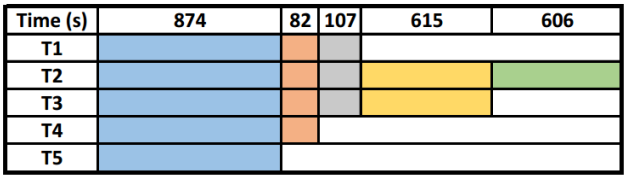


Figure 2: The heuristic supply policy beginning with all tanks open (HEURISTIC-I)

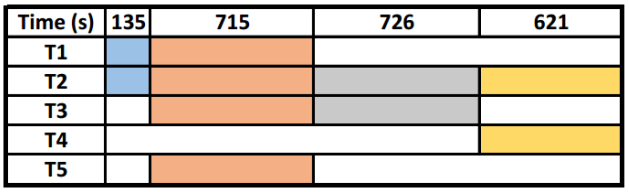


Figure 3: Gantt chart showing optimal supply policy (OPT)

Now if in the real scenario, we have only 1750 s to complete the supply, even the optimal supply policy would not meet the requirement in the given time. In such a case, the active time of every network configuration in the optimal operational policy has to be brought down by a factor of 1750/2197. Such a scheme of operation would reduce the supply to every ST by the same factor i.e.~ 0.8 and ensure equitable operation.

Table 1: Water supplied to the STs

Tank Number	**Demand (l)**	**Supplied (l)**		
		HEURISTIC-I	HEURISTIC-II	OPT
T1	6.40	6.44	6.27	6.37
T2	18.70	18.81	18.58	18.82
T3	11.35	11.39	11.41	11.33
T4	7.48	7.53	7.43	7.50
T5	7.48	7.45	7.41	7.51

5. Conclusions

The paper presented the determination of supply time in WDNs with operational policies with the help of experiments conducted on a lab scale network. The optimal supply scheme required 4 - 69 % less time in comparison to heuristic supply strategies. Also, on implementing the operational policies developed from measured flow rates, the actual water supplied met the demands for all consumers. The results are encouraging for implementing the operational scheme on intermittent WDNs retrofitted with automated ON/OFF valves.

Acknowledgements

This work was partially supported by the Department of Science and Technology, Govt. of India under the Water Technology Initiative (Project. No. DST\TM\WTI\2K13\144) and the IIT Madras Interdisciplinary Laboratory for Data Sciences (CSE\14-15\831\RFTP\BRAV).

References

Amrutur, B., Kumar, M. M., Kumar, K. S., Patel, L., Sundaresan, R., Vaidhiyan, N. K., 2016. Wateropt: a method for checking near-feasibility of continuous water supply. In: Proceedings of the 2nd International Workshop on Cyber-Physical Systems for Smart Water Networks. IEEE, pp. 7–12.

Bhave, P., Gupta, R., 2006. Analysis of water distribution networks. Narosa Publishing House Pvt. Ltd.

Chandapillai, J., Sudheer, K., Saseendran, S., 2012. Design of water distribution network for equitable supply. Water resources management 26 (2), 391–406.

Mulholland, M., Latifi, M., Purdon, A., Buckley, C., Brouckaert, C., 2013. Multi-objective optimisation of the operation of a water distribution network. In: Kraslawski, A., Turunen, I. (Eds.), 23rd European Symposium on Computer Aided Process Engineering. Vol. 32 of Computer Aided Chemical Engineering. Elsevier, pp. 709 – 714.

Nichita, C., Oprea, M., 2007. An agent-based model for water quality control. In: Ple¸su, V., Agachi, P. S. (Eds.), 17th European Symposium on Computer Aided Process Engineering. Vol. 24 of Computer Aided Chemical Engineering. Elsevier, pp. 1217 – 1222.

Sankar, G. S., Kumar, S. M., Narasimhan, S., Narasimhan, S., Bhallamudi, S. M., 2015. Optimal control of water distribution networks with storage facilities. Journal of Process Control 32, 127–137.

Shi, H., You, F., 2016. Energy optimization of water supply system scheduling: Novel MINLP model and efficient global optimization algorithm. AIChE Journal 62 (12), 4277–4296.

Solgi, M., Haddad, O. B., Seifollahi-aghmiuni, S., Loáiciga, H. A., 2014. Intermittent Operation of Water Distribution Networks Considering Equanimity and Justice Principles. Journal of Pipeline Systems Engineering and Practice 6 (4), 1–11.

Vairavamoorthy, K., Gorantiwar, S. D., Pathirana, A., 2008. Managing urban water supplies in developing countries - Climate change and water scarcity scenarios. Physics and Chemistry of the Earth 33 (5), 330–339.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50325-6

Sustainable Industrial Water and Energy Nexus Integration for an Industrial Park

Jamileh Fouladi, Patrick Linke

Texas A&M University at Qatar, Education City, Doha, Qatar
jamileh.fouladi@qatar.tamu.edu

The water-energy nexus and integration has been recently proposed to minimize water-energy footprint of an industrial park. Their sustainable design requires careful consideration of energy and water management aspects. This work represents an integrated management across the water and energy nexus. We propose an optimization-based methodology, which considers main decision across the water-energy nexus for industrial clusters. Building on our previous work in industrial water networks, a representation is generated that accounts for the links between water and energy. In this work, the Tool for the Reduction and Assessment of Chemical and other environmental Impacts (TRACI) metrics have been used in order to quantify some of the environmental emissions. Global warming and eco-toxicity impacts are the ones, which are calculated for the optimized design. Pareto front curves are generated for a case study in order to capture the trade-offs between the minimum total annual cost and environmental sustainability metrics.

Keywords: Water-Energy Nexus, Sustainability Metrics, Optimization.

1. Introduction

Global population is expected to reach 8.5 billion by 2030. As a consequence, the water and energy demands are increasing steadily. Industrial clusters have high energy and water footprints especially in the Gulf Cooperation Council (GCC) region. Therefore, solutions need to be developed for high performance with respect to sustainability dimensions, especially regarding economic and environmental impacts. In this work, waste heat utilization has been added to the previous published superstructure (Alnouri et al, 2014). Waste heat utilization means the use of waste heat that can be converted to power based on the design electricity demand, while the non-reusable heat is discarded to the cooling systems. There are different sustainability metrics and methods available nowadays which can cover these dimensions. Sustainability metrics mainly are divided into three different categories; economical, environmental, and social impacts. Concerns that are addressed by these metrics include economics, environmental, and safety concerns. Process design for sustainability involves complicated decision-making scenarios and this makes it difficult to explore the trade-offs using existing methods. Over the years, researchers have suggested many metrics and indicators but most of the metrics cover only one of the aspects. Accordingly, in this work the objective is to design a systematic approach to explore efficient strategies and design for integrated water and energy management. Moreover, to introduce a methodology for designing sustainable chemical and petroleum processes during early stages of design in order to capture the trade-offs between economic and environmental concerns.
Research contributions mainly focused on reducing the water and energy footprints of existing industrial park with process water-energy integration interactions especially

considering process cooling requirements and cooling systems, desalination systems, and treatment units. The method consider an optimization model to systematize the search for the representation.

2. Problem Statement

In this work an existing water network problem has been expanded from Alnouri et al (2014, 2015) by considering process cooling requirements (Air Coolers (AC), Cooling Towers (CT), Cooling Seawater (CS)), and desalination plants. Moreover, each plan has a surplus waste heat which can be converted to power and used in any energy sinks (Figure 1).

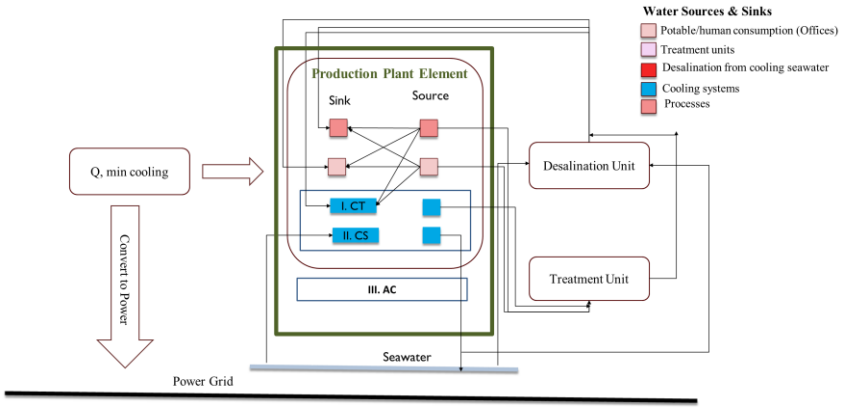


Figure 1. Water-Energy Integration

The main optimization objective is the total annual cost (TAC) minimization (which covers by economic metrics), subjected to some environmental emissions and metrics. In this work, the TRACI approach is used to calculate the environmental emissions and impacts. TRACI is one of the environmental impact assessment software available with different characterization factors to assess life cycle assessment and sustainability metrics. Characterization factors measure the potential impacts. The impact categories include the followings: Ozone Depletion, Global Warming, Acidification, Eutrophication, Smog Formation, Human Health (Particulate, Cancer, and Non-cancer), Eco-toxicity, Fossil Fuel Use, Land Use, and Water Use. The water-energy integration and optimization problems will not change the industrial city design and layout. Therefore, safety and health metrics are not affected by these types of simulations and can their change can be negligible.

3. Implementing Methodology

Figure 2 shows the methodology in order to calculate the sustainability metrics for the proposed water-energy networks. As it is shown in Figure (2), the main objective of the optimization problem is minimizing the total annual cost of the design.

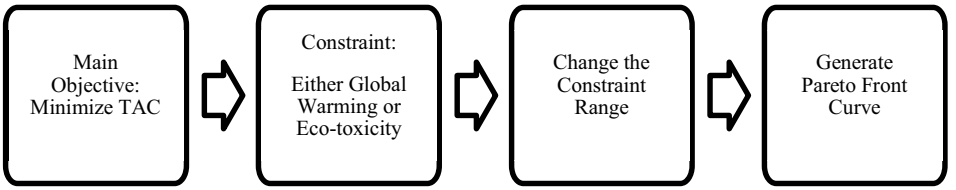


Figure 2. Implemented methodology

Meanwhile, an emission cap is subjected to the objective as a constraint. Once the optimum design with the minimum total annual cost is obtained, the next trial the optimum design for the minimum selected emission will be found. Therefore, the boundaries and range of constraint can be determined by getting the maximum and minimum values for the corresponding emission.

In this work, the environmental waste and emissions for the proposed problem were analyzed as shown in Figure 3. The environmental emissions in this problem based on the components can be divided as solid waste, atmospheric impact, and aquatic impact. However, solid waste do not change significantly since the treatment options are not the main differences in different trials. So, only two categories have been quantified; global warming and eco-toxicity impacts.

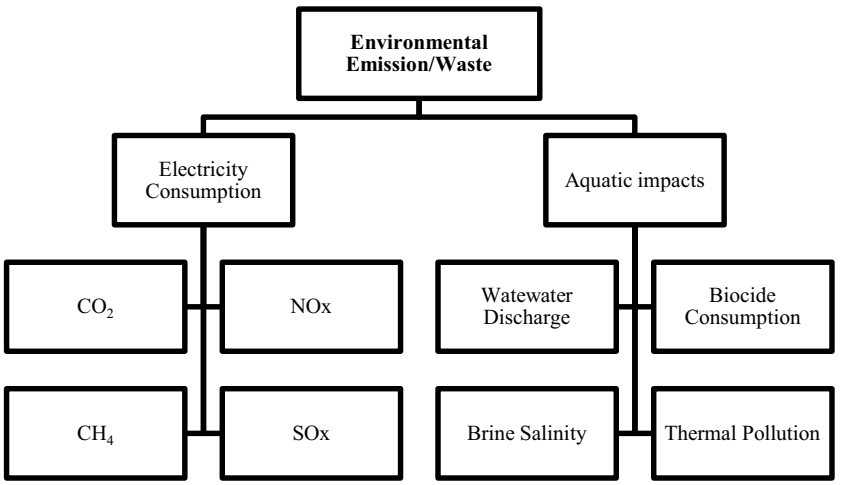


Figure 3. TRAICI environmental emissions corresponding to the defined problem

In order to quantify emission impacts the following equation has been used (EPA, 2012):

$$I^i = \sum_{xm} CF^i_{xm} * M_{xm} \quad (1)$$

Where:

I^i= The potential impact of all chemicals (x) for a specific impact category of concern (i)

CF^i_{xm}= The characterization factor of chemical (x) emitted for impact category (i)

M_{xm}= The mass of chemical (x) emitted

4. Case Study

The proposed method is illustrated with a case study of three plants; Ammonia, Methanol, and Gas-to-liquid (GTL). Water sources and sinks data with selected for four contaminants (TDS, Organics, Sulphate, Oil and Grease) have been collected (Martinez, 2014). It was assumed that there is a single central and decentral treatment as well as central desalination units. The technology used in the problem is Nano-Filtration since it satisfies all the sink contaminants composition.

5. Results and Discussion

Different scenarios were solved to compare the differences between applying energy concept with and without waste heat to power utilization and see the trade-off between economic and sustainability metrics.

5.1. Without Waste Heat Utilization

Different scenarios have been optimized without the waste heat utilization (cooling requirements) in order to get the trades-off between the total annual cost and different categories in sustainability metrics. The total annual cost of the design increases as the environmental emissions (Global Warming (GW)) decreases. Obtained trends show the more sustainable the solution is, the more expensive the design gets. The cooling system design changes from air coolers to cooling seawater from point F to A. The global warming impact reduces in case of cooling seawater as the power consumption is the minimum. Moreover, the water network design changes by changing the cooling systems, which means the amount of reused water, discharged water and other flow rates change. In general the trends show, the total annual cost increases by decreasing the global warming impact, since more treatment is required and therefore more water reuse.

Table 1: Total annual cost versus global warming impact before waste heat utilization

	Scenarios					
	A	**B**	**C**	**D**	**E**	**F**
Cooling System	Both AC	Both AC	AC+CT	AC+CT	CT+CS	CS

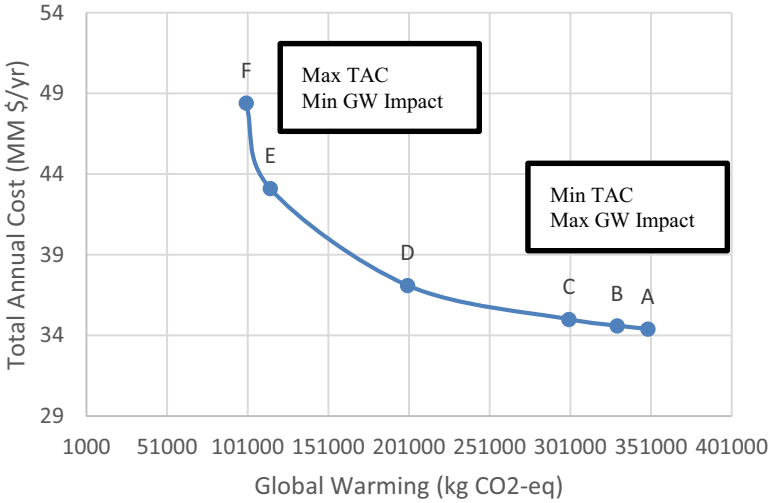


Figure 4. TAC versus GW impact (without waste heat utilization)

For the eco-toxicity influence (Aquatic impact), in case of the optimal solution, which is selecting Air Coolers there is no difference in the impact by doing different trials since the contaminants included in the case study are not in the Aquatic eco-toxicity categories in TRACI metrics and in the Air Cooler design there is no flow for water. If we force the design to select other cooling options, it can be noticed that as the cooling system option changes from the Air Coolers to Once-Through Cooling Seawater the eco-toxicity increases with total annual cost. Once-Through Cooling Seawater has the maximum impact because of the Biocide components (Seawater Chlorination).

5.2. With Waste Heat Utilization

In the next section, the sustainability metrics were quantified and compared for the scenario with waste heat utilization. In this scenario, the global warming effect decreased by almost 90 percent while the total annual cost is reduced by 10 percent. The remaining 10 percent of global warming impact is emission related to the external utility which will not be affected by the problem design. As it is shown in Figure 5 different Pareto front curve is generated after waste heat utilization. As it is clear by waste heat utilization the annual cost decreases a little. However, the change in the emissions impact is significant (around 90%), which means the result is lower cost, lower emission. Figure 4 indicates that the obtained trend shows that in the case of air coolers the maximum benefits is obtained for the global warming impact. It can be justified since the air coolers have the maximum power consumption compared to other cooling system options. The change in the total annual cost is not significant since the power-generated cost from the waste heat is almost the same as the external power cost. By waste heat utilization, the cooling load reduces which reduces the cooling systems cost.

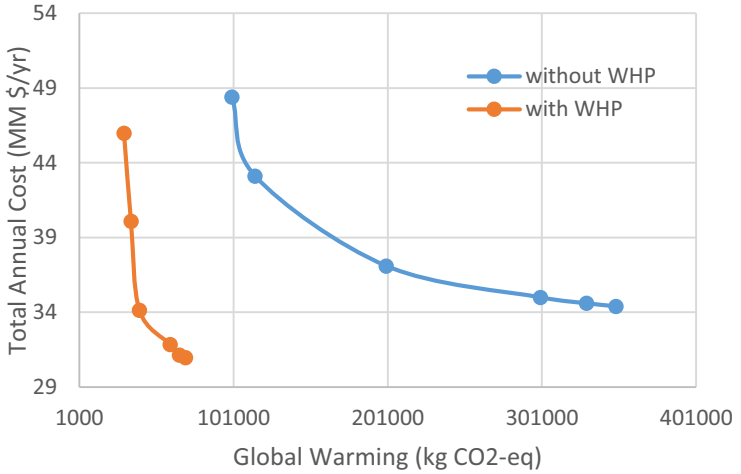


Figure 5. Global warming impact comparison before and after waste heat utilization

6. Conclusions

This work represents a systematic approach for integration of industrial parks across the water-energy nexus. The objective of this paper is to develop a framework for optimizing energy and water resources from processes that have a surplus of energy at various qualities by accounting for different sustainability metrics. Furthermore, a case study was used to illustrate the approach and compare different scenarios for developing sustainable water-energy strategies.

Acknowledgements

This publication was made possible by NPRP grants no. NPRP 7-724-2-269 from the Qatar National Research Fund (a member of Qatar Foundation).

References

D. Y. Martínez, A. Jiménez-Gutiérrez, P. Linke, K.J. Gabriel, M.M. El-Halwagi, (2014). Water and Energy Issues in Gas-to-Liquid Processes: Assessment and Integration of Different Gas-Reforming Alternatives. ACS Sustainable Chemistry & Engineering, 2(2), 216-225. doi:10.1021/sc4002643

Senior Design Report, Ammonia Plant. Qatar University, (2015)

Senior Design Report, Methanol Plant. Texas A & M University at Qatar, (2015)

S.Y. Alnouri, P. Linke, M.M. El-Halwagi, 2014, Water integration in industrial zones: A spatial representation with direct recycle application. Clean Technol. Environ. Policy, 10.1007/ S10098-014-0739-2

S.Y. Alnouri, P. Linke, M.M. El-Halwagi, 2015, J. Clean. Prod. 89, 231e250

S.Y. Alnouri, P. Linke, M.M. El-Halwagi, 2015, J Clean Prod. 89:231–50

United States Environmental Protection Agency (EPA), 2012, Tool for the Reduction and Assessment of Chemical and other Environmental Impacts (TRACI).

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50326-8

Stochastic Optimization Tools for Water-Heat Nexus Problems

Sumit Kr. Bishnu[a,b] , Jamileh Fouladi[b], Patrick Linke[b], Mahmoud El-Halwagi[a]

[a] *The Artie McFerrin Department of Chemical Engineering, Texas A&M University,College Station, Texas-77843, USA*
[b] *Department of Chemical Engineering, Texas A&M University,Doha-23874,Qatar*

Abstract

This aim of this work is to provide a framework for application of Simulated Annealing (SA) Algorithm for solving Water - Heat Nexus problems for Industrial Cities. The combined water-heat network synthesis problem has been formulated into a Mixed Integer Non-Linear Programming (MINLP) and this work describes the various components of the framework needed to handle it. The SA algorithm converges asymptotically across several runs. Apart from the best solution, this tool stores multiple solutions with varying degree of performance. The application of this algorithm has been illustrated with a case study and the benefits of metaheuristics search techniques have been highlighted.

Keywords: Simulated Annealing, Meta- Heuristics, Water - Energy Networks, MINLP.

1. Introduction

Synthesis of efficient water or heat networks in industries is an important aspect of planning for a sustainable and cost effective enterprise, as they are important constituent of all industry. These networks can be more efficient if the nexus between water and energy resources are studied and benefits exploited. Often these problems are presented as MINLP and solved using deterministic solvers, which reportedly struggle with complexities, involved (Gao and You, 2015).The aim of the paper is to apply SA for solving these network problems and highlight their benefits. Though it has been used to solve problems like molecular design [2] and reactor networks [3], it has not been used in solving water-energy network problems and this work will be the first attempt to solve these problems using SA. It would provide a basic framework for executing SA and can later be utilized for solving similar optimization problems with more resources involved.

2. Simulated Annealing

SA is a probability based metaheuristic method that is used for finding optimum values for large solution space problems. It mimics the actual annealing process for determining the optimal solution. For water-energy network problem, an initial solution is provided and then transitions are made on this solution to generate multiple solutions around its neighbourhood for given artificial temperature. This process mimics the part where the system is allowed to attain the maximum number of states and equilibrates the system at each temperature stage. If the performance of the solution is better than the previous solution, it is accepted. If it is worse, it is not rejected straightaway but

rejected with a probability and gives the system a chance to go towards going towards adverse performing regions. Compared to deterministic algorithms that are gradient based and process the full MINLP model simultaneously, the approach utilized to develop the stochastic algorithm implementation handles the equations, inequalities, objective function with different modules. The advantage being its ability to able to handle non-linearities in the problem effectively and since the search mechanism is independent from other components, the tool has the freedom to explore the solution space thoroughly without being stuck.

3. Water – Energy Nexus Problem

The water-heat network problem illustrated in this work includes water network formulation adapted from Alnouri et al [4] with addition of cooling systems for waste heat. Figure 1 shows the combined network and the synergy between them. The problem consists of an industrial infrastructure with sources, sinks, on-site decentral treatment units and a centralized treatment units for water networks. To discharge the waste heat, options of cooling tower, air cooler and once through seawater system have been incorporated. Cooling Tower and once through seawater system require water for their operation. Usage of cooling water incorporates a new sink and source in the water network when they are utilized and hence the nexus. The sinks can be supplied with external supply of freshwater, treatment units and water from sources. Water from sources can be discharged to environment, sent to treatment units or can be recycled back to sinks. There are concentration constraint on sinks and environment. Together with all the information about these facilities and spatial information about the industrial city, a MINLP problem is formulated with an objective of minimizing total cost and finding the optimal configuration of the combined network.

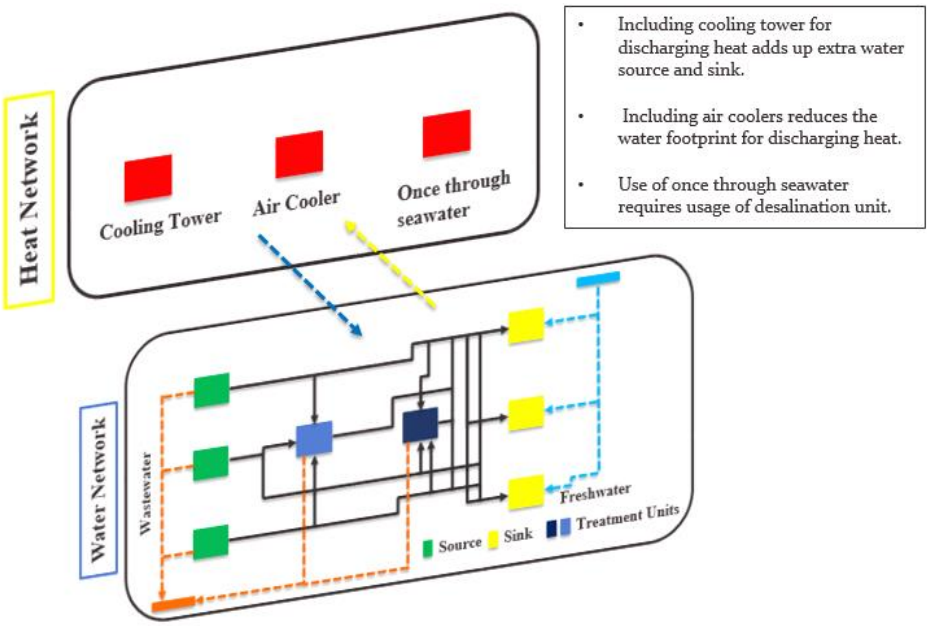


Figure 1: Water – Energy Nexus

4. Implementation

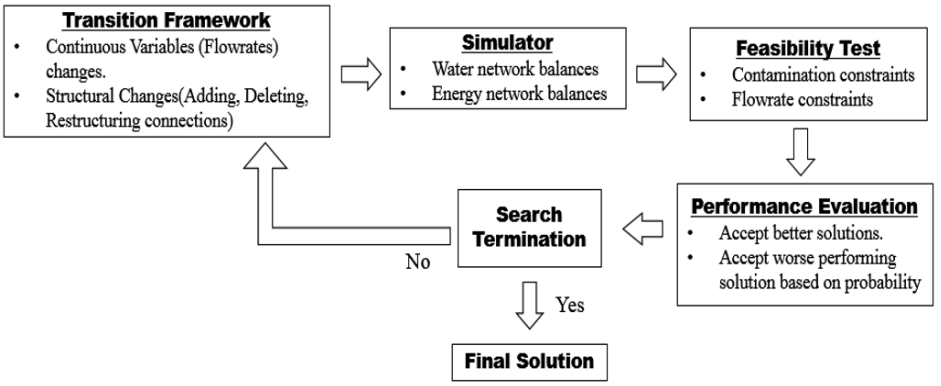


Fig 2: S.A Implementation: Modular Approach

The implementation begins with an initial solution, which in this case is a linear structure in which the requirement of sinks are met by freshwater and heat being disposed from cooling tower. Random changes are made on the given structure using the transition framework to generate new solutions which involves assigning values to different connections across various facilities. As the superstructure of this network has a large number of possible connections, the search becomes complicated exponentially with the problem size. Due to constraint of space, this transition framework will be described in detail in the complete manuscript. This solution is then simulated to calculate the freshwater and wastewater discharges. The simulator consists of mass balance equations at sources, sinks, treatment units and wastewater discharges and heat balances for the three cooling options. These equations are transformed into set of linear equations as many of the variables are varied by the solver in one module. The new solution is simulated to find the freshwater and wastewater discharges and feasibility checks are performed. The solution is accepted if it better than the previous one. If it has lower performance, then it is not rejected outright but has a probability assigned to it according to which it may be accepted or rejected. Traditional MINLP optimization packages often struggle to handle larger size case studies just for single resource problem [5], the tool presented in this paper is able to handle multiple resources in single problem and stores multiple optimal solutions which are similar in performance but different in structure. This gives the user insight into the other solutions and options to implement any of them. In case of traditional package, the user gets only one structure as final answer and no insight into other solution.

5. Case Study

The case study consists of three plants of Ammonia, Methanol and GTL, which discharge four major contaminant, namely TDS, Organics, Ammonia and Nitrogen. Tables 1 & 2 present the source and sink data used for the case study and Table 3 presents the waste heat data from the three different plant.

Table 1: Source Data

	Flowrate(tons/d)	TDS(ppm)	Organics(ppm)	Ammonia(ppm)	Nitrogen(ppm)
P1S1	45	50	4	1	50
P1S2	154	2500	20	2.5	25
P1S3	400	550	15	25	40
P2S1	281	500	100	0.5	5
P2S2	115	2500	20	2.5	25
P2S3	500	550	15	25	40
P3S1	16648	500	46	0.5	5
P3S2	147	550	15	25	40

Table 2: Sink Data

	Flowrate(tons/d)	TDS(ppm)	Organics(ppm)	Ammonia(ppm)	Nitrogen(ppm)
P1D1	2571	500	4	0.5	21
P1D4	840	200	4	0.5	5
P2D1	1912	500	4	0.5	21
P2D4	500	200	4	0.5	5
P3D1	7115	500	4	0.5	21
P3D3	163	200	4	0.5	5

Table 3: Waste Heat Data (MW)

Plant	Cooling Tower
Ammonia	237
Methanol	104.8
GTL	982.67

The objective function of optimization problem is minimization of cost that includes the freshwater cost, capital and operating cost of treatment unit and the piping cost for water network. For the heat network, the cost includes capital and operating cost of cooling tower, air cooler and once through seawater cooling option.

6. Results

SA algorithm has been implemented using MATLAB using a desktop PC with Intel® Core ™ i7-2620M, 2.7 GHz, 16.00 GB RAM, and a 64-bit Operating System. The optimal solution has been obtained after running several runs with different markov chain length with sizes varying from 50 -100 and it takes around 4-6 hours for a run to converge. The lowest cost solution has an objective function value of \$33.1 x 10^6 per year. Apart from the best solution, the program stores all the solutions are better than initial configuration. The result for the least cost structure has been listed in tables 4-6. All the flowrates are in tons/hr.

Table 4: Source to Treatment Unit Flowrates

	P1D1	P1D4	P2D1	P2D4	P3D1	P3D3
P1S1	0	0	0	0	0	0
P1S2	0	0	0	0	0	0
P1S3	0	0	0	0	0	0
P2S1	0	0	0	0	0	0
P2S2	0	0	0	0	0	0
P2S3	0	0	0	0	0	0
P3S1	0	0	0	0	0	0
P3S2	0	0	0	0	0	0

Table 5: Source to Treatment Unit Flowrates

	TR(Ammonia)	TR(Methanol)	TR(GTL)	TR(Central)
P1S1	0	0	0	0
P1S2	0	0	0	6.4167
P1S3	0	0	0	16.67
P2S1	0	0	0	11.708
P2S2	0	0	0	4.792
P2S3	0	0	0	20.833
P3S1	0	0	0	539.977
P3S2	0	0	0	6.125

Table 6: Treatment Unit to Sink Flowrates

	P1D1	P1D4	P2D1	P2D4	P3D1	P3D3
TR(Ammonia)	0	0	0	0	0	0
TR(Methanol)	0	0	0	0	0	0
TR(GTL)	0	0	0	0	0	0
TR(Central)	107.125	35	79.67	20.833	296.45	6.792

Requirement of water at all the sinks is met by treated water from sources. The treatment unit utilized is reverse osmosis. The best solution doesn't require extra fresh water from an external source and there are no connections from sources and sinks.

The waste heat generated from the three plants is directed to Air Coolers to be discharged into the environment. The reason being the other two options required a lot more water for removal of heat which in turn added to the cost.

The problem was solved using various initial points and varying markov chain to test the robustness of the platform. Figure 3 shows various runs converging asymptotically to solutions with similar objective function value thereby establishing the robustness of the solver.

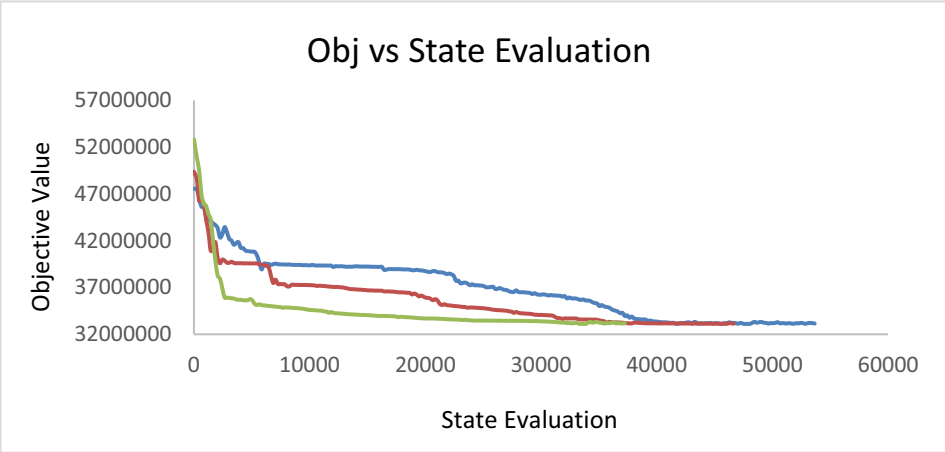


Fig 3: Objective Function vs State Evaluation

7. Conclusions

The solution procedure proposed in this work provides an alternative approach for solving Water Network problem. Various aspect of an optimization problem (equality constraint, inequality constraints and the objective function) has been dealt with separately compared to deterministic methods where they are considered together while solving the problem. The method converges asymptotically to a solution and provides multiple solution options apart from the best solution obtained. The work provides a framework for applying SA to solve multi - resource optimization problems. Future works will target problems like Water-Heat- Carbon Dioxide network optimization.

8. Acknowledgement

This publication was made possible by NPRP grants no. NPRP 7-724-2-269 from the Qatar National Research Fund (a member of Qatar Foundation).

References

[1] Gao. J & You. F (2015). “Optimal Design and Operations of Supply Chain Networks for Water Management in Shale Gas Production: MILFP Model and Algorithms for the Water – Energy Nexus.” AiChe Journal.**61**, 4, 1184 – 1208.

[2] Berhane H. Gebreslassie, Diwaker,U. (2016). "Homogenous multi-agent optimization for process systems engineering problems with a case study of computer aided molecular design." Chemical Engineering Science 159,194-206.

[3] E. Marcoulaki, Linke,P, A. Kokossis (2001). "Design of Separation Trains and Reaction-Separation Networks Using Stochastic Optimization Methods." Chemical Engineering Research and Design 79(1): 25-32.

[4] Alnouri, S.Y, Linke, P, El-Halwagi, M.M(2015). A synthesis approach for industrial city water reuse networks considering central and distributed treatment systems. Journal of Cleaner Production.89, 231-250.

[5] Shi. H & You. F (2016). “Energy Optimization of Water Supply System Scheduling: Novel MINLP Model and Efficient Global Optimization Algorithm.” AiChe Journal. **62**, 12, 4277-4296.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50327-X

Optimization of sugarcane bagasse based industrial cluster for economic and environmental benefits

D. Varshney[a], P. Mandade[a], Yogendra Shastri[a*]

[a]*Department of Chemical Engineering, Indian Institute of Technology Bombay, Mumbai, India - 400076*
[*]*yshastri@iitb.ac.in*

Abstract

This work proposes an optimization model for the design of an industrial cluster based on sugarcane bagasse in the Indian context. Three different products, namely, ethanol, electricity, and pellets, along with the possibility of selling raw bagasse and burning it in boiler are considered. An economic optimization model to maximize the net present value of the cluster is formulated. Simultaneously, a multi-objective optimization model to minimize the environmental impacts is formulated. The model was applied to a case of a sugar mill processing 181 Mg/h sugarcane. Economic optimization recommended electricity production as the optimal option at the current product prices. This was also desirable from the GHG emission reduction standpoint. However, electricity production led to highest water footprint. Ethanol price has to increase by 100% over the target price for economic feasibility. Pelletization was preferred only from a water footprint perspective.

Keywords: Bagasse, Industrial Ecology, Optimization, India, Ethanol

1. Introduction

Sugarcane bagasse is a highly promising source of biomass in the Indian context for the establishment of a bio-based economy. Bagasse can be used to produce a number of value added products such as pellets, electricity, ethanol, paper board, and specialty chemicals such as vanillin. Each of these products varies in terms of costs, technology readiness level, per unit revenue, and market capacity. Moreover, the environmental impacts of these alternatives also vary considerably. Utilization of sugarcane bagasse, therefore, poses a sustainability challenge with food, energy and water nexus. Region specific factors are critical and must be considered in the analysis. Such studies have been reported in literature for some sugar producing countries (Furlan et al, 2012, Ensinas et al, 2013). However, such an analysis has not been conducted in the Indian context. This work addresses this knowledge gap by proposing an industrial ecology approach and optimizing the design of an industrial cluster based on sugarcane bagasse. An existing sugar refinery providing a fixed quantity of bagasse is considered. Three different products of bagasse, namely, pellets, electricity, and ethanol, are considered. An economic optimization model is formulated to understand the profitable products and the sensitivity of those results to product market prices. Simultaneously, the global warming potential (GWP) and water footprint for the same products are calculated to understand the environmental trade-offs.

The article is arranged as follows. The optimization model formulation is given in Section 2. Section 3 discusses the key results from the simulation studies, and the important conclusions are summarized in Section 4.

2. Optimization model formulation

An existing sugar mill with a boiler to provide process heat in the form of steam is assumed. The procurement of sugarcane to the facility and the operations part of the conventional sugar mill are not considered since it is assumed that those are already set-up and operational. Similarly, transportation of products produced from bagasse (or bagasse as a product) is not the part of the system boundary. For a fixed capacity of the sugar mill, the bagasse availability is also known.

Bagasse generated can potentially be split into five streams, three of which include further processing to produce value added products, namely, ethanol, pellets, and electricity (Fig. 1). Of the remaining two streams, one stream goes to the boiler to meet the energy demand of the mill as well as various production processes within the industrial cluster. This ensures that the industrial cluster is energy self-sufficient. The second stream represents bagasse that is directly sold in the market without any processing. The energy production operations in the clusters (boiler and electricity) are further supported by trash collected from sugarcane farms as feedstock. It is assumed that only the trash on farms supplying sugarcane to the mill is available, which sets an upper bound on the total trash that can be utilized.

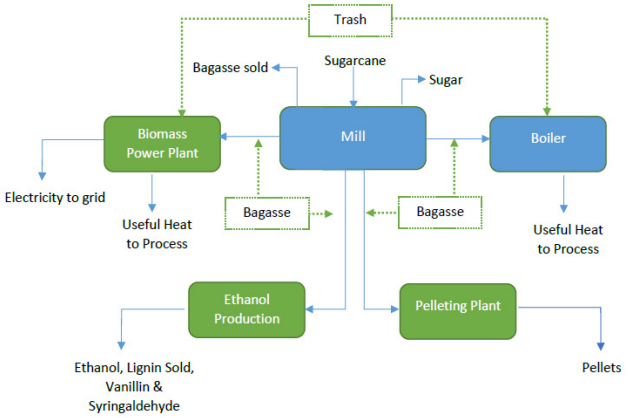


Figure 1: Alternatives for the utilization of bagasse and trash considered in the cluster. Blue boxes indicate existing facilities in the sugar mill while green boxes indicate new facilities

The ethanol production process uses the organosolvent pretreatment, which produces relatively pure lignin that can be used as feed to the boiler. The cellulose and hemicellulose separated after pretreatment are used for ethanol production. Pelletization involves three major steps: drying (up to 10% on wet basis), grinding (size reduction) and densification (Mani *et al*., 2006). Individual pellet density ranges from 1000 to 1200 kg/m^3 whereas the bulk density of pellets is 600 to 700 kg/m^3. Pellets coming out of pellet mill (at about 70°C to 90°C) are cooled, packaged and sold. Biomass Integrated Gasification Combined Cycle (BIG-CC) configuration is considered for electricity production using bagasse and trash as the feedstock. Combustion gases drive the gas turbine to generate electricity through generator. The exhaust gases from gas turbine are used to generate steam in Heat Recovery Steam Generator. The steam drives

the steam turbine to generate more electricity and useful heat, which can be delivered to other processes of the facility (Ensinas *et al*., 2007).

Two different optimization models have been formulated, capturing the economic and environmental impact associated with the proposed industrial cluster. These models are explained in the following sections.

2.1. Economic optimization model formulation

The model constraints include the mass and energy balances relevant to each unit process/operation described in the previous section. These constraints are formulated by considering the performance data from literature. Only selected constraints are reported here for brevity. A key mass balance constraint is the distribution of bagasse to five possible routes as shown in Fig. 1, which is modeled as follows:

$$total_{bagasse} = bgs_{eth} + bgs_{elec} + bgs_{pel} + bgs_{boiler} + bgs_{market} \quad (1)$$

where, $total_{bagasse}$ is the total available bagasse (Mg/h) obtained after sugarcane juice extraction, bgs_{eth} is the amount of bagasse (Mg/h) sent to ethanol production, bgs_{elec} is the amount of bagasse (Mg/h) sent to the electricity generation unit, bgs_{pel} is the amount of bagasse (Mg/h) sent to the pelleting plant, bgs_{boiler} is the amount of bagasse (Mg/h) sent to the boiler, and bgs_{market} is the amount of bagasse (Mg/h) that is directly sold in the market. Additional key constraint is the distribution of collected trash to electricity generation unit and the sugar mill boiler:

$$total_{trash} = trash_{elec} + trash_{boiler} \quad (2)$$

where, $total_{trash}$ is the total trash collected and transported (Mg/h) to the integrated facility, $trash_{elec}$ is the amount of trash (Mg/h) sent as a feed to electricity generation unit, and $trash_{boiler}$ is the amount of trash (Mg/h) sent to the boiler. Another important energy constraint is the overall heat demand of the facility:

$$C_{p,bagasse} * bgs_{boiler} + C_{p,lignin} * lignin_{boiler} + C_{p,trash} * trash_{boiler} + heat_{useful,elec} \geq heat_{mill} * Feed + bgs_{eth} * heat_{ethanol} \quad (3)$$

where, $C_{p,i}$ is the specific heat of combustion of material i (bagasse, trash, or lignin), $heat_{useful,elec}$ is the heat supplied by the power plant, $heat_{mill}$ is the heat demand of existing sugar mill, $heat_{ethanol}$ is the heat demand of ethanol generation facility, bgs_{boiler} is the amount of bagasse sent to boiler, $lignin_{boiler}$ is the amount of lignin sent to boiler, $trash_{boiler}$ is the amount of trash sent to boiler, and $Feed$ is the amount of sugarcane processed in the facility. The capital and operating costs for each of these processes are considered, and the corresponding cost functions constitute another important set of constraints. The key decision variables are the distribution of bagasse and trash to the various streams, as described the constraints in Eq. 1-3. The objective function is the maximization of Net Present Value (NPV) of the integrated facility where the NPV equation is linearized. This is a Mixed Integer Linear Programming (MILP) problem with 144 continuous variables, 4 binary variables

(selection of processing options), and 144 constraints. The model is formulated in GAMS® (General Algebraic Modelling System) solved using the CPLEX® solver.

2.2. Environmental optimization model formulation

The multi-objective optimization problem considering two environmental impact categories, namely, net greenhouse gas emissions (GHG) and water footprint, is formulated. Only the direct water footprint is considered. The model also considered the avoided GHG emission due to replacement effect, such as the replacement of petroleum based gasoline with ethonol from bagasse. Since the overall GHG emissions and water footprint numbers for all the processing options are directly used, a detailed mass balance for each step is not necessary. This model, therefore, is much simpler than the economic optimization model previously described. It is linear programming model with the decision variables as the amount of bagasse diverted to each of the five streams. ε constraint method is used to solve the problem and obtain the Pareto front. The model is not discussed here for brevity.

3. Results and discussion

The model was applied to a case of sugar refinery processing 181 Mg/h of sugarcane. 50% retention of trash on farm was assumed as base case. The selling prices for electricity (₹6/kWh) and pellets (₹10/kg) were as per the market values while the selling price of ethanol (₹39 /l) was as per the target price set by the government (₹ represents Indian Rupees). The pelletization rate was 4 Mg/h and a minimum size of 6.5 MW power plant was enforced. The prices of various chemicals and inputs were taken from literature. Table 1 shows the values of GHG emissions and water requirement for various products, including GHG offset due to replacement of an existing product.

Table 1: life cycle Input data for the different products produced from bagasse

Products	Bagasse requirement (kg)	GHG emissions; GHG offset (kg of CO_2/unit of Product)	Water requirement (lit /unit of Product)
Ethanol (lit)	6.636	0.31; 0.357	247
Electricity (kw)	5.14	0.051; 0.81	224
Pellets (kg)	2	0.27; 0	73

3.1. Economic optimization

The model was first used to determine the Break Even Selling Price (BESP) for all the three commodities individually when integrated with the existing sugar mill. BESP of ethanol, pellets and electricity was ₹64/l, ₹9/kg, and ₹3/kWh, respectively. This indicated that ethanol production was not profitable. In contrast, electricity production was profitable; primarily due to the consideration that bagasse (feedstock) was free.

When the optimization model was solved considering all potential uses of bagasse, electricity generation was recommended as the optimal solution and no other processing route was chosen. With the feed rate of 181 Mg/h and 50% trash retention, all of the 22.17 Mg/h available bagasse and 12.62 Mg/h trash was sent to the power plant facility. Boiler got no feed of bagasse and trash as the heat generated by the power plant was enough to meet the energy demands. The size of the plant was 37.84 MW and hence

capable of providing around 281 million units (kWh) of electricity annually to the grid generating revenues of ₹169 crores at an initial investment of ₹712.6 crores with annual profit of ₹105.75 crores (1 crore is equal to 10 Million, and 1 ₹ is about 0.015 USD).

The sensitivity of the result to the market prices of various products was studied. Electricity market is well established and therefore the prices are relatively stable. Therefore the cost of electricity was fixed at the base value. The selling prices of ethanol and pellet were varied and the resulting optimal product portfolio was determined. The results (Fig. 2) indicated that ethanol production became profitable only if the selling price was more than ₹76/l. Similarly, pelletization became profitable if the selling price was more than ₹14/kg. Note that since there was an upper cap on the pellet market, electricity was still produced for a pellet price higher than ₹14/kg.

Trash retention is often recommended as a sustainable farming practice since it reduces the requirement of chemical fertilization and increases yield due to natural soil enrichment. Therefore, a scenario with 100% trash retention on farm was studied. This reduced the availability of trash. The results (Table 2) indicated that the optimal strategy in such a case was to use bagasse in the boiler to meet the energy demands of the mill and sell the remaining bagasse directly to the market. Production of any of the processed products was economically infeasible. This indicated the importance of trash collection and utilization for this system. Cases for various trash retention values between 0% and 100% were simulated and the results are not reported here.

Fig. 3 shows the results of the environmental impact optimization model. There was a strong trade-off between water footprint and net GHG emission. For mimium water footprint, pelletization was the only product produced. However, greater importance to GHG emission resulted in the production of ethanol and then electricity. This was primarily due to the avoided GHG emissions of conventional products (gasoline and coal based electricity). For minimam GHG emission, only electricity production was recommended.

Table 2: Impact of trash retention on optimal industrial cluster design

Variable	Base Case	100% trash retention
Feed	181 Mg/h	181 Mg/h
Sugarcane yield	72.4 Mg/ha	82.9 Mg/ha
Bagasse processing route	Power Plant (37.84 MW)	Market & Boiler
Net Present Value (NPV)	₹366.06 crores	₹104.24 crores

4. Conclusions

This work developed an optimization model to design an industrial cluster based on sugarcane bagasse in India. The economic optimization results indicated that electricity production was the most profitable option for a sugar mill processing 181 Mg/h sugarcane. Trash availability from sugarcane farms was shown to be important to achieve economic feasibility of the electricity production option. Ethanol and pellet market prices needed to increase significantly to achieve economic feasibility of those products from bagasse. The net GHG emission and water footprint had a trade-off and the desirable products changed with varying weights given to each of these two

objectives. The ongoing work is focusing on integrating the economic and environmental impact optimization problems. Although results presented here provide some insights into the trade-offs, the integrated problem will provide quantitative results.

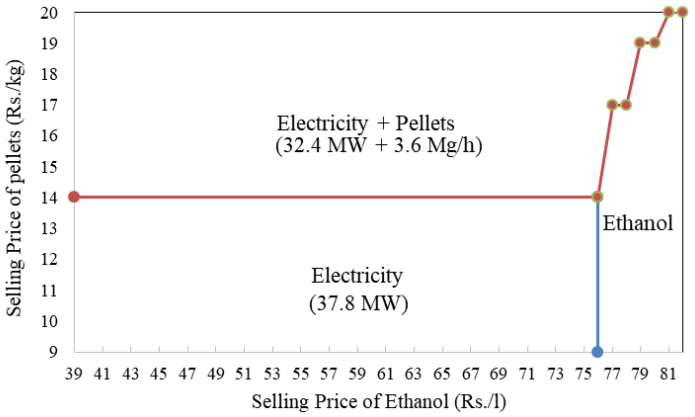


Figure 2: Impact of product prices on optimal product portfolio of the industrial cluster. Points on orange line indicate electricity+pellets region and points on blue line indicate ethanol region

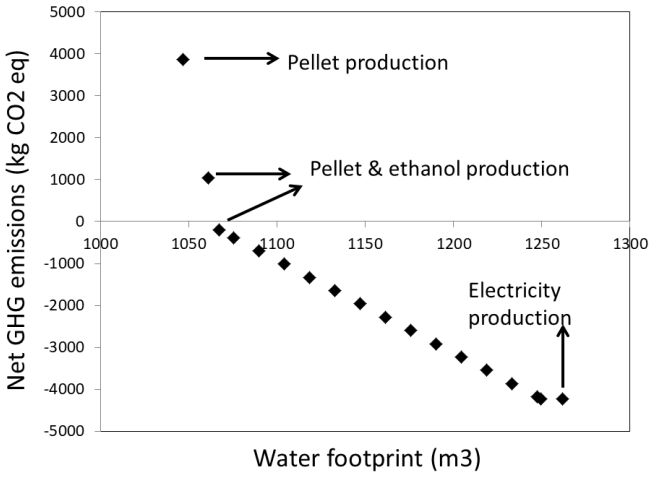


Figure 3: Net GHG emission and water footprint trade-off and the product portfolio for Pareto solutions. For the intermediate points, electricity and ethanol are produced in varying proportions

References

A.V. Ensinas, S.A. Nebra, M.A. Lozano, and L.M. Serra, 2007, Energy Conversion and Management, 48, 2978–2987

A.V. Ensinas, V. Codina, F. Marachal, J. Albarelli, and M.A. Silva, 2013, Chemical Engineering Transactions, 35, 523-528.

F. Furlan, C. Costa, G. Fonseca, R. Soares, A. Sechhi, A. Cruz, and R. Giordano, 2012, Comput. Chem. Eng., 43, 1-9.

S. Mani, S. Sokhansanj, X. Bi, and A. Turhollow, 2006, Applied Engineering in Agriculture, 22(3), 421-426.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50328-1

Simulation of Reaction in a Fluidized Bed Incinerator with Mixing Ratio of Double Based Propellant and Water

Jiheon Lee[a*], Raymoon Hwang [a], Hyunsoo Kim [b], Jungsoo Park [b], Min Oh [c], Il Moon [a]

[a]*Yonsei University, Seoul 03722, Republic of Korea*
[b]*Agency of Defence Development, Daejeon, Republic of Korea*
[c]*Hanbat University, Daejeon, Republic of Korea*
wlgjs622@yonsei.ac.kr

Abstract

Recently, there have been various studies how to incinerate explosive waste safely due to environmental problems and safety problems. Explosive waste requires a new incineration mechanism, because of the pollution gas generated during the conventional treatment process, such as rotary kiln or outdoor exploration. This study focuses on the fluidized bed incinerator technology which can burn the target material using only a small amount of air at a relatively low temperature condition. In this study, we simulated the process of burning Double Based Propellant (DBP) mixed with water in the incinerator safely. The fluidized bed incinerator was modeled as a cylinder with a diameter of 0.5 m and a height of 2.0 m, and a case study was carried out by changing the mixing ratio of the injected slurry. As a result, we confirmed the optimal mixing ratio with water for burning DBP without explosion, and confirmed that DBP combustion and decomposition reaction occurring inside the incinerator can be safely simulated. Based on this, it is considered that the design of the actual incinerator will provide a new direction for research on the explosive waste treatment process in Korea in the future.

Keywords: Fluidized bed, Double based propellant, water composition, simulation.

1. Introduction

Methods of disposing of explosive wastes, such as ammunition and propellants, have been used for incineration or for exploitation in the open air. However, it has various problems such as environmental problems caused by combustion gases and safety problems of processing facilities. In Korea, about 3,000 tons of explosive waste are generated annually, and 9,700 tons of waste propellant is being loaded untreated due to lack of proper treatment facilities. In the past, research on the incineration process using a rotary kiln has been carried out, but the safety problem has not been solved and the incineration process using a fluidized bed incinerator has been proposed. The treatment process using the fluidized bed incinerator has a higher rate of complete combustion than the conventional method and the combustion gas emission is remarkably low, and the efficiency of operation is also high due to the characteristics of the fluidized bed. Also, by utilizing the effect of mixing with the charged particles in the reactor, the heat energy generated during combustion is efficiently dispersed, thereby reducing the possibility of high temperature or overpressure. In this study, the optimum conditions of

the flow phenomenon and mixing effect in the incinerator were determined through the previous study of the cold model of the fluidized bed reactor. Based on this, the process of burning the waste propellant in the incinerator was designed, and the computational fluid dynamics program was used to simulate and analyze the particle behavior, the internal temperature change, and the gas component change.

2. Simulation scheme and condition

2.1 Design of fluidized bed

The fluidized bed incinerator simulated in this study was selected based on previous studies to determine optimal fluidization conditions. [1] The fluidized-bed incinerator shown in Figure 1 is designed as a cylinder with a height of 2.0 m and a diameter of 0.5 m and a fluidization of 440 K is introduced at a speed of 0.5 m / s under the incinerator. Inside, sand particles with a diameter of 260 μm were charged to 30% of the height of the incinerator. After fluidization reaches a steady state, slurry particles are injected at the side of 1.1 m in height and burned. The slurry is a mixture of water and Double based propellant (DBP) at a ratio of 1: 1 for the stability of process and reaction, and its size is 3 mm in diameter. DBP is a mixture of Nitrocellulose 52% and Nitroglycerin 48%. [8]

2.2 Double based propellant combustion reaction

DBP has a total of seven reaction mechanisms until complete decomposition. When water evaporates as heat is applied, conditions are set for DBP to ignite, and degradation occurs from the surface, resulting in gas generation and combustion in a chain. The two reactions occurring in the condensed phase are the degradation and gasification reactions of DBP, respectively [9] In the gas phase, there are two secondary reactions to NO2, a reaction to NO-carbon, and three secondary reactions to aldehyde and NO.

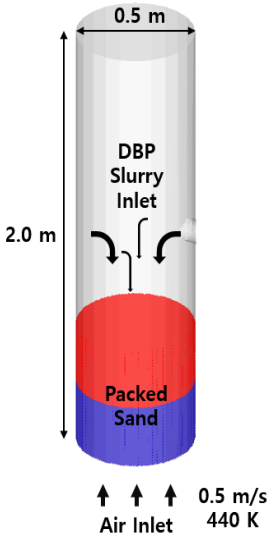


Figure 1: Schemes of fluidized bed incinerator

Table 1: Thermochemical parameters of double based propellant

Parameter	Value
E_c	167.4, kJ/mole
E_{NO_2}	20.9, kJ/mole
E_{CH_2O}	104.7, kJ/mole
$E_{(CHO)_2}$	83.7, kJ/mole
E_{NO}	209.3, kJ/mole
B_c	1.0×10^{17}, 1/s
B_{NO_2}	1.3×10^{6}, 1/s
A_{NO_2}	1.0×10^{5}, $m^3/kg{\cdot}s$
A_{CH_2O}	5.0×10^{8}, $m^3/kg{\cdot}s$
$A_{(CHO)_2}$	5.0×10^{8}, $m^3/kg{\cdot}s$
A_{NO}	5.0×10^{9}, 1/s
Q_c	-135, kcal/kg
Q_{NO_2}	1040, kcal/kg
Q_{CH_2O}	389, kcal/kg
$Q_{(CHO)_2}$	-539, kcal/kg
Q_{NO}	1900, kcal/kg
ρ_c	1620, kg/m^3
C_p	0.35, kcal/kg·K
C_c	0.40, kcal/kg·K

$$Double\ based\ Propellant \rightarrow 2.49NO_2 + 2.36CH_2O + 1.26(CHO)_2 + 0.17CO + minor\ residue \quad (1)$$

$$NO_2 + 0.56CH_2O + 0.16(CHO)_2 \rightarrow NO + 0.38CO + 0.5CO_2 + 0.5H_2O + 0.22H_2 \quad (2)$$

$$NO_2 + 0.56CH_2O + 0.16(CHO)_2 \rightarrow NO + 0.38CO + 0.5CO_2 + 0.5H_2O + 0.22H_2 \quad (3)$$

$$CH_2O + CH_2O \rightarrow CO + 0.5C_2H_4 + H_2O \quad (4)$$

$$(CHO)_2 + (CHO)_2 \rightarrow 4CO + 2H_2 \quad (5)$$

$$NO + 0.16CO + 0.12C_2H_4 + 0.12H_2 \rightarrow 0.5N_2 + 0.4CO_2 + 0.36H_2O \quad (6)$$

$$C + NO \rightarrow CO + NO_2 \quad (7)$$

Eq.(2) and Eq.(3) are the same reaction but different reactions occurring in the condensed phase and gas phase, respectively. The reaction heat of the DBP decomposition reaction in the condensed phase is about 1,100 cal/g [2, 3, 4, 5, 6, 7, 8]. The units are the same in kg/m^3·s. The thermochemical parameters of the reactions are shown in Table 1. All values were calculated in previous studies or calculated using the measured values. [8]

2.3 Fluidization condition

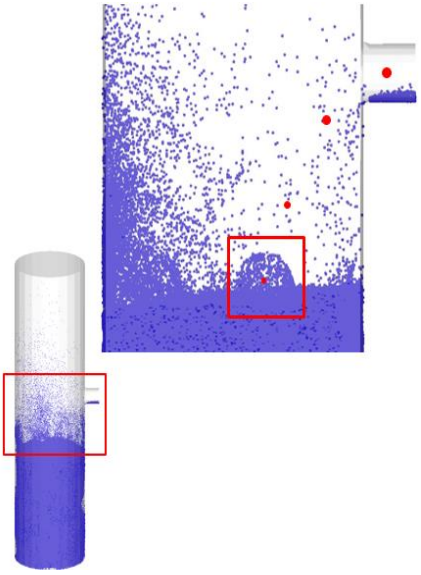


Figure 2: Sequence of slurry injection

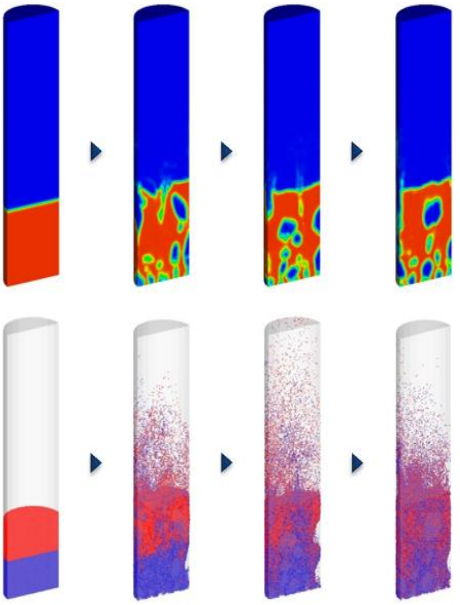


Figure 3: Fludization and mixing effect of fluidized bed incinerator

Before injecting DBP to make safe combustion conditions, the inside of the reactor must reach a steady state where flow phenomenon occurs. Figure 2 shows that all the sand particles are mixed and flow phenomenon occurs actively after 15 seconds. [1]

3. Simulation result and analysis

3.1. Particle movement analysis

Figure 3 simulates the inside of the reactor when the slurry is injected. After the slurry is injected, the water evaporates as it drops, and its size gradually decreases. Thereafter, the reaction occurs instantaneously between the sand particles. This can be confirmed by the fact that the sand particles are spread around the slurry. Thus, it was confirmed that the sand particles absorb the gas generation impact caused by combustion and protect the inner wall of the reactor. And the disappearance time of the particles is 0.28 seconds after the first injection, which is a very short time. Therefore, DBP decomposition reaction is very fast reaction.

3.2. Temperature profile

In Figure 4, temperature changes were observed in the reactor from 15.28 seconds when the slurry particles disappeared, and temperature changes were irregular spherical hot spots. In each case, the temperature of the hot spot was 901, 964, 1003K for each case, and then the gas exited over the incinerator as heat energy was dispersed. It takes about 0.7 seconds for gas to escape to the incinerator. As the water ratio increased, the temperature of the hot spot decreased.

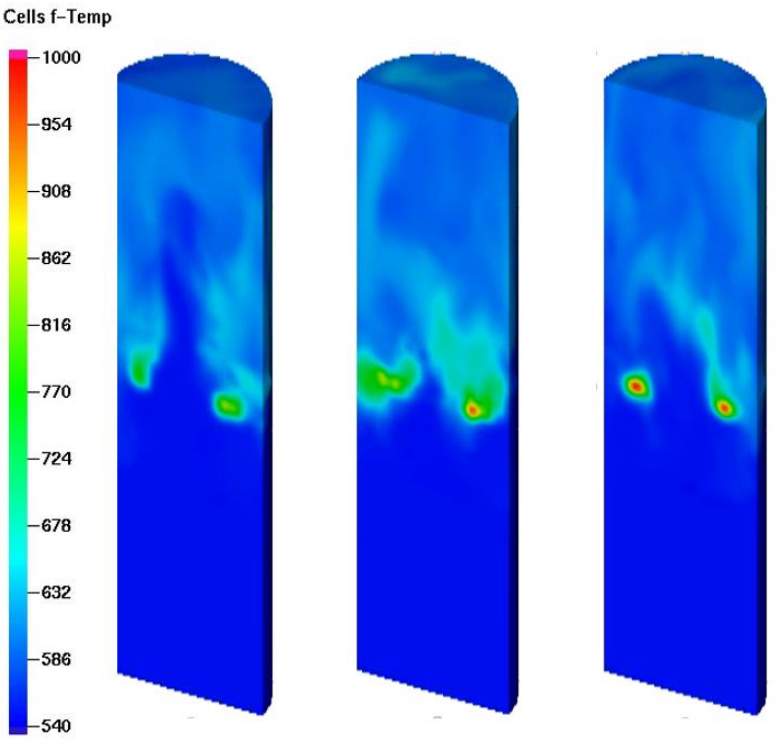


Figure 4: Hot spot profile, (a)5:5, (b)6:4, (c)7:3 (DBP:Water)

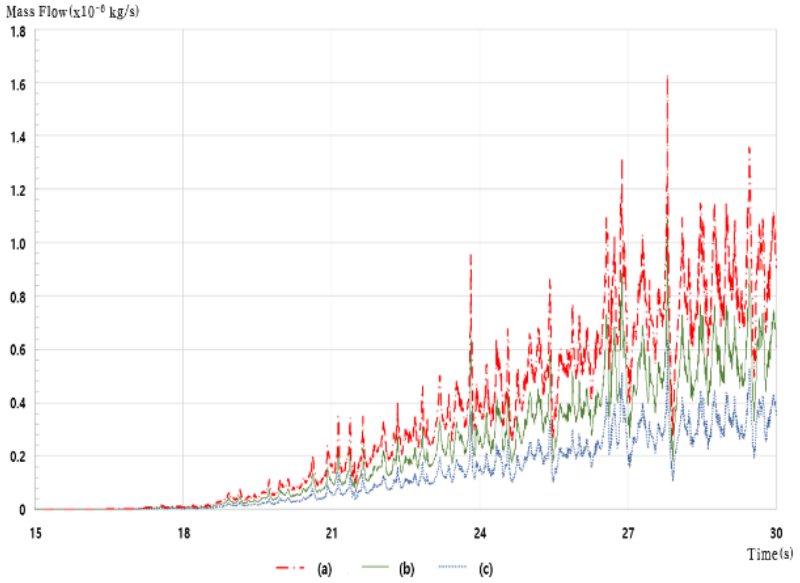


Figure 5 Flow rates of NO by time. (a)5:5, (b)6:4, (c)7:3

3.3 Flue gas analysis

Figure 5 shows the fluctuation of air flow at the time of reaction. This is because the reaction occurs rapidly and a large amount of gas is generated in a short time and the air is pushed up irregularly to the top. As a result of the analysis of the components, the chemical species generated at the time of disappearance of the hot spot starts to escape to the upper part of the reactor. Only N_2, O_2, H_2O and CO_2 should be observed at the upper part of the incinerator if complete combustion actually occurs. However, since incomplete combustion products such as NOx and aldehyde are observed, it can be seen that the gas is pushed up before complete combustion occurs. Also, as the water ratio decreases, the incomplete combustion products decrease, because the rate at which the reactants burn is faster.

4. Conclustion

In this study, CFD is used to explain the mechanism of the combustion of pulsed propellant without explosion in a fluidized bed incinerator. The maximum temperature of the generated hot spot was about 1000 K. As the water ratio increased, the temperature of the hot spot was lowered and the reaction rate was slowed down. The results of all cases are within the range that does not affect the physical properties of the reactor or packed particles. It was found that the feed ratio (water: DBP), the superficial velocity, and the size of the sand particles were found to have a great influence on the design of the operation condition, and energy optimization for this need to be further performed. As a result of the simulation, some of the gas produced after the reaction escapes to the top of the incinerator without being completely combusted, which can be treated by designing an additional downstream separation purification process.

If the simulated fluidized bed incinerator model is designed according to the actual process size, it can process 2,100 tons of waste propellant per year. These studies are

expected to contribute greatly to the disposal of accumulated pulmonary propellants in Korea.

5. Acknowledgement

This study was supported by CEMRC and Agency of Defence Development.

References

[1] Jiheon Lee, et al. “Calculation of Optimum Operating Conditions by Cold Model Simulation of Fluidized Bed Reactor Incinerator in Explosive Waste Incineration Process,” Theories and Applications of Chemical Engineering, 22 (1), p.727,April 2016.
[2] T.S. Roh, Tseng, Yang,. "Effects of Acoustic Oscillations on Flame Dynamics of Homogeneous Propellants in Rocket Motors," Journal of Propulsion and Power Vol. 11, No. 4, July 1995.
[3] Song, Y. H., Beer, J. M., and Sarofim, A. F., “Reduction of Nitric Oxide by Coal Char at Temperatures of 1250-1750K,“ Combustion Science and Technology. Vol. 25, pp.237-240, 1981.
[4] Lengelle, G., Bizot, A., Duterque, J., and Trubert, J. F., “Steady-State Burning of Homogeneous Propellants,“ Fundamentals of Solid-Propellant Combustion, edited by K. K. Kuo and M. Summerfield, Vol. 90, Progress in Astronautics and Aeronautics, AIAA, New York, pp.361-407, 1984.
[5] Bizot, A., and Beckstead, M. W., “A model for Double-Base Propellant Combustion,“ Proceedings of the 22nd Symposium on Combustion, The combustion Inst., Pittsburgh, PA, pp. 1827-1834, 1988.
[6] Faddoul, F., Most, J. M., and Joulain, P., “ Combustion Kinetic of a Homogeneous Double Base Propellant,“ Dynamics of Deflagrations and Reactive System-Flames, edited by A. L. Kuhl, Vol. 131, Progress in Astronautics and Aeronautics, AIAA, Washington, DC, pp. 275-296, 1989.
[7] Lengelle, G., Duterque, J., Godon, J. C., and Trubert, J. F., “Solid Propellant Steady Combustion: Physical Aspects,“ AGARD Lecture Series 180, 1991.
[8] Kubota. N., “Determination of Plateau Burning Effect of Catalyzed Double-Base Propellant,“ Proceedings of the 17th Symposium on Combustion,, The Combustion Inst., Pittsburgh, PA, pp. 1435-1441, 1979
[9] Aoki, I., and Kubota, N., “Combustion Wave Structures of High and Low-Energy Double-Base Propellants,“ AIAA Journal, Vol. 20,pp.100-105, 1982

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50329-3

A high-performance molecular reconstruction method with parameter initialization based on PCA

Kexin Bi, Tong Qiu*

Department of Chemical Engineering, Tsinghua University, 100084 Beijing, China
Corresponding author's E-mail: qiutong@tsinghua.edu.cn

Abstract

Molecular reconstruction of naphtha is the first step of naphtha cracking, and a detailed and accurate feedstock composition is required for molecular management in ethylene preparation. PCA method extracts key variables from naphtha properties, and then parameter initialization and range limitation can be regressed based on industrial data. Relatively accurate result is obtained in simulated annealing model during naphtha composition's optimization process due to the precise distribution graph of carbon number. A smaller optimization limitation range is indispensable for the high-performance simulation because the local optimal solution is avoided.

Keywords: Molecular reconstruction, Gamma distribution, Parameter initialization, PCA

1. Introduction

As the first step of molecular management[1], molecular reconstruction means molecular composition modeling and properties prediction of crude oil. For naphtha cracking process, the accurate molecular reconstruction result of naphtha expedites high-quality products and brings economic interests for refineries[2]. Analytical techniques to obtain detailed composition of crude oil, such as GC, GC×GC, GC-MS,[3] is improving rapidly nowadays, but these operations are time-consuming and the output is excessively specific. So a high-performance molecular reconstruction model need to be developed.

During the past decade, a wide variety of molecular reconstruction methods of naphtha have been proposed. Lumped components modeling methods are adopted by simulation software such as SPYRO[4] and CRACKER[5], but the flexibility and extendibility of these methods are rather limited, and accuracy of the result can't be ensured. Recent proposed stochastic methods solve these problems, and stochastic Monte Carlo modeling approach is adopted extensively in studies. Van Geem (2006) applied Shannon's entropy criterion in molecular reconstruction model with linear constraints.[6] Steven P (2011) imposed probability density function(PDF) on both the carbon number distribution in each homologous series of components and on the structural attribute distributions to reduce the number of parameters.[7] Zhang (2012) combined the model mentioned above and demonstrated the feasibility to obtain missing information for models.[8]

For some optimization algorithms, the initial values and ranges of variables are vital to get global optimal solution in small amount of calculation time. In some studies, a series of carbon number distribution probabilities are randomly generated to describe the detailed molecular composition information, and simulated annealing algorithm is applied in optimization process to meets all the property conditions. Nevertheless, using the variables with strong randomicity may result in a wrong solution because of the uncertainty of fraction characterization. These deficiencies make stochastic methods error-prone, time-consuming, and hard to apply into simulation software.

Based on analyses of naphtha fractions, parameter initialization and limitations of optimization can be determined by principal component analysis (PCA) regression model. In this paper, naphtha properties and composition information of western part of China is collected to characterize the carbon number distribution of homologues. PCA method extracts a few uncorrelated variables from Multi-collinear property variables. The extracted principle components are applied to regress the mathematical expectations and variances of the distribution. Assisted with simulated annealing algorithm, the distribution graph can be portrayed accurately and then molecular composition which satisfy the measured properties can be obtained.

2. Parameter initialization using PCA regression method

Stochastic molecular reconstruction process uses PDF to character the carbon number distribution. Some mixture properties obtained from industry analysis are applied to optimize the model, including the PIONA (P for paraffins, I for isoparaffins, O for olefins, N for naphthenes, A for aromatics), the Engler distillation curve (T for ASTM D86 distillation curve), the mixture's molecular weight (Mw) and the H/C ratio (HC).

To fit the carbon number distribution, a reasonable PDF should be picked up for main homologous series (PIONA) in naphtha. Gamma distribution is a rather flexible PDF with two parameters (shape parameter k and scale parameter θ). It is widely used to regress the boiling point curve and carbon number curve of crude oil: [9]

$$f(x;k,\theta) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\Gamma(k)\theta^{k}} \tag{1}$$

With $\mathrm{E}(\mathrm{x}) = k\theta$ and $\mathrm{Var}(\mathrm{x}) = k\theta^2$, and x is the carbon number of homologues series.

The origin distribution function need truncation and normalization before application.

Optimization parameters in simulated annealing algorithm can be set as P,I,N weight percentages (olefins is approximately equal to 0) and Gamma distribution parameters. Before optimization, initial value and range of parameters need to be assigned. A novel method of parameter initialization is proposed through PCA on the basis of already given properties of feedstock. Figure 1 explains the process of PCA regression.

3. Optimization model of naphtha molecular reconstruction

During the optimization process with simulated annealing algorithm, Shannon's entropy[10] is applied as a criterion. Shannon's entropy, proposed by Claude Shannon in 1948, can be introduced into naphtha molecular reconstruction to maximize the amount of "information":[11]

$$S(x_i) = -\sum_{i=1}^{N} x_i \ln x_i \quad \text{with} \sum_{i=1}^{N} x_i = 1 \tag{2}$$

Before optimization, weight percentage of specific component need to be calculated. For example, weight percentage of N-Pentane can be described as:

$$P(series = P, Cnum = 5) = P(series = P) \times P(Cnum = 5 \mid series = P) \tag{3}$$

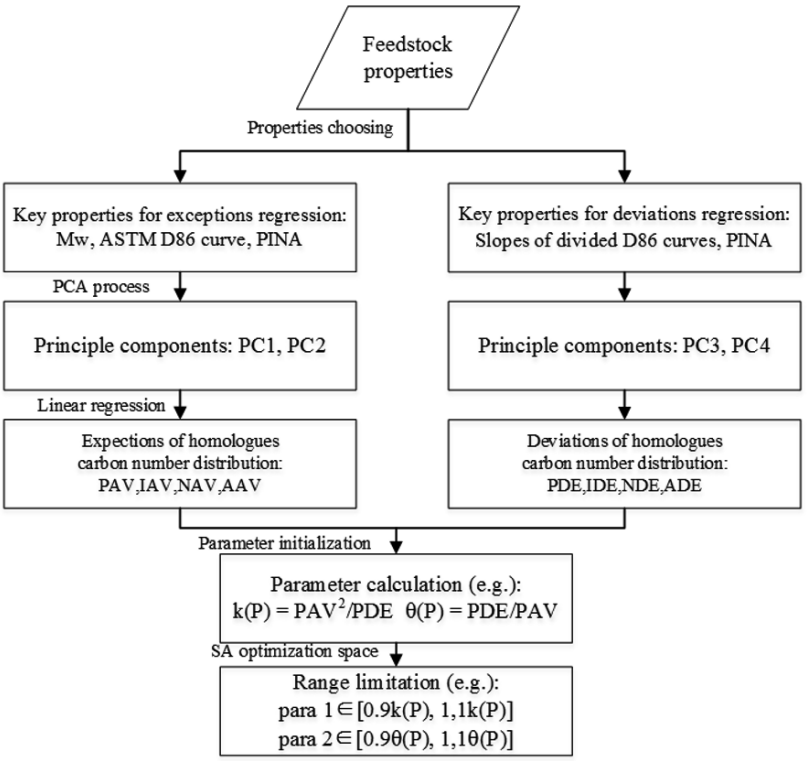


Figure 1. Parameter initialization method using PCA regression for simulated annealing algorithm

To summarize, the simulated annealing optimization model uses a combined function including Shannon's entropy and mixture properties. Properties and related objective functions are listed in Table 1.

Table 1. Properties and corresponding objective functions in simulated annealing algorithm

Properties	Objective functions
Molecular Weight (Mw)	$\exp(k_1 \times \frac{Mw^0 - Mw}{Mw^0})$
H/C ratio (HC)	$\exp(k_2 \times \frac{HC^0 - HC}{HC^0})$
PIONA	e. g. $\exp(k_3 \times \frac{P^0 - P}{P^0})$, $\exp(k_3 \times \frac{I^0 - I}{I^0})$
Boiling Point	e. g. $\exp(k_4 \times \frac{T^0 - T}{T^0})$

The overall objective function in simulated annealing algorithm can be concluded as:

$$f = \sum_{i=1}^{N} x_i \ln x_i + \sum \exp(k_j \times \frac{property^0 - property}{property^0}) \quad (4)$$

with $property \in \{PIONA, ASTM, Mw, HC\}$.

The weighting factors, $k_{j,}$ is assigned as 50 to balance the effects of different factors in objective function above.

4. Result and discussion

Once naphtha properties are entered, expectations and variances of carbon number distribution of homologues PINA can be calculated. Table 2 lists the key properties of naphtha for calculation. Initialization and range limitation for parameters related to PDF in simulated annealing algorithm are presented in Table 3, and the calculation method can be found in Figure 1. As a comparison, parameters obtained from average statistics are listed and the reconstruction result is also calculated by simulated annealing algorithm.

Table 2. Naphtha properties used for calculation

Property	Mw	HC	P (wt%)	I (wt%)	O (wt%)	N (wt%)	A (wt%)
Value	94.31	2.1	32.46	30.75	0	29.19	7.6
Property	IBP (°C)	10% (°C)	30% (°C)	50% (°C)	70% (°C)	90% (°C)	FBP (°C)
Value	35.4	60.3	82.7	100.8	119	145.9	163.5

Table 3. Parameter initialization for carbon number distribution (paraffin as an example)

parameter	k (P)	θ (P)	k (I)	θ (I)	k (N)	θ (N)	k (A)	θ (A)
				Average statistics				
initial value	10.03	0.57	7.87	0.66	9.18	0.42	9.30	0.39
lb	5.22	0.31	4.29	0.43	6.34	0.31	5.49	0.21
ub	14.84	0.83	11.46	0.88	12.01	0.54	13.12	0.56
				PCA regression				
initial value	7.38	0.70	5.82	0.78	9.47	0.39	10.75	0.30
lb	6.65	0.63	5.24	0.70	8.53	0.35	9.67	0.27
ub	8.12	0.77	6.40	0.86	10.42	0.43	11.82	0.33

The calculation process is executed by Matlab R2017a, and the simulation time is counted by Matlab on an Inter Core i7 processor of 2.80 GHz. Figure 2 demonstrates the calculation results of two methods compared to analysis data, and Table 4 concludes the calculation error and time-consumption of two methods

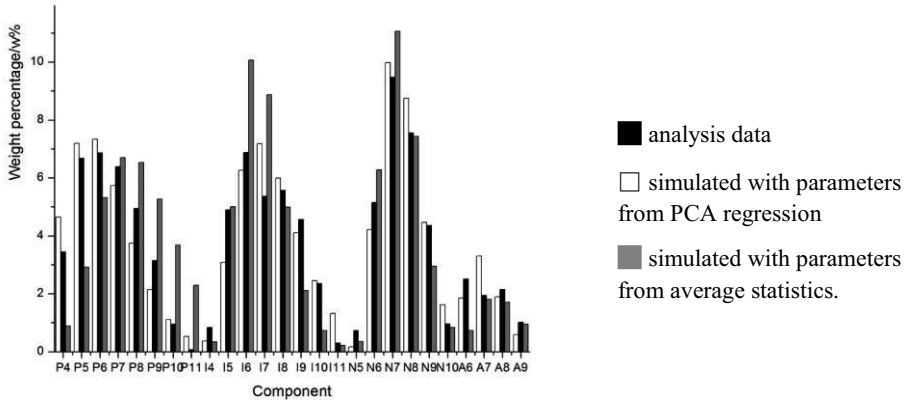


Figure 2. Comparison of analysed and calculated naphtha components' weight percentage.

Table 4. Comparison of calculation quality between two simulation process

	Manhattan Distance (wt%)	Average deviation (wt%)	Time-consumption(s)
simulated with PCA regression	19.42	0.7469	13.04
simulated without PCA regression	36.71	1.4119	186.59

Obviously, the calculation process with parameter initialization and range limitation provided by PCA regression demonstrates a lower deviation and simulation time. The reason can be traced back to Table 3. The fluctuation of the carbon number distribution in paraffin homologue of different calculation process is represented in Figure 3 after inverse transformation of data pretreatment.

Without PCA regression, the carbon number distribution may change from monotonically decreasing to monotonically increasing, as shown in curve (1) and curve (7) in Figure 3. In consequence, the large parameter space could make simulated annealing algorithm fall into local optimal solution. As a comparison, parameters initialization and range limitation using PCA regression cause the zenith and nadir of the carbon number distribution graph relatively fixed and the fluctuation range of the function shape narrow, as shown curve (3), (4), and (5) in Figure 3. With the shrinking of optimization searching area and the precise initialized parameters, the iteration process in simulated annealing algorithm displays faster rate of convergence. What's more, most of unrealistic local optimal solutions are excluded due to the minimal fluctuation range of distribution graph. Compared with the imprecise method using average statistics, the absolute deviation of molecular reconstruction result is reduced by 47.1% and the calculation time is shortened by 93.0% after using these ranges of parameters obtained by principal component regression.

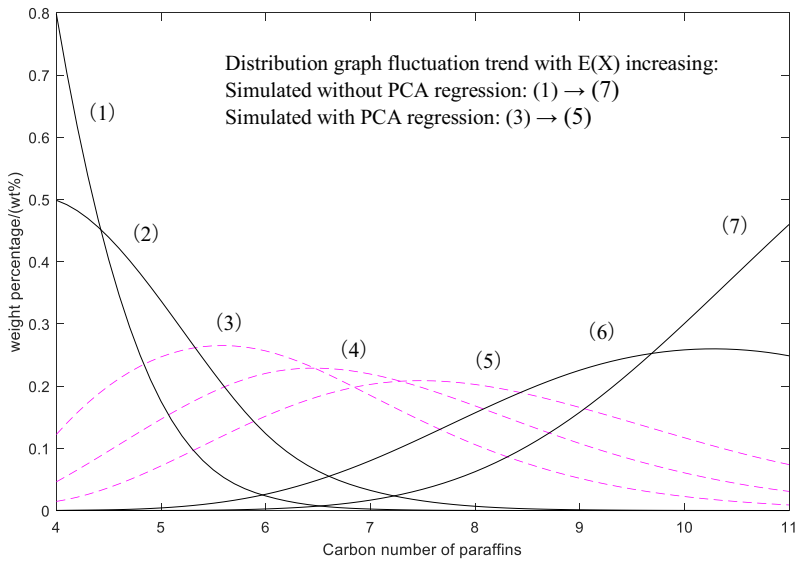


Figure 3. The fluctuation of carbon number distribution shape for paraffin. Optimization space shrinks from (1) → (7) to (3) → (5) with PCA regression of parameters

5. Conclusion

Parameter initialization and range limitation based on PCA regression is applied for optimization process of molecular reconstruction of naphtha. Calculation accuracy is promoted and simulation time is shortened for simulated annealing algorithm with Shannon's entropy and mixture properties as objective function. Lesser fluctuation of carbon number distribution is the main reason, causing the optimization algorithm convergence faster and the local optimal solution hard to appear. The molecular reconstruction model is extendable for naphtha due to multi-regional naphtha calibration data application for the regression process.

Accurate and fast molecular reconstruction method is essential to molecular management in the refining process, such as ethylene thermal cracking process. This high-performance simulation process has potential to be applied for component analysis of feedstock and product in smart manufacturing, which may create higher timely profit in molecular management strategy. [12] Data set and molecular reconstruction model of specific area can be constructed for accuracy promotion and process acceleration.

Acknowledgements

The authors gratefully acknowledge the National Natural Science Foundation of China for its financial support (Grant No. U1462206).

References

[1] Bogle, I. D. L. (2017). A Perspective on Smart Process Manufacturing Research Challenges for Process Systems Engineers. Engineering, 3(2), 161-165.
[2] Young, R. E. (2006). Petroleum refining process control and real-time optimization. IEEE control systems, 26(6), 73-83.
[3] Boduszynski, M. M. (1988). Composition of heavy petroleums. 2. Molecular characterization. Energy & Fuels, 2(5), 597-613.
[4] Bozzano, G., Dente, M., Sugaya, M., & McGreavy, C. (1998). The characterization of residual hydrocarbon fractions with model compounds. Retaining the essential information. ACS National Meeting, Boston, 43, 653–657.
[5] Joo, E., Park, S., & Lee, M. (2001). Pyrolysis reaction mechanism for industrial naphtha cracking furnaces. Industrial & engineering chemistry research, 40(11), 2409-2415.
[6] Van Geem, K. M., Hudebine, D., Reyniers, M. F., Wahl, F., Verstraete, J. J., & Marin, G. B. (2007). Molecular reconstruction of naphtha steam cracking feedstocks based on commercial indices. Computers & chemical engineering,31(9), 1020-1034.
[7] Pyl, S. P., Hou, Z., Van Geem, K. M., Reyniers, M. F., Marin, G. B., & Klein, M. T. (2011). Modeling the composition of crude oil fractions using constrained homologous series.Industrial & Engineering Chemistry Research, 50(18), 10850-10858.
[8] Zhang, L., & Chen, B. (2012). Applications of Shannon's entropy theory to naphtha pyrolysis simulation. Chemical Engineering & Technology, 35(2), 281-286.
[9] Wu, Y., & Zhang, N. (2010). Molecular characterization of gasoline and diesel streams. Industrial & Engineering Chemistry Research, 49(24), 12773-12782.
[10] Guiasu, S., & Shenitzer, A. (1985). The principle of maximum entropy. The mathematical intelligencer, 7(1), 42-48.
[11] Shannon, C. E. (1948). A mathematical theory of communication, Part I, Part II. Bell Syst. Tech. J., 27, 623-656.
[12] Li, D. (2016). Perspective for smart factory in petrochemical industry. Computers & Chemical Engineering, 91, 136-148.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50330-X

On Integration of Model Predictive Control with Safety System: Preventing Thermal Runaway

Zhihao Zhang[a], Zhe Wu[a], Helen Durand[a], Fahad Albalawi[b], Panagiotis D. Christofides[a,b,*]
[a]*Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA, 90095-1592, USA*
[b]*Department of Electrical and Computer Engineering, University of California, Los Angeles, CA 90095-1592, USA*
pdc@seas.ucla.edu

Abstract

In this paper, the methyl isocyanate (MIC) hydrolysis reaction in a CSTR is used as an example to demonstrate the performance of a Lyapunov-based model predictive control (LMPC) system integrated with the activation of a safety system. Initially, in the presence of small disturbances, it is demonstrated that the closed-loop system state under LMPC is able to stay close to the operating steady-state and within the stability region. However, in the presence of a significant disturbance in the feed, the control system cannot maintain the closed-loop state inside the stability region, and thus, it has to be integrated with the reactor safety system to avoid thermal runaway.

Keywords: Predictive control; Process control; Process safety; Reaction runaway

1. Introduction

Process operational safety is a critical element of chemical plant operation due to the severe consequences for both lives and property when safe process operation is not maintained. Despite the very significant efforts to develop safe process/plant design and operation procedures, accidents continue to occur. Innovative operating strategies such as integrating directly the actions of process control and safety systems are required to prevent unsafe process operation which may lead to catastrophic events. In particular, coordinating the control system actions with the ones of the safety systems, while maintaining their independence for redundancy purposes, would represent a significant paradigm shift in both control and safety system design that has the potential to impact process operational safety.

Motivated by the above considerations, in the present work, we investigate how the activation of the safety system should be accounted for in conjunction with model predictive control (MPC) design because the actions of the safety system change the process dynamics, and thus, they should be accounted for in the MPC implementation. To demonstrate for the first time the integration of MPC with a safety system, we focus on the methyl isocyanate (MIC) hydrolysis reaction in a CSTR; MIC is the principal chemical involved in the Bhopal disaster (Toro et al., 2016).

2. Process description

The exothermic hydrolysis reaction of methyl isocyanate to the corresponding amine and carbon dioxide is given as follows:

$$CH_3NCO_{(l)} + H_2O_{(l)} \longrightarrow CH_3NH_{2(aq)} + CO_{2(aq)}$$

The MIC hydrolysis reaction in a CSTR is considered (Ball, 2011). By applying mass and energy balances, the dynamic model of the process can be described as follows:

$$\begin{aligned} m\frac{dC_A}{dt} &= -mk_0 e^{\frac{-Ea}{RT}} C_A + F(C_{A0} - C_A) \\ mC_P\frac{dT}{dt} &= (-\Delta H)mk_0 e^{\frac{-Ea}{RT}} C_A + FC_P(T_0 - T) - L(T - T_j) \end{aligned} \tag{1}$$

where C_A is the concentration of MIC in the reactor, m is the total mass of the mixture in the reactor, and T is the temperature of the reactor. The concentration of reactant MIC, the temperature and the mass flow rate of the feed are denoted as C_{A0}, T_0 and F, respectively. The reacting liquid has a constant heat capacity of C_P. k_0, E_a and ΔH are the reaction pre-exponential factor, activation energy and the enthalpy of the reaction, respectively. The CSTR is equipped with a cooling jacket where L is the heat transfer coefficient and T_j is the cooling jacket temperature. The reactor is simulated at the conditions reported for the Bhopal catastrophe (Toro et al., 2016). Process parameter values are listed in Table 1.

Table 1: Parameter values of the CSTR.

$T_0 = 293\ K$	$F = 57.5\ kg/s$
$m = 4.1 \times 10^4\ kg$	$E_a = 6.54 \times 10^4\ J/mol$
$k_0 = 4.13 \times 10^8\ /s$	$\Delta H = -8.04 \times 10^4\ J/mol$
$C_P = 3 \times 10^3\ J/(kg\ K)$	$R = 8.314\ J/(mol\ K)$
$L = 7.1 \times 10^6\ J/(s\ K)$	$C_{A0} = 29.35\ mol/kg$
$T_{js} = 293\ K$	$C_{As} = 10.1767\ mol/kg$
$T_s = 305.1881\ K$	

3. LMPC design and thermal runaway

3.1. LMPC control design

The control objective is to stabilize the reactor at its steady-state by adjusting the manipulated input (denoted by u) which is the cooling jacket temperature T_j. The LMPC scheme is formulated as the following optimization problem:

$$\min_{u \in S(\Delta)} \int_{t_k}^{t_{k+N}} (\|\tilde{x}(\tau)\|_{Q_c}^2 + \|u(\tau)\|_{R_c}^2) d\tau$$

$$\text{s.t.} \quad \dot{\tilde{x}}(t) = f(\tilde{x}(t), u(t)) \tag{2b}$$

$$\tilde{x}(t_k) = x(t_k)$$

$$u(t) \in U, \forall t \epsilon [t_k, t_{k+N})$$

$$\frac{\partial V(x(t_k))}{\partial x} f(x(t_k), u(t_k)) \leq \frac{\partial V(x(t_k))}{\partial x} f(x(t_k), h(t_k)) \tag{2e}$$

where $\tilde{x}$ is the predicted state trajectory, $S(\Delta)$ is the set of piecewise constant functions with period Δ, and N is the number of sampling periods in the prediction horizon. The

optimal input trajectory of the LMPC optimization problem is $u^*(t)$, which is calculated over the entire prediction horizon $t\epsilon[t_k, t_{k+N})$. The control action computed for the first sampling period in the prediction horizon $u^*(t)$ is applied over the first sampling period, and the LMPC problem is resolved at the next sampling period. The objective function Eq.(2a) is minimizing the integral of $\|\tilde{x}(\tau)\|_{Q_c}^2 + \|u(\tau)\|_{R_c}^2$ over the prediction horizon. The constraint of Eq.(2b) is the deviation form of Eq.(1) that is used to predict the states of the closed-loop system. Eq.(2c) defines the initial condition $\tilde{x}(t_k)$ of the optimization problem which is the state measurement $x(t_k)$ at $t = t_k$. Eq.(2d) defines the input constraints applied over the entire prediction horizon. The constraint of Eq.(2e) is to decrease $V(x)$ such that $x(t)$ will move towards the origin at least at the worst-case rate achieved by the Lyapunov-based controller $h(x)$, which will be defined later. The explicit Euler method with an integration time step of $h_c = 10^{-2}\ s$ was applied to numerically simulate the dynamic model of Eq.(1) under the LMPC. The nonlinear optimization problem of the LMPC of Eq.(2) was solved using the IPOPT software package with the following parameters: sampling period $\Delta = 1\ s$; prediction horizon $N = 10$. $Q_c = [30; 05]$ and $R_c = 1$ are chosen such that the term related to the states and the term related to the input are on the same order of magnitude in $\|\tilde{x}(\tau)\|_{Q_c}^2 + \|u(\tau)\|_{R_c}^2$. The manipulated input is the cooling jacket temperature T_j, which is bounded as follows: $280\ K \le T_j \le 300\ K$. The CSTR is initially operated at the steady-state $[C_{As} T_s] = [10.18\ mol/kg\ \ 305.19\ K]$, with steady-state $T_{js} = 293\ K$. Therefore, the states and the input of the closed-loop system are represented in deviation form as $x^T = [C_A - C_{As}\ \ T - T_s]$ and $u = T_j - T_{js}$, such that the equilibrium point of the system is at the origin of the state-space. The Lyapunov function is designed using the standard quadratic form $V(x) = x^T P x$, where the positive definite matrix P is as follows: $[200\ 33; 33\ 40]$. The stability region Ω_ρ is characterized as $\Omega_\rho = \{x \in R^2 | V(x) \le \rho\}$. For the system of Eq.(1), the stability region Ω_ρ with $\rho = 8000$ is found based on the above Lyapunov function V and the following controller $h(x)$ (Lin and Sontag, 1991) for the system of form $\dot{x} = f(x,u) = f(x) + g(x)u$:

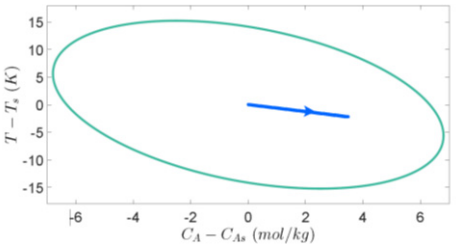
$$h(x) = \begin{cases} -\dfrac{L_f V + \sqrt{L_f V^2 + L_g V^4}}{L_g V^2} L_g V & \text{if} \quad L_g V \neq 0 \\ 0 & \text{if} \quad L_g V = 0 \end{cases} \tag{3}$$

3.2. Simulation results

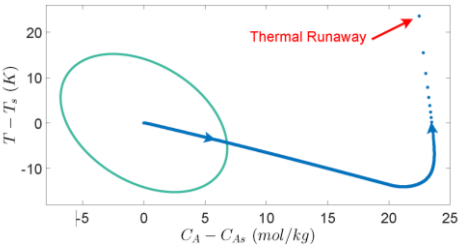
A small feed disturbance (i.e., change of feed concentration from $29.35\ mol/kg$ to $35\ mol/kg$) is initially considered and Figs. 1a and 1b demonstrate that the closed-loop system under the LMPC is robust to the small disturbance by stabilizing the system state at another steady-state within the stability region.

However, when there exists a large disturbance (i.e., the change of feed concentration is from $29.35\ mol/kg$ to $70\ mol/kg$) due to, for example, device failure, it is shown in Fig. 1c that the state goes out of the stability region and the manipulated input hits its lower bound to cool down the reactor as much as possible. However, after 150 seconds of implementation of maximum cooling, the reactor temperature starts to increase significantly. The reason for this increasing value of the temperature is that when the reactor temperature rises, the exothermic reaction rate also increases, causing further increase in temperature, which is a dangerous phenomenon called thermal runaway. Therefore, it can be concluded that in the presence of large disturbances, the reactor

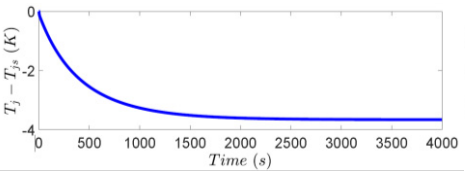
may operate in an unsafe region due to the restriction of the control actuator, which motivates the development of a safety system to maintain reactor safety.



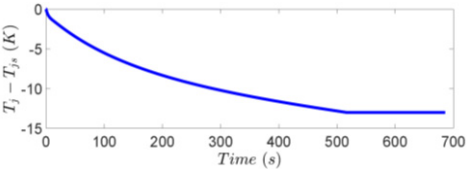
(a) State-space profile under small disturbance



(c) State-space profile under large disturbance



(b) Input trajectory under small disturbance



(d) Input trajectory under large disturbance

Figure 1: (a) and (b) demonstrate that the LMPC can stabilize the closed-loop state at another steady-state when there is a small disturbance. (c) and (d) demonstrate that the LMPC fails to keep the closed-loop state inside the stability region when there is a large disturbance.

4. Integration of MPC with safety system

In this section, the safety system is first designed using two different safety mechanisms: (a) safety relief valve; (b) cool water injection. Then, the entire process control/safety system which integrates the safety system with the LMPC is developed to guarantee closed-loop safety and stability. Finally, the MIC reaction example is used to demonstrate the application of the proposed control/safety scheme.

4.1. Components of safety system

4.1.1. Safety relief valve

The valve is activated to reduce the temperature and pressure of the reactor by discharging material when the temperature or pressure is extremely high in the reactor. In industry, reaction runaway may occur due to different failures, such as mis-charging reactant, loss of cooling temperature and so on. Since the above unsafe operating conditions are unpredictable and uncontrollable, a suitable and correctly sized relief system is crucially important as a backup method to prevent fatal accidents (Hace, 2013). The size of the relief valve is carefully chosen. Specifically, if the relief valve is under-sized, high pressure and equipment failure may occur; if the relief device is over-sized, the relief system may become unstable during the operation and too much material may be wasted (Crowl and Tipler, 2013).

4.1.2. Cool water injection

A direct cool solvent injection can cool down the reaction mixture temperature. It has been demonstrated in both simulation and experiment (Vernières-Hassimi and Leveneur, 2015) that cool water injection can lower the reactor temperature very fast.

4.1.3. Safety system for simulation

In our simulation, high temperature is the trigger of the opening of the relief valve. Specifically, the valve opens once the temperature is higher than $320\ K$. To simplify the development, we assume that the relief recharge flow is in liquid phase. The relief valve size is $0.004\ m^2$ and the relief flow is determined by the equation in (Hace, 2013):

$$G_{relief} = 0.9 \times 144 \times \frac{dP}{dT} \times \left(\frac{32.2}{778.16} \times \frac{T}{C_P}\right) \tag{4}$$

Cool water is injected at 280 K, and the flow rate is the same as the relief valve flow rate, thus the total mass in the reactor remains unchanged.

4.2. Logic integrating control and safety systems

LMPC integrated with the activation of the safety system is developed to help the closed-loop system state to avoid thermal runaway when the LMPC fails to maintain the state inside the stability region in the presence of large disturbances. A schematic of different regions and an example closed-loop trajectory is shown in Fig. 2, where different control schemes will be activated in the following three regions.

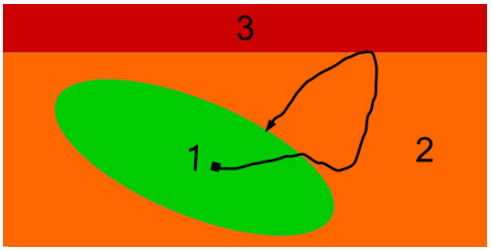


Figure 2: A schematic showing the stability region (green), unsafe operating region (orange), and the thermal runaway region (red), together with an example trajectory starting from the origin.

Region 1: When the closed-loop state is inside the stability region, the LMPC is implemented to stabilize the system at the origin or at another steady-state if there continuously exist small disturbances. At this stage, the safety system is not activated.

Region 2: If large disturbances are introduced into the reactor, the state comes out of the stability region. In order to ensure process operational safety, the manipulated input (i.e., T_j) is set to its lower bound, namely the lowest cooling jacket temperature, since the control system fails to work outside of the stability region.

Region 3: If large disturbances keep affecting the reactor and the maximum cooling is not able to lower the temperature, then the reactor temperature reaches a high value (i.e., the lower boundary of Region 3). Safety actions are taken in Region 3. Specifically, the relief valve opens immediately after the state enters Region 3 and stays open until the state goes back to Region 1. Meanwhile, cold water is injected into the reactor, cooling down the reactor. Injection stops once the relief valve is closed (state goes back into Region 1). At the same time, the jacket temperature stays at its lower bound to apply maximum cooling.

4.3. Simulation results

In Fig. 3, it is demonstrated that in the presence of a large disturbance, LMPC integrated with the safety system via the above logic succeeds to avoid thermal runaway and to drive the state back to the origin. At the beginning of the simulation, a large disturbance

is introduced into the reactor, resulting in the failure of LMPC to keep the system state within the stability region. After about 600 seconds, since the heat generated by the reaction is much more than the heat that the cooling system can remove, the concentration of the reactant is accumulated to such a great extent that the temperature starts to increase rapidly and reaches the safety limit of 320 K. Once the temperature exceeds the safety limit, the relief valve opens to discharge hot fluid from the reactor and an additional stream is employed to feed fresh water into the reactor. Liquid relief flow rapidly decreases the total internal energy and the reactant concentration in the reactor. Cool water promptly cools down the reactor temperature and it dilutes the reactant concentration. The safety system is activated for about 10 seconds to drive the closed-loop state back to Region 1. Once the closed-loop state goes back to Region 1, the LMPC replaces the safety system to stabilize the system state to the origin. Inside Region 1, a well-designed LMPC is guaranteed to stabilize the system state to the origin when there is no disturbance. It should be noted that if the large disturbance still exists after the closed-loop system state goes back into Region 1, then the overall process control system with the safety system is implemented again to avoid thermal runaway as discussed above.

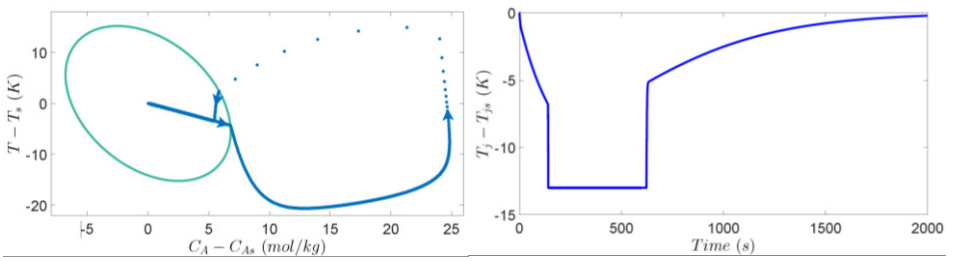


Figure 3: State-space plot and input plot of LMPC integrated with the safety system.

5. Conclusion

In this work, an LMPC system integrated with the activation of a safety system was developed for the MIC reaction in a CSTR to avoid thermal runaway. Specifically, in the presence of large disturbances, it was demonstrated that the closed-loop system under LMPC integrated with a safety system achieved process safety by avoiding thermal runaway and driving the state back into the stability region.

References

R. Ball. Oscillatory thermal instability and the bhopal disaster. Process Safety and Environmental Protection, 89:317–322, 2011.

D. Crowl and S. Tipler. Sizing pressure-relief devices. Chemical Engineering Progress, 109:68–76, 2013.

I. Hace. The pressure relief system design for industrial reactors. Journal of Industrial Engineering, 2013:1–14, 2013.

Y. Lin and E. Sontag. A universal formula for stabilization with bounded controls. Systems & Control Letters, 16:393–397, 1991.

J. Toro, I. Dobrosz-Gómez, and M. García. Dynamic modeling and bifurcation analysis for themethyl isocyanate hydrolysis reaction. Journal of Loss Prevention in the Process Industries, 39:106–111, 2016.

L. Vernières-Hassimi and S. Leveneur. Alternative method to prevent thermal runaway in case of error on operating conditions continuous reactor. Process Safety and Environmental Protection, 98:365–373, 2015.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50331-1

Estimating the Spatial Temperature Distribution in a Steam Methane Reforming Furnace Using Bayesian Modelling

Anh Tran[a], Madeleine Pont[a], Marquis Crose[a], Panagiotis D. Christofides[a,b*]

[a]*Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA, 90095-1592, USA*
[b]*Department of Electrical Engineering, University of California, Los Angeles, CA, 90095-1592, USA*
pdc@seas.ucla.edu

Abstract

In this work, we propose a parallelized Bayesian model identification approach to derive a data-driven model for the outer reforming tube wall temperature (ORTWT) distribution in a steam methane reforming furnace as a function of the fuel distribution and total fuel flow rate. The parallelization ensures that the data-driven models for all ORTWT can be derived simultaneously from the reformer data and independently from one another. The identification of the models is based on Bayesian variable selection, Bayesian model averaging, sparse nonlinear regression, reformer geometry, and theories of thermal radiation so that the model building process for each ORTWT can systematically identify predictors and simultaneously collect a corresponding library of sub-models. A collection of data-driven models for all ORTWT is defined as a data-driven model for the ORTWT distribution. The Bayesian model identification is executed on the Hoffman2 cluster at UCLA to construct the data-driven model for the ORTWT distribution from the training data set, and the results from the resubstitution and forecast tests of the data-driven model are used to demonstrate the effectiveness of the proposed approach.

Keywords: Steam methane reforming; Bayesian Statistics; Sparse nonlinear regression.

1. Introduction

Hydrogen gas is used in a wide range of applications. Among the commercial processes for the production of hydrogen gas, steam methane reforming (SMR) is the most common process (Darvishi et al., 2017). SMR is a catalytic endothermic process in which methane reacts with superheated steam to produce hydrogen, carbon dioxides, and carbon monoxide in the presence of nickel-based catalyst pellets at high temperature, e.g., 1023-1223 K, and high pressure, e.g., 20-40 bar, conditions to ensure that the process is economically viable and a desired methane conversion is achieved (Darvishi et al., 2017).

SMR is carried out inside furnaces. A typical steam methane reforming furnace, i.e., a reformer, has two closed, thermally coupled domains, i.e., the tube side and furnace side. In the tube side, there are hundreds of tubular packed-bed reforming reactors, which are designed to sustain the severe operating condition required for a profitable hydrogen production, and in the furnace side, fuel is burned in excess air to create the high temperature environment to facilitate the production of hydrogen in the tube side. The reformer considered in this work belongs to a hydrogen plant with a designed daily production of 2.8×10^6 Nm^3 of hydrogen (Tran et al., 2017b). The reformer is the most energy intensive unit at the hydrogen plant, and approximately 60% of the production cost is attributed to the procurement of feedstock; therefore, improving the thermal

efficiency of the reformer is expected to result in a significant financial benefit for the hydrogen plant.

The thermal efficiency of the reformer can be qualitatively measured by the hydrogen composition of the synthesis gas stream exiting the reformer, controlled by the total fuel mass flow rate and constrained by the design temperature of the reforming tube wall (Tran et al., 2017a). Therefore, the ORTWT distribution is frequently used as a control variable for the real-time reformer optimizer whose objectives are to retain the expected service life of the reforming tubes and to maintain the reformer safe operation to avoid unplanned plant shutdowns and capital/production losses. In practice, the task of finding the reformer input to deliver the maximum fuel flow rate without the risk of rupturing the reforming tubes is challenging and is an issue of interest in the hydrogen manufacturing industry. Therefore, an accurate and computationally efficient model for the relationship between the ORTWT distribution and the reformer inputs is required to design a real-time reformer optimizer.

Although this relationship can be modeled by using first principles modeling and computational fluid dynamic (CFD) modeling, first principles models and CFD models for the ORTWT distribution are often unsuitable for designing a real-time optimizer. First principles models overlook the influence of flow field on transport variables, which often yield solutions with poor accuracy, and CFD models are computationally expensive. Therefore, data-driven modeling is an alternative solution because data-driven models can be computationally inexpensive and can have reasonable accuracy. Motivated by this, we propose a model identification approach based on Bayesian statistics, sparse nonlinear regression, principles of thermal radiative heat transfer and the reformer geometry to derive a data-driven model for the ORTWT distribution.

2. Bayesian model general structure

This section aims to elaborate on the motivation for implementing Bayesian statistics in the model building process for the ORTWT distribution; to simplify our discussion, we consider a model building process for an arbitrary *ith* ORTWT.

We have found that in the reformer the dominant mode of heat transfer is thermal radiation, and more than 95% of the energy absorbed by the tube side is transferred to outer wall of reforming tubes by thermal radiation (Tran et al., 2017b). In addition, the first principle model for thermal radiative heat transfer indicates that the intensity of thermal radiation decreases drastically with increasing distance between the heat source and the heat sink. This analysis allows us to deduce that the *ith* ORTWT depends only on the fuel flow rates of the nearby burners. Nevertheless, the size of the neighborhood of the *ith* reforming tube is an ambiguous hyperparameter because there is no evidence suggesting that a data-driven model that uses a larger neighborhood has higher forecast accuracy compared to a data-driven model that uses a smaller neighborhood.The Bayesian model identification is designed to search for all data-driven models for the *ith* ORTWT with a high goodness of fit (which are denoted $\tilde{T}_{i,k}$ and referred to as sub-models in this manuscript) and to combine them to create the prediction model for the *ith* ORTWT (denoted $\hat{T}_i$) as follows,

$$\hat{T}_i = \sum_{k=1}^{K_i} \omega_{i,k} \tilde{T}_{i,k} \quad \text{where} \quad \sum_{k=1}^{K_i} \omega_{i,k} = 1. \tag{1}$$

where $\omega_{i,k}$ is the weighting factor of $\tilde{T}_{i,k}$. The prediction model for the *ith* ORTWT is expected to have a higher forecast accuracy than any individual sub-model (Hoeting et al., 1999).

3. Sparse nonlinear regression

We use sparse nonlinear regression to derive the sub-models for the *ith* ORTWT so that the sub-models can capture any nonlinearity in the relationship between the *ith* ORTWT and the reformer input. In this effort, we propose a library of transformations based on the expected response of the *ith* ORTWT to changes in the reformer input. For instance, if the *jth* burner is in the neighborhood of the *ith* reforming tube, increasing the fuel flow rate of the burner is expected to cause the *ith* ORTWT to increase. In addition, a non-zero fuel flow rate of the burner is expected to cause the *ith* ORTWT to raise above the ambient air temperature. Thus, we propose eight non-negative, monotonically increasing transformations as follows $f_1\left(\vec{F}^n\right)=\left[F_1^n,\ldots,F_{96}^n\right]^T$, $f_2\left(\vec{F}^n\right)=\left[(F_1^n)^2,\ldots,(F_{96}^n)^2\right]^T$, $f_3\left(\vec{F}^n\right)=\left[(F_1^n)^3,\ldots,(F_{96}^n)^3\right]^T$, $f_4\left(\vec{F}^n\right)=\left[\sqrt[2]{F_1^n},\ldots,\sqrt[2]{F_{96}^n}\right]^T$, $f_5\left(\vec{F}^n\right)=\left[\sqrt[3]{F_1^n},\ldots,\sqrt[3]{F_{96}^n}\right]^T$, $f_6\left(\vec{F}^n\right)=\left[\sqrt[4]{F_1^n},\ldots,\sqrt[4]{F_{96}^n}\right]^T$, $f_7\left(\vec{F}^n\right)=\left[\sqrt[5]{F_1^n},\ldots,\sqrt[5]{F_{96}^n}\right]^T$, $f_8\left(\vec{F}^n\right)=\left[\exp(F_1^n),\ \ldots,\exp(F_{96}^n)\right]^T$, where $\vec{F}^n$ is the *nth* fuel distribution. Next, we use the library of transformations to formulate the generalized form of $\tilde{T}_{i,k}$ as follows,

$$\tilde{T}_{i,k}=\sum_{g=1}^{8}\left(\vec{\alpha}_i^{kg}\right)^T\cdot f_g\left(\vec{F}\right)+\alpha_i^k \tag{2}$$

where $\vec{\alpha}_i^{kg}$ is the parameter vector associated with the *gth* transformation function of $\tilde{T}_{i,k}$ and α_i^k represents ambient air temperature of $\tilde{T}_{i,k}$. Then, we use a maximum likelihood estimator (MLE) formulated as a constrained L_1 minimization to estimate the parameters of $\tilde{T}_{i,k}$ as follows,

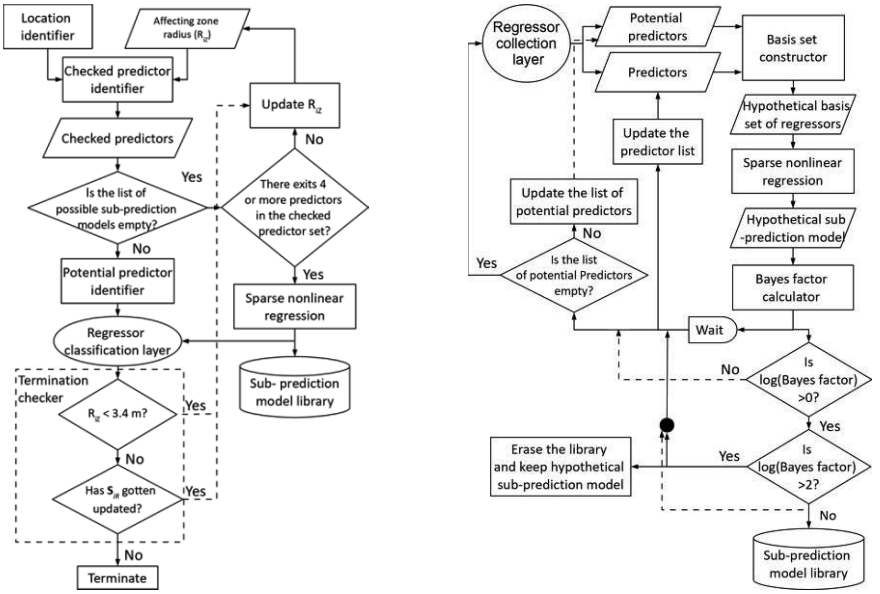
$$\min_{\substack{\alpha_i^k\in[298.15,348.15]\\ \alpha_{ij}^{kg}\in[0,\infty)}}\sum_{n=1}^{N}\frac{\left(T_i^n-\tilde{T}_{i,k}^n\right)^2}{2(\sigma^n)^2}+\lambda_i\sum_{g=1}^{8}\left\|\vec{\alpha}_i^{kg}\right\|_1 \tag{3a}$$

$$\text{subject to:}\qquad \bar{F}^0=\frac{F_{tot}^{typ}}{96} \tag{3b}$$

$$\sum_{g=1}^{8}\alpha_{il}^{kg}f\left(\bar{F}^0\right)=\sum_{g=1}^{8}\alpha_{ij}^{kg}f\left(\bar{F}^0\right)\qquad \text{if } d_{il}=d_{ij} \tag{3c}$$

$$\left(\frac{d_{ij}}{d_{il}}\right)^{\beta_u}\sum_{g=1}^{8}\alpha_{ij}^{kg}f\left(\bar{F}^0\right)\geq\sum_{g=1}^{8}\alpha_{il}^{kg}f\left(\bar{F}^0\right)\geq\left(\frac{d_{ij}}{d_{il}}\right)^{\beta_l}\sum_{g=1}^{8}\alpha_{ij}^{kg}f\left(\bar{F}^0\right)\qquad \text{if } d_{il}<d_{ij} \tag{3d}$$

where N is the total number of training data sets, T_i^n is the *ith* ORTWT in the *nth* training data, d_{ij} and d_{il} are the distances from the *ith* reforming tube to the *jth* and *lth* burners, respectively, σ^n is the standard deviation of the measurement noise in the *nth* training data, λ_i is the LASSO parameter of the model building process for the *ith* ORTWT, β_u = 6.0 and β_l = 1.0 are hyperparameters of the constraints and are chosen by trial and error, respectively, and F_{tot}^{typ} is the total fuel flow rate typically reported in the SMR literature. In the formulation tot of the MLE, Eqs.(3c) and (3d) are proposed in an effort to integrate the inverse square law of thermal radiative heat transfer into the sub-



(a) Regressor collection layer (b) Regressor classification layer

Figure 1: Framework for the Bayesian model identification

models. Specifically, Eq.(3c) allows the sub-models to realize that when two burners are in the neighborhood of the *ith* reforming tube, are separated from the *ith* reforming tube by the same distance and have the same fuel flow rate, the two burners should have the same impact on the *ith* ORTWT. Using this logic, Eq.(3d) allows the sub-models to realize that a closer burner to the *ith* reforming tube is expected to have a higher, but finite, impact on the *ith* ORTWT compared to that of the farther burner. The proposed sparse nonlinear regression method is integrated into the Bayesian model identification so that the impact of incorporating a burner into the predictor set for the *ith* ORTWT on the model goodness of fit can be quantified, which allows it to judge if this burner is an important predictor.

4. Framework for the Bayesian model identification

In this section, the framework for the Bayesian model identification is presented. Specifically, it is structured to have two layers, namely, the regressor collection layer and the regressor classification layer as shown in Figs. 1a and 1b, respectively. The regressor collection layer makes use of the knowledge that thermal radiative heat transfer is the dominant mode, and it functions under an impression that the *ith* ORTWT is controlled by the fuel flow rates of the nearby burners, which it identifies by creating a cylindrical neighborhood with the radius of the affecting zone around the *ith* reforming tube. This allows the Bayesian model identification to reduce the predictor space and to define an optimized path to search for the important predictors for the *ith* ORTWT. In addition, we assume that the 4 nearest burners of the *ith* reforming tube are the default predictors for the *ith* ORTWT, and together they form the default predictor set, which is denoted by $\mathbf{S}_{iR}$ and is used as the input to the sparse nonlinear regression to create the default sub-model.

After $\mathbf{S}_{iR}$ has been identified, and the default sub-model, which is also considered to be the "current best" sub-model, has been created, the regressor collection layer considers

all new nearby burners to be potential predictors which are used as inputs to the regression classification layer. Then, the regression classification layer uses the potential predictor set to identify additional important predictors, additional sub-models and new "current best" sub-model for the *ith* ORTWT in a step-wise fashion by means of Bayes factors, Bayesian information criterion (BIC) and the modified Jeffery's interpretation of Bayes factor (Kass and Raftery, 1995). Next, the updated $\mathbf{S}_{iR}$ is returned to the regressor collection layer.

In the regressor collection layer, the termination checker uses the radius of the affecting zone and $\mathbf{S}_{iR}$ to decide to terminate the model building process of the *ith* ORTWT. Specifically, when the radius of the affecting zone is less than the critical value of 3.4 m, the model building process for the *ith* ORTWT is always continued to ensure that the Bayesian model identification thoroughly explores the minimum predictor space. However, when the radius of the affecting zone is greater than the critical value, the model building process is continued if at least one new potential predictor has been added to $\mathbf{S}_{iR}$ to prevent the Bayesian model identification from exploring unnecessary regions in the predictor space. In a case that the model building process of the *ith* ORTWT is continued, the radius of the affecting zone is expanded by a step increment of 1.5 m. Then, the process of constructing $\mathbf{S}_{iR}$ is repeated until all important predictors judged based on the training data set (denoted $\mathbf{T}_i$) are identified.

We continue this discussion by assuming that the model building process has identified K_i distinct sub-model for the *ith* ORTWT. Using Bayesian model averaging, the expected value of the *ith* ORTWT based on $\mathbf{T}_i$, $\vec{F}^n$ and F_{tot}^n is denoted $E\left(T_i^n|\mathbf{T}_i,\vec{F}^n,F_{tot}^n\right)$ and is computed as follows,

$$E\left(T_i^n\,|\,\mathbf{T}_i,\vec{F}^n,F_{tot}^n\right)=\sum_{k=1}^{K_i} pr\left(\tilde{T}_{i,k}\,|\,\mathbf{T}_i\right)E\left(T_i^n\,|\,\mathbf{T}_i,\vec{F}^n,F_{tot}^n,\tilde{T}_{i,k}\right) \tag{4}$$

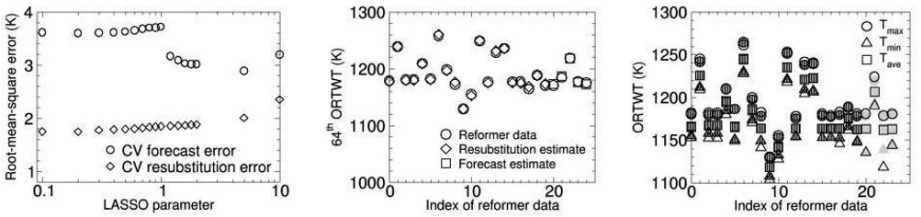
where $pr\left(\tilde{T}_{i,k}|\mathbf{T}_i\right)$ is the posterior probability of $\tilde{T}_{i,k}$ and $E\left(T_i^n|\mathbf{T}_i,\vec{F}^n,F_{tot}^n,\tilde{T}_{i,k}\right)$ is the expected value of the *ith* ORTWT based on $\mathbf{T}_i$, $\vec{F}^n$, F_{tot}^n and $\tilde{T}_{i,k}$. We compare Eq.(1) to Eq.(4) and deduce that $\omega_{i,k}$ is equivalent to $pr\left(\tilde{T}_{i,k}|\mathbf{T}_i\right)$, which can be expressed in terms of Bayes factors and approximated using the BIC. As a result, we can form an explicit algebraic model for the *ith* ORTWT as a function of the total fuel flow rate and fuel distribution.

5. Results

In this work, we partition our reformer database, which consists of 24 reformer CFD solutions generated by subjecting the reformer CFD model with varying fuel flow rates and fuel distributions, into two groups, i.e., the training data set and testing data set (Tran et al., 2017a,b). The training data set is randomly assigned 80% of the reformer database, and the testing data is assigned the remaining 20%. We propose 18 LASSO parameters, i.e., $\mathbf{S}_{\boldsymbol{\lambda}}$ = {0, 0.1, ⋯, 1, 1.2, ⋯, 2, 5, 10}, and we use leave-out-one (LOO) cross validation (CV) to select the optimal parameter of each model building process. In the remainder of this section, we present the result of the model building process for the 64*th* ORTWT as an illustrative example.

The results of LOO cross validation used to determine the optimal LASSO parameter of the model building process for the 64*th* ORTWT are shown in Fig. 2a. We note that the

magnitude of LASSO parameters controls the goodness of fit and complexity of the prediction model for the 64*th* ORTWT. Specifically, the model building processes that use small LASSO parameters yield high goodness-of-fit prediction models which might



(a) (b) (c)

Figure 2: Fig. 2a displays the results of the LOO cross validation for selecting λ_{64}^{op}. Fig. 2b displays the results of the resubstitution and forecast tests of $\hat{T}_{64}$. Fig. 2c. displays the results of the resubstitution (denoted by stripe symbols) and forecast (denoted by gray-filled symbols) tests of the prediction model for the ORTWT distribution.

be overfitting the CV training data sets, and the model building processes that use large LASSO parameters yield CV prediction models with low level of complexity which might be underfitting the CV training data sets. Fig. 2a reveals that the optimal LASSO parameter (denoted λ_{64}^{op}) is 5 which yields CV prediction models with the smallest forecasting error. Then, we use λ_{64}^{op} and $\mathbf{T}_i$ to derive $\hat{T}_{64}$. Fig. 2b shows that $\hat{T}_{64}$ yields estimates with small resubstitution error and small forecast error, specifically, the maximum and average forecasting errors are 4.62 K and 1.85 K, respectively. Finally, we combine the prediction models for all ORTWTs to create the prediction model for the ORTWT distribution. Fig. 2c demonstrates the high fidelity of the prediction model by showing that the resubstitution estimates and forecasting estimates of the ORTWT distribution are consistent with the reformer data.

6. Conclusions

In the present work, we used Bayesian statistics, BIC, sparse nonlinear regression, theories of thermal radiation and reformer geometry to develop a Bayesian model identification for the ORTWT distribution as a function of the fuel distribution and total fuel mass flow rate. We demonstrated the effectiveness of the proposed method by showing that the prediction model for the ORTWT distribution generates resubstitution estimates and forecasting estimates with small errors.

References

P. Darvishi and F. Zareie-Kordshouli. A rigorous mathematical model for online prediction of tube skin temperature in an industrial top-fired steam methane reformer. *Chemical Engineering Research and Design*, 126:32–44, 2017.

J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky. Bayesian model averaging: a tutorial. *Statistical Science*, 14:382–401, 1999.

R. E. Kass and A. E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90:773–795, 1995.

A. Tran, A. Aguirre, M. Crose, H. Durand, and P. D. Christofides. Temperature balancing in steam methane reforming furnace via an integrated CFD/data-based optimization approach. *Computers & Chemical Engineering*, 104:185–200, 2017.

A. Tran, A. Aguirre, H. Durand, M. Crose, and P. D. Christofides. CFD modeling of an industrial-scale steam methane reforming furnace. *Chemical Engineering Science*, 171:576–598, 2017.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50332-3

Study and Application of the Computing Architecture of Petrochemical Cyber-Physical System (PCPS)

Defang Li*

China Petrochemical Corporation, Beijing, 100728, China
hansheng.suo@pcitc.com

Abstract

Cyber-Physical System has become the key infrastructure supporting the development of smart manufacturing. In this paper, the computing architecture of Petrochemical Cyber-Physical System (PCPS) is defined in combination with the features of petrochemical industry. This paper divides PCPS into unit PCPS, system PCPS and System of Systems PCPS, and describes the work pattern of PCPS achieving multiscale Smart Factory business and optimization from the perspectives of system architecture and computing architecture. Finally, it introduces the application scenarios of PCPS computing architecture in the construction of Sinopec Smart Factory.

Keywords: Petrochemical Cyber-Physical System (PCPS); computing architecture; Smart Factory.

1. Background

Smart manufacturing, characterized by accelerating integration of new-generation information communication technology and manufacturing industry, has become the major trend of global manufacturing industry development (Yuan et al., 2017), and is gradually permeating into the whole process of manufacturing industry, hence promoting a great transformation in production process. Not only researchers around the world devote to smart manufacturing and have made progress both in theory and applications, governments also have given strong support. For instance, Jim Davis et al. think smart manufacturing uses the integration of next generation operations technology and information technology to realize significant untapped market opportunities, and they utilize the concept of smart manufacturing into furnace temperature balancing in steam methane reforming process (Davis, 2017; Korambath et al., 2016). Moreover, US Dept of Energy also leveraging high performance computing to drive advanced manufacturing R&D (Johnson, 2017).The core of achieving integration refers to the Cyber-Physical Systems (CPS) capable of connecting real physical world and virtual network world (Rho et al., 2016; Branicky, 2008).

China Petrochemical Corporation (Sinopec for short) took a lead in carrying out the construction of Smart Factory in 2013(Li, 2015), and put forward industry-oriented Petrochemical Cyber-Physical System (PCPS for short), taking PCPS as the infrastructure of petrochemical Smart Factory, and carrying out pilot construction in four enterprises.

By combining the construction of Sinopec Smart Factory, the system architecture and computing architecture of PCPS of petrochemical Smart Factory are further discussed in this paper, and the application practices of unit PCPS, system PCPS and System of Systems (SoS) PCPS in the construction of Sinopec Smart Factory are summarized.

2. System Architecture of PCPS

2.1. Definition of PCPS

Petrochemical Smart Factory, as the entire industry chain link oriented towards petrochemical production, deeply integrates new-generation information technology, process and equipment operation technology of existing petrochemical production process as well as people, achieves high horizontal, longitudinal and end-to-end integration of factory, improves four key capabilities of comprehensive perception, predictive warning, collaborative optimization and scientific decision-making, enhances the operation and management level of factory in a more delicate and dynamic manner, and promotes the innovation of new manufacturing and business model(Li, 2016).

PCPS as the infrastructure of petrochemical Smart Factory, aims to construct a new-generation petrochemical production environment featuring ubiquitous perception and ubiquitous smart service, and connects the ubiquitous sensor, smart hardware, control system, computing facility and information terminal into a smart network via PCPS to enable interconnectivity among enterprises, people, equipment and service, so as to develop, integrate and make use of various information resources, knowledge and wisdom to the largest extent (Wang, 2017).

2.2. PCPS system architecture

Unlike traditional pyramid layered information hierarchical architecture, PCPS adopts an architecture based on distribution. High-level PCPS, formed by interconnected integration and flexible networking of low-level PCPS, has high openness, interconnectivity and flexibility, achieving flexible expansion and ubiquitous perception of system. On each level, embedded intelligence, responsive control and predictive analysis as well as distributed computing measurement and control and predictive analysis can be deployed, and distributed computing measurement and control system and distributed edge computing system also can be deployed. Unlike strict layered structure of traditional industrial control system, the system architecture of PCPS is as shown in Figure 1.

(1) Unit PCPS

Unit PCPS, facing process production unit, process node or logistics node, can finish specific tasks through physical hardware, own embedded software system and communication module, and coordinate with other PCPSs to complete complex production and operation.

(2) System PCPS

System PCPS, facing factory, finishes the production process in the factory through the introduction of network and coordination with unit PCPS.

(3) SoS PCPS

SoS PCPS, facing the whole value chain of petrochemical industry, finishes the collaboration of multiple production lines or factories, and achieves optimal collaboration of total process and whole value chain of product lifecycle.

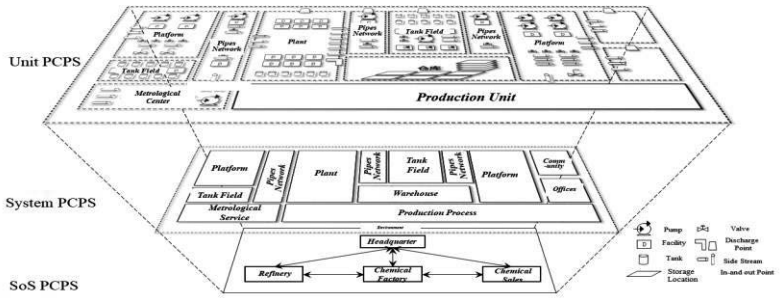


Figure 1. PCPS system architecture

2.3. PCPS unit structure

PCPS completes target, strategy, perception, computing, decision-making and execution through its basic model unit, achieves mutual collaboration, timely adjusts operational process, production process and management status, and achieves the interconnectivity, global optimization and overall coordination of devices, factories and enterprises. The model structure of basic unit of PCPS is given in this section.

The basic model structure of PCPS can be divided into two parts, namely domain layer and collaborative control layer. The domain layer is used for solving problems in professional fields, while the collaborative control layer aims to control the domain layer, enabling the domain layer to coordinate with domain layers of other PCPSs. PCPS model consists of communication network, planning goal, goal decomposition, strategy control and computing unit (perception, simulation calculation, decision-making and execution), knowledge library, database, and other modules.

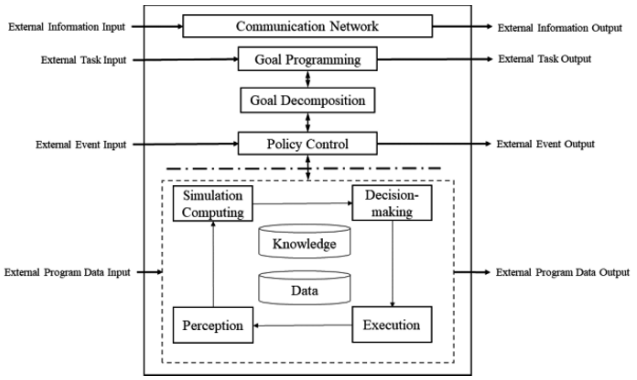


Figure 2. PCPS unit structure

PCPS system adjusts task planning through monitoring and protection, each PCPS is responsible for one specific task, and the knowledge library of each PCPS can be built more specifically, obtaining higher search efficiency. In the PCPS system, each PCPS improves its own performance through information interaction and mutual learning, thus improving the problem solving ability of entire self-adaption coordinative protection system. The coupling relationship between PCPSs is loose, so the entire system has more flexible structure and better real-time response to the changes of environment, as shown in Figure 2.

3. Computing Architecture of PCPS

3.1. Logical structure of PCPS

The logical structure of PCPS is divided into five levels, namely perception level: which perceives the data and information of physical system through control system, IoT equipment, new sensor and intelligent instrument; communication network level: which provides communication paths with service quality guarantee for devices through sensing network, industrial wireless network, wired network and other transmission forms; computing processing level: which provides centralized integration, intelligent analysis, visualization, real-time computing and other computing power, processes the data of perception level, and provides the obtained knowledge to the strategy control level; strategy control level: which serves the optimization of corporate production and operating activities, and makes strategies to control the physical system, including energy optimization, equipment operation optimization, and production coordination optimization; application level: which provides applications facing Smart Factory, smart oil and gas field, smart supply chain, and so on, and meets different production and operation demands of enterprises.

3.2. PCPS computing architecture

PCPS computing architecture includes the edge side and the cloud with each part having its own independence and the ability to achieve local autonomy. And the overall function of PCPS is realized through their mutual collaboration.

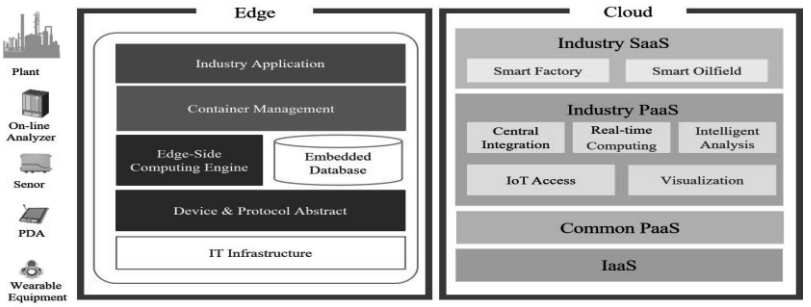


Figure 3. PCPS computing architecture

(1) PCPS computing architecture – the edge side

In a bid to decrease pressure on cloud computing, cut down storage overhead, speed up the response of enterprise field applications, and reduce the usage of network bandwidth, PCPS expands the edge layer, which boasts storage, computing, and container management capabilities, between the cloud and terminal devices, thereby deploying the timeliness-demanding computing services and related data on the edge side in form of containers. The software stack on the PCPS edge side can be embedded into industrial equipment (such as gateways, control systems, and sensors) to provide such capabilities as distributed intelligence, homogeneous software architecture, open development architecture, authentication and secure communications, and edge application stores based on container technology.

(2) PCPS computing architecture – the cloud

As a driving engine, the PCPS cloud drives deep integration of technical services such as artificial intelligence, big data, and Internet of Things, and industrial services in the

process industry to build an innovative industrial ecosystem; provides industry SaaS applications which cover smart factories and smart oil and gas fields; industrial PaaS has five core competencies, namely, centralized integration, real-time computing, intelligent analysis, Internet of Things (IoT) access, and visualization; PaaS layer provides effective support for the operation, development and maintenance of the upper level intelligent applications and services; and IaaS layer provides resources such as computing, storage and network.

4. PCPS Application Practices

According to the development characteristics of the industry and itself, Sinopec started the construction of Smart Factory in four enterprises, namely, Sinopec Beijing Yanshan Company, Sinopec Maoming Petrochemical Company, Sinopec Zhenhai Refining & Chemical Company and Sinopec Jiujiang Company in 2013, enabling the PCPS to be deeply applied to the existing petrochemical production process. The specific architecture is as shown in Figure 4 in which the PCPS is divided into three levels, namely, SoS PCPS, system PCPS, and unit PCPS. SoS PCPS is mainly responsible for decision-making and procurement optimization, the results of which are transmitted to enterprises to conduct planning and scheduling optimization. Then the planning and scheduling optimization results are broken down to each device of the enterprises (atmospheric and vacuum, catalytic cracking, continuous catalytic reforming, hydrogen cracking, etc.), so that the real-time optimization system in the unit PCPS is utilized to achieve optimization objectives. The application practices and challenges of PCPS stated in this paper are based on the actual application of four enterprises.

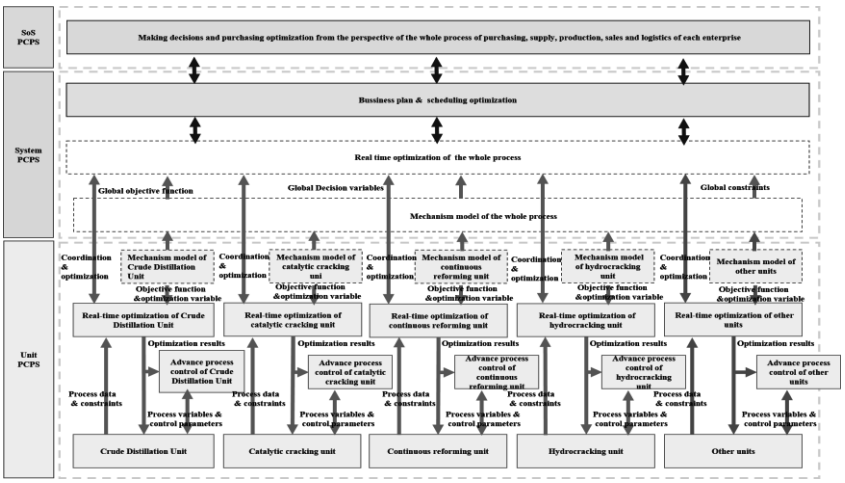


Figure 4. PCPS application architecture of smart factory which has 3 levels

The PCPS application in Smart Factory realizes the overall connection of SoS PCPS, system PCPS and unit PCPS. Sinopec Zhenhai Refining &Chemical Company is taken for instance. When market price fluctuates, SoS PCPS optimizes procurement and sales plans based on market price changes. System PCPS develops a factory-wide plan and scheduling program, which is broken down to each device. Finally, the real-time optimization system of unit PCPS calculates the optimal operating parameters of each device (atmospheric and vacuum, catalytic cracking, continuous catalytic reforming,

hydrogen cracking, etc.) according to the scheduling program and feeds back the calculation results to SoS PCPS. Where the calculation results are consistent with the procurement and sales objectives set by SoS PCPS, the advanced control system performs the optimization results of the real-time optimization system, enabling the device to operate under optimal operating parameters.

Thanks to PCPS, the ethylene unit of Sinopec Zhenhai Refining & Chemical Company achieves integrated application of Real Time Optimization and Advanced Process Control, and realizes the overall connection of SoS PCPS, system PCPS and unit PCPS based on the whole-process mechanism model. With maximum economic benefits as the objective, data reconciliation, model regression and other technologies are utilized to ensure the accuracy of the basic data; the sequence control, steady-state test and other technologies are adopted to improve the feasibility of computing results; and the price system model of chemical products is employed to ensure the timely update of raw material and product value. Without increasing the investment in major equipment, the real-time optimization system fully exerts the operation potential of the existing production facilities and effectively achieves the objectives of output increase, energy conservation and consumption reduction, thus creating new benefits for Sinopec Zhenhai Refining & Chemical Company. After the system is put into operation, the system is stable and reliable in operation. The ethylene benefit increases by RMB 26.34 per ton with the annual benefit increasing RMB 30.91 million.

5. Conclusions and Prospects

This paper presents the system architecture and computing architecture of PCPS. The application practices are conducted in four pilot enterprises of Sinopec Smart Factory with great application effects obtained. Along with the development of petrochemical industry and in-depth application research of PCPS in future, the PCPS will be applied and popularized in many pilot enterprises, and PCPS theory, application research results and experience will be extended to other industrial sectors, enabling CPS to realize more favorable development.

References

M. Branicky, 2008, CPS initiative overview, Proceedings of the IEEE/RSJ International Conference on Robotics and Cyber-Physical Systems, Washington D. C., USA: IEEE.

J. Davis, 2017, Smart manufacturing, Encyclopedia of Sustainable Technologies, (1), 417-427.

M. Johnson, 2017, Leveraging high performance computing to drive advanced manufacturing R&D at the US department of energy, IEEE International Conference on Big Data, Boston MA, USA:IEEE.

P. Korambath, J. Wang, A. Kumar, et al., 2016, A smart manufacturing use case: furnace temperature balancing in steam methane reforming process via kepler workflows. Procedia Computer Science, (80), 680-689.

D. Li, 2016, Perspective for smart factory in petrochemical industry, Computers & Chemical Engineering, (91), 136-148.

D. Li, D. Jiang, H. Suo, et al., 2015, Overview of smart factory studies in petrochemical industry, Computer Aided Chemical Engineering, (37), 71-76.

S. Rho, A. Vasilakos, W. Chen, 2016, Cyber physical systems technologies and applications, Future Generation Computer Systems, (56), 436-437.

J. Wang, 2017, Theoretical research and application of petrochemical Cyber-physical Systems, Frontiers of Engineering Management, (4), 242-255.

Z. Yuan, W. Qin, J. Zhao, 2017, Smart Manufacturing for the Oil Refining and Petrochemical Industry, Engineering, (3), 179-182.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50333-5

Research and Exploration on Intelligent Scheduling System in Petrochemical Industry

Baihua Jiang[a*], Yan Zhao, Wenyuan Gu

[a]*Petro-CyberWorks Information Technology Co., Ltd, 12th floor, Nanxincang building, No.22 Dongsi, Dongcheng district, Beijing, 100007, China*
hansheng.suo@pcitc.com

Abstract

Scheduling is not only a key component of petrochemical factory production management, but also the core of the implementation of production plan, the coordination of production activities and the maintenance of on-site safety. Based on the characteristics of petrochemical industry scheduling and the design conception of Cyber-Physical System (CPS), a new method of comprehensive application of industrial big data algorithm, expert rules engine and abnormal operation condition identification model is proposed, the architecture of the Intelligent Scheduling System, which has four major capabilities: comprehensive perception, real-time analysis, early forecasting and warning and collaborative execution, is constructed. Intelligent Scheduling System has achieved successful application in the construction of Sinopec Smart Factory. Also, good practice effect and satisfied economic benefit have been obtained.

Keywords: Petrochemical industry; intelligent scheduling system; smart factory.

1. Introduction

As the control centre of refining enterprises, production scheduling plays an important role in realizing safe, stable and efficient production of petrochemical factory. Starting with the research work of Conway, Maxwell and Miller's (1967), the theory of production scheduling has been developing for more than half a century and the satisfied achievements have been gained (Cecchini et al., 2012). The national strategy such as Made in China 2025 and Industry 4.0 has promoted further transformation and upgrading of traditional industries. Also, intelligent scheduling system in petrochemical industry has gradually realized informatization and intelligence (Huang et al., 2006), a series of base layer intelligent software such as HOLLIAS is formed.

In recent years, due to the changes of social demand, production scale, raw material composition and product structure, the existing production scheduling system in petrochemical factory cannot meet the requirements of work. The abnormal events are disposed by human experience and the emergency disposal scheme cannot be generated and pushed automatically. The lack of effective early warning mechanism makes the emergency events cannot be detected and prevented in advance. The yield cannot be predicted, and the optimal production plan cannot be chosen.

Aiming at the weaknesses of the existing system, especially the above three problems, Intelligent Scheduling System is carried out to find a solution. Based on the characteristics of petrochemical industry scheduling and the design conception of Cyber-Physical System (CPS), a new method of comprehensive application of industrial big data algorithm, expert rules engine and abnormal operation condition identification model is

proposed, the architecture of the Intelligent Scheduling System, which has four major capabilities: comprehensive perception, real-time analysis, early forecasting and warning and collaborative execution, is constructed. Intelligent Scheduling System has achieved successful application in the construction of Sinopec Smart Factory. Also, good practice effect and satisfied economic benefit have been obtained.

Nomenclature

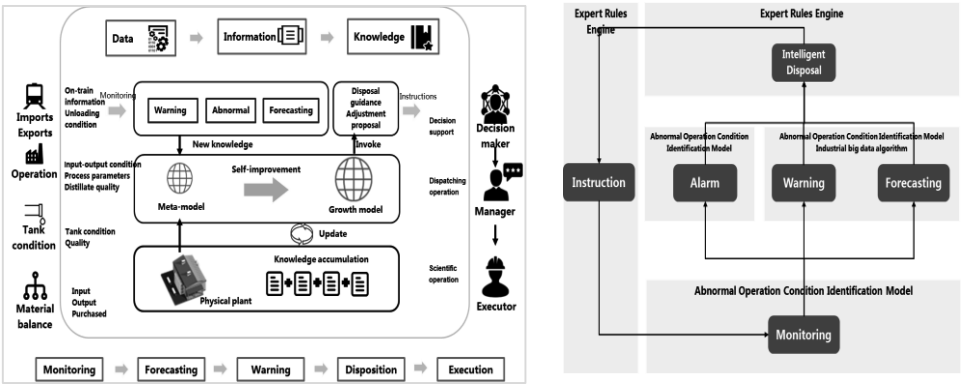
T_n	least squares method fitting real time value
n	nargin (take 12 consecutive samples in every 15 seconds)
f	frequency (s)
y	least squares method fitting function
m	slope
b	y-intercept
Y	least squares method simplified fitting function
M	slope of simplified fitting function
B	y-intercept simplified fitting function
K	the upper limit of rapid increase
Q	the lower limit of rapid decrease
H	parameter change value

2. Design concept

Studies are conducted on intelligent scheduling system based on CPS design concept. Proposed by U.S. NSF (National Science Foundation) in 2006, the design concept of CPS system is the core technology of the fourth industrial revolution, performing in-depth symmetric management on contents of physical activities through data analysis, modelling and control at virtual terminals of the network.

Proceeding with system-level CPS, this study constructs a set of closed-loop enabling system (Intelligent Scheduling System) based on automatic flows of data between information space and physical space with 4 abilities: status awareness, real-time analysis, scientific decision-making, and precise execution (Li and Qiu, 2015; Sun et al., 2016), so as to improve the scientific decision-making level and enhance the proactive problem-solving ability (Cui et al., 2012). Specifically, the identification rules base, prediction algorithm base, and early warning algorithm base are set up for petrochemical factories' management of material incoming and product delivery, tank farm management, equipment operation management, and material balance, so as to implement monitoring, early warning and prediction on the whole production process of the factory. Operation rules are refined and operation logic is solidified, with disposal rules base for abnormalities being created, so as to automatically push suggestions and assist production directors in making scientific decisions. A closed-loop management system integrating production instructions simultaneously is established, and the closed-loop management of the process from plans to instructions, execution, and feedback is initially

implemented, with the efficiency of coordinated command being enhanced. The design schematic of intelligent scheduling system is shown in figure 1.



(a)The overall design scheme (b) The application of key technology

Figure 1. The design schematic of intelligent scheduling system

3. Key technology

3.1. Industrial big data

With the process data of production equipment as the basis, the intelligent scheduling system optimizes operation process and improves quality and efficiency by exploiting the potential of equipment through utilization of big data technology plus mechanism model and algorithm analysis, and it is mainly applied in the yield prediction models which solves the problem that product yield cannot be predicted in advance (Ouyang and Jiang, 2008; Papadopoulos et al.,2012).

In this study, the industrial big data analysis technology is employed to build a prediction model for the yield of ethylene and propylene. Meanwhile, in the current production conditions, with the increase of the yield of ethylene and propylene as the target, the adjustable parameters that can be optimized are calculated so as to ensure that the yield of ethylene and propylene is the highest in current working conditions. The idea of yield prediction is shown in figure 2.

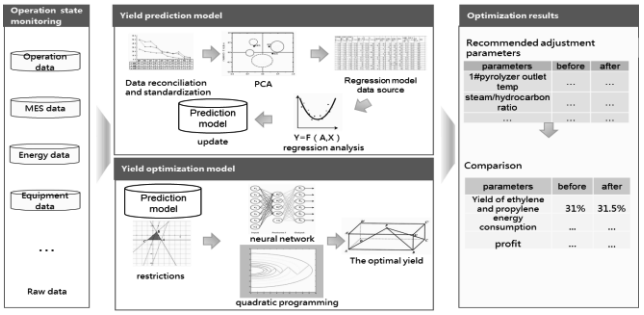


Figure 2. Yield prediction

3.2. Enterprise expert rules engine

The intelligent scheduling system constructs an intelligent disposal experience base by drawing on the experience of dispatching experts, refines business control rules, solidifies business experience logic, sets up an intelligent disposal suggestion model, supports the disposal and analysis of production alarming, early warning, and prediction, pushes scientific guidance and suggestive plans automatically, and assists dispatchers in making scientific decisions. It solves the problem that the solution cannot be automatically generated and pushed. The schematic of enterprise expert rules engine is shown in figure 3.

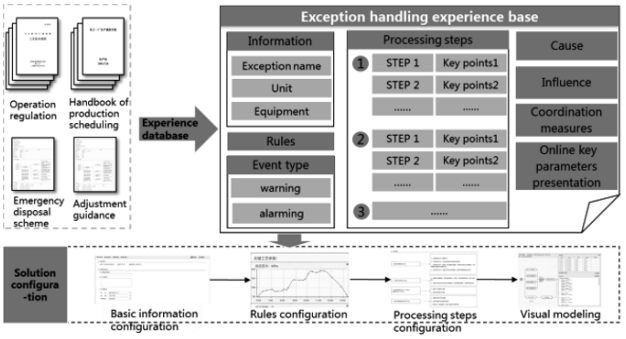


Figure 3. Enterprise expert rules engine

3.3. Abnormal condition identification model

Intelligent Scheduling System establishes abnormal working condition identification models and turns the passive monitoring mode into proactive perception of abnormal information in the factory's operation by comparing it with defined normal working conditions, identifying anomalies, and focusing on anomalies, so as to solve the problem of that emergency events cannot be detected and prevented in advance.

Alarms of operation parameters are the most common abnormal conditions in equipment operation and management (Laird et al., 2011; Hada et al.,2013), and the equipment cannot work normally when such parameters as temperature, pressure, liquid level, and flow exceed the range specified on the technique card. Through real-time monitoring over parameters, the system sets up abnormal working condition identification models to identify potential anomalies in advance, with countermeasures being pushed automatically. In the following, early warnings for rapid changes in parameters of the equipment, which is shown in figure 4, are taken as an example to introduce the abnormal working condition identification model.

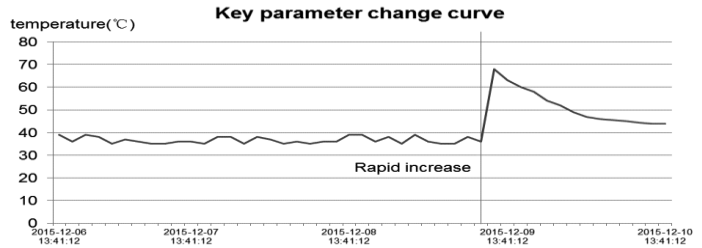


Figure 4. Early warnings for rapid changes

Model rule: When M> K, it means that the real-time value rises rapidly; and when M<Q, it means that the real-time value declines rapidly.

$$y = m_1T_1 + m_2T_2 + \cdots + m_nT_n + b \quad (1)$$

$$Y = MT + B \quad (2)$$

$$H = n \times f \times M \quad (3)$$

Intelligent countermeasures: When the system detects a rapid change in the real-time value, it will push: "XX technique parameter rises/falls rapidly by H within n * f seconds. Please check it timely."

4. Case study

Taking Sinopec Zhenhai Refining & Chemical Company as an example, the previous scheduling system failed to realize the automatic generation and push of pre-warning and disposal plan. After the successful application of Intelligent Scheduling System, real-time state perception monitoring is achieved by collecting data, data models of every factories and units as well as the abnormal operation condition identification model are set up for real-time analysis. Industrial big data analysis is used for early warning and forecasting. Also, scientific decision-making and precision execution are realized by the use of expert rules engine. The company, based on the intelligent scheduling system, has established an online instruction circulation system fully covering the production work of the entire company, and changed the management of planning, dispatching and operation from a segmented mode to a whole-process on-line and closed-loop mode, and its production management efficiency has been improved significantly, with labour time being reduced by 20 people/day and the working efficiency being increased by 15 %. The tank management pages are shown as below:

(a) Tank condition monitoring

(b) Tank abnormal disposal

Figure 5. Tank management page

5. Conclusions

Intelligent Scheduling System has been applied to 4 subsidiaries of Sinopec as demonstration pilots. There were remarkable results including the level-improving of automation, visualization and digitization in production operations management as well as the improvement of reforming promotion, quality and efficiency.

The establishment of the Intelligent Scheduling Control Centre has achieved production management delayering so that dispatching instructions can be given directly. It is also realized the cooperation of internal and external operation has been achieved. As a result,

the operation pass rate has increased to 100 % from 90.7 % and the team number in demonstration pilots has decreased by 13 % with significantly improved productivity.

Because of the application of industrial big data analysis in catalytic cracking process, through analysis of the root cause for major warnings, the warning time could be 1 minute to 2 minutes earlier before failure emerging and the operators would get enough valuable time for taking steps to avoid risks. Moreover, the application of industrial big data analysis in continuous catalytic reforming process has increased the yield of high-value-added product, with annual economic benefit being increased by more than ¥7,000,000.

Intelligent Scheduling System is the successful application of CPS design concept. It has realized information and data sharing among various levels. Moreover, material flow, cash flow and information flow have been unified as well as intelligent scheduling has been achieved within the whole plant. By building data models for factory and plant, productive process can be automated and forecasted. With these models, the optimum operation pattern that fits in current status can be obtained for production process, which lay a good foundation for effective explosion of “Prospective Unmanned Factory”.

References

R. L. Cecchini, I. Ponzoni, and J. A. Carballido,2012,Multi-objective evolutionary approaches for intelligent design of sensor networks in the petrochemical industry,Expert Systems with Applications, 39,3,2643-2649.

R. W. Conway, W. L. Maxwell, and L. W. Miller,1967,Theory of Scheduling. Addison-Wesley Publishing Co. Reading, Mass.-London-Don Mills, Ont, etc.

Q. Cu, Y. Fu, and F. Zou, 2012,Research on intelligent scheduling system in aluminum industrial production, International Conference on Automatic Control and Artificial Intelligence. IET, 1720-1723.

S. Hada, R. H. H. Iii, M. R. Eden,2013,Design of Ionic Liquids Using Property Clustering and Decomposition Techniques, Computer Aided Chemical Engineering, 32,12,955-960.

D. P. Huang, Y. Li, J. Ding, C. L. Wang, J. R. Wu, and N. B. Gao, 2006,Data Communication for Intelligent Management System of Petrochemical Distillation Device. Computing Technology & Automation, etc.

C. D. Laird, A. V. Wong, and J. Akesson, 2011, Parallel Solution of Large-Scale Dynamic Optimization Problems,Computer Aided Chemical Engineering, 29,2,813-817.

J. Li and B. H. Qiu,2015, Industrial big data : the revolutionary transformation and value creation in INDUSTRY 4.0 Era. Machinery Industry Press, etc.

F. Ouyang and H. Jiang, 2008,Exploration and Practice on Course Construction of Mathematical Model for Complex Reaction System in Petrochemical Industry, Higher Education in Chemical Engineering, etc.

A. I. Papadopoulos, M. Stijepovic, P. Linke, P. Seferlis, and S. Voutetakis,2012,Multi-level Design and Selection of Optimum Working Fluids and ORC Systems for Power and Heat Cogeneration from Low Enthalpy Renewable Sources,Computer Aided Chemical Engineering, 30,66-70.

B. L.Su, X. L, B. Wan, C. Wang, X. H. Zhou, and X. L. Chen, 2016,Definitions of predictability for Cyber Physical Systems, Journal of Systems Architecture, 63,C,48-60.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50334-7

Naphtha Pyrolysis Process Modeling Based on Ensemble Learning with LSSVM

Xianpeng Wang, Yang Zhang, Zan Wang, Lixin Tang*

Liaoning Key Laboratory of Manufacturing System and Logistics, Institute of Industrial & Systems Engineering, P.O.BOX #135, Heping District, Shenyang 110819, China
qhjytlx@mail.neu.edu.cn

Abstract

Ethylene and propylene are the most important basic resources for chemical industry, and the demand of them dramatically increases worldwide. However, the pyrolysis process is too complex to build a precise mechanism model. Traditional mechanism models based on molecular reactions are also very complex and one single simulation of them is generally time-consuming. In addition, the off-line lab test of yields of the components of outputs has great time delay. So it is very difficult to achieve a precise closed-loop control for pyrolysis process because the yields of ethylene and propylene cannot be tested online or predicted precisely with traditional mechanism models. To handle this problem, this paper tries to establish some data-driven models for the pyrolysis process based on ensemble learning methods, such as AdaBoost.R2 and AdaBoost.RT, and process data. The parameters used in the ensemble learning models are determined by the evolutionary algorithm instead of traditional method based on experience. Their performances are evaluated based on practical process data.

Keywords: pyrolysis process, yield prediction, ensemble learning method, evolutionary algorithm

1. Introduction

1.1. Naphtha pyrolysis process

As one of the most important monomers in the petrochemical industry, ethylene is the base for organic material productions. The key equipment for the naphtha pyrolysis in the ethylene plant is the cracking furnace, whose schematic production process is illustrated in Fig. 1. First, the naphtha feed is put into the convection section and preheated, and then the dilution steam gets in and mixes with the heated naphtha feed. After completely vaporized in the convection section, their mixture enters the U-type cracking coil of the radiant heating section and the pyrolysis reaction occurs immediately. Finally, the products mainly consisting of ethylene and propylene flow out to the subsequent sections to be separated. In the operation of the cracking furnace, the key control parameters (or variables) are the coil outlet temperature (COT), the steam-to-naphtha ratio, and the outlet pressure, because they are closely related to the reaction depth that determines the yielding rate of ethylene and propylene. The main task of an operator is to determine the real-time values of these key control variables so as to maximize the yields of ethylene and propylene. However, in most pyrolysis processes there is no online testing equipment for the yields of ethylene and propylene. The output products need to be sent to the lab for precise test of the yield of each component. That is, the off-line lab test of yields of the components of final outputs has great time delay.

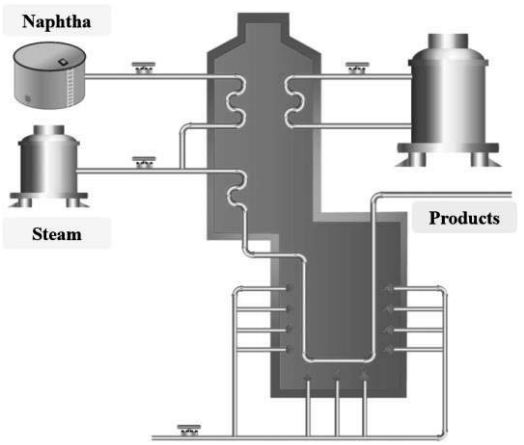


Figure 1. Schematic production process of the steam cracking furnace

Although there have been many kinds of reduced mechanism model based on molecular reactions for the pyrolysis process, there is no reaction kinetics model that can precisely describe the pyrolysis process because once activated thousands of reactions will occur concurrently in the cracking coil at a very short time. Moreover, one single simulation of the traditional mechanism models based on molecular reactions is also very time-consuming. So it is very difficult to achieve a precise closed-loop control for pyrolysis process because the yields of ethylene and propylene cannot be tested online or predicted precisely with traditional mechanism models.

1.2. Literature review

Though there have been many literatures on the modeling of the naphtha pyrolysis, few papers focused on the yield prediction of the naphtha pyrolysis process.

With respect to the mechanism modelling methods, they can be classified into three categories according to the review of Gao *et al*. (2008), namely the free-radical reaction kinetics model, the molecular reaction kinetics and the experiential models. Although the free-radical reaction kinetics model, e.g., Joo and Park (2001), had the most precise yield prediction, it was too time-consuming to be used in practical real-time control. The experiential models such as Hao *et al*. (1999) were less complex, but they were based on experience, especially for specific kinds of naphtha feed and cracking furnaces. The molecular kinetics models, e.g., Kumar and Kunzru (1985), were often adopted in practical production, however, one single run of such models was still very time-consuming.

1.3. Contribution of this paper

To handle the problem that traditional mechanism models cannot be used to establish real-time optimization or closed-loop control for the pyrolysis process, this paper tries to establish the data-driven models for the pyrolysis process based on ensemble learning methods, such as AdaBoost.R2 and AdaBoost.RT. Since the parameters used in these modeling methods have great impact on the prediction precision, we prefer to use evolutionary algorithm, instead of traditional method based on experience, to obtain the appropriate parameter settings. Finally, the performances of the two kinds of ensemble learning methods are evaluated based on practical process data.

2. Sub-learning machine based on LSSVM

For the sub-learning machine in the ensemble learning modelling, we prefer to adopt the least square support vector machine (LSSVM) because it has good balance of simplicity and precision. In the LSSVM, the key parameters are the regularization parameter γ and the kernel width parameter σ. Since the two key parameters have significant impact on the performance of the LSSVM model, in this paper we adopt an improved particle swarm optimization (IPSO) to determine their values. Different from the standard PSO, the main feature of the IPSO is that it incorporates an elite set consisting of the best solutions found in search history to guide the search direction of particles and replace non-promising particles in the population to improve the search diversity of PSO.

In the IPSO, a solution is represented as $X_i=(\gamma_i, \sigma_i)$ and the fitness of each particle is calculated as $fitness(X_i)=\sqrt{\sum_{k=1}^{l}(\hat{y}_k-y_k)^2/l}$, where $\hat{y}_k$ is the prediction value of the output (yield of ethylene or propylene) according to sample data k consisting of the input (the steam-to-naphtha ratio, the coil outlet temperature, and the outlet pressure).

As mentioned above, the elite set has two functions: (1) guiding the search direction of particles and (2) replacing nonpromising particles in the population. The first function is implemented by revising the update equations of particles as follows.

$$v_{ij}^{t+1}=w\cdot v_{ij}^{t}+c_1r_1\cdot(p_{ij}^{t}-x_{ij}^{t})+c_2r_2\cdot(g_j^{t}-x_{ij}^{t})+c_3r_3(e_j^{t}-x_{ij}^{t}) \tag{1}$$

$$x_{ij}^{t+1}=x_{ij}^{t}+v_{ij}^{t+1} \tag{2}$$

where w is the inertia parameter, c_1, c_2 and c_3 are respectively the cognitive, social and elite learning parameters, r_1, r_2 and r_3 are random numbers uniformly distributed in (0,1), p_{ij}^{t} is the jth position value of the personal best, g_j^{t} is the jth position value of the global best, and e_j^{t} is the jth position value of a solution randomly selected from the elite set. Since the elite solutions are the best 20 solutions found in the search history and they are randomly selected in implementation, the incorporation of the elite set can help to improve the search diversity of PSO while keeping good solution quality.

3. Experimental results of AdaBoost based on LSSVM

Ensemble learning is one of the hottest research areas in machine learning because it can enforce the prediction precision or classification precision of several learning machines, even if they are weak learning machines. Besides, the experimental results showed that the ensemble of several sub-learning machines could significantly improve the robustness of the prediction or classification model.

In this section, we combined the LSSVM in the two famous versions of AdaBoost, i.e., AdaBoost.R2 and AdaBoost.RT, and tested their performance for the modelling of the pyrolysis process to predict the yields of ethylene and propylene. In the two ensemble learning methods, a set of 5 sub-learning machines (i.e., LSSVMs) were used.

In the experiment, we collected a total of 108 data samples, among which 88 samples randomly selected from the 108 samples were used as training data while the left 20 samples were used as the testing data to verify the performance of the two ensemble learning methods.

3.1. AdaBoost.R2 based on LSSVM for yields prediction

The yield prediction of ethylene for the training and testing data were given in Figure 2 and Figure 3, respectively. And the yield prediction of propylene for the training and testing data were given in Figure 4 and Figure 5, respectively.

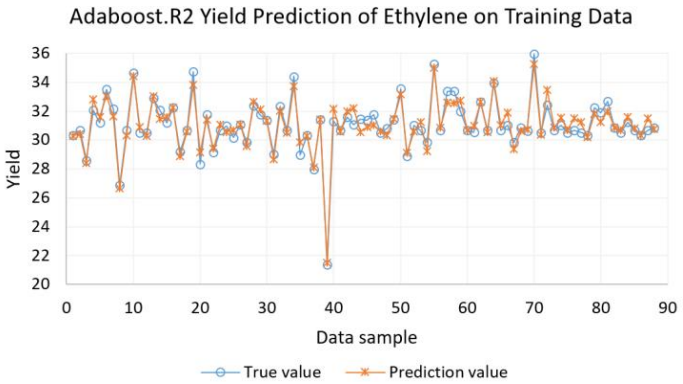


Figure 2. Yield prediction of ethylene with AdaBoost.R2 on training data

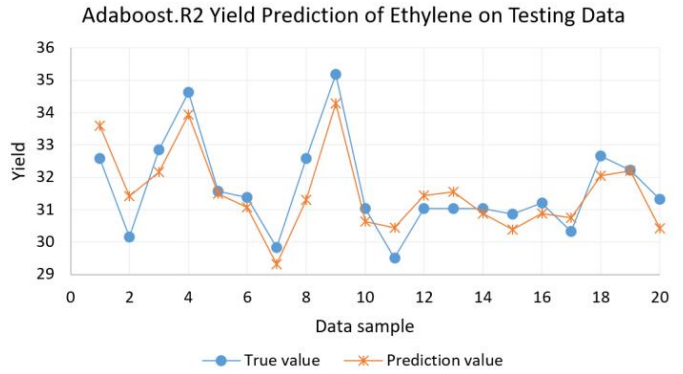


Figure 3. Yield prediction of ethylene with AdaBoost.R2 on testing data

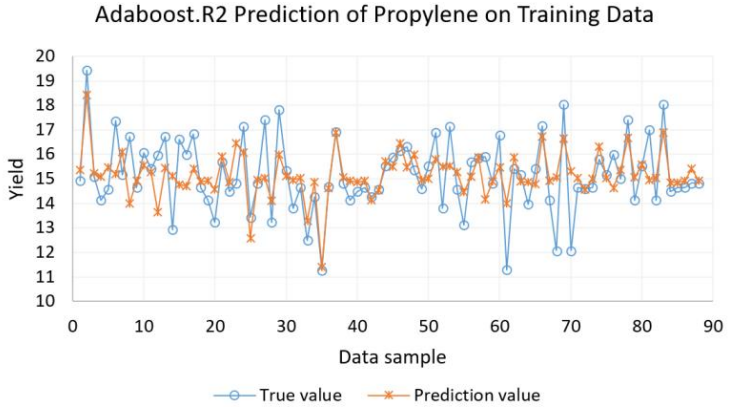


Figure 4. Yield prediction of propylene with AdaBoost.R2 on training data

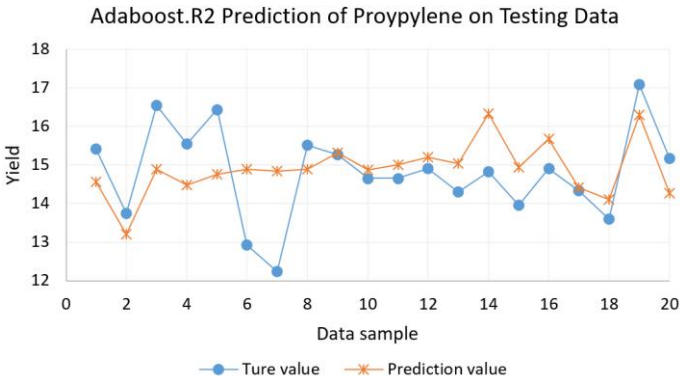


Figure 5. Yield prediction of propylene with AdaBoost.R2 on testing data

3.2. AdaBoost.RT based on LSSVM for yields prediction

The yield prediction of ethylene for the training and testing data were given in Figure 6 and Figure 7, respectively. And the yield prediction of propylene for the training and testing data were given in Figure 8.

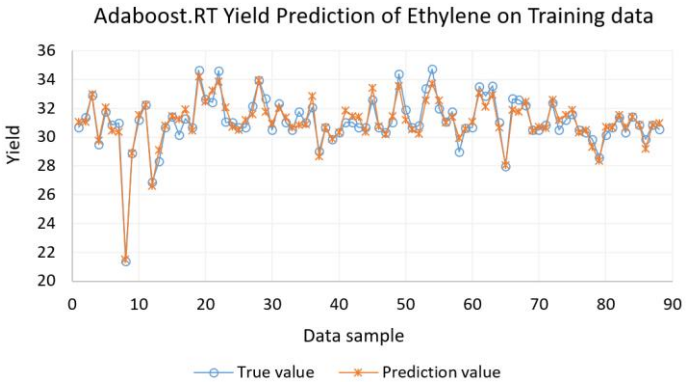


Figure 6. Yield prediction of ethylene with AdaBoost.RT on training data

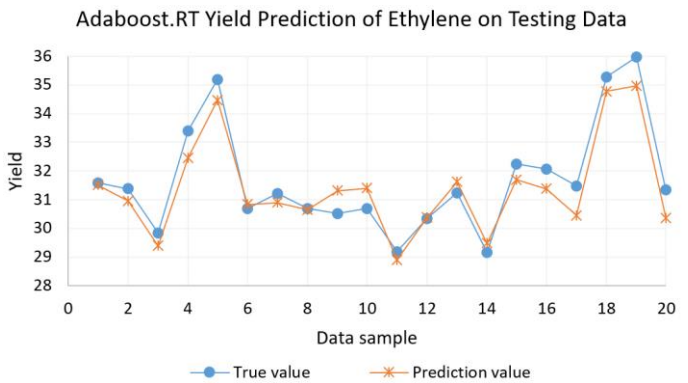


Figure 7. Yield prediction of ethylene with AdaBoost.RT on testing data

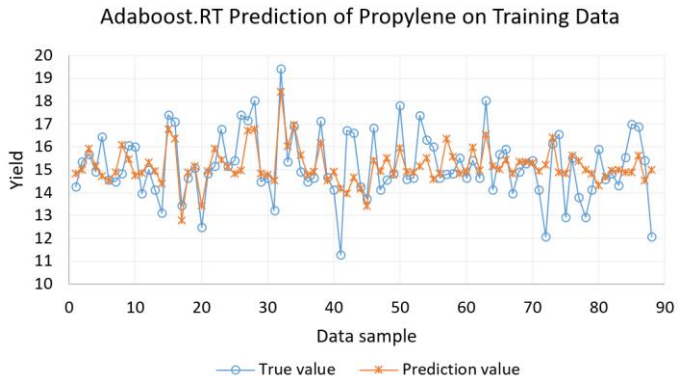


Figure 8. Yield prediction of propylene with AdaBoost.RT on training data

4. Conclusions

In this paper, two kinds of ensemble learning methods based on LSSVM were tested for the yield prediction of pyrolysis. The results show that the two models can both predict the yields precisely and the AdaBoost.R2 is a little better. These models can be used in the online prediction and real-time optimization of the pyrolysis process.

Acknowledgements

This research was supported by the National Key Research and Development Program of China (2016YFB0901900), the Major Program of National Natural Science Foundation of China (71790614), the Fund for the National Natural Science Foundation of China (61374203, 61573086), the Fund for Innovative Research Groups of the National Natural Science Foundation of China (71621061), the Major International Joint Research Project of the National Natural Science Foundation of China (71520107004), and the 111 Project (B16009).

References

X.D. Gao, B.Z. Chen, X.R. He, T. Qiu, J.C. Li, C.M. Wang, L.J. Zhang. Multi-objective optimization for the periodic operation of the naphtha pyrolysis process using a new parallel hybrid algorithm combining NSGA-II with SQP. *Comput Chem Eng* 2008; 32: 2801-2811.

H. Hao, G.H. Xiong, S.F. Zhang, B.Z. Chen. The calculation of propylene yield by the thermal cracking of hydrocarbon. *Chem Eng* 1999; 27: 54-56.

E. Joo, S. Park. Pyrolysis reaction mechanism for industrial naphtha cracking furnaces. *Ind Eng Chem Res* 2001; 40: 2409-2415.

P. Kumar, D. Kunzru. Modeling of naphtha pyrolysis. *Ind Eng Chem Process Des Dev* 1985; 24: 774-782.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50335-9

Online System Identification for the Real Time Control of the Plasma Parameters

Junmo Koo[a], Damdae Park[a], Sangwon Ryu[b], Gon-Ho Kim[b], Youn-Woo Lee[a*]

[a]*School of Chemical and Biological Engineering, Seoul National University, Build.302, 1, Gwanak-ro, Gwanak-gu, Seoul (08826), Republic of Korea*
[b]*Department of Nuclear Engineering, Seoul National University, Build.30, 1, Gwanak-ro, Gwanak-gu, Seoul (08826), Republic of Korea*
ywlee@snu.ac.kr

Abstract

In recent years, the development of AI has been accelerated so that the semiconductor manufacturing process has also been regarded as a very important issue. Many research results related to the plasma etch process which is one of the most important processes in the semiconductor manufacturing process have been presented in various forms. Recently, our research group has also presented a case of successful real-time multivariable control of plasma parameters based on optical emission spectroscopy.

Although the above results are expected to bring about positive effects in various fields, they do not consider the variability of the system. Plasma related systems are so sensitive that online system identification is inevitable. In this paper, we performed the online system identification of the system composed of the plasma parameter through recursive parameter estimation method. Through the recursive algorithm using forgetting factor and input and output information which was gathered in real-time, we detected that the system was changing in real-time and updated the model parameters. We expect the results introduced in this paper contributes not only to the semiconductor manufacturing process but also to all the processes using plasma.

Keywords: Online System Identification, Recursive Parameter Estimation, Real-time Control, Plasma Parameter Control

1. Introduction

Due to the ever smaller feature size according to Moore's law, the semiconductor industry is demanding higher quality control. In addition, the development of AI has been accelerated so that the semiconductor manufacturing process has also been regarded as a very important issue. Many research results related to the plasma etch process which is one of the most important processes in the semiconductor manufacturing process have been presented in various forms. Recently, our research group has also presented a case of successful real-time multivariable control of plasma parameters based on optical emission spectroscopy (OES).

Even though those studies show excellent results, it is difficult to directly apply them to the semiconductor manufacturing environment for a variety of reasons. One of the reasons is that real-time detection of system changes is essential in systems dealing with sensitive materials such as plasma, but some studies have been carried out without consideration of them. Especially, in the previous studies conducting the plasma control through model predictive control (MPC), which is widely used in recent years, the importance of the

online system identification should not have been excluded because MPC has a high model dependency.

In this paper, we performed the online system identification of the system composed of the plasma parameter through recursive parameter estimation method. We utilized an OES based single input single output (SISO) plasma parameter controller. OES is a default plasma monitoring sensor for every plasma reactor so that there is no concern about installation. In addition, OES, a non-invasive tool, has the advantage of less impact on the system than other invasive sensors such as hairpin probe. This study is of great value as the first attempt to control the plasma parameter with non-invasive sensor and online system identification.

This paper is organized as follows. In Section 2, the experiment environment and the system definition of our electron density control system are explained briefly where the monitoring sensor is OES. In Section 3, the theoretical backgrounds which contain a recursive parameter estimation algorithm and the Box-Jenkins model structure used in our system model are explained. In Section 4, the results of the online system identification are illustrated, along with the MPC control results with the unchanged model applied.

2. Estimation of the plasma parameters from OES: System definition

In plasma etch, it is important to control plasma parameters such as electron density, electron temperature and so on. There have been many studies and literature dealing with the control of plasma parameters. Our research group have also published a paper about the control of plasma parameters monitored by optical emission microscopy (OES) which is a non-invasive diagnostic tool. Optical fiber attached on the viewport of the chambers measures the emission in plasmas. An emission intensity from a particular p^{th} state to k^{th} state is described as

$$\Phi_{pk} = n_0 n_e \int_{E_{thr}}^{\infty} \sigma_{pk}(E) \left(\frac{2E}{m_e}\right)^{\frac{1}{2}} f(E) dE \tag{1}$$

where n_0 is the number density of ground state atoms, n_e is the electron density, $\sigma_{pk}(E)$ is the excitation cross section from level p into level k as a function of electron energy E, $f(E)$ is the electron energy distribution function (EEDF), m_e is the electron mass.

In order to measure plasma parameters with OES, it would be useful to have a separate line-ratio techniques that is only a function of the electron density (J. B. Boffard et al., 2004). In a similar way, we utilized several emission lines to measure electron density.

In our experiment, we used a capacitively coupled plasma reactor that can carry 300 mm wafers. It is powered by three radio frequency (RF) generators and the gap betwen top and bottom electrodes is 25 mm. The reference plasma condition is 20 mT of pressure, 400 sccm of Ar flow rate and 16 sccm of SF_6 flow rate. The input variable defined in our system is 60 MHz RF power and the output variable is electron density. With OES and Eq. (1), the electron density is measured in real-time.

3. Recursive parameter estimation for Box-Jenkins model

The estimated model parameter that minimizes the weighted error criterion is as follows:

$$\widehat{\theta_t} = \underset{\theta}{argmin} \sum_{k=1}^{t} \beta(t,k)[y(k) - \varphi^T(k)\theta] \tag{2}$$

where $\widehat{\theta_t}$ is the estimated model parameter at time *t*, $y(k)$ is the measured output, $\varphi(k)$ is the regression vector, θ is the present model parameter and $\beta(t,k)$ is the weighting sequence which has the following property:

$$\beta(t,k) = \lambda(t)\beta(t-1,k), \qquad 0 \leq k \leq t-1$$

$$\beta(t,t) = 1 \tag{3}$$

where $\lambda(t)$ is the forgetting factor the value range of which is recommended to be between 0.98 and 0.995. When solving Eq.(2), we have

$$\widehat{\theta_t} = \hat{\theta}_{t-1} + \bar{R}^{-1}(t)\varphi(t)[y(t) - \varphi^T(t)\hat{\theta}_{t-1}] \tag{4.a}$$

$$\bar{R}(t) = \lambda(t)\bar{R}(t-1) + \varphi(t)\varphi^T(t) \tag{4.b}$$

which is a recursive algorithm where

$$\bar{R}(t) = \sum_{k=1}^{t} \beta(t,k)\,\varphi(k)\varphi^T(k) \tag{5}$$

The model structure we estimated in real-time is a linear, time-varying, discrete-time Box-Jenkins (BJ) polynomial model which can be generally described by

$$y(k) = \frac{B(k,z)}{F(k,z)}u(k) + \frac{C(k,z)}{D(k,z)}e(k) \tag{6}$$

where $u(k)$ is the input variable, $e(k)$ is a zero-mean, white noise with variance σ^2 and $B(k,q), C(k,q), F(k,q)$ and $D(k,q)$ are model parameters. The model parameters are calculated in real-time based on Eq. (2) ~ (5). Specifically in our system, $u(k)$ represents 60 MHz RF power and $y(k)$ represents the electron density. The recursive parameter estimation is supported in the MATLAB's commands and SIMULINK.

4. Results and Discussion

In order to evaluate the performance of the recursive parameter estimation algorithm, it is reasonable to compare the situation where the output variable is controlled. The setpoint tracking test of electron density was performed. The setpoint was calculated through the wall fluorination which is one of the most popular performance variable in the semiconductor etch process. The controller for the electron density was designed in MPC method which has been widely used as an advanced process controller recently.

First of all, the setpoint value of the electron density was calculated, then the MPC was activated to track the setpoint with manipulating the 60 MHz RF power. While this control operation was being operated, the model parameters were not changed. The unchanged

model (discrete-time output error (OE) model structure) was obtained by referring to previous research results (J. Koo et al., 2017). As shown in figure 1, the electron density setpoint tracking results were biased by distinct errors.

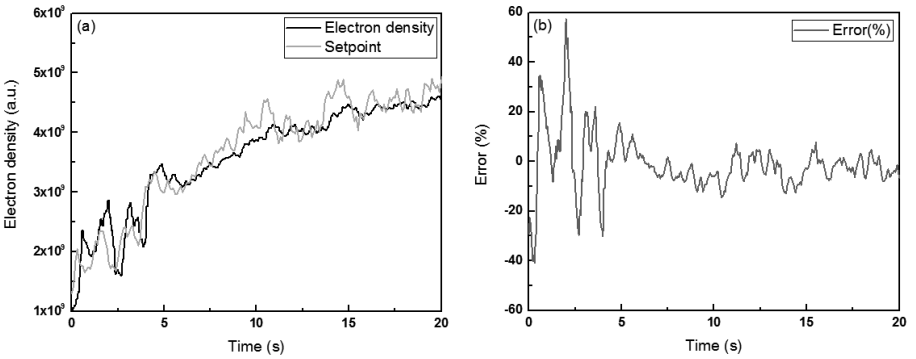


Figure 1. The setpoint tracking results with MPC controller applying constant model parameters (a) and the error percentage of the control results.

The bias can occur for a variety of reasons such as drift of the system caused by fluorine spread from the etcher wall, OES measurement drift error, system linearity issue and so on. The results clearly show that online parameter estimation of the system is inevitable.

The recursive parameter estimation algorithm, Eq. (2) ~ (5), was then applied to the same dataset. The dataset, electron density and 60 MHz RF power for $y(k)$ and $u(k)$ respectively, was used to the model parameter estimation simulation. The initial value of the each parameter was set from the unchanged model. That is, B_0 and F_0 had specific value from the unchanged model and C_0 and D_0 were 1 at time 0s. The forgetting factor, $\lambda(t)$, was set to be 0.995. The recursive parameter estimation simulation results are shown in figure 2.

The subscript number after the parameter in figure 2 refers to the degree of the parameter plus 1, that is, B_8 means $B(z) = b_8 z^{-7}$ and other parameters have the structure like $C(z) = c_1 + c_2 z^{-1}$, $F(z) = f_1 + f_2 z^{-1}$ and $D(z) = d_1 + d_2 z^{-1}$. All of the parameters b_1 to b_7 are 0 because of the time delay issue of the system.

The comparison of the measured electron density and the estimated electron density from the recursive parameter estimation algorithm are shown in figure 3.

It is clear that the electron density estimates in figure 3-(a) are fitting the actual measurements perfectly. The error percentage of that is shown in figure 3-(b) which shows nearly 0 % (within +/- 2 %) after 5 s. Within 5 s, there is an inaccuracy because the number of data for parameter estimation is small.

Based on these results, the plasma parameter control can be expected to be much more effective and accurate when an adaptive MPC controller with online system identification is implemented.

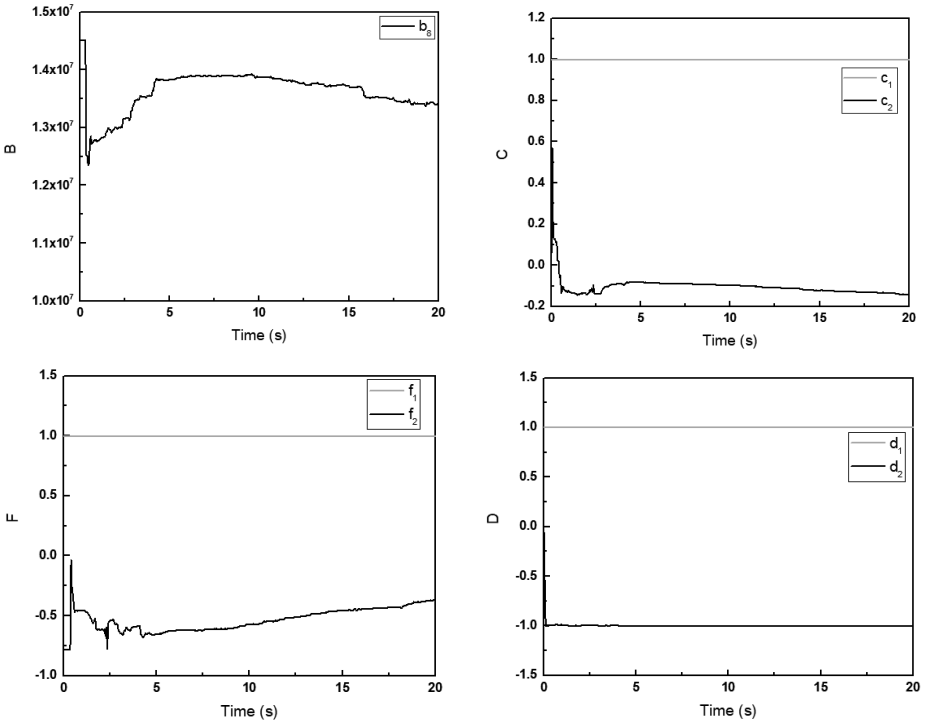


Figure 2. The model parameters estimated from the recursive parameter estimation algorithm.

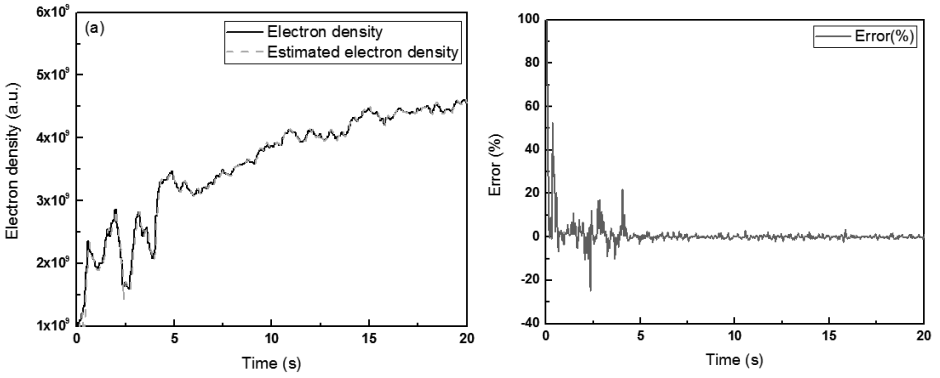


Figure 3. (a) The comparison of the measured electron density and the estimated electron density calculated by the recursive parameter estimation algorithm and (b) error percentage of that.

5. Conclusions

Through this research, we performed the online system identification to improve the efficiency of real-time control of the plasma parameters. The control objective was tracking the setpoint of the electron density in Ar and SF_6 plasma condition by manipulating 60 MHz RF power which was calculated by MPC controller. In the control results without online system identification, there was bias caused by several reasons such

as system drift, measurement drift and so on. The results justified the necessity of online system identification. The online system identification simulation of Box-Jenkins polynomial model was then performed using the dataset used in the control with the constant model parameters. As a result, it was confirmed that the estimation of the electron density follows the actual values well. Its error percentage was just only 2% which means the control performance will be maximized when applied to the adaptive MPC controller. These results are of great value as the first attempt to perform online system identification of the plasma parameter system using OES and recursive parameter estimation in the semiconductor manufacturing process. We hope that these successful results of applying established methodologies to the plasma related systems will contribute to maximize the sophistication and the efficiency of all processes using plasma as well as semiconductor manufacturing processes.

Acknowledgement

This research was respectfully supported by the IT R&D program of MOTIE/KEIT (10049155, Development of equipment control algorithm based on plasma monitoring for efficiency improvement of 10 nm etch process) and by Engineering Development Research Center (EDRC) funded by the Ministry of Trade, Industry & Energy (MOTIE). (No. N0000990)

References

D. E. Seborg, T.F. Edgar, D.A. Mellichamp, 2004, Process Dynamics and Control, 2nd ed., John Wiley & Sons

D. L. Guarin and R. E. Kearney, 2016, Identification of a time-varying, box-jenkins model of intrinsic joint compliance, IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 25, no. 8, pp. 1211–1220

E. J. Klein, W. F. Ramirez, 2002, Optimal parameter estimation for the online control of ion milling etch depth. AIChE journal, *48*(2), 311-323.

J. B. Boffard, C.C. Lin, C.A. DeJoseph Jr, 2004, Application of excitation crosssections to optical plasma diagnostics, J. Phys. D: Appl. Phys. 37, R143.

J. Koo, D. Ha, D. Park, H. Roh, S. Ryu, G. Kim, K.H. Baek, C. Han, 2017, Design of optical emission spectroscopy based plasma parameter controller for real-time advanced equipment control, Computers & Chemical Engineering, 100, 38-47

L. Ljung, 1999, System Identification: Theory for the User, 2nd ed., Upper Saddle River, NJ, USA: Prentice-Hall.

L. Ljung, 2009, Experiments with identification of continuous time models. IFAC Proceedings Volumes, 42(10), 1175-1180

V. Laurain, M. Gilson, R. Tóth, and H. Garnier, 2010, Refined instrumental variable methods for identification of LPV Box–Jenkins models, Automatica, vol. 46, no. 6, pp. 959–967

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50336-0

An Ontology Based Cyber-infrastructure for the Development of Smart Eco Industrial Parks

Aravind Devanand[a], Li Zhou[a], Iftekhar A. Karimi[a*], Markus Kraft[b]

[a]*Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117585*
[b]*Department of Chemical Engineering and Biotechnology, University of Cambridge, New Museums Site, Pembroke Street, Cambridge, CB2 3RA, United Kingdom*
cheiak@nus.edu.sg

Abstract
This paper provides insight into the methodology employed for the development of a smart system called the J-Park simulator (JPS). JPS provides a virtual representation of an Eco-Industrial Park (EIP) and contains information pertaining to various aspects of an EIP. An ontology based approach is used to store this information in a structured and machine-readable form. Ontology represents data in a structured form and can be used for the creation of knowledge bases. These knowledge bases can then be utilised to conduct case studies on JPS. In one such case study the most suitable locations for modular nuclear plants in an EIP is identified. The architecture for the case study and how it can be implemented in JPS is demonstrated in the paper.

Keywords: ontology, knowledge base, modular nuclear power

1. Introduction

The ever-increasing price, demand and environmental cost for energy have driven mankind invest into developing more efficient and economical ways for utilizing energy. This line of thinking has led to the formation of industrial clusters called Eco Industrial Parks which achieves higher economic and environmental performance through collaboration over a wide range of sectors like energy, water, material etc. This concept maximizes the collective benefit of the cluster rather than focus on the profits of individual entities.

The functioning and efficiency of an EIP could be significantly improved by applying the concepts of Industry 4.0. J-Park simulator (JPS) (Zhou et al., 2017) is envisaged as a smart system which provides a virtual representation of an Eco Industrial Park (EIP) and contains information pertaining to its various aspects. The current version of JPS is inspired from the industrial park in Jurong island, Singapore. JPS studies the interactions between various domains that exist in an EIP through a variety of models. The sheer quantity of data and the complex nature of interactions make this a daunting task. We have used an ontology-based knowledge base for overcoming this difficulty. Ontology is a formal and explicit representation of knowledge in a domain expressed in a machine-readable form. It behaves as a road map with several interconnected nodes which makes data access efficient. Data pertaining to each physical object in the EIP could be represented by a virtual 'avatar' which is capable of communicating with each other. This paper focuses on the power generation aspect of an EIP and provides a case study which demonstrates the capabilities of JPS.

Power generation facilities are an integral part of an EIP. Conventional power generation technologies contribute significantly to global carbon emissions, so it stands against the design principles of an EIP. These emissions could be reduced drastically by employing greener alternate technologies like nuclear power. In the case study, we have considered a scenario in which the entire power demand for the EIP is supplied from modular nuclear power plants which are located strategically. The locations of the plants should be near the consumer locations so that transmission losses can be reduced. The major concern about nuclear power is nuclear fallout in case of an accident. This concern is quantified by conducting a risk analysis that takes into consideration a variety of factors like population density, cooling water availability etc. Based on these parameters a multi-objective MINLP model was formulated in GAMS. The model is then implemented in JPS where it conducts a risk benefit analysis on all available land sites and returns the locations of the most suitable ones. The data required for the model were taken from the knowledge bases through the usage of SPARQL queries. Once the model is executed the results are updated in the knowledge base and the locations are displayed in a Google map. The entire process is automated through the use of a series of agents capable of performing specific tasks.

2. Knowledge base development for power plants

2.1. OntoEIP

OntoCAPE is an existing ontology for the computer aided process engineering domain (Morbach et al.,2007). OntoCAPE is capable of representing knowledge in the chemical engineering domain. However, an EIP contains many other critical aspects, such as electrical engineering, transport system, civil engineering etc. Consequently, it is necessary to extend OntoCAPE to represent other domain knowledge, further to indicate cross-domain interdependencies. This extension of OntoCAPE ontology is called OntoEIP.

OntoEIP uses the concept of abstraction layering to represent concepts wherein the ontology is divided into different layers. Each level contains information required for its representation and all other information are passed to lower levels. The large-scale EIP system is decomposed into subsystems by applying a bottom up approach where the lower levels reflect information of more specific nature. In our case, an EIP is modelled by a five-layered system model with Industrial network, Plant network, Process, Unit and molecular levels arranged top-down. The industrial network level has a broad information about the electrical, energy, material and road network of an EIP while the plant level will have information specific to a plant. As we go down the layers the information becomes more detail-oriented and more specific to the application. For example, the molecular level will have information regarding the chemical structure and interactions between molecules.

2.2. Extension of OntoEIP for power plants

The upper levels of the OntoEIP ontology provides a framework for describing most of the objects and properties in an EIP. More specific details relevant to a sector have to be added to the lower layers separately. In the case of a power plant, it is defined as a sub-class of the technical system in the existing ontology. This power plant technical system has different aspect systems. Each of the aspect system limits the information in it to a particular view point. The realization aspect takes into consideration the physical and

design properties of a power plant. The function aspect considers the processes and the unit operations happening inside the plant. Finally, the performance aspect has different performance evaluators like carbon emissions, cost comparisons etc. to analyse the performance of a plant. It is important to ensure that the new classes and properties are consistent with the existing ontology. Also, the classes and properties should be unique and should not be presented elsewhere in the ontology. This ontology was developed in Protege which uses the Ontology Web language (OWL) and it's an extension of the OntoEIP ontology.

2.3. Creation of knowledge base

The ontology described above is capable of holding information about power plants. An instance of the power plant ontology provides information about a particular plant that is under consideration. An ontology with a set of instances can be called as a ontological knowledge base as it represents knowledge in a structured manner with classes, properties and instances (Zhang et al., 2017).In this study, the information used for the knowledge base creation was taken from the global energy observatory website (http://globalenergyobservatory.org/). A python script was used to mine the data from the website and the apache Jena library of java was used to add it into the ontology. A set of OWL files were created using the java code and each file has information regarding one power generation facility. These ontologies were then published online to keep the information decentralized.

The power plant knowledge base is then visualised using the google map java script API. The final visualization of the website is available at www.theworldavatar:82/ppalt and is shown in the figure below. It is a google map with location markers for all the power plants in the world. A single click on any power plant will reveal design information of that power plant while a double click will open a new tab with the ontological knowledge base of the power plant displayed in the xml schema format. Similar to this, the JPS has knowledge bases for chemical plants, road network etc.

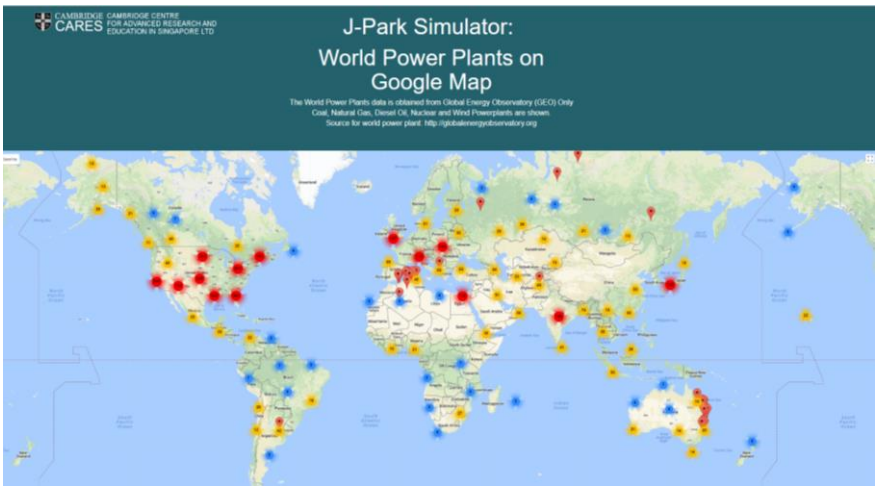


Figure 1 World power plants knowledge base

2.4. Architecture for case study in JPS

The knowledge bases created for representing the information in an EIP can be utilised for several purposes. One possible application is conducting case studies to analyse the

outputs from various scenarios that might occur in an EIP. These case studies should comply with the design principles of the JPS (Kleinelanghorst et al., 2016) so that it can be implemented easily. The architecture consists of agents capable of acquiring and changing the data presented in the knowledge base. Figure 2 shows how the case study can be implemented in the JPS.

Any case study or scenario analysis will have a model which can predict the output of that scenario. This model requires information that has to be acquired from the knowledge bases. An Input/output agent has the capability to search through the knowledge bases and acquire relevant information for the model. This I/O agent can be a SPARQL query, which is a query language developed for RDF framework. Once the information is fed to the model, it will be executed, and the output will be generated. This output has to be updated in the ontological knowledge bases. This is done by another agent which is a JavaScript that utilises the apache Jena library. Finally, the output will be visualized on the website using various tools.

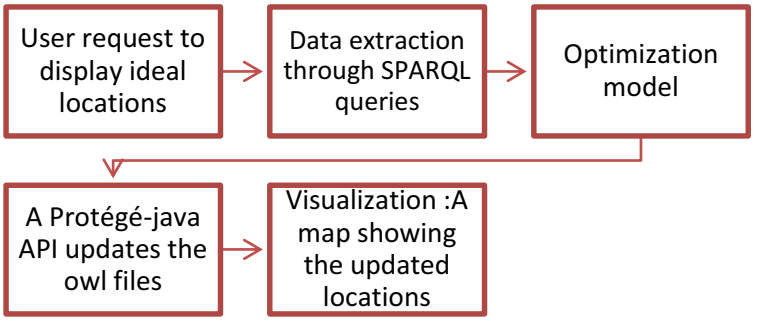


Figure 2 Architecture for case studies

3. Case study

3.1. Problem statement

The objective of the problem is to design a standalone power system for an EIP. An EIP consists of several industries and it is spread over a large area. Most of these EIPs are located in remote areas away from residential sites. A standalone power system can deliver power to the EIP with better efficiency and lower losses than from connecting it to an external grid. The entire area of an EIP can be divided into two categories. The first category includes all the areas which consume power like industries, storage facilities etc. These are defined as load points in this study. The second category includes the empty areas like unused spaces, roads, gardens etc. These are available sites for new power plants and are called potential sites. As a result, the entire area of the EIP can be divided into grids and each lot in the grid represents either a load point or a potential site. The suitability of a potential sites depends on a lot of factors like the population density, cooling water availability etc.

The objective function evaluates each potential site and selects the sites with least cost and risk associated with them. The objective function for the problem is defined below.

$$z = \min w_1 \left[\sum_{j=1}^{n} \sum_{i=1}^{m} \alpha d_{ij} x_{ij} + \sum_{j=1}^{n} f_j y_j\right] + w_2 \sum_{j=1}^{n} \beta_j y_j \quad (1)$$

where n represents the number of areas available for power plants and m the number of load points to which the power has to be delivered. d_{ij} represents the distance between the

potential plant location to the load point and x_{ij} represents the fraction of power delivered by the j^{th} plant to the i^{th} load point. y_j is a binary variable which shows whether the potential area is selected or not, f_j is the capital cost associated with the plant and β_j is the normalized risk factor . w_1 and w_2 are weighing parameters to balance the risk cost ratio while α is the power loss to cost conversion factor.

Subject to the constraints,

$$\sum_{i=1}^{m} x_{ij} = 1 \tag{2}$$

$$y_j = [0,1] \tag{3}$$

$$s_j y_j \geq \sum_{i=1}^{m} (1+\gamma) D_i x_{ij} \tag{4}$$

$$\sum_{j=1}^{n} s_j y_j \geq \sum_{i=1}^{m} (1+\gamma) D_i \tag{5}$$

Where D_i and s_j represents the demand and power supply from the load point and plant respectively. γ represents the excess power factor.

The risk factor can be written as,

$$R_j = \sum_{i=1}^{m} \left[\frac{A_i \rho_j}{d_{ij}} + \mu d_j^f \right] \forall\, d_{ij} \leq r \tag{6}$$

$$\beta_j = \frac{R_j}{\max(R_j)} \tag{7}$$

$$n_{jmax} = \frac{A_j}{A_u} \; where \; 0 \leq n_j \leq 12 \tag{8}$$

$$f_j = u_j n_j C \tag{9}$$

Where A_i represents the area of the land lot, ρ_j the population density, d_j^f distance to nearest cooling water source. A_u represents the minimum area required for a unit of the modular power plant and n_j the number of units that can be placed in a potential site. C is the fixed cost associated with a single unit. The capital cost of a modular plant decreases with the increase in capacity and the cost factor u_j takes that into consideration.

The information regarding the modular nuclear reactor is taken from the website of a company called NuScale Power. The other necessary information like the land lot properties and load point data is taken from the knowledge base using SPARQL queries.

3.2. Results and Discussion

The architecture proposed in the paper is utilized to implement the case study in the JPS. Three parameters in the model are user defined. They are the excess power factor, risk factor and cost factor. The user inputs the values of these models in the JPS website and clicks on the run button. This acts as the stimulus and initiates the following events. A

series of SPARQL queries will be executed and they will through the relevant knowledge bases to collect information. Information regarding the land, its area, location, ownership etc is stored in the Landlots.owl file. But not all the land-lots will be a suitable site for a modular nuclear power plant. The SPARQL query acts as a preliminary filter and removes impossible options based on the area and present purpose of the land-lot. The possible choices of the sites are then fed to the model. The model also requires information regarding the consumer power consumptions and their location. This is obtained from the electrical network owl file. Once the model is executed the results are written to an excel file. The Java agent then creates a series of owl files each of them representing a single power plant. The information in these owl files are then visualized using the Google map JavaScript API. The visualization of the power plants is available in JPS website and a snapshot is given in the figure below.

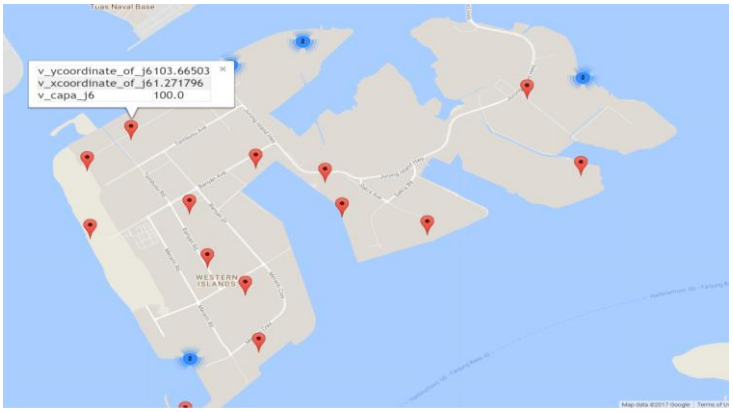


Figure 3 Visualization of the optimization model

4. Conclusion

An ontology based approach for building knowledge bases were discussed in the paper. It can be established that an ontology based knowledge management system can provide more efficient and hassle-free communication between different entities in an EIP. In the future real-time information from sensors could be fed into the knowledge bases and more complex case studies could be conducted in JPS.

References

Kleinelanghorst MJ, Z. L., Sikorski J, Shyh EFY, Aditya K, Mosbach S, et al. (2016). J-park simulator: roadmap to smart eco-industrial parks. Computational Modelling Group, University of Cambridge(Technical report 174).

Morbach, J., Yang, A., & Marquardt, W. (2007). OntoCAPE—A large-scale ontology for chemical process engineering. Engineering Applications of Artificial Intelligence, 20(2), 147-161.

Zhang, C., Romagnoli, A., Zhou, L., & Kraft, M. (2017). Knowledge management of eco-industrial park for efficient energy utilization through ontology-based approach. Applied Energy,204(SupplementC),1412-1421.

Zhou, L., Pan, M., Sikorski, J. J., Garud, S., Kleinelanghorst, M. J., Karimi, I. A., & Kraft, M. (2017). System Development for Eco-industrial Parks Using Ontological Innovation. Energy Procedia, 105(Supplement C), 2239-2244.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50337-2

Incorporating Automation Logic in the Online Scheduling of Batch Chemical Plants

Blake C. Rawlings[a*], Venkatachalam Avadiappan[b], Stéphane Lafortune[a], Christos T. Maravelias[b], John M. Wassick[c]

[a]*Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor, MI, USA*
[b]*Department of Chemical and Biological Engineering, University of Wisconsin, Madison, WI, USA*
[c]*The Dow Chemical Company, Midland, MI, USA*
bcraw@umich.edu

Abstract

We present an approach to online scheduling of chemical plant operations that incorporates precise knowledge of the dynamics enforced by the automation system. The scheduling problem is solved using a discrete-time state-space resource task network formulation. Rescheduling is triggered when the schedule currently being executed conflicts with the dynamical behavior allowed by the automation system. We present a case study to show that better operation is achieved using our approach compared to iterative rescheduling alone.

Keywords: Automation, Online Scheduling, Formal Methods

1. Introduction

Scheduling of operations in a chemical plant is the activity of deciding when, and in which order, to perform various tasks, often by solving an optimization problem. Chemical production scheduling problems occur in a wide variety of applications, from batch production of fine chemicals to continuous production of bulk chemicals. Before a scheduling problem can be solved, an abstract framework is needed to represent the different elements of the problem, such as the production facility and the associated production recipe. The resource task network (RTN) enables this representation. The RTN representation has two components: resources and tasks. Resources include equipment, materials, utilities, etc., and a task is an operation that consumes and generates resources through the period of its execution. Many scheduling models employ a discrete time grid, where the time horizon is discretized into periods of uniform length, and tasks start and end at the grid points.

In a dynamic environment, disturbances or new information may lead to suboptimality or infeasibility of the computed schedule, necessitating rescheduling (Gupta and Maravelias, 2016). The revision of an existing schedule in response to such disturbances or new information is called online scheduling.

This work is sponsored in part by the Department of the Army through the Digital Manufacturing and Design Innovation Institute. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Department of the Army.

One source of new information that may necessitate rescheduling is the plant's automation system, which makes the low-level discrete decisions that drive the plant's behavior. The key issue that we seek to address in this work is the mismatch between the model of the plant used in the scheduling problem and the actual behavior of the plant, enforced by the automation system. For example, for a batch reaction which is modeled as a single task with a fixed duration in the scheduling model, the automation system may include a sequence of 10 or more distinct steps, with transitions between those steps dependent upon the state of other pieces of equipment in the plant (which, in turn, have their own sequences of steps). Thus, a schedule computed using the coarser model may not actually be feasible when executed in the plant, even in the absence of disturbances. A similar issue arises when integrating scheduling and control (Engell and Harjunkoski, 2012; Baldea and Harjunkoski, 2014).

To address this problem, we develop a methodology that incorporates knowledge about the dynamics of the automation system to guide rescheduling. In the past, researchers have focused on extensions of the RTN model and advances in the state-space scheduling model. Wassick and Ferrio (2011) have proposed a number of extensions to the conventional RTN model. Nie et al. (2014) proposed a discrete-time RTN formulation for rescheduling. Subramanian et al. (2012) showed a general procedure to transform the mixed-integer programming (MIP) scheduling model into a state-space model, as is used in model predictive control. Gupta and Maravelias (2017) presented an extended state-space state task network (STN) model suitable for implementation in an online scheduling setting. Building upon these previous works, we develop a generalized discrete-time state-space RTN model that accounts for disturbances detected by analyzing the automation logic online. In the state-space RTN model, the status of the process is indicated by the state variables, and information obtained by analyzing the automation logic is represented by additional constraints involving those state variables. To monitor the execution of a schedule, we construct a finite transition system that models the plant's automation logic as in Rawlings et al. (2017), and augment that model with historic data on the amount of time the system spends in each step. With this detailed model of the plant's dynamics, we treat the current schedule as a specification of the desired system behavior, and apply model checking to determine whether that specification is violated (not to compute the schedule itself, as in Kim et al. (2009)). When it is, we extract relevant information to explain the infeasibility, and reschedule.

We illustrate the applicability of our methods through a case study that involves a batch reactor and postprocessing units, multiple final products, maintenance requirements, and operator availability restrictions. We show that the integration of scheduling and automation leads to better quality closed-loop solutions, with higher throughput than in solutions obtained through iterative rescheduling.

2. Accounting for Automation Logic when Scheduling

We now present our approach to accounting for decisions made by the automation system when scheduling (and rescheduling) chemical plant operations. An outline of the approach is shown in Figure 1.

2.1. Checking the Feasibility of a Schedule Subject to the Automation Logic

The scenario that we consider is periodic rescheduling at a fixed interval using a relatively coarse model, along with event-triggered rescheduling based on a more

detailed model of the automation logic. A rescheduling event is triggered whenever the more detailed model detects that the current schedule is no longer feasible (for example, because it does not account for some constraint imposed on the plant's behavior by the automation system).

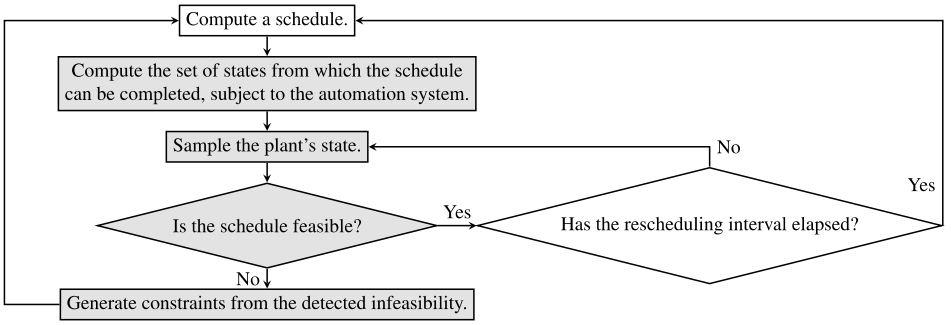


Figure 1. Rescheduling in response to a conflict with the automation logic (in addition to periodic rescheduling with a fixed interval). The new steps developed in this paper are highlighted in gray.

Because the automation system deals primarily with making discrete decisions like transitions between operating steps (whereas the control system generally makes the continuous decisions like flow rates, etc.), it is naturally modeled as a finite transition system. To this model, we add additional details related to scheduling, such as the minimum amount of time spent in each step historically and counters to keep track of batch number, with the result still being a finite transition system. Given the more detailed model, formal falsification can be applied as in Rawlings et al. (2017) to determine whether the schedule is feasible from the current state of the plant. In this case, the specification, written in computation tree logic, has the form

$$\mathrm{AG}(\textit{allotted time not exceeded} \wedge \mathrm{EF}(\textit{schedule completed}) \tag{1}$$

which requires that every task always (AG) stay within its allotted time, and that it remain possible to reach (EF) the state in which the schedule has been completed. For a finite transition system, any specification that has the form of Eq. (1) corresponds to a unique, maximal set of states in which the specification is satisfied (Rawlings and Ydstie, 2017). This leads to the strategy shown in Figure 1, where, every time a new schedule is computed, the set of states in which the schedule is feasible is computed, and then, if the plant ever enters a state not included in that set, the schedule is known to be infeasible.

2.2. Revising an Infeasible Schedule

When the current schedule is proven to be infeasible, it is necessary to guide the rescheduling problem so that it produces a schedule which avoids the infeasibility. Without additional constraints derived from the detailed model of the automation logic, the scheduling formulation might produce the same infeasible schedule. To produce these additional constraints, we modify the specification in Eq. (1) by relaxing the requirements until it becomes feasible. From the relaxation required to make the specification feasible, we can determine which part of the schedule has been violated, and apply the corresponding constraint to the rescheduling problem to avoid that infeasibility.

3. Case Study

To demonstrate our approach, we now apply it to a small example. The process is shown in Figure 2. The process consists of a reactor, which can produce two different products ("premium" or "regular"), a solvent recovery unit, and a check tank.

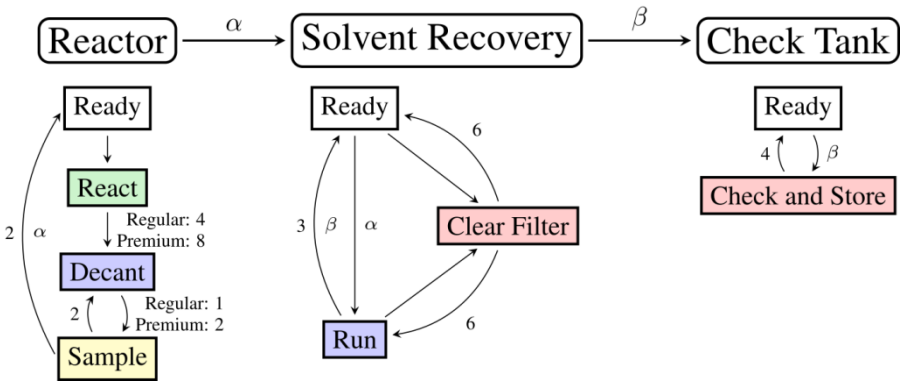


Figure 2. The example process. At the top, the flow sheet. Below each unit on the flow sheet, the sequence of steps that the automation logic implements (color coded to match Figure 3). Transitions between processing steps that require a certain amount of time are labeled with the time. Transitions that occur simultaneously (as enforced by the automation system) are labeled with the same Greek letter. The steps shown in red occupy the operator's attention, and cannot occur at the same time.

For both products, the reactor executes three steps: reaction, decanting, and sampling. Depending on the result of the sampling step, the decanting and sampling may need to be repeated. The amount of time required to complete some of the steps depends on which recipe (premium or regular) is being executed, as shown in Figure 2.

The solvent recovery unit processes each batch produced by the reactor. Over time, the filter in the solvent recovery unit becomes fouled, and requires routine cleaning, which is performed by the operator. Processing the regular product fouls the filter faster than processing the premium product.

The check tank receives the final product from the solvent recovery unit so that it can be checked for quality and sent to the appropriate storage area. The analysis is more involved than the sampling in the reactor, and requires the operator's attention.

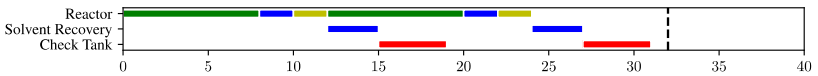
We consider the problem of scheduling batches to maximize profit with a fixed deadline, i.e., a batch must be completed (including passing through solvent recovery and the check tank) before the deadline to count toward the profit. The parameters for the problem are shown in Table 1. Note that the scheduling formulation does not account for the individual steps in the reactor, and instead treats a batch as a single task, so that each product is assumed to require the amount of time to complete the reaction, decanting, and sampling steps. This also assumes that the decanting and sampling will not have to be repeated, which is the most likely scenario.

The automation logic in Figure 2 was implemented in the PLC programming language Structured Text and converted to a model using st2smv, as in Rawlings et al. (2017). This model was augmented to track the number of batches produced, the elapsed time, and to account for the operator only being able to perform one task at a time.

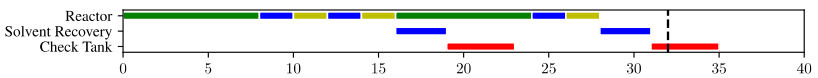
Table 1. Parameters for the scheduling problem.

Task	Time	Profit	Other details
React, decant, sample (premium)	12	100	
React, decant, sample (regular)	7	20	
Solvent recovery (premium)	3	-	Consumes 1 “unit” of filter
Solvent recovery (regular)	3	-	Consumes 2 “units” of filter
Clear filter	6	-	Restores filter capacity to 2 “units”
Check tank	4	-	

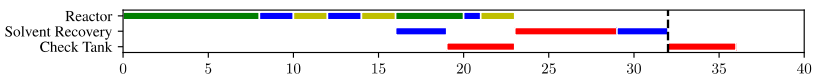
Consider the problem described above, with a deadline of 32 hours. Figure 3(a) shows the nominal optimal schedule, in which two batches of the premium product are completed before the deadline. The nominal schedule is feasible as long as neither batch requires a second decant-sample cycle. If such a disturbance does occur, however, the processing time for that batch will be extended by four hours. Figure 3(b) shows that if a disturbance occurs in the first batch, simply pushing the schedule back will result in the second batch not being completed before the deadline. Figure 3(c) demonstrates that, because the filter in the solvent recovery unit does not have enough capacity to process one premium and one regular batch without maintenance, replacing the second premium batch with a regular batch at time 16 will not work, either.



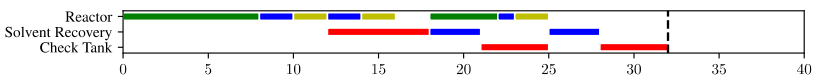
(a) The nominal optimal schedule.



(b) The naïve response to the delay is to simply push the remaining tasks back.



(c) An alternative is to replace the second batch of the premium product with a batch of regular product, which requires clearing the filter.



(d) If the filter is cleared at time 12, then the second batch can be completed before the deadline.

Figure 3. Schedules from the example in Section 3 (color coded to match Figure 2).

If the delay is detected at time 12, when the automation system receives the signal to return to the decant step after the first sampling step, it is already known that the nominal schedule is no longer feasible, because the first batch will not be completed before time 16. This comes from the fact that at time 12, the relaxed specification (where b is the batch number) AG($b = 1 \Rightarrow t < 15$) is falsified, but the further relaxation AG($b = 1 \Rightarrow t < 16$) is not. Solving the rescheduling problem at time 12, with the added constraint that the end time of the first batch is greater than or equal to 16 (i.e., the binary state variables in the RTN formulation that correspond to the first batch being assigned to the reactor between times 12 and 16 are all fixed to 1), produces the

schedule shown in Figure 3(d). The filter is cleared proactively starting at time 12 to overlap with the second cycle of decanting and sampling in the reactor, which allows the regular batch to be completed before the deadline. Note that, because the rescheduled operation finishes at the deadline, with no time to spare, it was only possible because the delay was predicted at time 12. Attempting to reschedule at time 13 (when it would be known that the nominal schedule is no longer feasible even without using the approach from this paper) would be too late, as the second batch could not be completed before time 33, after the deadline.

4. Conclusions

We have shown that knowledge of the automation system's dynamics in a chemical plant can be incorporated in the rescheduling problem to improve the throughput compared to simple periodic rescheduling. This is an early step toward bridging the gap between the detailed models used when analyzing automation systems and the higher-level models of the plant's operations used in a typical scheduling problem. One key question to answer is whether the constraints that are applied when rescheduling after detecting infeasibility are sufficient to push the solution away from the problematic area. Applying our approach to a real industrial process and investigating embedding the automation system's dynamics directly in the scheduling formulation (thereby avoiding schedules that conflict with the automation logic altogether) are the subject of current and future work.

References

M. Baldea, I. Harjunkoski, 2014, Integrated Production Scheduling and Process Control: A Systematic Review, Computers & Chemical Engineering, vol. 71, pp. 377–390

S. Engell, I. Harjunkoski, 2012, Optimal Operation: Scheduling, Advanced Control and Their Integration, Computers & Chemical Engineering, vol. 47, no. 20, pp. 121–133

D. Gupta, C. T. Maravelias, 2016, On Deterministic Online Scheduling: Major Considerations, Paradoxes and Remedies, Computers & Chemical Engineering, vol. 94, no. 2, pp. 312–330

D. Gupta, C. T. Maravelias, 2017, A General State-Space Formulation for Online Scheduling, Processes, vol. 5, no. 4, p. 69

J. Kim, J. Kim, I. Moon, 2009, Error-Free Scheduling for Batch Processes Using Symbolic Model Verifier, Journal of Loss Prevention in the Process Industries, vol. 22, no. 4, pp. 367–372

Y. Nie, L. T. Biegler, J. M. Wassick, C. M. Villa, 2014, Extended Discrete-Time Resource Task Network Formulation for the Reactive Scheduling of a Mixed Batch/Continuous Process, Industrial & Engineering Chemistry Research, vol. 53, no. 44, pp. 17112–17123

B. C. Rawlings, J. M. Wassick, B. E. Ydstie, 2017, Application of Formal Verification and Falsification to Large-Scale Chemical Plant Automation Systems, Computers & Chemical Engineering, In press (available online)

B. C. Rawlings, B. E. Ydstie, 2017, Falsification of Combined Invariance and Reachability Specifications in Hybrid Control Systems, Discrete Event Dynamic Systems, vol. 27, no. 2, pp. 463–479

K. Subramanian, C. T. Maravelias, J. B. Rawlings, 2012, A State-Space Model for Chemical Production Scheduling, Computers & Chemical Engineering, vol. 47, no. 20, pp. 97–110

J. M. Wassick, J. Ferrio, 2011, Extending the Resource Task Network for Industrial Applications, Computers & Chemical Engineering, vol. 35, no. 10, pp. 2124–2140

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50338-4

Manufacturing Sustainability Enhancement: A Model Predictive Control Based Approach

Majid Moradi-Aliabadi[a], Yinlun Huang[a*]

[a] *Department of Chemical Engineering and Materials Science, Wayne State University, Detroit, MI 48202, USA*
yhuang@wayne.edu

Abstract

From the sustainability science point of view, a manufacturing system's sustainability status change involves a series of system state transitions in an economics-environment-society-based three-dimensional space. To ensure the achievement of the sustainability goal effectively in this multi-stage process under uncertainty, continuous sustainability assessment and decision making is needed. In this paper, we formulate this type of sustainability problem as a terminal target control problem. A multi-objective optimization approach is applied for generating sustainability goal options for decision makers to select or adjust. A model predictive control (MPC) approach is then employed to derive sustainability actions for implementation in different sustainability improvement stages. The introduced methodology can provide a stage-wised guideline to industries in defining sustainable development strategies and deriving sustainability actions systematically. A comprehensive case study on biodiesel manufacturing is illustrated to demonstrate methodological efficacy.

Keywords: Sustainable manufacturing, MPC strategy, sustainability action

1. Introduction

Industries across the world face increased costs in materials, energy, and compliance coupled with higher expectations of customers, investors and local communities. Thus, making manufacturing economically sound, environmentally clean, and socially responsible is no longer just nice-to-have, but an industry's imperative. Sustainable manufacturing demands (i) triple-bottom-line based practice and (ii) continuous performance improvement. The former reflects the nature of sustainability, which is a multi-objective optimization problem, while the latter is a dynamic control problem. In short, sustainable manufacturing is a multi-objective optimal control problem. Thus, an important research direction is to study how system control science can be effectively used to guide sustainability study in a holistic way.

Over the past decades, various approaches have been proposed to support decision-making for sustainable manufacturing, most of which use Multi Criteria Decision Making (MCDM) techniques (Azapagic and Perdan, 2005; Yue et al., 2013). The collaborative profitable pollution prevention (CP3) idea was introduced by Piluso and Huang (2009), which aid in decision-making for the study of sustainable development of industrial zones. All these methodologies that are applied at different scales focus on the first aspect of the definition of sustainable manufacturing that is a multi-objective problem.

In this work, we formulate manufacturing sustainability problems as a terminal target control problem, which is investigated by resorting to sustainability fundamentals,

system control and engineering science. A multi-objective optimization approach is applied for generating sustainability goal options for decision makers to select or adjust. A model predictive control (MPC) approach is then employed to derive sustainability actions for implementation in different sustainability improvement stages.

2. Methodology

Evaluation of a manufacturing system's sustainability performance requires selection of sets of indicators. The indicators are organized to measure the economic, environmental, and social sustainability of the system over time to determine if desirable progresses are made. Figure 1(a) shows a sustainability valuation model. The sustainability status of any system can be described by a time-variant vector function, $\vec{S}(k)$, that is defined by Moradi-Aliabadi and Huang (2016) as follows:

$$\vec{S}(k) = \langle E(k), V(k), L(k) \rangle, \tag{1}$$

where $E(k)$, $V(k)$, $L(k)$ are, respectively, the composite economic index, the composite environmental index, and the composite social index. Note that each composite index is normalized with the value range of 0 to 1, which is derived from a set of measured indicators in each of the three sustainability categories.

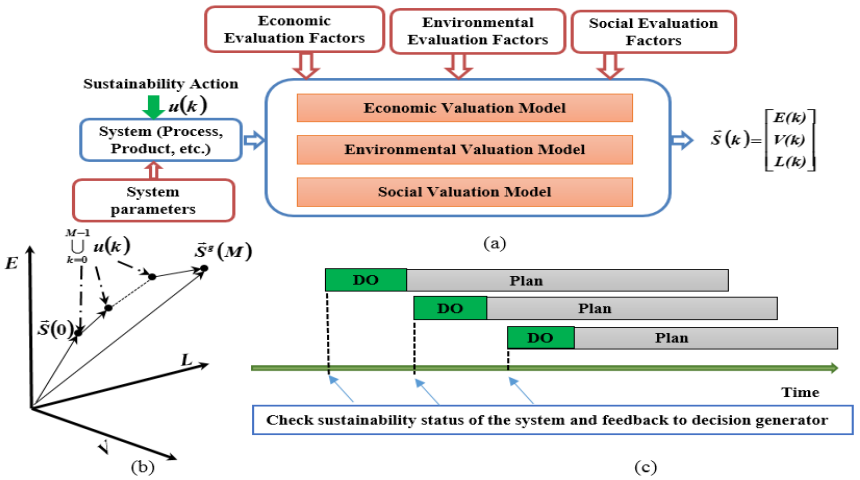


Figure 1. (a) Sustainability valuation model, (b) terminal target problem and (c) Plan-Do-Check-Act cycle.

At any time instant k, $\vec{S}(k)$ is a vector from the origin of the space (0, 0, 0) to the point, $(E(k), V(k), L(k))$ as shown in Fig. 1(b). As an illustration, $\vec{S}(0)$ is a vector pointing to the system's sustainability status at time $k = 0$. The sustainability status of the system is changed when sustainability action $u(k)$ is taken by the decision makers at that time. Thus, the sustainability status at a given time is determined by the pervious sustainability status and the action received by the system in the last period of time. In the most general term, this can be represented by:

$$\vec{S}(k+1) = f(\vec{S}(k), u(k)), \tag{2}$$

2.1. Sustainability goal setting

Setting a sustainability goal is a challenging task for decision makers because of the trade-off among economic (E), environmental (V), and social (L) sustainability objectives. Here we formulate the sustainability problem as a Multi-Objective Optimization (MOO) problem, aiming at generating the Pareto optimal solutions (or Pareto set) for decision makers to choose the most preferred solution. A Pareto optimal solution that is also called non-dominated solution is defined as follows: assume that the objective functions, f_i, $i = 1, 2, \ldots, p$, are for maximization, a feasible solution $\vec{x}$ of a MOO problem is a Pareto optimal solution, if there is no other feasible solution $\vec{x}'$ such as $f_i(\vec{x}') \geq f_i(\vec{x})$ for every i with at least one strict inequality. In this work, the ε-constraint method is adopted for generating the Pareto set. This method optimizes one objective function, while the other objective functions are considered as constraints as shown below,

$$\begin{aligned} \max \quad & f_1(\vec{x}) \\ s.t. \quad & f_2(\vec{x}) \geq \varepsilon_2 \\ & f_3(\vec{x}) \geq \varepsilon_3 \\ & \vec{x} \in S \end{aligned} \tag{3}$$

Pareto optimal solutions are obtained through changing the values of the constrained objective functions (ε_i). Here S denotes a feasible decision space.

2.2. MPC strategy for sustainability improvement

The MPC strategy for industrial sustainability development is implemented by following the steps below:

Step 1**.** The sustainability status of the system, $\vec{S}(t)$, is measured/estimated at each instant t using the sustainability valuation model.

Step 2**.** The set of future sustainability actions, $U^*(t) = \{u^*(t|t), u^*(t+1|t), \ldots, u^*(t+M-1|t)\}$, is calculated for the entire planning window, M, by optimizing a determined criterion. For example, in a terminal target problem shown in Fig. 1(b), we are interested in minimizing the error between the desired sustainability target position, $\vec{S}^g(M)$, and the actual sustainability status of the system, $\vec{S}(M)$, at the end of the final time M.

Step 3**.** The first sustainability action, $u^*(t|t)$, is implemented, whilst the other sustainability actions are rejected, because at the next sampling instant, $\vec{S}(t+1)$ is already known and Step 1 is repeated with this new value.

The MPC strategy for sustainability performance improvement can be viewed as a sustainability management system (SMS), which is based on the well-known Plan-Do-Check-Act cycle for continuous improvement (see Fig. 1(c)).

2.3. Decision generator

In this work, sustainability performance improvement is through implementing appropriate technologies. Here we use the sustainability performance improvement methodology developed by Moradi-Aliabadi and Huang (2016). In the second step of the MPC strategy described above, the decision generator needs to identify sets of technologies (among N technology candidate) by solving an optimization problem for implementation in M stages to achieve a pre-set sustainability goal, $S^g(M)$. For the terminal target problem, the following optimization model should be solved at each sample time t.

$$
\begin{aligned}
&\underset{U_t}{argmin}\ \ J=\left(\vec{S}^{g}(M)-\hat{\vec{S}}(M)\right)^{T} W\left(\vec{S}^{g}(M)-\hat{\vec{S}}(M)\right)\\
&s.t.\\
&\quad \hat{\vec{S}}(0)=\vec{S}(t)\\
&\quad \hat{E}(k+1)=\hat{E}(k)+\sum_{j=1}^{N} y_{k,j}\Delta E(T_j,P) \qquad k=0,1,\ldots,M-1\\
&\quad \hat{V}(k+1)=\hat{V}(k)+\sum_{j=1}^{N} y_{k,j}\Delta V(T_j,P) \qquad k=0,1,\ldots,M-1\\
&\quad \hat{L}(k+1)=\hat{L}(k)+\sum_{j=1}^{N} y_{k,j}\Delta L(T_j,P) \qquad k=0,1,\ldots,M-1\\
&\quad \hat{E}(k+1)\geq \hat{E}(k) \qquad k=0,1,\ldots,M-1\\
&\quad \hat{V}(k+1)\geq \hat{V}(k) \qquad k=0,1,\ldots,M-1\\
&\quad \hat{L}(k+1)\geq \hat{L}(k) \qquad k=0,1,\ldots,M-1\\
&\quad Cost\left(u(k)\right)\leq B^{up}(k) \qquad k=0,1,\ldots,M-1
\end{aligned}
\tag{4}
$$

where the objective function, *J,* is the weighted Euclidian distance; *W* is a diagonal weighting matrix that its diagonal elements (w_{11}, w_{22}, w_{33}) are the weights associated to economic, environmental, and social composite indexes respectively; $\Delta E(T_j,P)$, $\Delta V(T_j,P)$, and $\Delta L(T_j,P)$ are, respectively, the economic, environmental, and social sustainability improvements of system *P* when technology T_j is adopted; $y_{k,j}$ is a binary variable ("1" if technology *j* is selected at stage *k*; otherwise "0"). The equality constraints describe the sustainability status transition of the system, the inequality constraints show the sustainability space, and the last constraint is a budget limit for project implementation. The hat notation is used for the sustainability status of the system predicted by the models. There are several logical constraints associated with this optimization model that are not discuss here for the sake of brevity. Details about the constraints and methodology can be found in Moradi-Aliabadi and Huang (2016).

3. Case study on biodiesel manufacturing

The developed methodology is used to investigate the sustainability performance improvement problem of an alkali-catalyzed biodiesel manufacturing plant. The plant seeks cost-effective technologies to reduce waste, improve energy efficiency and product quality. A three-year (i.e., $M = 3$) sustainability performance improvement goal is set, and 10 candidate technologies (T_1, T_2, …, T_{10}) are available for selection. The technology specifics and sustainability related data can be found in Moradi-Aliabadi and Huang (2016).

To simplify the description of methodological application, only the economic (*E*) and environmental (*V*) performance improvement is illustrated. The economic sustainability indicators are the value added (E_1) and the gross margin per direct employee (E_2). The indicators in the environmental sustainability category are the total raw materials used per pound of product produced (V_1), the hazardous solid waste per unit value added (V_2), and the energy intensity (V_3). The budget limits for technology adoption are \$300k, \$200k, and \$100k for the first, second and third stages, respectively. The weighting factors (w_{ii} in matrix W in Eq. (4)) are set equal that shows equal importance of the categorized sustainability improvement.

Table 1. Pareto optimal solutions in *E-V* space

Plan	Pareto optimal solution	Final sustainability state vector	Overall sustainability $\left\|\vec{S}(3)\right\|$
P_1	$\{T_6\},\{T_1,T_2,T_9\},\{T_4\}$	$\vec{S}_1(3)=0.78\vec{i}+0.54\vec{j}$	0.67
P_2	$\{T_6\},\{T_2,T_7,T_9\},\{T_1\}$	$\vec{S}_2(3)=0.76\vec{i}+0.57\vec{j}$	0.67
P_3	$\{T_6\},\{T_2,T_7,T_9\},\{T_{10}\}$	$\vec{S}_3(3)=0.71\vec{i}+0.60\vec{j}$	0.66
P_4	$\{T_2,T_4,T_5,T_8\},\{T_1,T_7\},\{T_9\}$	$\vec{S}_4(3)=0.68\vec{i}+0.62\vec{j}$	0.65
P_5	$\{T_5,T_7,T_8\},\{T_1,T_2,T_9\},\{T_{10}\}$	$\vec{S}_5(3)=0.65\vec{i}+0.65\vec{j}$	0.65
P_6	$\{T_3,T_4,T_5,T_7\},\{T_2,T_9,T_{10}\},\{T_8\}$	$\vec{S}_6(3)=0.60\vec{i}+0.66\vec{j}$	0.63

In this case, the initial sustainability status of the system is $\vec{S}(0)=\langle 0.50,\ 0.40\rangle$. To provide sufficient decision support to the decision makers for sustainability goal achievement, a set of Pareto optimal solutions are derived using the ε-constraint method. The derived solutions are summarized in Table 2. Each listed plan (P_i) is a solution that can be selected for future sustainability improvement of the biodiesel manufacturing process. The finial sustainability status of the system and the overall sustainability that is defined as the norm of sustainability status vector are shown in the third and fourth columns of the table. As the results show, the overall sustainability performances using different plans are similar (see the values in the fourth column). However, the range of each composite sustainability index is considered large. The Pareto front curve in the *E-V* space that shows a trade-off opportunity between the economic and environmental objectives is plotted and shown in Fig. 2(a). The Pareto optimal solutions (P_1 to P_6) are marked on the curve in the figure and any of these points can be selected by the decision maker. Note that all the points above the Pareto curve are infeasible solutions; the points below the curve are feasible but suboptimal. Here, the decision makers set the final goal for the categorized sustainability to $S^g(M)=\langle 0.71,0.6\rangle$ that is located on the Pareto curve.

The measured initial sustainability status of the system, $\vec{S}(0)=\langle 0.50,0.40\rangle$, is used as a feedback to the *decision generator*, where the MPC strategy is implemented, to generate an optimal policy for the next three years. At $t=0$, the *decision generator* outputs the optimal policy, $U^*(t=0)=\{\{T_2,\ T_4,\ T_5,\ T_8\},\{T_1,\ T_7\},\{T_9\}\}$, for the next three years by solving the terminal optimization problem shown in Eq. (4). The sustainability development path (see the dotted arrowed line in Fig. 2(a)), and the dynamic sustainability value of the system after implementation of this optimal policy is shown in the Fig. 2(a) and (b), respectively. As the results show, the plant can achieve the preset sustainability goal after three years. The investments for taking sustainability actions are \$289k, \$190k, and \$60k in the first, second, and third years, respectively.

Based on the MPC strategy, the first sustainability action, $\{T_2,\ T_4,\ T_5,\ T_8\}$, is implemented at $t=0$ and other sustainability actions are rejected. After the first stage of sustainability enhancement, technologies T_2, T_4, T_5, and T_8 are removed, and the sustainability status of the system, $\vec{S}(1)=\langle 0.56,0.52\rangle$, is measured again. The assessment result is sent to the decision generator. If the plant operation is not affected by the environment and other conditions, the sustainability actions will not be changed in the follwing sample times based on the principle of optimality. In this case, the predicted sustainability status of the system will be same as the actual one, $\vec{S}(1)$, after the first stage of sustainability enhancment. However, if there is a disturbance (assume a disturbance $v=\langle -0.02,\ 0\rangle$

because of a change in demand or product price) that perturbs the sustainability status of the system, the optimal policy for the next two years will be different to reach the same sustainability goal. Based on the MPC strategy, the new sustainability status of the system will be $\vec{S}(1)=\langle 0.54, 0.52\rangle$; this information is sent to decision generator as a new initial point, which requires a derivation of a new policy for the next two years. In this scenario, the results show that the final sustainability goal is achievable, if the budget limit be increased from \$200k to \$300k for the second stage. The optimal policy would be $U^*(t=1)=\{\{T_6\},\{T_9\}\}$ that is different from the previous one ($\{T_1, T_7\}$). The sustainability development path is shown in Fig. 1(a) (see the solid arrowed line) for this scenario.

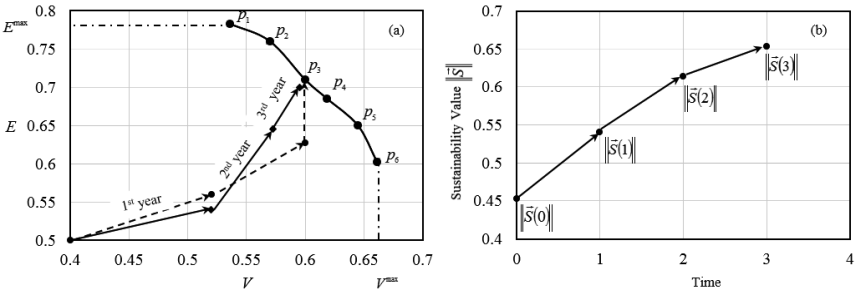


Figure 2. (a) Pareto front curve and sustainability development path with disturbance (solid path) and without disturbance (dotted path), and (b) dynamic of sustainability value of the system.

4. Conclusions

In this paper, we formulate industrial sustainability enhancement problem as a dynamic control problem. As sustainability seeks solutions for triple-bottom-line-based development, a multi-objective optimization approach is applied to derive a set of optimal sustainability alternatives for the decision makers to set the most desirable sustainability goal(s). The MPC strategy is effectively used to derive optimal sustainability improvement strategies and implement them in a dynamic domain. The introduced methodology is general in application, and can provide a structured step-by-step guideline for industries to define sustainability goals derive sustainability actions systematically.

Acknowledgment

This work is in part supported by the US NSF (Grant No. 1434277 and 1604756).

References

A. Azapagic, S. Perdan, 2005. An integrated sustainability decision-support framework Part II:Problem analysis. International Journal of Sustainable Development and Word Ecology, 12(2): 112-131.

D. Yue, M. A. Kim, F. You, 2013. Design of sustainable product systems and supply chains with life cycle optimization based on functional unit: general modeling framework, mixed-integer nonlinear programming algorithms and case study on hydrocarbon biofuels. ACS Sustainable Chem. Eng., 1 (8), 1003−1014.

C. Piluso, Y. Huang, 2009. Collaborative profitable pollution prevention: an approach for the sustainable development of complex industrial zones with uncertain information. Clean Technologies and Environmental Policy, 11(3), 307-322.

M. Moradi-Aliabadi, Y. Huang, 2016. Vector-based sustainability analytics: a methodological study on system transition toward sustainability. Ind. & Eng. Chem. Res., 55(12), 3239-3252.

M. Moradi-Aliabadi, Y. Huang, 2016. Multistage optimization for chemical process sustainability enhancement under uncertainty. ACS Sustainable Chemistry & Engineering 4 (11), 6133-6143.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50339-6

Process Fault Detection in Heat Recovery Steam Generator using an Artificial Neural Network Simplification of a Dynamic First Principles Model

Parag Patil[a], Babji Srinivasan[a], Rajagopalan Srinivasan[b*]

[a]*Department of Chemical Engineering, Indian Institute of Technology Gandhinagar,Gandhinagar, 382355, India*
[b]*Department of Chemical Engineering, Indian Institute of Technology Madras, Chennai,600036, India*
raj@iitm.ac.in

Abstract

A combined cycle power plant (CCPP) is a complex system with a Gas Turbine, Steam Turbine and a Heat Recovery Steam Generator (HRSG) working together. These three units work together and make the process highly interdependent. The onset of any fault in one of the above units would results in a significant reduction in overall efficiency and potentially lead to catastrophic accidents. Such failures can occur due to process faults because of large abrupt variations of operating conditions and structural faults due to corrosion, uneven stresses due to frequent cyclic operations. Conventionally, the identification of such leakage locations is made via visual inspection which is a time consuming and tedious. In the present work, we discuss a fault diagnosis strategy for an actual industrial HRSG present in a CCPP. Various steady state models at different loads of CCPP as well as a dynamic model are developed. Various structural faults in the form of leakages are incorporated in the heat exchangers. An Artificial Neural Network (ANN) model is developed based on data from the above simulations to detect the leaking heat exchangers.
Keywords: Combined Cycle Power Plants, Heat Recovery Steam Generators, Fault Diagnosis, Artificial Neural Networks

1. Introduction

Combined Cycle Power Plants (CCPP) are popular for electricity production in power industry. Unlike other electricity production options such as Gas Turbine and Steam Turbine alone, CCPP can offer up to 60% efficiency (Hoang and Pawluskiewicz, 2016). A CCPP consists of a topping cycle (Gas Turbine) and a bottoming cycle (Steam Turbine) with an HRSG connecting the two. Because of their strong interdependency (Dev et al., 2015), a failure or an inception of fault can quickly propagate and result in drastic reduction of the efficiency of the power plant and catastrophic accidents. There are several reasons for faults in CCPPs. Process faults occur due to large variations in the temperatures/pressures of feed water or exhaust gas; additionally, structural faults such as leakages occur due to corrosion of the water/steam tubes, uneven stresses generated by cyclic operations of CCPP, (Fontaine and Galopin, 2007) etc. Trips due to the sequential propagation of faults are often seen in CCPPs.
An HRSG is a critical part of a CCPP. The heat from exhaust gas coming out of the Gas Turbine is utilized by HRSG to produce steam. This steam is fed to Steam Turbine to generate additional electricity. A typical HRSG has a series of heat exchangers

consisting of superheaters, evaporators, and economizers. HRSGs can be classified by the direction of the exhaust gas flow (vertical or horizontal) and numbers of steam circuits (single, double or triple pressured). In the horizontal arrangement, the exhaust gas passes through vertical tubes whereas in the vertical arrangement the exhaust gas passes through horizontally arranged tubes (Ganapathy, 2015).

HRSGs are prone to structural faults leading to leakages in the tubes arising from over and prolonged heating, varying cyclic operations due to discontinuous electricity load demand and retention of active corroding materials such as NaOH used for cleaning of tubes of HRSG. The detection of the location of leaking and fractured tubes within such large structures is conventionally carried out by non-technical methods (Bain and Christophersen, 2003). The operators in the control rooms usually get an indication of leaks by observing changes in pressure drops in water/ steam lines and exhaust gas temperature instabilities. Further, visual inspections are done during shutdown/maintenance periods to identify the location of leaking sections. Online fault detection and isolation techniques can help to identify the process and structural faults in a timely fashion.

Although fault diagnosis and health monitoring of HRSGs have not received much attention in the literature, various mathematical and data-driven modelling approaches to enhance the performance have been reported. Kumar and Raju (2012) reported that there is a significant effect of heat exchanger arrangement in HRSG on the overall efficiency of CCPP. Feng et al. (2014) provided a generalized model for analyzing the performance of HRSG. Asgari et al. (2013) utilized the operating data to optimize performance and condition monitoring. Nowak and Rusin (2016) developed an Artificial Neural Network-based model for fast prediction of stresses in steam turbine casing. Sharma and Singh (2016) indicated the potential for performance improvement with exergy and economic analysis. Various authors have also analyzed the dynamic behavior of HRSGs during start-up and shutdowns. Benato et al. (2015) developed a dynamic model of triple pressure HRSG and showed that the most stressed component in transient operations are High-Pressure Superheaters and High-Pressure Drum. Alobaid et al. (2014) simulated the startup and shutdown of HRSG. Zhou et al. (2017) performed CFD analysis of an HRSG and showed that unburnt fuel causes structural damage to tubes due to sudden pressure rise.

In the present work, the prime focus is on leakages arising from structural faults. A fault diagnosis strategy is developed for a triple pressure vertical HRSG present in an industrial CCPP. The steady-state models of HRSG at various loads of CCPP are simulated in Aspen Plus®. Verification and validation studies are conducted to ensure that the simulated data is in good congruency with the observed temperatures, pressures and flow-rates at various loads. Further, one of the steady-state models is extended to a dynamic model in Aspen Plus Dynamics®. Various failure modes of leakages are introduced in this dynamic model. Data from these are used to train an Artificial Neural Network (ANN) to classify the location of leakages.

2. HRSG in an Industrial CCPP

The considered HRSG is from an industrial CCPP with a capacity of 368 MW. The schematic diagram of a typical CCPP with vertical Heat Recovery Steam Generator is shown in Figure 1. The HRSG has three levels of high, intermediate, and low pressures. Typically, each pressure circuit in the HRSG has superheater(s), evaporator and economizer(s). In the HRSG in Figure 2, a Boiler Feed Pump draws water from Low-Pressure Drum and supplies it to Intermediate-Pressure and High-Pressure circuits.

High-Pressured Superheaters (HPSH-1,2), Reheaters (RH-1,2) and Low-Pressure Superheater (LPSH) deliver the superheated steam to respective Steam Turbines. The High and Intermediate Drums draw water from the economizers from their respective pressure circuits whereas Low-Pressure Drum draws water from Condensate Preheater (CPH). The evaporators in each circuit evaporate the water coming from the drums and feed the saturated steam to the drums again; they maintain this close circulation. Further, the superheaters in all circuits draw saturated steam from their respective drums and convert it into superheated steam suitable for Steam Turbine operation.
The exhaust gas from Gas Turbine enters HRSG at a temperature of 613.9 $^{\circ}$C with a flow rate of 2192 t/h (at 100% load of CCPP). The gas flows downstream through all

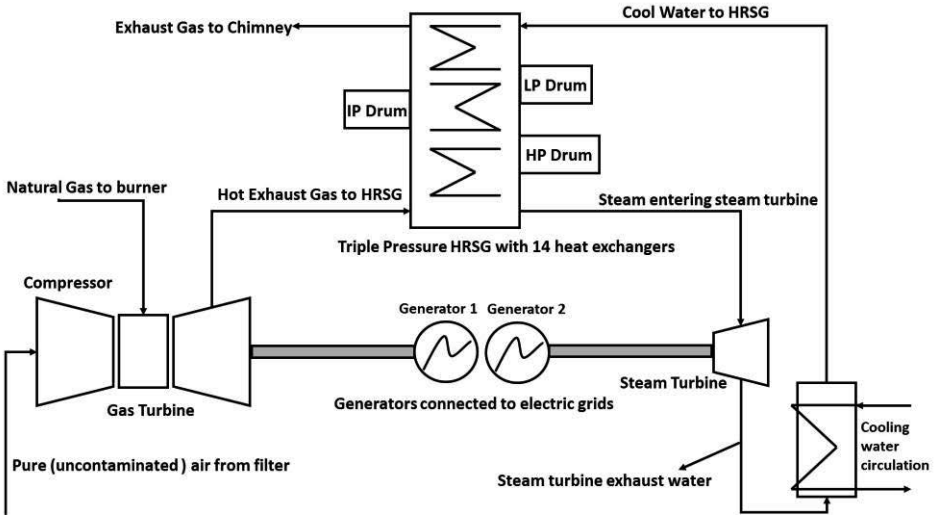


Figure 1. Schematic of CCPP with vertical HRSG

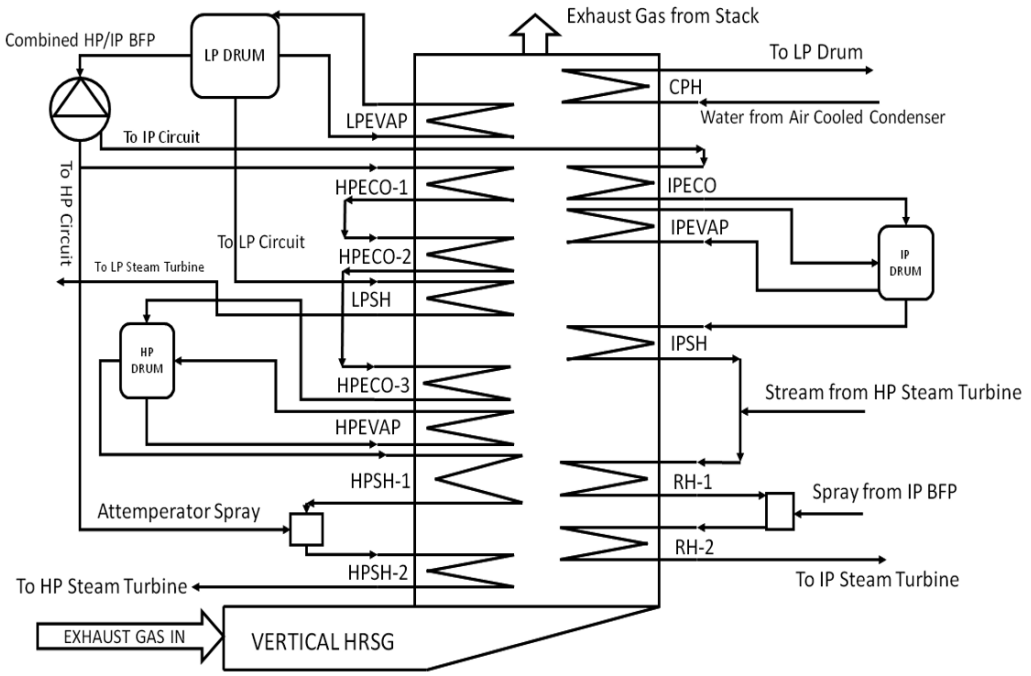
the heat exchangers as shown in Figure 2. Finally, it is relieved to the atmosphere through the stack. To extract maximum energy, the used steam from in higher pressured circuits is fed to lower pressure circuits. The low-pressure circuit outlet stream is cooled in an Air Cooled Condenser (ACC). The water is again circulated into these circuits. The lost water is compensated with the addition of make-up water to ACC.

3. HRSG Simulations and Fault Incorporation

Detailed steady-state HRSG models at loads of 100%, 80%, and 50% of CCPP were simulated in Aspen Plus®. The heat exchangers are cross flow with finned-tube bundles. These crossflow bundles are developed in Aspen Exchanger Design and Rating®. Real geometry data of the exchangers are used to develop these models. Peng-Robinson Equation of State is utilized to simulate the system with rigorous calculations. The performance of the developed models is in good congruency with the data observed from the real HRSG.
Further, the steady state HRSG model at 80% of the CCPP load was extended to a dynamic model in Aspen Plus Dynamics®. Various control loops are also included in this model to maintain pressures, liquid levels of drums and steam flow temperature in all the pressure circuits. They are tuned with a relay-feedback test to get the ultimate

gain and frequency in a feedback loop. The developed dynamic models were run for a period of 10 h.

Once the normal dynamic operation of the system has been successfully simulated, various heat exchanger leaks scenarios were created in Aspen Plus Dynamics®. The heat exchangers chosen to introduce the leakages in the steam lines are the three High-Pressure Economizers, two Reheaters, two High-Pressure Superheaters and an Intermediate-Pressure Superheater. A total of 104 leak fault cases were generated – 13 examples from each of the above eight heat exchangers. A set of 6 examples without any fault were also generated.



Legends			
HPSH	: High Pressure Superheater	**RH**	: Reheater
HPEVAP	: High Pressure Evaporator	**IPEVAP**	: Intermediate Pressure Evaporator
HPECO	: High Pressure Economizer	**IPECO**	: Intermediate Pressure Economizer
CPH	: Condenser Preheater	**BFP**	: Boiler Feed Pump
IPSH	: Intermediate Pressure Superheater	**LPEVAP**	: Low Pressure Evaporator

Figure 2. Vertical triple pressure HRSG under consideration

Time series data from 21 sensors for various temperatures, pressures, levels, and flow-rates across the HRSG are emulated in the simulation. Each simulation run corresponded to an operating period of 10 h, with samples obtained over 0.01 h. This data was utilized to train an ANN for HRSG fault diagnosis.

4. ANN-based identification of Leakage Locations in HRSG

From the dynamic simulation, the time series data for each of the 21 sensors is obtained. However, the range of each sensor is different and varies in magnitude; rescaling of this time series across each sensor is done in the range of (-1, 1). As per Table 1, the fault locations are divided into four types according to their accessibility from each manhole in the actual plant.

Table 1: Classes along with associated heat exchangers.

Location wise class	Heat Exchangers associated with leaks
Type 1 (No Faults)	--
Type 2 (Accessible from Manhole 1)	High-Pressure Superheater-1, 2, Reheater-1,2
Type 3 (Accessible from Manhole 2)	Intermediate-Pressure Superheater, High-Pressure Economizer-3
Type 4(Accessible from Manhole 3)	High-Pressure Economizer-1,2, Intermediate-Pressure Economizer

The multi-label classification problem is solved with a back propagation ANN. Out of the 75872 total examples, 60% were utilized for training the ANN. 20% examples were utilized for validation and testing each. The confusion matrix for the trained ANN is shown in Figure 3. This final optimized ANN was able to classify 99.8% of the testing examples successfully.

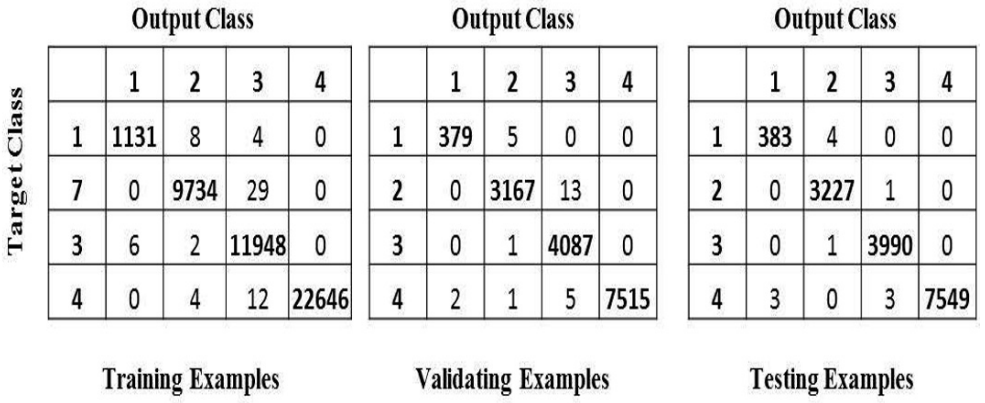


Figure 3. Confusion Matrices for trained ANN

5. Conclusions and Future Work

Fault diagnosis is an important need and challenge in HRSGs. In this work, we have focused on leakages in HRSG heat exchangers – the most commonly occurring faults. Steady State models for an industrial triple pressure HRSG were developed at various loads of CCPP. After verifying and validating the accuracy of these models, one of these steady-state models is utilized to develop a dynamic model. Faults are introduced in these dynamic models at various locations. The data from the dynamic simulation was used to develop a multi-label classification ANN to identify the location of leaking

heat exchangers (as per accessible manhole nearer to it). The developed neural network can classify all the testing faults with 99.8% accuracy. In future, all possible off-load conditions based dynamic models could be used to train the ANN to identify the leak locations in HRSG.

References

Alobaid,F., Karner,K., Belz, J., Epple, B. &Kim ,H., 2014, Numerical and experimental study of a heat recovery steam generator during start-up procedure, In Energy, Volume 64, Pages 1057-1070

Asgari, H.,Chen, X. & Sainudiin,R.,2013, Analysis of ANN-Based Modelling Approach for Industrial Systems. International Journal of Innovarion, Management and Technology, Volume 4, Issue 1, Pages 165-169

Bain, D. I., & D.L. Christophersen, 2003, Some Common Mechanisms Leading to Failures in Heat Recovery Steam Generators, NACE International.

Benato, A., Stoppato, A., Mirandola, A. & Destro,N., 2015, Dynamic simulation of a three pressure level heat recovery steam generator, The 28th International Conference on Efficiency, Cost, Optimization, Simulation and environemental effect of energy systems.

Dev, N., Samsher, Kachhwaha, S., & Attri, R., 2015, GTA modeling of combined cycle power plant efficiency analysis, In Ain Shams Engineering Journal, Volume 6, Issue 1,Pages 217-237

Feng, H., Zhong,W.,Wu, Y. , & Tong,S., 2014, Thermodynamic performance analysis and algorithm model of multi-pressure heat recovery steam generators (HRSG) based on heat exchangers layout, In Energy Conversion and Management, Volume 81, Pages 282-289.

Fontaine, P., & Galopin, J.-F., 2007, HRSG Optimization for cycling duty based on Euro Norm EN 12952-3., ETD Conference on Cyclic Operation of Power Plant, Volume 5, Issue 1, Pages 1–14.

Ganapathy, V. , 2015, HRSG Simulations, Steam Generators and Waste Heat Boilers for Process and Plant Engineers, CRC Press, Taylor and Francis Group, United States.

Hoang, T. & Pawluskiewicz,D., 2016, The efficiency analysis of different combined cycle power plants based on the impact of selected parameters. International Journal of Smart Grid and Clean Energy, Volume 5, Issue 2, Pages 77-85

Kumar, P., & Raju, V. , 2012, Off Design Performance Analysis of a Triple Pressure Reheat Heat Recovery Steam Generator. Int. J. Eng. Res. Technol., Volume 1, Issue 5, Pages 1–10.

Nowak, G. & Rusin,A., 2016, Using the artificial neural network to control the steam turbine heating process, In Applied Thermal Engineering, Volume 108, Pages 204-210.

Sharma, M., & Singh,O., 2016, Exergy analysis of dual pressure HRSG for different dead states and varying steam generation states in gas/steam combined cycle power plant, In Applied Thermal Engineering, Volume 93, Pages 614-622

Zhou, W., Cardenas, M. & Moyeda, D., 2017, Method for CFD facilitated pressure rise calculation due to deflagration in heat recovery steam generator., Proc. Safety Prog., Volume 36, Issue 4, Pages: 408–413.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50340-2

Statistics Pattern Analysis: A Statistical Process Monitoring Tool for Smart Manufacturing

Q. Peter He[a*], Jin Wang[a]

[a]*Department of Chemical Engineering, Auburn University, Auburn AL 36849, USA*
qhe@auburn.edu

Abstract

Statistical process monitoring (SPM) is an important component in the long-term reliable operation of any system and its importance can only become greater in the era of smart manufacturing (SM). Previously we proposed statistics pattern analysis (SPA) based on the idea of using various statistics to quantify process characteristics, and monitoring these statistics instead of process variables themselves to perform process monitoring. In this work we examine the SPA's capability in handling process characteristics including dynamics, nonlinearity and data non-Gaussianity, and compare its performance to representative state-of-the-art SPM methods. In addition, we discuss how SPA can help address new challenges presented by smart manufacturing big data.

Keywords: statistical process monitoring, fault detection, fault diagnosis, smart manufacturing, statistics pattern analysis.

1. Introduction

The goal of process monitoring is to detect the onset and identify the root cause of any change that causes a manufacturing environment to deviate from its desired operation. Process monitoring is an important component and key enabler for the long-term reliable operation of any process (Severson et al., 2016). It has been recognized that the importance of process monitoring can only become greater when the controlled systems are getting more complex, equipped with more sensors, operated under non-steady state, controlled at tighter margins, and getting yet closer to autonomous operations in the era of smart manufacturing (He and Wang, 2017). Principal component analysis (PCA), partial least squares (PLS), and their variants based multivariate SPM methods, which are the state-of-the-arts for industrial applications, have limitations when applied to the processes that are not operated at steady-state. This is mainly due to the underlying assumption for the PCA and PLS based SPM methods that the process data are assumed independent and identically distributed (i.i.d.) samples drawn from a multivariate Gaussian distribution. When a process exhibits strong dynamics, nonlinearity or non-Gaussian distribution, to name a few, the performance of SPM methods may deteriorate significantly, depending on how well the normal process operation data can be approximated by a multivariate Gaussian distribution. To address these limitations, many SPM methods, such as Kernel PCA (KPCA), and independent component analysis (ICA), have been developed recently. However, as discussed in Section 2, these methods have their own challenges and only achieved limited success (He and Wang, 2017). On the other hand, a feature based method that we recently proposed, termed statistics pattern analysis (SPA), has demonstrated superior performance in different applications such as fault detection, fault diagnosis and soft sensor or virtual metrology for both batch and continuous processes (He and Wang, 2011; He and Wang, 2017;

Shah et al., 2017; Wang and He, 2010). Since then, many extensions and variants of SPA have been proposed by others (Deng et al., 2016; Deng and Tian, 2013; He and Xu, 2016; Ma et al., 2011; Ning et al., 2014; Song et al., 2015; Zhang et al., 2015). In the rest of the paper, we first briefly review the SPA framework in Section 2. Section 3 compares SPA with several representative methods in addressing process characteristics using a simulated example; Section 4 discusses how SPA could help address the challenges of smart manufacturing big data and Section 5 draws conclusions.

2. Statistics Pattern Analysis (SPA) Framework

Recently we proposed statistics pattern analysis (SPA) based on the idea of using various statistics to quantify process characteristics, and monitoring these statistics instead of process variables themselves to perform process monitoring (He and Wang, 2011; Wang and He, 2010). In other words, SPA models the variance-covariance structure of the process statistics, instead of the variance-covariance structure of the process variables used in PCA and PLS based methods. Figure 1 shows the schematic plot of how SPA can be applied for continuous and batch process monitoring.

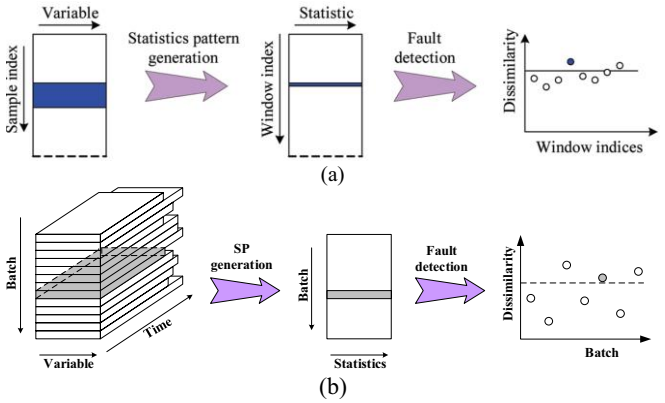


Fig. 1: SPA based fault detection for continuous (a) and batch (b) processes.

For continuous processes, statistics are computed using a window-based approach as shown in Figure 1 (a). The generated process variable statistics, also known as statistics patterns (SPs), are monitored for fault detection. For batch processes, SPs are computed for each batch as shown in Figure 1 (b); then the generated SPs are being monitored. Various simulated and real industrial case studies have demonstrated that SPA is a good platform for SPM (He and Wang, 2011; He and Wang, 2017; Wang and He, 2010). Different process statistics can be included in the SPA framework to capture the key process characteristics: such as those of each individual variable (such as mean, variance, and skewness), the interactions among different variables (such as correlation), as well as process dynamics (such as autocorrelation and cross-correlation). It is worth noting that SPA makes explicit use of higher order statistics (HOS) to measure the extent of nonlinearity and quantify the non-Gaussianity of the probability distribution of the process variables. In addition, process knowledge based landmark features (Wold et al., 2009); profile-driven features (Rendall et al., 2017); geometry based features (Wang et al., 2015) can be included in SPA as well. Therefore, the exact form or the number of statistics/features in SPA varies with applications.

3. SPA in Addressing Current Challenges in SPM

Several key characteristics, such as process dynamics, nonlinearity, and process data non-Gaussianity, are ubiquitous for manufacturing processes, and will become even more so for smart manufacturing. In this section we use a simple yet realistic example to evaluate how several representative methods address these challenges.

The simulated example is a stirred tank heater (Bequette, 1998) as shown in Figure 2, where the objective is to raise the temperature of the inlet stream to a desired value. Detailed descriptions and model assumptions can be found in (Bequette, 1998).

The material and energy balances yield the following two modelling equations:

$$\frac{dT}{dt} = \frac{F}{V}(T_i - T) + \frac{UA(T_j - T)}{V\rho c_p} \tag{1}$$

$$\frac{dT_j}{dt} = \frac{F_j}{V_j}\left(T_{ji} - T_j\right) + \frac{UA(T_j - T)}{V_j \rho_j c_{pj}} \tag{2}$$

where A is the heat transfer area, c_p heat capacity, F volumetric flowrate, ρ density, T temperature, t time, V volume, subscripts i, j, and ji for inlet, jacket, jacket inlet, respectively. To excite the system so that various characteristics such as nonlinearity, dynamics and non-Gaussianity are amplified, a sinusoidal disturbance is injected into :

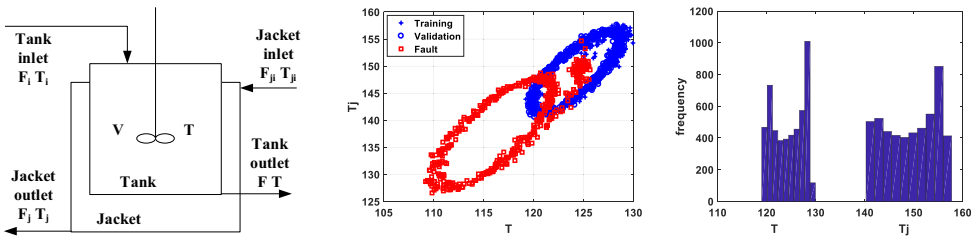
$$F_j = F_{js} + 0.5sin\left(\frac{t \cdot \pi}{10}\right) + n(0, 0.01) \tag{3}$$

where steady-state $F_{js} = 1.5$, and $n(0, 0.01)$ represents white noise with zero mean and standard deviation 0.01.

To illustrate whether the models, reflected as the scores (i.e., samples projected onto the models), can differentiate faulty behavior from normal operation, we also introduce a fault into the process where a leakage in jacket inlet stream is introduced at some point.

$$F_{jf} = F_j - 0.3 \tag{4}$$

The normal data are divided into training and validation subsets. The behaviors of normal and faulty operations are shown in Figure 2 (b) where the tank temperature T is plotted against the jacket temperature T_j. The histograms of T and T_j (Figure 2 (c)) show that the observations obtained from this nonlinear process are highly non-Gaussian.



(a) (b) (c)

Figure 2. (a) Jacketed stirred tank heater (Bequette, 1998); (b) Process behaviors under normal (blue pluses and circles) and faulty (red squares) operations; (c) Histograms of T and T_j.

It is worth noting that the disturbance and fault introduced are rather large for illustrative purposes. In addition, we only investigate the principal or dominant

components while ignoring residual or excluded components. The purpose is to study how each modelling approach would capture or transform certain process characteristics.

For process nonlinearity, kernel based methods, such as kernel PCA (KPCA) and its variants have been studied (Lee et al., 2004a). However, it has been reported that the performances of KPCA-based methods are greatly affected by the choice of kernel and parameters and often the resulted PCs (or mapped features) are still nonlinearly correlated (Wang et al., 2013). This is validated using the stirred tank heater example. Figure 3 (a) and (b) shows the normal and faulty data projected onto the first two PC's of KPCA using Gaussian and sigmoid kernels, respectively. For the same kernel but different parameters, the shapes could be significantly different too. It has been suggested that there is no proper way to guide the choice of kernel and determination of parameters other than cross-validation. Other nonlinear SPM methods such as neural networks and principal curves based methods are not investigated in this work.

To address process dynamics, time lag shifting methods, including dynamic PCA (DPCA) (Ku et al., 1995) and its variants were commonly used SPM methods. However, it has been reported that the resulted PCs are often still auto-correlated. This is also validated using the simulated example. As can be seen from Figure 3 (c), the PC's are still auto-correlated.

For process data non-Gaussianity, independent component analysis (ICA) (Lee et al., 2004b) has been adapted to address it and has gained some popularity in the SPM community. It has been argued that for such cases, the ICA mapped samples would conform to a Gaussian distribution. However, Figure 3 (d), which shows the first two IC's using the default nonlinearity setting of FastICA algorithm (Bingham and Hyvärinen, 2000), indicates that the non-Gaussianity of the process data was not removed or transformed.

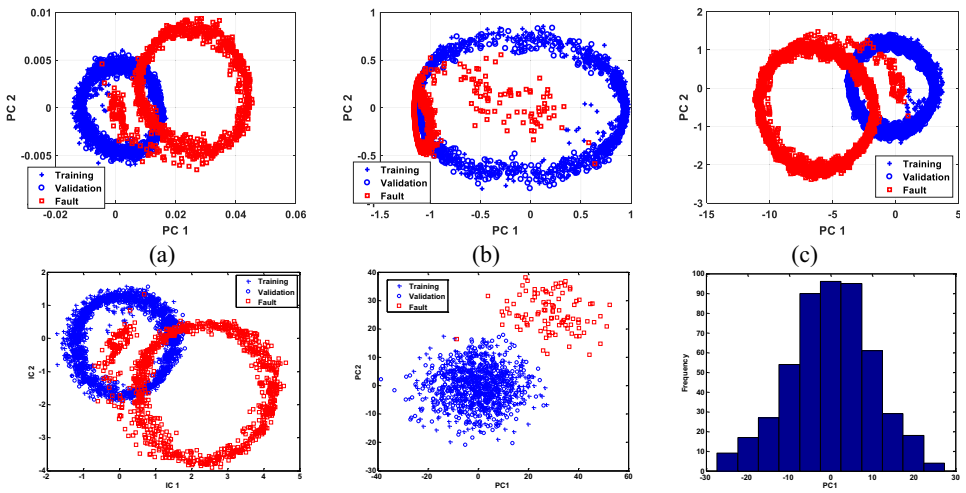


Figure 3. (a) KPCA – Gaussian kernel; (b) KPCA – sigmoid kernel; (c) DPCA; (d) ICA; (e) SPA; (f) Histogram of PC1 from SPA (normal data)

Next we apply SPA to the simulated case study. Window width was chosen to include about ten cycles of T and T_j. Statistics of mean, standard deviation, skewness, kurtosis,

auto- and cross-correlations with maximum lag of 20 are used. Figure 3 (e) shows the score plot of the statistics extracted from the data. It can be seen that the nonlinear dynamic data under normal operation has been transformed into a multivariate Gaussian distributed scores, which is also validated in the histogram of the first score of the statistics in Figure 3 (f). In other words, the SPA transformation has completely transformed the dynamic, nonlinear, and non-Gaussian original process measurements into independent Gaussian distributed statistics. In addition, the cluster of faulty samples are clearly separately from the normal samples.

This simulated example demonstrates that SPA has some advantages in dealing with these common process characteristics, which have contributed to its good performances in fault detection for both batch and continuous processes. Other process characteristics such as multimode, noises, and outliers are not studied in this work.

4. SPA in Addressing Smart Manufacturing Big Data

Smart manufacturing (SM) and big data from SM have drawn increased attention in the SPM community in the past few years (Qin, 2014; Severson et al., 2016). As detailed in (He and Wang, 2017), SPA has many advantages in addressing the 4V challenges of big data. In reducing the number of observations, SPA has been used to reduce an entire batch (or batch step) into batch (or batch step) features. For continuous processes, it has been shown a window-based SPA approach is efficient in significantly reducing number of observations. In reducing number of variables, SPA has been used to extract features from optical emission spectroscopy (OES) and UV-Vis spectra, which effectively reduce number of variables (equal to the number of wavelengths at which the intensities were measured) to much smaller number of features. SPA can help address big data variety as statistics extracted from different data sources can be conveniently integrated. Because SPA can significantly reduce problem size in both time/sample wise and variable wise, and it does not require data pre-processing, SPA has the potential to be used for monitoring real-time streaming data. SPA can also help address big data veracity as data uncertainty will have much less impact on extracted statistics (*e.g.*, mean) than variable themselves.

5. Conclusions

In conclusion, previous work has demonstrated the potential of SPA for SPM in the area of fault detection. But its capability in capturing and transforming process dynamics, nonlinearity, and process data non-Gaussianity and its potential in fault diagnosis and soft sensor or virtual metrology have not been well studied. In this work, we demonstrate that SPA can transform dynamic, nonlinear process data that exhibit strong non-Gaussianity into multivariate Gaussian statistics, which capture the key process characteristics and can be conveniently modelled by multivariate statistical methods such as PCA for various applications such as fault detection and diagnosis, and soft sensor or virtual metrology. Because SPA-based fault diagnosis method breaks down the contribution of a fault to different variable statistics, it provides extra information in addition to identifying the major fault-contributing variable(s), such as whether the fault is due to a change in the variable mean or variance. It is worth noting that SPA is not without its limitations, such as the amount of data required to estimate various statistics, tradeoff between robustness and sensitivity, which will be studied more systematically in the future. In addition, under what conditions that SPA would outperform or underperform other methods needs further investigation.

References

Bequette BW. 1998. Process dynamics: modeling, analysis, and simulation. Prentice Hall PTR Upper Saddle River, NJ.

Bingham E, Hyvärinen A. 2000. A fast fixed-point algorithm for independent component analysis of complex valued signals. *Int. J. Neural Syst.* **10**:1–8.

Deng X, Tian X. 2013. Nonlinear process fault pattern recognition using statistics kernel PCA similarity factor. *Neurocomputing* **121**:298–308.

Deng X, Tian X, Chen S, Harris CJ. 2016. Statistics local fisher discriminant analysis for industrial process fault classification. In: . *Control* , pp. 1–6.

He F, Xu J. 2016. A novel process monitoring and fault detection approach based on statistics locality preserving projections. *J. Process. Control.* **37**:46–57.

He QP, Wang J. 2011. Statistics Pattern Analysis - A New Process Monitoring Framework and Its Application to Semiconductor Batch Processes. *AIChE J.* **57**:107–121.

He QP, Wang J. 2017. Statistical process monitoring as a big data analytics tool for smart manufacturing. *J. Process. Control.*

Ku W, Storer RH, Georgakis C. 1995. Disturbance detection and isolation by dynamic principal component analysis. *Chemom. Intell. Lab. Syst.* **30**:179–196.

Lee J-M, Yoo C, Choi SW, Vanrolleghem PA, Lee I-B. 2004a. Nonlinear process monitoring using kernel principal component analysis. *Chem. Eng. Sci.* **59**:223–234.

Lee J-M, Yoo C, Lee I-B. 2004b. Statistical process monitoring with independent component analysis. *J. Process. Control.* **14**:467–485.

Ma H, Hu Y, Shi H. 2011. Statistics kernel principal component analysis for nonlinear process fault detection. In: . *Intell. Control. Autom.*, pp. 431–436.

Ning C, Chen M, Zhou D. 2014. Hidden Markov model-based statistics pattern analysis for multimode process monitoring: an index-switching scheme. *Ind. & Eng. Chem. Res.* **53**:11084–11095.

Qin SJ. 2014. Process data analytics in the era of big data. *AIChE J.* **60**:3092–3100.

Rendall R, Lu B, Castillo I, Chin S-T, Chiang LH, Reis MS. 2017. A Unifying and Integrated Framework for Feature Oriented Analysis of Batch Processes. *Ind. & Eng. Chem. Res.* **56**:8590–8605.

Severson K, Chaiwatanodom P, Braatz RD. 2016. Perspectives on process monitoring of industrial systems. *Annu. Rev. Control.* **42**:190–200.

Shah D, He QP, Wang J. 2017. A spectroscopic chemometric modeling approach based on statistics pattern analysis. In: . *Proc. 10th IFAC Int. Symp. Adv. Control. Chem. Process.*

Song Y, Jiang Q, Yan X. 2015. Fault diagnosis and process monitoring using a statistical pattern framework based on a self-organizing map. *J. Cent. South Univ.* **22**:601–609.

Wang J, He QP. 2010. Multivariate process monitoring based on statistics pattern analysis. *Ind. Eng. Chem. Res.* **49**:7858–7869.

Wang RC, Edgar TF, Baldea M, Nixon M, Wojsznis W, Dunia R. 2015. Process fault detection using time-explicit Kiviat diagrams. *AIChE J.* **61**:4277–4293.

Wang Z, Xu J, Gao D, Fu Y. 2013. Multiple empirical kernel learning based on local information. *Neural Comput. Appl.* **23**:2113–2120.

Wold S, Kettaneh-Wold N, MacGregor JF, Dunn KG. 2009. Batch process modeling and MSPC.

Zhang H, Tian X, Deng X, Cai L. 2015. A local and global statistics pattern analysis method and its application to process fault identification. *Chin. J. Chem. Eng.* **23**:1782–1792.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50341-4

Simultaneous Fault Detection and Identification in Continuous Processes via nonlinear Support Vector Machine based Feature Selection

Melis Onel[a,b], Chris A. Kieslich[c], Yannis A. Guzman[a,b,d], Efstratios N. Pistikopoulos[a,b,*]

[a] *Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX, 77843, USA*
[b] *Texas A&M Energy Institute, Texas A&M University, College Station, TX, 77843, USA*
[c] *Coulter Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, 30332, USA*
[d] *Department of Chemical and Biological Engineering, Princeton University, Princeton, NJ, 08544, USA*
stratos@tamu.edu

Abstract

Rapid detection and identification of process faults in industrial applications is crucial to sustain a safe and profitable operation. Today, the advances in sensor technologies have facilitated large amounts of chemical process data collection in real time which subsequently broadened the use of data-driven process monitoring techniques via machine learning and multivariate statistical analysis. One of the well-known machine learning techniques is Support Vector Machines (SVM) which allows the use of high dimensional feature sets for learning problems such as classification and regression. In this paper, we present the application of a novel nonlinear (kernel-dependent) SVM-based feature selection algorithm to process monitoring and fault detection of continuous processes. The developed methodology is derived from sensitivity analysis of the dual SVM objective and utilizes existing and novel greedy algorithms to rank features that also guides fault diagnosis. Specifically, we train fault-specific two-class SVM models to detect faulty operations, while using the feature selection algorithm to improve the accuracy of the fault detection models and perform fault diagnosis. We present results for the Tennessee Eastman process as a case study and compare our approach to existing approaches for fault detection, diagnosis and identification.

Keywords: Fault Detection and Identification, Process Monitoring, Data-driven Modeling, Feature Selection, Support Vector Machines.

1. Introduction

Maintaining a safe operation while minimizing the losses in productivity is one of the major goals in chemical processing. Timely detection, diagnosis and identification of faults occurring during the operation is critical and necessary, and this is continuously encouraging researchers to develop novel process monitoring techniques. Today, with the advances in automation and sensor technologies as well as novel smart manufacturing frameworks, real time process data acquisition has become effortless and created an immense opportunity for manufacturing and process industries by enabling

data-driven real time decision making through it. A fault is defined as an abnormal process behaviour where controllers lack the capability of reversing it. Faults can occur due to mechanical reasons such as equipment failure and/or wear as a result of equipment aging or process-based reasons stemming from severe process disturbances (Chiang et al., 2001). Once a fault is detected during an operation, revealing the root-cause by identifying the key process variables is of utmost importance for rapid counteraction to avoid abnormal event progression (Venkatasubramanian et al., 2003). The most prevalent data-driven process monitoring techniques include Principal Component Analysis (PCA), Dynamic PCA (DPCA), Partial Least Squares (PLS), Independent Component Analysis (ICA), and Fischer Discriminant Analysis (FDA) as dimensionality reduction step followed by Q and T^2 statistics, contribution plots, and discriminant analysis (Qin, 2012). The aim of dimensionality reduction is to ensure robustness of the analysis. Dimensionality reduction methods are classified into (a) feature extraction, and (b) feature selection. The dimensionality reduction methods adopted within the listed techniques are based on feature extraction, which project input process data onto another space and alter the original representation of input data, thus may cause loss of information. On the other hand, feature selection is the process of selecting the most informative and relevant original features (e.g. process variables) characterizing the system. Therefore, there is a prominent need for novel data-driven process monitoring methods incorporating feature selection techniques for dimensionality reduction. In this work, we modify a well-known, powerful machine learning algorithm formulation, Support Vector Machines (SVM) (Cortes and Vapnik, 1995), for simultaneous modelling and feature selection. Then, we present the application of the nonlinear (Kernel-dependent) SVM-based feature selection algorithm to process monitoring of continuous processes. Previous studies have used SVM for fault detection in chemical processes (Mahadevan and Shah, 2009; Yin et al., 2014; Xiao et al., 2016). Fault detection is achieved through two-class SVM models, where the feature selection algorithm further improves the model accuracy and also reveals diagnosis of the detected fault. We test the performance of the proposed data-driven framework through the 21 faults introduced in the Tennessee Eastman process data and provide comparisons to existing approaches. The methodology can be implemented as an online decision support tool for continuous process monitoring purposes.

2. Nonlinear Support Vector Machine-based Feature Selection Algorithm

Here, we define a supervised learning problem with l training instances, where $x_i \epsilon R^n$. Indices $i, j = 1,2, \dots, l$ correspond to instances, whereas indices $k, k' = 1,2, \dots, n$ correspond to input data features. Accordingly, the input data has become $x_i = (x_{i1}, x_{i2}, \dots, x_{ik}, \dots, x_{in})^{\mathrm{T}}$. From process monitoring perspective, instances represent distinct continuous operations while features are the collected process measurements from distinct variables. We formulate the fault detection problem in a classification setting with C −parameterized SVM (C −SVM) with hinge loss, l_2 −norm penalty, and linear Kernel as follows:

$$\begin{aligned} & \min_{w,b,\xi} \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{l}\xi_i \\ \text{s.t.} \quad & y_i(w.x_i + b) \geq 1 - \xi_i \qquad i = 1,2, \dots, l \\ & \xi_i \geq 0 \qquad\qquad\qquad\quad i = 1,2, \dots, l \end{aligned} \tag{1}$$

where $y_i \in \{-1,1\}$ denotes the class label of instance i. Eq.(1) is a convex nonlinear problem satisfying a first-order qualification, hence strong duality holds. When Eq.(1) is solved to global optimality, resulted optimal solution (w^*, b^*, ξ^*) is then used to determine the linear decision function $f(x) = w^*.x + b^*$, which yields the sign, thus class, of a new instance x.

One of the major advantages of exploiting Support Vector Machines in modelling is to be able to use Kernel functions, $K(x_i, x_j)$. These functions implicitly map the input feature space, where data is nonlinearly separable, onto a different, possibly higher feature space, where the data become linearly separable. Since most of the chemical processes are nonlinear in nature, we adopt nonlinear (Kernel-dependent) C-SVM formulation, where Kernel functions are introduced in the Lagrange dual formulation of Eq.(1). Next, we introduce binary variables $z \in \{0,1\}^n$ in Eq.(2), which control elimination or selection of feature k, and aim to minimize the number of selected features while maximizing model accuracy via traditional C −SVM formulation:

$$\min_{z} \max_{\alpha} \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j y_i y_j K(x_i \ z, x_j \ z)$$

$$\text{s.t.} \quad \sum_{i=1}^{l} \alpha_i y_i = 0 \tag{2}$$

$$\alpha_i \epsilon [0, C] \qquad i = 1,2, \dots, l$$

$$\sum_{k} z_k = m$$

$$z_k \in \{0,1\} \qquad k = 1,2, \dots, n$$

where α_i are dual variables. In Eq.(2), is the Hadamard product operator for componentwise multiplication, and m is the number of optimal subset of features, where the aim is to attain highest C −SVM model accuracy with minimum number of input features. Since Eq.(2) is a challenging and impractical problem to solve to global optimality, we propose an algorithmic solution procedure, which utilizes the Lagrangian sensitivity of the objective function value of Eq.(2) with respect to z_k at $(\alpha^*; z)$, where z_k is treated as a fixed parameter. The procedure is iterative where in each step, the features are eliminated individually according to the following criterion:

$$crit_k = -\frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i^* \alpha_j^* y_i y_j \frac{\partial K(x_i \ z, x_j \ z)}{\partial z_k}\bigg|_{z=1}$$

$$k_{worst} = arg_k \max crit_k \tag{3}$$

Eq.(3) delineates the nonlinear SVM-based feature selection algorithm adopted in this study. In particular, the algorithm is equivalent to well-known recursive feature elimination (RFE)-SVM classification algorithm when performing a linear classification that is using a linear Kernel. The presented algorithm (Kieslich et al., 2016) has been implemented in C++/Python environment using LibSVM library (Chang and Lin, 2011), and has been successfully utilized in our previous studies in bioinformatics.

3. Tennessee Eastman Process

The Tennessee Eastman process simulation, an extensively used benchmark study to evaluate and further develop process monitoring algorithms, was designed by the Eastman Chemical Company (Downs and Vogel, 1993). It is based on a real industrial process, in which the components, kinetics, and operating conditions have been modified for proprietary reasons (Chiang et al., 2001). The process includes five primary units: a reactor, condenser, compressor, separator, and a stripper. It describes the production of chemicals G and H from feedstocks A, C, D and E with byproduct F and inert compound B. The simulated process data used in this study is adopted from Chiang and Braatz, 2001, where the process contains 41 measured and 11 manipulated variables. The variables are sampled every 3 minutes and the dataset includes measurement samples from normal and 21 distinct faulty operations. For further information on the process and simulation, the interested reader can refer to Downs and Vogel, 1993 and Chiang et al., 2001.

4. Proposed Framework

In this study, we are building 21 fault-specific C −SVM binary classifiers for 21 different faults introduced in the process data. Thus, for each of the 21 model building, we combine data from normal and faulty operation from each of the 21 faults. The initial step in data-driven modelling is normalization of the input data. Normalization is performed on each of the 21 faulty-normal process data by subtracting the mean of measurements across the operations from each measurement and then dividing into the standard deviation of them.

4.1. Offline Phase Model Building

The initial step in the offline phase is to create balanced train and test sets. Imbalanced data sets may significantly increase risk of overfitting problem. Therefore we create train and test sets from each 21 datasets with 100 runs of 5-fold cross-validation. Next, we build binary classifier models iteratively for each feature subset. In each iteration we (i) tune C −SVM parameters using train and test data sets with the active set of features (whole feature set in the first iteration), (ii) train C −SVM classifiers with Gaussian Radial Basis Function, where the class probabilities are smoothened via taking median of probabilities with a window size of 3, (iii) calculate Lagrangian sensitivity of dual C −SVM objective function with respect to the parameter z_k to obtain a rank list of features, (iv) eliminate the most redundant, "worst" feature satisfying Eq.(3) from data. This procedure is repeated until we are left with 1 feature in the train and test sets. The result of this framework yields 1092 C −SVM classifiers, which includes 21 fault-specific C −SVM classifiers for each of 52 feature subsets. Particularly, in the first step, we tune C −SVM hyperparameters C and γ. Tuning is done via grid search for all combinations between 2^{-10} and 2^{10}. Hyperparameter combination yielding the highest average testing accuracy and recall is chosen in each iteration of the framework above. The final stage in the offline phase is selection of fault-specific end-models for fault detection. Among all feature subsets, the C −SVM classifier yielding the highest fault detection rate and accuracy is chosen to be the fault-specific end-model.

4.2. Online Phase Model Building

21 fault-specific end models that are built in offline phase are implemented in industrial setting to monitor the incoming online process data. When online process data is fed,

models generate a binary answer for detection of each fault. If a fault is detected, alarm rises and the optimal set of features of the corresponding end-model yield the root-cause analysis of the detected fault instantaneously.

5. Results

We present the results of the proposed data-driven process monitoring framework on the Tennessee Eastman process below (Table 1), where we have adopted a five consecutive fault alarm policy to report a fault. The performance of the end-models is evaluated via (i) fault detection rate, (ii) accuracy, and (iii) latency.

Table 1. Performance of the Fault Specific End Models.

Fault	Optimal Feature Subset Size	Fault Detection Rate	Accuracy	Latency (min)
1	2	99.88	99.90	6
2	8	98.13	98.44	48
3	4	100.00	83.33	3
4	1	100.00	100.00	3
5	25	100.00	100.00	3
6	1	100.00	100.00	3
7	1	100.00	100.00	3
8	6	99.38	96.88	3
9	6	99.50	83.33	3
10	9	98.25	83.85	3
11	29	96.63	86.46	3
12	9	100.00	96.15	3
13	2	95.00	89.48	3
14	3	100.00	100.00	3
15	7	98.25	82.81	3
16	26	100.00	87.92	3
17	7	97.63	84.69	3
18	24	90.38	91.56	183
19	3	100.00	83.33	3
20	4	100.00	83.75	3
21	1	100.00	99.90	3

Table 2 compares the performance of this study to the other existing data-driven process monitoring techniques (Mahadevan and Shah, 2009; Yin et al., 2014; Xiao et al., 2016) in terms of fault detection rate. This comparison demonstrates the power of the proposed data-driven process monitoring framework. Additionally, the average latency among the reported faults has been provided as 306.19, 145.58, 263.12, 151.00, and 98.50 min for PCA-T^2, PCA-Q, DPCA-T^2, DPCA-Q, and 1-class SVM respectively in Mahadevan and Shah, 2009; whereas it is significantly lower for the proposed framework (15.67 min for the Table 2 reported faults, and 13.86 min among all 21 faults).

6. Conclusions

In this study, we have presented a nonlinear SVM-based feature selection algorithm procedure and implemented it for data-driven process monitoring of continuous processes. The proposed framework establishes a promising decision support tool for

online fault detection and identification. This research was funded by U.S. National Institute of Health (NIH) grant P42 ES027704.

Table 2. Comparison through Fault Detection Rate. Best results of Xiao et al., 2016 is adopted.

Ref.	Mahadevan and Shah, 2009					Xiao et al, 2016	Yin et al, 2014	This Study
Fault	**PCA-T2**	**PCA-Q**	**DPCA-T2**	**DPCA-Q**	**One class SVM**	**One class SVM**	**Two class SVM**	**Two class SVM**
1	99.20	99.80	99.40	99.50	99.80	99.50	99.50	**99.90**
2	98.00	**98.60**	98.10	98.50	**98.60**	98.30	98.12	98.10
4	4.40	96.20	6.10	**100.00**	99.60	47.40	99.88	**100.00**
5	22.50	25.40	24.20	25.20	**100.00**	45.20	90.75	**100.00**
6	98.90	**100.00**	98.70	**100.00**	**100.00**	99.20	60.13	**100.00**
7	91.50	**100.00**	84.10	**100.00**	**100.00**	70.10	98.91	**100.00**
8	96.60	97.60	97.20	97.50	97.90	97.40	96.00	**99.38**
10	33.40	34.10	42.00	33.50	87.60	68.00	81.00	**98.25**
11	20.60	64.40	19.90	80.70	69.80	65.80	80.25	**96.62**
12	97.10	97.50	99.00	97.60	99.90	98.80	97.75	**100.00**
13	94.00	**95.50**	95.10	95.10	**95.50**	95.00	92.50	95.00
14	84.20	**100.00**	93.90	**100.00**	**100.00**	93.90	91.00	**100.00**
16	16.60	24.50	21.70	29.20	89.80	73.10	89.38	**100.00**
17	74.10	89.20	76.00	94.70	95.30	75.20	81.63	**97.62**
18	88.70	89.90	88.90	90.00	90.00	89.30	89.50	**90.38**
19	0.40	12.70	0.70	24.70	83.90	43.60	85.88	**100.00**
20	29.90	45.00	35.60	51.00	90.00	69.00	80.50	**100.00**
21	26.40	43.00	35.60	44.20	52.80	59.40	-	**100.00**

References

L.H. Chiang, E.L. Russell, R.D. Braatz, 2001, Faut Detection and Diagnosis in Industrial Systems, Springer.

C. Cortes and V. Vapnik, 1995, Support Vector Networks, Machine Learning, 20, 273-297.

J.J. Downs and E.F. Vogel, 1993, A Plant-wide Industrial Process Control Problem, Computers and Chemical Engineering, 17, 245-255.

C.A. Kieslich, P.Tamamis, Y.A. Guzman, M. Onel, C.A. Floudas, 2016, Highly Accurate Structure-based Prediction of HIV-1 Coreceptor Usage Suggests Intermolecular Interactions Driving Tropism, PLoS One, 11.

S. Mahadevan and S.L. Shah, 2009, Fault Detection and Diagnosis in Process Data Using One-class Support Vector Machines, Journal of Process Control, 19, 1627-1639.

S.J. Qin, 2012, Survey on Data-driven Industrial Process Monitoring and Diagnosis, Annual Reviews in Control, 36, 220-234.

V. Venkatasubramanian, R. Rengaswamy, K. Yin, S.N. Kavuri, 2003, A Review of Process Fault Detection and Diagnosis Part I: Quantitative Model-based Methods, Comp Chem Eng, 27, 293-311.

Y. Xiao, H. Wang, W. Xu, J. Zhou, 2016, Robust one-class SVM for fault detection, Chemometrics and Intelligent Laboratory Systems, 151, 15-25.

S. Yin, X. Gao, H.R. Karimi, X. Zhu, 2014, Study on Support Vector Machine-Based Fault Detection in Tennessee Eastman Process, Abstract and Applied Analysis, Hindawi.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50342-6

Feature-based Virtual Metrology for Semiconductor Manufacturing

Kerul Suthar[a], Devarshi Shah[a], Jin Wang[a], Q. Peter He[a*]

[a]*Department of Chemical Engineering, Auburn University, Auburn AL 36849, USA*
qhe@auburn.edu

Abstract

In semiconductor manufacturing, virtual metrology (VM), a.k.a. soft sensor, is the prediction of wafer properties using process variables and other information available for the process and/or the product (*i.e.*, machine data) without physically conducting property measurement. VM has been utilized in semiconductor manufacturing for process monitoring and control. In this work, we discuss the shortcomings of some of the commonly used VM methods and propose a feature-based VM framework. An industrial case study is used to demonstrate the effectiveness of the proposed method.

Keywords: semiconductor manufacturing, virtual metrology, process monitoring, statistics pattern analysis.

1. Introduction

In semiconductor manufacturing, a wafer undergoes hundreds of different steps to yield the final product. After a processing step, typically a few (1–3) wafers within a lot are measured at the metrology station, and this sampled metrology data represent the whole lot. However, this methodology using the traditional off-line metrology tools (*e.g.*, ellipsometer, atomic force microscope (AFM)) becomes insufficient when the device dimensions continue to decrease and the lot-to-lot process control is being increasingly replaced with the wafer-to-wafer (W2W) control. W2W control requires metrology measurements of every wafer and it has been proposed to use the integrated metrology (IM) sensors at the processing tool to provide such measurements (Lensing and Stirton, 2006). However, issues such as impact on throughput, increase in cycle time, and higher cost make IM less attractive in many process environments. On the other hand, virtual metrology (VM) technology has been suggested as an alternative to 100% wafer measurement to support W2W control (Gill et al., 2010; He et al., 2012; Hung et al., 2007). Because machine data are usually sampled much more frequently compared to metrology data, and machine data are instantly available compared to delays often associated with metrology tools, an accurate VM can significantly improve process monitoring and control performance by providing real-time predicted metrology data.

One of the most important factors that need to be considered when implementing any VM for industrial applications is the level of data pre-processing required, which affects the deployment and maintenance of the VM. Generally speaking, the more automated the data pre-processing step is, the more sustainable the method is likely to be in a production environment. To address this challenge, in this work, we propose a feature-based VM framework, which is based on the statistics pattern analysis (SPA) process modeling and monitoring framework we proposed previously (He and Wang, 2011; Wang and He, 2010). The most significant difference between the proposed VM approach and other existing approaches is that instead of extracting correlations between

process variables and metrology measurements, the proposed method extracts the correlations between process features and metrology measurements to build VM models. By doing so, the proposed method can not only eliminate most data pre-processing steps, but also provide superior prediction performance. In addition, the proposed feature-based VM framework is advantageous in handling big data challenges. The rest of the paper is organized as follows: Section 2 provides a brief review on existing VM approaches; Section 3 presents the feature-based VM framework; Section 4 uses an industrial case study to demonstrate the performance of the proposed VM method, which is compared with several other VM approaches; finally Section 5 presents the conclusion and discussion.

2. A Brief Review of Existing VM Approaches

VM is not unique to the semiconductor industry, which essentially serves the same purposes as the soft sensor, a term used in the process industry. VM or soft sensor makes use of secondary variables that are measured online frequently to predict the product quality variables that are not measured online or not measured frequently, which could be modeled using either model-based approaches or data-driven approaches. For industrial processes, data driven approaches are usually easier to develop and to implement online, therefore they are potentially more attractive. Due to the limited space, only some of the data-driven VM approaches applied to semiconductor manufacturing processes are reviewed in this work.

Among data-drive approaches, the commonly used ones are time series analysis (TSA), Kalman filter (KF), multiple linear regression (MLR), principal component regression (PCR), and partial least squares (PLS).

2.1. Time series analysis (TSA)

Because the metrology data are generally sequential in time, auto-regressive-integrated-moving average (ARIMA) models can be identified, e.g., following the procedure proposed by Box and Jenkins. Once the model structure is determined and parameters are estimated using the historical metrology data, the model can be used to predict the future values of the metrology data. It is worth noting that TSA only utilizes the metrology data for model building and prediction, while the process or machine data are completely ignored. This could be the reason for the poor prediction performance of TSA in the industrial case study presented in Section 4.

2.2. Kalman filter (KF)

Kalman filter was proposed in the early 1960s and has been extensively used for the state estimation of dynamic systems. It has also been formulated for VM (Gill et al., 2010):

$$\boldsymbol{K} = \boldsymbol{P}_{old}\boldsymbol{C}^T(\boldsymbol{C}\boldsymbol{P}_{old}\boldsymbol{C}^T + \boldsymbol{R})^{-1} \quad (1)$$

$$\boldsymbol{x}_{new} = \boldsymbol{x}_{old} + \boldsymbol{K}(y - \boldsymbol{C}\boldsymbol{x}_{old}) \quad (2)$$

$$\boldsymbol{P}_{new} = \boldsymbol{P}_{old} - \boldsymbol{K}\boldsymbol{C}\boldsymbol{P}_{old} \quad (3)$$

$$y_{est} = \boldsymbol{C}\boldsymbol{x}_{new} \quad (4)$$

where $\boldsymbol{K}$ is the Kalman gain, $\boldsymbol{P}$ the state error covariance matrix, $\boldsymbol{R}$ the measurement noise covariance matrix, $\boldsymbol{x}$ the independent or process variables, y the dependent or metrology variable.

2.3. Multiple linear regression (MLR)

Multiple linear regression (MLR) aims to model the relationship between multiple explanatory or independent variables from machine data and a response or dependent variable of metrology data by fitting a linear equation to the historical data, which takes the following form:

$$y = \sum_{i=1}^{p} \beta_i x_i + v \quad (5)$$

where y represents the metrology measurement, $x_i (i = 1,2, \cdots p)$ the process variables, v the intercept and β_i the coefficients. The coefficients are estimated by minimizing the sum of squares of the differences between the actual and modeled metrology measurements, and the obtained model is used to predict metrology measurement when a new set of process variables are available. The potential issue with MLR for VM is that, more often than not, the process variables are (highly) correlated and the collinearities among x_i can cause severe problems for MLR – the estimated coefficients $\hat{\beta}_i$ can be very unstable, which makes predictions by the regression model unstable or poor.

2.4. Principal component regression (PCR)

Principal component regression (PCR) is a regression analysis technique based on principal component analysis (PCA). Due to limited space, the technical details of PCR and PCA are not reviewed here. In short, instead of regressing the dependent variable (i.e., the metrology measurements) on the explanatory or independent variables (i.e., the process variables) directly as in MLR, the principal components (PCs) of the explanatory variables are used as regressors in PCR. Compared to MLR, PCR has the advantage of addressing the multicollinearity problem. In addition, PCR handles noisy process variables better as usually only a subset of all the PCs are used to build the model. However, the PCs are derived without any reference to the dependent variables. In other words, PC's explain the most variation in independent variables, which may not be (highly) related to the variation in the dependent or metrology variables. Due to this reason, the performance of PCR for VM is not guaranteed.

2.5. Partial least squares (PLS)

Partial least squares (PLS) has all the benefits of PCR while also taking the variation of dependent variables into account. Specifically, PLS models the inner relation that correlates the scores of independent variables with the scores of dependent variables. Therefore, PLS usually has better prediction performance than PCR, which explains why PLS and its variants are the most commonly used VM methods in industrial applications.

Other methods, such as artificial neural networks (ANNs) and support vector machine (SVM) have also been used for VM, which are not reviewed here.

It is worth noting that all existing VM methods discussed above make predictions by extracting linear or nonlinear correlations between process variables and metrology measurements. One drawback of utilizing process variables is that because of the characteristics associated with batch processes, such as unequal batch and/or step length, and unsynchronized or misaligned batch trajectories, various data preprocessing steps are usually required for the existing VM methods. These preprocessing steps, including trajectory alignment/warping and data unfolding, are often performed off-line and could be difficult to automate. To address this limitation, in the following section, we present the proposed feature-based VM framework, where batch statistics and other features are used as the regressors to predict metrology measurements, which eliminates the data preprocessing step. In addition,

we show that the feature-based VM framework provides superior prediction performance compared to the traditional VM methods in Section 4.

3. Feature-based Virtual Metrology (FVM)

The feature-based VM (FVM) framework is developed based on statistics pattern analysis (SPA), a process monitoring framework we proposed previously. In SPA, various statistics are used to quantify process characteristics, and monitoring these statistics instead of process variables themselves to perform process monitoring. SPA has been applied for fault detection, fault diagnosis and virtual metrology. In this work, we extend the features to not just statistics, but also other features such as integrals, derivative, etc. In the FVM framework, we hypothesize that the batch behavior can be better characterized by the *process features* than by the *process variables*. Therefore, in the FVM framework, process features instead of process variables are used as input variables to build the VM model. Figure 1 provides a schematic diagram of the FVM framework which consists of two steps. In the first step, various features are extracted from batch trajectories such as various statistics that characterize individual variables (such as the mean, variance, autocorrelation), the interactions among different variables (such as the cross-correlations), as well as other features that characterize the process (such as batch and step durations, the time integrals of power input). In the second step, a regression method, such as PLS, is used to extract the relationships between the features and the metrology measurements. As can be seen from Figure 1, unequal batch/step length and unsynchronized batch trajectories will have no effect on the FVM framework.

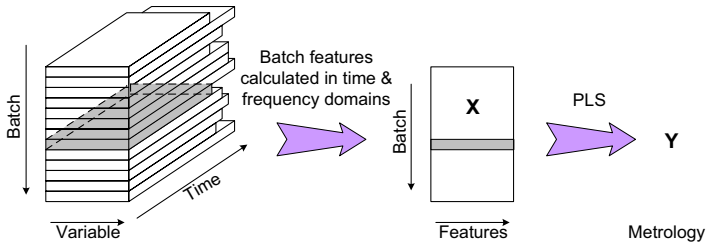


Figure 1. The schematic diagram of the feature-based virtual metrology

4. Application to an Industrial Case Study

In this work, a dataset collected from a plasma etch system at one of Texas Instruments' wafer fabs (Gill et al., 2010) is used to compare the proposed feature-based VM and other VM methods. The dataset contains the recorded values of 18 Optical Emission Spectroscopy (OES) signals collected every 0.1 second for 1121 wafers. The dataset also contains the metrology measurement values of the sheet resistance, which is one of the most important electrical- test parameters used in the semiconductor manufacturing industry to assess the electrical quality of a product. The goal of VM is to predict the end-of-batch sheet resistance using the OES signals.

One independent variable (*i.e.*, one OES signal) of several wafers is shown in Figure 2, which shows the typical characteristics of a semiconductor machine data: unequal run times or process durations; large variations between wafers and unsynchronized trajectories. To apply traditional VM methods such as PLS on this type of data, several

data pre-processing steps have to be taken, including trajectory alignment or time warping to make trajectories equal length, and trajectory unfolding to flatten the 3-D structure into 2-D matrix.

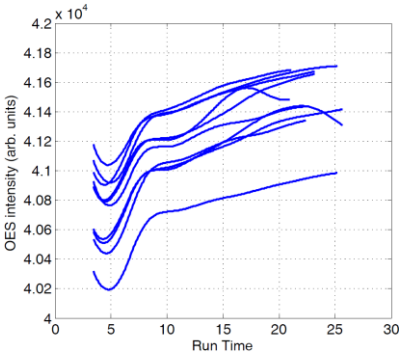


Figure 2. A sample OES signal of several wafers

4.1. Settings and performance measures for the VM methods

In this work, we compare the proposed feature-based VM (FVM) with the following VM modeling approaches: time series analysis (TSA); Kalman filter (KF); multiple linear regression (MLR); principal component regression (PCR); partial least squares (PLS) and recursive PLS (RPLS).

All methods were optimized using cross-validation. For feature-based VM, time integral of the OES signals, along with univariate statistics including mean, variance, skewness and kurtosis were included as the features. PLS and RPLS were used as the regression methods for static and recursive FVM methods. The prediction performance of the VM methods are quantified by mean absolute percentage error ($MAPE$) and the coefficient of determination (R^2) as defined below.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right| \times 100\% \quad (6)$$

where n is the total number of samples, y_i the actual metrology value of the output, and $\hat{y}_i$ the VM predicted value of the output.

$$R^2 = 1 - \frac{SS_{err}}{SS_{tot}} \quad (7)$$

where $SS_{err} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$, $SS_{tot} = \sum_{i=1}^{n}(y_i - \bar{y})^2$, and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

4.2. Performance comparison of static VM methods

In this subsection, the static FVM is applied to the dataset discussed previously to predict the sheet resistance using the OES data. The performance is compared with other VM methods. For all VM methods, 70% of the data (784 wafers) were utilized for model building and the rest 30% of the data (337 wafers) were used for testing. The VM models were not updated even when new data becomes available. Table 1 compares the *MAPE* and R^2 of the static FVM to those of TSA, MLR, PCR, and PLS. It can be seen that FVM outperforms all other methods in terms of both *MAPE* and R^2. Compared to other methods, the *MAPE* of FVM is reduced by 24.7%~58.5%, while the R^2 of feature-based VM is improved by 20.5%~67.1%.

Table 1. Comparison of different static VM methods

Model	MAPE	%change	R^2	%change
TSA	8.46	53.8%	0.40	-45.2%
MLR	8.72	58.5%	0.24	-67.1%
PCR	7.19	30.7%	0.55	-24.7%
PLS	6.86	24.7%	0.58	-20.5%
FVM	5.50	0%	0.73	0%

4.3. Performance comparison of recursive VM methods

Here the recursive FVM is applied to the dataset. The performance is compared with other recursive VM methods. For recursive VM, the initial VM model is built based on the training data of 784 wafers, and is updated when new data becomes available. It can be seen that the recursive FVM outperforms other recursive VM methods in terms of both *MAPE* and R^2.

Table 2. Comparison of different recursive VM methods

Model	MAPE	%change	R^2	%change
RPLS	5.67	13.4%	0.70	-9.1%
KF	5.07	1.4%	0.72	-6.5%
FVM	5.00	0%	0.77	0%

5. Conclusions

A feature-based VM (FVM) framework is proposed to address the challenges presented in semiconductor VM applications, such as unequal batch/step duration and/or unsynchronized trajectories; and large number of variables caused by data unfolding. Because FVM does not require any data preprocessing steps, it is uniquely suited for automatic online applications. The performance of FVM is compared with several commonly used VM algorithms using an industrial data set. Both static and recursive versions of FVM outperforms those of commonly used static and recursive VM methods, respectively.

References

Gill B, Edgar TF, Stuber J. 2010. A novel approach to virtual metrology using Kalman Filtering. *Future Fab Int.* **35**:86–91.

He QP, Wang J. 2011. Statistics Pattern Analysis - A New Process Monitoring Framework and Its Application to Semiconductor Batch Processes. *AIChE J.* **57**:107–121.

He QP, Wang J, Gilicia HE, Stuber JD, Gill BS. 2012. Statistics pattern analysis based virtual metrology for plasma etch processes. In: . *Am. Control. Conf.*, pp. 4897–4902.

Hung M-H, Lin T-H, Cheng F-T, Lin R-C. 2007. A novel virtual metrology scheme for predicting CVD thickness in semiconductor manufacturing. *IEEE/ASME Trans. mechatronics* **12**:308–316.

Lensing K, STIRTON B. 2006. Integrated metrology and wafer-level contro. *Semicond. Int.* **29**.

Wang J, He QP. 2010. Multivariate process monitoring based on statistics pattern analysis. *Ind. Eng. Chem. Res.* **49**:7858–7869.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50343-8

CFD-Based Design of Focal Brain Cooling System for Suppressing Epileptic Seizures

Kei Hata[a], Takuto Abe[a], Takao Inoue[b], Koichi Fujiwara[a], Takatomi Kubo[c], Toshitaka Yamakawa[d], Sadahiro Nomura[b], Hirochika Imoto[b], Michiyasu Suzuki[b], Manabu Kano[a*]

[a]*Department of Systems Science, Kyoto University, Kyoto 606-8501, JAPAN*
[b]*Department of Neurosurgery, Yamaguchi University, Ube 755-8505, JAPAN*
[c]*Department of Information Science, Nara Institute of Science and Technology, Ikoma 630-0192, JAPAN*
[d]*Department of Computer Science and Electrical Engineering, Kumamoto University, Kumamoto 860-8555, JAPAN*
manabu@human.sys.i.kyoto-u.ac.jp

Abstract

Epilepsy is a group of neurological disorders which is caused by excessive neuronal activities in cerebrum and characterized by recurrent seizures. A quarter of patients have intractable epilepsy and do not become seizure-free with medication. We are developing an implantable and wearable focal brain cooling system, which enables the patients to lead ordinary daily life. The system cools the epileptic focus, where the excessive neuronal activities begin, by delivering cold saline to a cranially implanted cooling device. In this research, we developed a whole system model through the first principles and animal experiments. The results of system design have shown that a cooling device with more complex channel structure achieves higher temperature uniformity in the brain with lower flow rate of saline. The optimal structure was derived by taking account of the trade-off between pressure drop and temperature uniformity. In addition, the results have demonstrated that the cooling duration is less than 10 minutes for the average temperature 2 mm below the cooling device (inside the brain) to reach 25 °C; it is short enough to cool the brain after seizure is predicted by existing electroencephalogram (EEG)-based algorithms. Moreover, the frequency of battery charging would be once in several days for most patients.

Keywords: epilepsy, focal brain cooling, computational fluid dynamics, Pennes bioheat equation, medical systems engineering

1. Introduction

Epilepsy is a chronic disorder, the hallmark of which is recurrent, unprovoked seizures. There are many different types of seizures, e.g. convulsion, consciousness disturbance, and unusual feeling. The seizures may increase the risk of injury and affect the social life of the patients. The epidemiological studies have reported that the epileptic patients consist of approximately 1% of the population, and that one-quarter of them do not become seizure-free by the existing treatments. Anti-epileptic drug therapy is the first option for most patients, and it provides seizure freedom for two-thirds of them. The remaining one-third is usually treated by surgery. The only surgical option for seizure freedom is the resection of the epileptic focus, where the excessive neuronal activity

begins. The resection cannot be applied to the patients whose epileptic foci are identified in cerebral hemispheres, difficult to localize precisely, or cannot be resected without irreversible brain dysfunction. These patients are treated by other surgical options such as callosotomy, vagus nerve stimulation (VNS), and responsive neurostimulation. They provide palliative treatments and rarely result in seizure freedom. Thus, a new effective treatment of epilepsy needs to be developed.

The potential treatments include focal brain cooling, which cools the epileptic focus. The stimulation-evoked seizures during intraoperative brain mapping were suppressed within 5-10 seconds by pouring the cold Ringer's lactate solution over the surface of the stimulated cortex (Sartorius and Berger, 1998). In addition, abnormality in the interictal electroencephalogram (EEG) was suppressed by applying the cold saline onto the brain surface near the epileptic focus (Karkar et al., 2002). The suppression effect of the focal brain cooling would result from the cooling of layers II and III in gray matter (Hiraishi et al., 2013). The target temperature in this region would be 25-30 °C (Imoto et al., 2006). In the experiments with rats, the penicillin G solution or the cobalt powder was applied on the brain surface and then a Peltier device was placed on the area of application. The Peltier device in some rats was activated and that in the other rats was not. As a result, the rats without the activation of the Peltier device had frequent seizures, while the rats whose brain surface was cooled to 20 or 15 °C had less frequent seizures, and the rats whose brain surface was 10 °C had no seizures (Fujii et al., 2012).

The aim of our project is the world's first clinical application of an implantable and wearable focal brain cooling system, with which the patients can lead ordinary daily life. The target patients are those who have their epileptic focus in their gyrus, because the surgical treatment of neocortical epilepsy is difficult (Rothman et al, 2005) and the gyrus is easier to cool by pressing the cooling device on the brain surface. The clinical application is preceded by the animal experiments and the system design. The present research aims to design the focal brain cooling system by using newly developed models: a whole system model and a pressure drop model.

2. Focal brain cooling

The use of a Peltier device is a potential way for the focal brain cooling. There is a safety risk in touching the Peltier device onto the brain, because it requires large electrical current and is not made of biocompatible materials. Figure 1 shows our focal brain cooling system. A patient wears a jacket with two pockets on his/her chest. One pocket contains the Peltier device in contact with a tank of saline. The tank is connected to a titanium cooling device implanted in the skull. The location of the implantation is just above the epileptic focus. The cooling device measures 30 x 30 x 7 mm, and the thickness of the top and bottom covers is 1 mm, and that of the middle layer is 5 mm. The saline circulates throughout the system; it is cooled in the tank by the Peltier device and pressurized by a pump attached on the chest. The Peltier device and the pump are powered by a mobile battery attached on the other side of the chest.

For determining when to activate and deactivate the system, a seizure prediction algorithm will be constructed in the future. The algorithm will use EEG, brain temperature, intracranial pressure, and oxygen saturation and concentration of oxyhemoglobin in cerebral blood flow. These indices are measured by sensors implanted in the skull (Yamakawa et al., 2014). Epileptic seizures could be predicted 90 minutes prior to their onset with an existing EEG-based algorithm (Iasemidis, 2005).

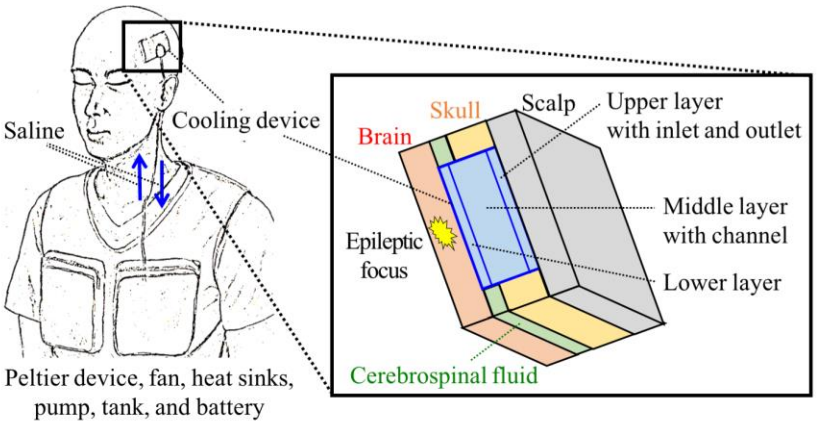


Figure 1. Focal brain cooling system.

3. Model of focal brain cooling system

The developed model consists of a whole system model and a pressure drop model. Figure 2 shows the hemispheric head model, which is the main part of the whole system model. It consists of a titanium cooling device, saline, and biological tissue. The biological tissue is further divided into soft tissue, skull, cerebrospinal fluid (CSF), and brain (Nelson and Nunneley, 1998). The channel structures in the middle layer of the cooling device are shown in Table 1. The tubes for inflow and outflow of the saline are implanted under the layer of the soft tissue.

The temperature of the titanium device was calculated by the heat conduction equation. The velocity, pressure, and temperature of saline were calculated by regarding flow as laminar flow. In addition, the temperature of biological tissue was calculated by the modified Pennes bioheat equation, which was based on the original bioheat equation proposed by Pennes (1948). The original equation describes the energy conservation, assuming that the heat convection rate by blood flow is proportional to the difference between the tissue temperature and the body core temperature. We modified the original equation in two ways. First, the metabolic heat production was neglected because its rate is much smaller than the heat convection rate by the blood flow (Orr and Eberhart, 1998). Second, the blood perfusion rate was set at 0 in the CSF, the skull, and the soft tissue; the blood perfusion rate in the brain was determined by an animal experiment so that the experimentally obtained temperature distribution would fit the theoretically obtained one.

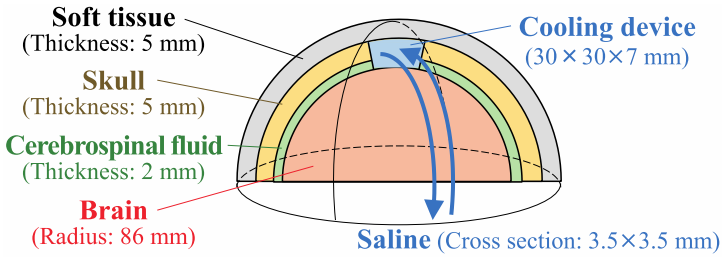


Figure 2. Geometrical representation of head model.

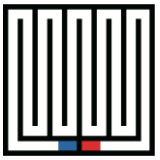
Since the whole system model imposes heavy computational burden, a simpler model called the pressure drop model was constructed. This model defines the relationship between the flow rate of the saline and the pressure drop for the circulation of saline; the tube was approximated to be cylindrical and 2 m in length. Computational fluid dynamics (CFD) simulations were conducted with COMSOL Multiphysics 5.1.

4. Comparison of channel structures

An optimization problem was formulated for finding the optimal channel structure that minimizes the pressure drop for the circulation of saline under the following constraints; 1) the average temperature 2 mm below the cooling device is 25.0 °C, and 2) the difference between the maximum and the minimum temperatures on the bottom of the cooling device at steady state is 2.3 °C. The temperature and the pressure drop were calculated with the whole system model and the pressure drop model, respectively.

The results in Table 1 have demonstrated that the most complex channel structure (device C) can achieve the desired temperature uniformity with the lowest flow rate and pressure drop. The more complex device than device C might result in higher temperature uniformity or lower pressure drop, which results in the use of a smaller pump. The upper limit of the channel complexity is given by the technical difficulty or the cost of manufacturing.

Table 1. Comparison of channel structures in cooling device. The blue and red regions represent the inlet and the outlet of saline.

Device	A	B	C
			
Width [mm] (channel, fin)	13.5, 1.0	3.83, 1.0	1.9, 1.0
Temperature on bottom of cooling device [°C] (max, ave, min)			
	19.9, 17.8, 17.6	19.7, 17.7, 17.4	19.6, 17.8, 17.3
Temperature 2 mm below cooling device [°C] (max / ave / min)			
	29.2, 25.0, 24.3	29.1, 25.0, 24.3	29.1, 25.0, 24.3
Flow rate [L/min]	1.0	0.1	0.06
Pressure drop [kPa]	20	2	1

5. Duration for focal brain cooling

Two unsteady-state analyses were performed for device C. The results shown in Fig. 3 have indicated that it took less than 10 minutes for the average temperature 2 mm below the cooling device to reach 25 °C. The temperature on the hot and the cold sides of the Peltier device was lower than 40 °C and higher than 5 °C, respectively.

The duration required for cooling the brain is much shorter than the time from seizure prediction to seizure onset, which is around 90 minutes with an existing EEG-based

algorithm. The maximum temperature 2 mm below the cooling device does not reach 25 °C after 20 minutes cooling. Since the maximum temperature is marked in the peripheral area of the cooling device, the cooling device should be larger than the target region for cooling.

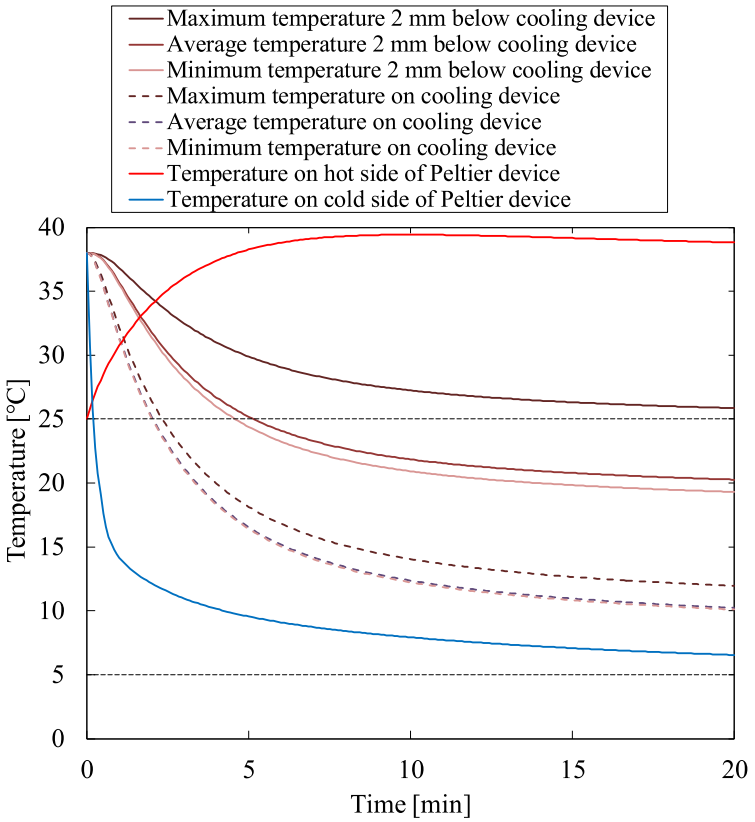


Figure 3. Temporal change of temperature in brain and Peltier device.

6. Frequency of Battery Charging

The cooling time per seizure is the sum of the time required for lowering and maintaining the temperature. The former was determined to be 10 minutes or less, and the latter is 10 minutes based on the experiences in Yamaguchi University School of Medicine. It was found that the battery has to be charged every two or three seizures. Since the quartile frequency of seizures with awareness impairment in temporal lobe epilepsy is 0.10, 0.16, 0.47 times in a day (Wiebe et al., 2001), most patients would need to charge the battery once in several days.

7. Conclusions

In the present work, the physical model of the focal brain cooling system was developed. It is characterized by the hemispheric head model, which utilizes the heat conduction equation, the basic equations of fluid dynamics, and the modified Pennes bioheat equation. To determine a constant in the modified Pennes bioheat equation, an

animal experiment was performed. To reduce the computational burden, the pressure drop model was constructed, which describes the relationship between the flow rate and the pressure drop of the saline. With these models, three types of cooling devices were compared. The results have shown that the more complex channel structure results in the desired temperature uniformity in the brain with the lower flow rate and pressure drop of saline.

The results have also shown that it takes less than 10 minutes for the average temperature 2 mm below the cooling device to reach 25.0 °C. It is much shorter than the duration from seizure prediction to seizure onset by using an existing EEG-based algorithm. In addition, the frequency of battery charging would be once in several days for most patients. The present work demonstrates the feasibility of the focal brain cooling system in clinical application.

References

M. Fujii, T. Inoue, S. Nomura, Y. Maruta, Y. He, H. Koizumi, S. Shirao, Y. Owada, I. Kunitsugu, T. Yamakawa, T. Tokiwa, S. Ishizuka, T. Yamakawa, M. Suzuki, 2012, Cooling of the epileptic focus suppresses seizures with minimal influence on neurologic functions, Epilepsia, 53, 3, 485-493.

T. Hiraishi, H. Kitaura, M. Oishi, M. Fukuda, S. Kameyama, H. Takahashi, A. Kakita, Y. Fujii, 2013, Significance of horizontal propagation of synchronized activities in human epileptic neocortex investigated by optical imaging and immunohistological study, Epilepsy Res., 104, 1-2, 59-67.

L. D. Iasemidis, D. S. Shiau, P. M. Pardalos, W. Chaovalitwongse, K. Narayanan, A. Prasad, K. Tsakalis, P. R. Carney, J. C. Sackellares, 2005, Long-term prospective on-line real-time seizure prediction, Clin. Neurophysiol., 116, 3, 532–544.

H. Imoto, M. Fujii, J. Uchiyama, H. Fujisawa, K. Nakano, I. Kunitsugu, S. Nomura, T. Saito, M. Suzuki, 2006, Use of a peltier chip with a newly devised local brain-cooling system for neocortical seizures in the rat, J. Neurosurg., 104, 1, 150-156.

K. M. Karkar, P. A. Garcia, L. M. Bateman, M. D. Smyth, N. M. Barbaro, M. Berger, 2002, Focal cooling suppresses spotaneous epileptiform activity without changing the cortical motor threshold, Epilepsia, 43, 8, 932-935.

D. A. Nelson and S. A. Nunneley, 1998, Brain temperature and limits on transcranial cooling in humans: Quantitative modeling results, Eur. J. Appl. Physiol., 78, 353-359.

C. S. Orr and R. C. Eberhart, 1998, Bioheat transfer in blood perfused tissues and clinical applications of hypothermia, Annual Reviews of Heat Transfer, 9, 1-78.

H. H. Pennes, 1948, Analysis of tissue and arterial blood temperatures in the resting human forearm, J. Appl. Physiol., 1, 2, 93-122.

S. M. Rothman, M. D. Smyth, X. F. Yang, G. P. Peterson, 2005, Focal cooling for epilepsy: An alternative therapy that might actually work, Epilepsy Behav., 7, 2, 214-221.

C. J. Sartorius, M. S. Berger, 1998, Rapid termination of intraoperative stimulation-evoked seizures with application of cold ringer's lactate to the cortex, J. Neurosurg., 88, 349-351.

S. Wiebe, W. T. Blume, J. P. Girvin, M. Eliasziw, 2001, A randomized, controlled trial of surgery for temporal lobe epilepsy, N. Engl. J. Med., 345, 5, 311-318.

T. Yamakawa, T. Inoue, Y. He, M. Fujii, M. Suzuki, M. Niwayama, 2014, Development of an implantable flexible probe for simultaneous near-infrared spectroscopy and electrocorticography, IEEE Trans. Biomed. Eng., 61, 2, 388-395.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018,San Diego,California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50344-X

Modeling of spreading and drying of aqueous polymer coatings on pharmaceutical tablets during film coating

Charalampos Christodoulou,[a] Luca Mazzei,[a] Salvador García-Muñoz,[b] Eva Sorensen[a*]

[a]*Department of Chemical Engineering, University College London, Torrington Place, London WC1E 7JE, UK*
[b]*Eli Lilly and Company, Lilly Research Laboratories, Indianapolis, IN 46285, USA*
**e.sorensen@ucl.ac.uk*

Abstract

Pharmaceutical tablets are film coated for several reasons, the most common being the enhancement of taste and appearance, the protection of sensitive ingredients and the control of active ingredient release. In this work, we developed a mathematical model that simulates coating spray impact, film formation and drying on pharmaceutical tablet surfaces. To describe spray impingement and liquid film formation we developed a 1-D film-spreading model. To predict the drying of the liquid film, we used the "mixture modeling" theory and the "lubrication approximation" theory to simplify the equations describing the liquid-particle system. The current model allows predicting the thickness of an evaporating coating layer applied on a curved tablet. We performed variance-based sensitivity analysis to study the effect of spray parameters on coating spreading and initial liquid film thickness and to investigate the influence of tablet and film properties on drying.

Keywords: pharmaceutical, film coating, drying, modeling.

1. Introduction

Coating is an important step in the manufacture of pharmaceutical tablets. Film coatings are usually applied on tablets by spraying polymer solutions or dispersions on their surface (Felton, 2013). After impinging on the surface, the droplets spread forming a liquid film. As the solvent evaporates, a polymeric coating develops, covering the tablet core (Fig.1). In our work, to model droplet impingement we assumed the spray to be uniform and we divided it into arrays of droplets that impinge successively on the substrate orthogonally to its surface. We assumed that over a time interval of magnitude $t_i = \overline{D}_0/\overline{U}_0$ the droplets in the first array impinge and spread without being affected by subsequent droplets. Here $\overline{D}_0$ and $\overline{U}_0$ denote the initial diameter and velocity of a single droplet, respectively. We estimated the initial wetted area by calculating the spreading of a single droplet and by assuming that all the droplets of the first array behave independently and identically as they impact on the dry surface. Then, subsequent droplet arrays impingement makes the wetted area spread further.

To model the drying of the liquid coating film, we employed the "mixture modeling" approach which is based on the assumption that the mean velocity fields of the solid (polymer particles) and liquid (solvent) phases relax to local equilibrium (Schwartz et al., 2001). The model (briefly presented in Section 2) yields both the spreading rate and the thickness of the liquid film as well as the drying rate and final thickness of the polymer coating.

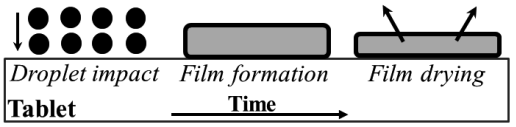


Figure 1. Outline of the coating application process during pharmaceutical coating.

After the tablet leaves the spray-zone, the film spreading caused by spray impingement stops. At this stage, we can ignore inertial forces and simplify the momentum and mass balance equations governing the thin film behavior using the lubrication approximation. In our model, the main equations that describe coating drying are a height evolution equation that calculates the film thickness during solvent evaporation and a convection-diffusion equation that predicts the polymer mass fraction in the film. The final outputs of the model are the time required by the film to dry completely and the final thickness of the dry polymer film.

Additionally, we performed variance-based sensitivity analysis (Sobol sampling; Process Systems Enterprise, 2017) to study the effect of spray parameters (droplet size, impact velocity, and coating viscosity, density and surface tension) on film spreading rate and thickness and to investigate the influence of coating and tablet properties (particle diffusion coefficient, evaporation rates, tablet curvature) on film drying times.

2. Mathematical model

In this section, we present the mathematical model we developed to simulate film application and drying during the coating process. The impact of the first array of droplets leads to tablet wetting. The initial wetted area will spread further when the subsequent droplets impinge (Fig.2). We assumed that the tablet curvature has negligible effect on the wetted area spreading while the tablet is under the spray. The liquid film height calculated from the spreading model is then considered as an initial condition for the film drying equations that give us the evolution of the film thickness during solvent (water) evaporation.

2.1. Droplet impact on dry substrate

To model the impingement of the first droplet array on the dry tablet we developed a 1D spreading model for a single droplet that before impact has the average velocity and size of the spray droplets. Assuming that all the droplets of the first array do not interact and behave in the same way, we calculate the initial wetted area. The spreading model is based on the mechanical energy equation, written as follows:

$$\frac{dE_k}{dt} + \frac{dE_p}{dt} + \Phi = 0 \tag{1}$$

where E_k and E_p are the kinetic and potential energies of a spreading droplet of volume V, respectively, and Φ is the rate of viscous dissipation of kinetic energy. The kinetic and potential energies are:

$$E_k = \frac{1}{2}\rho \int_V \boldsymbol{u} \cdot \boldsymbol{u}\, dV \;\; ; \;\; E_p = \sigma(A_d - \pi R_d cos\theta_e) \tag{2}$$

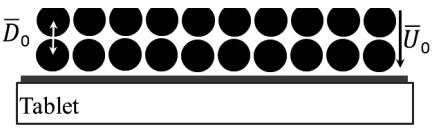


Figure 2. Spray (droplet arrays) impingement on the tablet during film-coating.

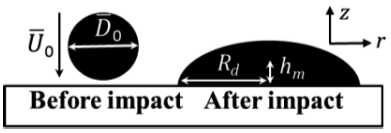


Figure 3. Single droplet impingement on a dry tablet. Droplet shape before and after impact.

where $\boldsymbol{u}$ is the velocity field in the spreading droplet and ρ, σ, A_d, R_d and θ_e are the droplet density (assumed to be constant), surface tension, air-droplet area, wetted area radius and contact angle, respectively. The kinetic energy loss rate is given by:

$$\Phi = \int_V \boldsymbol{\tau} : \nabla \boldsymbol{u} \; dV \; ; \; \boldsymbol{\tau} = \mu(\nabla \boldsymbol{u} + \nabla \boldsymbol{u}^T) \tag{3}$$

where $\boldsymbol{\tau}$ is the deviatoric stress tensor and μ is the viscosity of the droplet (assumed to be constant, but dependent on the solid concentration). To use Eqs. (2) and (3), we need to assume a functional form for the field **u** within the droplet. Following our previous approach (Christodoulou et al. 2017), we assumed irrotational axisymmetric stagnation point flow. The velocity components (u_r, u_z) in cylindrical coordinates are then given by:

$$u_r = -\frac{1}{2}\frac{1}{h_m}\frac{dh_m}{dt}r \; ; \; u_z = \frac{1}{h_m}\frac{dh_m}{dt}z \tag{4}$$

where h_m is the position of the center of mass of the droplet. By assuming that the droplet has a spherical cap shape after impact, we reduced the number of unknown variables to one (the wetted area radius, R_d) and we solved the 1D mechanical energy equation (Eq.1). We show the shape of the droplet before and after impact in Fig.3. At any given time $t<ti$, the equivalent radius of the wetted area is taken equal to $R_i^2 = N_d \cdot R_d^2$, where N_d is the number of droplets in each array, which depends on the spray mass flowrate Q and on the characteristic time t_i ($Q \cdot t_i = N_d \cdot \rho \cdot V$).

2.2. Spray impingement – film formation

After the first array impacts on the dry tablet surface, the subsequent droplets that impinge will contribute to the spreading of the initial wetted area. We assumed that at this stage the wetted area is made up of disconnected cylindrical films. The spreading of these films is governed by the mechanical energy equation:

$$\frac{dE_{kf}}{dt} + \frac{dE_{pf}}{dt} + \Phi_f + \frac{dW_d}{dt} = 0 \tag{5}$$

where E_{kf} and E_{pf} are the total kinetic and potential energies in each film, while Φ_f is the rate of kinetic energy dissipation in each film. The last term on the left hand side of Eq. (5) is the rate of work done on the surface of each film by the impinging droplets (of mass flowrate Q_f), which is given by:

$$\frac{dW_d}{dt} = -\frac{1}{2}Q_f \bar{U}_0^2 \tag{6}$$

We calculated the energy terms in a similar way as in Eqs. (2) and (3). Following Madejski (1976), we assumed that the components of the velocity field are given by:

$$u_{r,f} = \frac{2}{H_f}\frac{1}{R_f}\frac{dR_f}{dt}rz \; ; \; u_z = -\frac{2}{H_f}\frac{1}{R_f}\frac{dR_f}{dt}z^2 \tag{7}$$

where dR_f/dt is the rate of expansion of the cylindrical film and R_f and H_f are the wetted area radius and film height, respectively. We considered the spreading to be completed when the wetted area covers the entire tablet surface; this occurs when the disconnected cylindrical films get into contact with one another.

The film formation equations can provide information about the liquid coating film (thickness and wetted area on the tablet) that is applied on the tablet during one pass through the spray zone.

2.3. Film drying

Pharmaceutical coatings are multicomponent liquids. We assumed that the liquid film applied on the surface of the curved tablet surface after spray impingement (Sections 2.1 and 2.2) consists of a carrier fluid and polymer particles. To describe the liquid-particle system we used the mixture modeling approach that is based on three equations: a continuity and a momentum balance equation written for the mixture (which is treated as an effective fluid) and a continuity equation written for the solid phase:

$$\rho \frac{D\boldsymbol{u}}{Dt} = -\nabla p + \nabla \cdot \boldsymbol{\tau} + \rho \boldsymbol{g} \tag{8}$$

$$\partial_t \rho + \nabla \cdot (\rho \boldsymbol{u}) = 0 \ ; \ \partial_t (\rho w) + \nabla \cdot (\rho \boldsymbol{u} w) + \nabla \cdot \boldsymbol{J_p} = 0 \tag{9}$$

where $\boldsymbol{J_p}$ and w are the particle mass flux with respect to the mass averaged velocity of the mixture and the particle mass fraction, respectively. Because the coating dries more quickly in the thinner regions close to the corners of the tablet (Schwartz et al., 2001), the liquid-particle system will develop concentration gradients during drying. Following Weidner et al. (1996), we assumed Fickian diffusion of polymer particles in the coating film and wrote a linear relation between the species flux and mass fraction:

$$\boldsymbol{J_p} = -\rho \mathcal{D} \nabla w \tag{10}$$

where $\mathcal{D}(m^2 s^{-1})$ is the particle diffusion coefficient given by $\mathcal{D} = \mathcal{D}_0 \exp[-A(w - w_{t=0})]$ (Weidner et al., 1996). For the parameters $\mathcal{D}_0$ and A we used the characteristic values for coating film drying found in Weidner et al. (1996). This assumption is quite general and therefore future work should focus on specifying these parameters for a pharmaceutical system.

After the initial stage of spray impact (Sections 2.1 and 2.2), the thin film flow and drying are significantly slower. This allows us to use the lubrication approximation for the mixture continuity and momentum balance equations. The film height evolution equation for a Newtonian, incompressible coating reads (Schwartz et al., 2001):

$$\partial_t h = -\nabla \cdot \boldsymbol{F} - J_e \text{ , where } \boldsymbol{F} = \int_0^h \boldsymbol{u}\, dz = -\frac{h^3}{3\mu} \nabla p \tag{11}$$

In Eq. (11), $\boldsymbol{F}$ is the coating (mixture) flux rate on the surface of the tablet, h is the height of the film and p and J_e are the pressure in the film and the evaporation velocity, respectively:

$$p = -\sigma \kappa - \sigma \Delta h \ ; \ J_e = J_{e0}(1 - w)^{\nu} \tag{12}$$

where κ is the curvature of the tablet and Δh is the curvature of the surface of the liquid coating film. Characteristic values for J_{e0} and ν can be found in the literature (Weidner et al. 1996; Schwartz et al. 2001). The lubrication theory also allows us to simplify the continuity equation for the solid phase (Eq. 9b). We derived the following equation that agrees with the one reported in the work of Schwartz et al. (2001):

$$\frac{Dw}{Dt} \equiv \partial_t w + \boldsymbol{u} \cdot \nabla w = \frac{J_e}{h} w + \frac{1}{h} \nabla \cdot (h' \mathcal{D} \nabla w) \tag{13}$$

where h' is the polymer particle "height" given by $h' = \varepsilon_p h$, where ε_p is the volume fraction of the polymer particles.

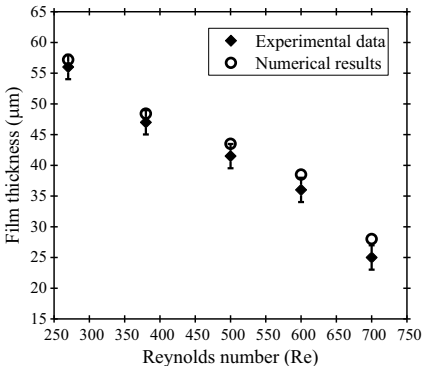


Figure 4. Validation for the liquid film thickness. Experiments by Kalantari and Tropea (2014).

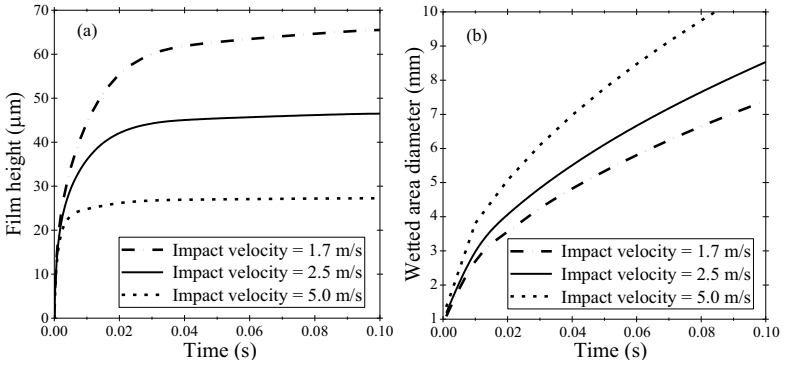


Figure 5. Effect of droplet impact velocity on a) film height and b) film spreading.

Solution of Eqs. (11) to (13) provides information about the film thickness at different positions on the tablet and about the drying time. We implemented the entire model in gPROMS, employing the DASOLV solver (backward differentiation formulae) of the Modelbuilder platform (Process Systems Enterprise Ltd., 2017).

3. Results

In this section, we present the numerical results that describe the film behavior on the tablet after spray impingement in the spray zone. We validated the model predictions for the liquid coating film thickness during spraying with experimental data from the literature and studied the effect of process parameters on the polymer film thickness.

In Fig. 4, we compare our numerical predictions for the liquid film thickness after spray impact with experimental data from the literature. Kalantari and Tropea (2014) experimentally investigated viscous liquid spray impact on rigid surfaces. Fig. 4 shows that higher spray impact Reynolds numbers ($Re \equiv \rho \overline{U}_0 \overline{D}_0 / \mu$) lead to thinner liquid films. The numerical results are in good agreement with the experimental data. In Fig. 5, we illustrate the effect of the mean droplet impact velocity on the film thickness and spreading. Higher impact velocities lead to thinner films and faster spreading.

In Fig.6, we present the model prediction for the film thickness evolution during drying in coating process conditions. The simulation shows that the tablet curvature ($\kappa = -1/L$, where L is the tablet surface curvature radius $\cong$ 13.5mm) affects the final dry film thickness since the film is thinner at the corner. This prediction agrees with the results of Weidner et al. (1996) and Schwartz et al. (2001).

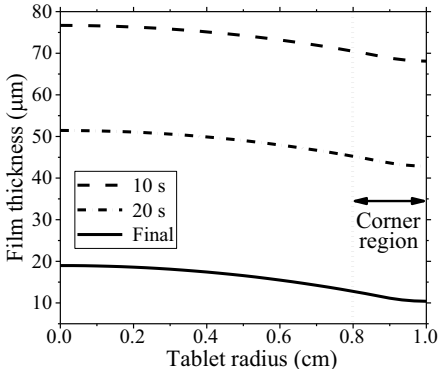


Figure 6. Film thickness during drying (simulation with J_{e0} =2.0μm/s, ν =0.5, A=15, $\mathcal{D}_0$=2·10^{-5}cm^2/s).

To determine the effect of spray parameters on film thickness and time required to cover a tablet surface we performed variance-based sensitivity analysis for a tablet with diameter 1cm. The total sensitivity indices appear in Table 1. The droplet mean velocity is the most important parameter followed by the mixture viscosity. For the drying model, only the tablet curvature significantly affects coating uniformity (thickness difference) and parameters J_{e0} and v influence drying times. The rest of the parameters have very small total sensitivity indices (<0.001).

Table 1. Variance-based sensitivity indices for spray impingement.

Factor	Film thickness (S_{Hf})	Spreading time (S_{tf})
Density, surface tension (σ ,ρ)	0.006	0.001
Droplet mean diameter ($\bar{D}_0$)	0.008	0.007
Droplet mean velocity ($\bar{U}_0$)	0.822	0.940
Coating viscosity (μ)	0.164	0.052

4. Conclusions

The mathematical model developed in this work calculates the liquid coating spreading after spray impingement on a tablet and the subsequent polymer film drying on its surface. The numerical results are in good agreement with experimental data from the literature (Kalantari and Tropea, 2014). The sensitivity analysis yields useful insight into the effect of model parameters on coating application. Future work will include investigation of the non-Newtonian behavior of coatings and a detailed evaporation model that accounts for polymer particle coalescence during drying.

References

Christodoulou, C., Sorensen, E., García-Muñoz, S. and Mazzei, L., 2018. Mathematical modelling of water absorption and evaporation in a pharmaceutical tablet during film coating. Chem. Eng. Sci., 175, pp.40-55.

Felton, L.A., 2013. Mechanisms of polymeric film formation. Int. J. Pharm., 457(2), pp.423-427.

Kalantari, D. and Tropea, C., 2014. Liquid spray impact onto flat and rigid walls: Formation and spreading of accumulated wall film. Fluid. Dyn. Mater. Process, 10, pp.37-61.

Madejski, J., 1976. Solidification of droplets on a cold surface. International Journal of Heat and Mass Transfer, 19(9), pp.1009-1013.

Process Systems Enterprise Ltd. gPROMS User Guide. London, UK, 2017.

Schwartz, L.W., Roy, R.V., Eley, R.R. and Petrash, S., 2001. Dewetting patterns in a drying liquid film. J. Colloid Interface Sci., 234(2), pp.363-374.

Weidner, D.E., Schwartz, L.W. and Eley, R.R., 1996. "Role of surface tension gradients in correcting coating defects in corners", J. Colloid Interface Sci., 179(1), pp.66-75.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018,San Diego,California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50345-1

Towards the optimal design of a minimum set of clinical trials for the identification and characterization of VWD

Beatrice Taverna[a], Alessandra Casonato[b], Fabrizio Bezzo[a], Federico Galvanin[c]

[a]*CAPE-Lab, Department of Industrial Engineering, University of Padova*
[b]*Department of Medicine,University of Padua Medical School*
[c]*Department of Chemical Engineering, University College London*
fabrizio.bezzo@unipd.it

Abstract

Von Willebrand disease is one of the most severe inherited bleeding disorders in humans, characterized by qualitative and/or quantitative defects of the von Willebrand factor protein. Diagnosis is difficult due to the high heterogeneity of the disease. Pharmacokinetic models have been recently proposed and applied to help in the disease characterization and diagnosis. However, the complexity of the models requires long and invasive dynamic non-routine tests to be carried out on the subjects to achieve a statistically satisfactory estimate of the individual metabolic parameters. In this work, it is demonstrated how the use of basal clinical tests and a shorter dynamic clinical test may allow for the identification of a mechanistic model of the disease. An existing mechanistic model of von Willebrand disease has been modified to account for the basal tests, where new model equations are derived using response surface methodology. Results show a good agreement between the model response and the clinical data.

Keywords: von Willebrand disease; model identification; basal clinical tests

1. Introduction

Von Willebrand disease (VWD) is one of the most frequent human bleeding disorders, caused by an alteration of von Willebrand factor (VWF), a multimeric glycoprotein present in the bloodstream. VWF has a fundamental function in the hemostatic process. It mediates platelet aggregation and thrombus growth, and binds, transports and protects the coagulation factor VIII. VWD is caused by defects of VWF, with symptoms ranging from sporadic nosebleeds and mild bleeding from small lesions to acute thrombocytopenia and/or prolonged bleeding episodes (Lillicrap, 2007). Diagnosis of VWD may be complicated due to the existence of different VWD types (1, 2A, 2B, Vicenza) depending on the nature of VWF defect. Pharmacokinetic (PK) models have been recently proposed for the classification of the disease, elucidating the critical pathways involved in the disease characterization (Galvanin et al, 2013). However, the complexity of the models requires the execution of the long (at least 24 h) and invasive 1-desamino-8-D-argine vasopressin (DDAVP) test (Casonato et al, 2011) on the patients to achieve a statistically satisfactory estimate of the individual metabolic parameters. Hence, there is the need to propose alternative easier and less costly clinical investigations to identify the PK model of VWD. A possible solution is to take advantage of basal clinical trials, *i.e.* the VWF antigen (VWF:Ag), VWF collagen-binding (VWF:CB) and VWF propeptide (VWFpp), which can be carried out taking a single blood sample. In this

work it will be shown how basal clinical trials can be used effectively to estimate a subset of the kinetic parameters in a PK model of VWD. Specifically, response surface methods will be used to develop regression models based on a set of available clinical data. Furthermore, an analysis of the dynamic Fisher Information Matrix (FIM) (Bard, 1974) will be then used to design a shorter DDAVP test for the estimation of the PK parameters. The procedure will be validated on clinical data with the aim of demonstrating the potential for an identification approach based on shorter and less costly clinical tests.

2. Clinical data

Our research work has been carried out using clinical data derived by 20 VWD patients belonging to 2B and Vicenza VWD categories, and 42 normal subjects with HnonO and HO blood group. Patients and normal subjects were studied in accordance with the Helsinki Declaration, after obtaining their written informed consent, and the approval of the Hospital ethical board for the study. Clinical data and experimental methods have been supplied by the Hospital of Padova and they mainly come from the standard diagnostic procedure for VWD, which consider both basal and dynamic tests. The basal tests that have been considered in our study are: i) VW:Ag, to evaluate the concentration of VWF in the bloodstream. Plasma VWF:Ag was measured with a home-made ELISA method, using a horseradish peroxidase (HRP)-conjugated anti-VWF antibody. A pool of normal plasma was used to construct the reference curve. The results are given in U/dL, taking the first reference curve dilution as 100 U/dL; ii) VWF:CB, to evaluate the ability of VWF to bind collagen, and explore the presence of large, hemostatically more efficient, multimers. VWF:CB is assessed by ELISA using type III collagen diluted in acetic acid. The results are given in U/dL, taking the first reference curve dilution as 100 U/dL; iii) VWFpp to quantify the elimination rate of VWF from the bloodstream. VWFpp was measured using a home-made ELISA method, using CLB-Pro 35 and CLB-Pro 14.3-HRP antibodies. From these tests, two other basal quantities are derived: the VWFpp ratio, defined as VWFpp/VWF:Ag, which gives an indirect measure of VWF survival and the VWF:CB ratio, expressed as VWF:CB/VWF:Ag, which is suggestive of the amount of large VWF multimers in the bloodstream (Casonato et al, 2011). The dynamic test used is the DDAVP trial, which consists in the collection of blood samples before and 15, 30, 60, 120, 180, 240, 480 min and 24 h after administration of DDAVP subcutaneously at a dose of 0.3 μg kg^{-1}. The changes in VWF:Ag and VWF:CB plasma concentrations at different time points after the DDAVP are then analysed to get important information on the three most important VWF pathways, which are: VWF release, VWF proteolysis and VWF elimination.

3. The PK model of VWD

The PK model taken as reference in this research work has been recently proposed by Ferrari et al (2018). As reported in Figure 1, this PK model of VWD assumes that after DDAVP administration, high molecular weight (HMW) and ultralarge molecular weight (UL) VWF multimers are released from the endothelial cells. Then, the HMW and UL multimers are cleaved into low molecular weight (LMW) multimers by the metalloproteinase ADAMTS-13 and then eliminated from the bloodstream. The model is described by a system of differential and algebraic equations where each subject is characterized by three main PK parameters, namely the VWF release rate k_0 [h^{-1}], the proteolysis rate k_1 [h^{-1}] and the elimination rate k_e [h^{-1}]. The amount of released VWF is represented by parameter D [U/dL]. It is important to notice that, for a given subject,

parameter k_0 quantifies the rate of release, while D quantifies the amount of VWF released from the endothelial cells.

The model differential and algebraic equations are:

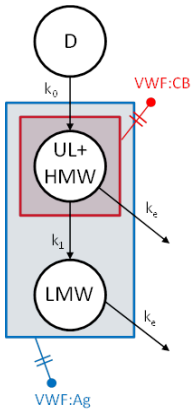


Figure 1. *The simplified PK model of VWD by Galvanin et al.*

$$\frac{dx^{UL+HMW}}{dt} = k_0 D\, e^{-k_0(t-t_{max})} - k_1\left(x^{UL+HMW} - x_b^{UL+HMW}\right) - k_e\left(x^{UL+HMW} - x_b^{UL+HMW}\right) \quad (1)$$

$$\frac{dx^{LMW}}{dt} = k_1\left(x^{UL+HMW} - x_b^{UL+HMW}\right) - k_e\left(x^{LMW} - x_b^{LMW}\right) \quad (2)$$

$$y^{AG} = \frac{x^{UL+HMW} + x^{LMW}}{V_d} \quad (3)$$

$$y^{CB} = \frac{x^{UL+HMW}}{V_d} \quad (4)$$

where: V_d = 40 mL/kg$_{bw}$ is the approximate distribution volume. Basal conditions are assumed at $t(0)$, i.e. $x(0) = [x_b^{UL+HMW} x_b^{LMW}] = [y_b^{CB} V_d y_b^{AG} V_d - y_b^{CB} V_d]$.

The model is able to correctly characterize the three most important VWF pathways, which are: VWF release (identified by k_0), VWF proteolysis (identified by k_1) and VWF elimination (identified by k_e). However, to estimate the model parameters, which means to quantify the rate of VWF release (k_0), proteolysis (k_1) and elimination (k_e) and to describe the time courses of VWF:Ag and VWF:CB, the DDAVP test needs to be carried out. The set of parameters to be estimated from the available measurements is $\boldsymbol{\theta} = [k_0\ k_1\ k_e\ D\ t_{max}]$. A maximum likelihood parameter estimation has been carried out using the commercial software gPROMS®. Measurements are assumed to be normally distributed with a standard deviation of 2 U/dL. The parameter set $\boldsymbol{\theta}$ is determined for each individual by iteratively solving an SQP optimization problem. A crucial step in the parameter estimation exercise is setting the value of the first guess $\boldsymbol{\theta}_0$ for $\boldsymbol{\theta}$. In this study, $\boldsymbol{\theta}_0$ was obtained from a preliminary parameter estimation for each group of subjects (healthy, VWD type 2B, and VWD type Vicenza) using average concentration profiles obtained from subjects belonging to the same group. Then the parameter estimation routine was initialized with $\boldsymbol{\theta}_0$, and the set of parameters was estimated for each individual subject based on his/her own concentration profiles.

4. Proposed methodology

The PK model of Ferrari et al (2018) shows two significant limitations. Firstly, it requires the execution of the invasive DDAVP test to achieve a statistically satisfactory estimation of the PK parameters; secondly, it does not consider the information that can be obtained from basal tests. Therefore, to overcome these issues, it is reasonable: (i) to investigate methods which allow to calculate the PK parameters from basal clinical trials through regression models; (ii) to include the correlations derived by a regression procedure in the set of model equations, leading to a first level and to a second level of model modification; (iii) to suggest a shorter DDAVP test for the estimation of the remaining (reduced) set of parameters.

The PK model parameters, which are directly calculated from basal clinical trials are the elimination kinetic constant k_e and the proteolytic kinetic constant k_1. It is possible to demonstrate that these parameters require the longest DDAVP execution time for being identified with the maximum information. To achieve this conclusion, information content analysis has been carried out on the simplified PK model of VWD in order to quantify the DDAVP execution time that is required to maximize the information content. The metric that is used to evaluate the information content of the clinical test is the trace of Fisher Information Matrix (FIM), which is defined as $I_d = \sum_{t_{sp}} tr[H_\theta]$. Approximatively, the FIM can be mathematically expressed as the product of the sensitivity of the *j*-th output variable with respect to the *k*-th parameter in the conditions investigated in the i-th experiment divided for the variance $[\mathrm{H}]_{\mathrm{kl}} \cong \sum_{\mathrm{i}=1}^{\mathrm{N_{exp}}} \sum_{\mathrm{j}=1}^{\mathrm{N_m}} \left[\frac{1}{\sigma_{\mathrm{ij}}^2}\left(\frac{\partial \hat{y}_{\mathrm{ij}}}{\partial \theta_{\mathrm{k}}}\frac{\partial \hat{y}_{\mathrm{ij}}}{\partial \theta_{\mathrm{l}}}\right)\right]$ (Bard, 1974). The remaining PK parameters k_0, D and t_{max} are all related to the description of the VWF release from the endothelial cells. However, it was demonstrated (Ferrari et al, 2018) that k_0 and D are highly correlated and are difficult to estimate separately without dynamic experiments. Furthermore, t_{max} is not strictly related to any basal clinical quantity, but it is function of the intrinsic release mechanism. Hence, regression methods cannot be applied to obtain k_0, D and t_{max}, but a shorter DDAVP test may be used for their estimation.

4.1 The response surface methodology (RSM)

The regression method used in this research work is known as response surface methodology (RSM). RSM is a technique developed by Box-Behnken (1960), used to define black-box models, which allows to investigate possible correlations between inputs and outputs in systems where nothing is known of the inside. Basal clinical data (§2) and the estimation of the PK model parameters (k_0, k_1, k_e, D, t_{max}) (§3) were available for each patient considered in the pool and they constitute the experimental data set. The experimental data set has been used for the construction of basal state correlations, which relate the PK model parameters k_e and k_1 to basal clinical trials thorough a response surface methodology (RSM). The first objective of RSM is to hypothesize an analytical form of the response surface, which manages to approximate the experimental data. Based on the Akaike information criterion (Akaike, 1974), a linear response surface model with interactions appears to be the best candidate for representing the real system. The fitting procedure has been executed in OriginProand the sum of squared errors (SSE) and the R-square (R^2) statistics were used to assess the goodness of fit. Two basal state correlations have been defined, which allow to estimate the elimination kinetic constant k_e (Eq. 5) and the proteolytic kinetic constant k_1 (Eq. 6) from basal clinical trials (VWF:Ag, VWF:CB and VWFpp) and the related basal quantities (VWFpp ratio and VWF:CB ratio, indicated in the equations as VWF:R):

$$k_e = A + B \cdot VWFppratio + C \cdot VWF{:}R + D \cdot VWFppratio \cdot VWF{:}R \quad (5)$$

$$k_1 = A + B \cdot VWF{:}Ag + C \cdot VWFppratio + D \cdot VWF{:}Ag \cdot VWFppratio \quad (6)$$

In Figure 2, it can be observed that the fitting of the experimental data of HnonO subjects using (Eq.5) appears to be good and this is confirmed also by the goodness of Fit statistics reported in Table 1, where the R^2 is higher than 90% for all the considered categories. The same observation can be derived considering the fitting of the experimental data using (Eq.6). The two correlations (Eq. 5 and 6) have been included in the PK model of VWD. From Table 2, it appears that the modified PK model can estimate both the elimination kinetic constant k_e and the proteolytic kinetic constant k_1 in an accurate way:

the relative error based on the difference between the kinetic parameters calculated with the modified PK model and estimated with the original PK model is lower than 10% (except for k_1 for the HO category). Hence, in the modified PK model the proteolytic kinetic constant k_1 and the elimination kinetic constant k_e are calculated from basal clinical trials, whereas only the release kinetic constants k_0, D and t_{max} require the DDAVP execution. Information content analysis has been performed on the modified PK model of VWD. The metric used to evaluate the dynamic information content is the FIM trace. The position in time at which the trace reaches the peak corresponds to the maximum information content that can be achieved in the DDAVP. Looking at the dynamic profiles of four randomly selected subjects (Figure 3), an information peak can be observed at around three hours of DDAVP execution.

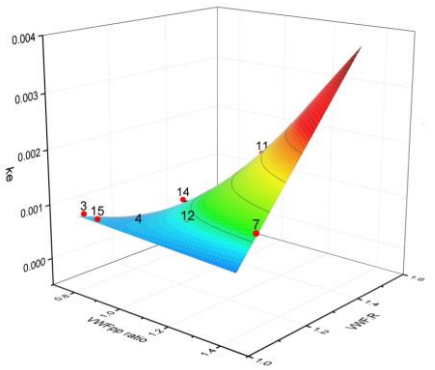


Figure 2. *Fitting of the HnonO category experimental data with the linear RSM presented in Equation 5 (k_e as function of VWFpp ratio and VWF:CB ratio)*

Table 1. *Summary of goodness of fit statistics for k_e as in Eq. 5.*

	Reduced chi-sqr	Residuals sum of squares	R^2	$\acute{R}^2$
HnonO	2.0E-08	8.1E-08	0.926	0.750
HO	9.3E-08	4.6E-07	0.910	0.724
2B	1.4E-07	2.7E-07	0.979	0.896
Vicenza	3.2E-07	3.2E-07	0.929	0.455

Table 2. Average k_1 *and* k_e *calculated and estimated values.*

	k_e calculated by modified model	k_e estimated through the original model	relative error	k_1 calculated by modified model	k_1 estimated through the original model	relative error
HnonO	6.73E-04	7.04E-04	0.046	2.55E-04	2.37E-04	0.0759
HO	0.001858	0.00152	0.082	4.53E-04	6.25E-04	0.2752
2B	0.003332	0.00323	0.031	0.00484	0.00471	0.0268
Vicenza	0.008406	0.00818	0.027	0.00155	0.00149	0.0345

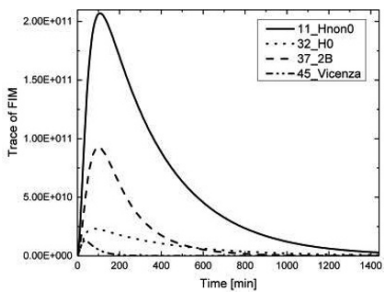


Figure 3. *Trace of FIM executed with the modified PK model of VWD.*

The amount of information introduced by basal data is sufficient to reduce the DDAVP execution down to 180 minutes. The DDAVP sampling schedule considered in this work is: [0, 15, 30, 60, 120, 180], which is a simple reduction in time of the original DDAVP protocol. Optimized sampling procedures could be obtained, *e.g.* by exploiting Model Based Design of Experiments (MBDoE) techniques (Castaldello et al, 2017).

The modified VWD model with the time reduced DDAVP has been validated considering four subjects that were excluded from the original pool (named as patients: 1, 19, 39, 50). The VWF:Ag and VWF:CB profiles of response of the subjects have been generated by simulation of the original model of VWD (Ferrari et al, 2018) and compared with those obtained with the modified model proposed in this work. As an example, in Figure 5, the VWF:Ag profiles of patient 1 (Figure 4a) and 19 (Figure 4b) have been reported.

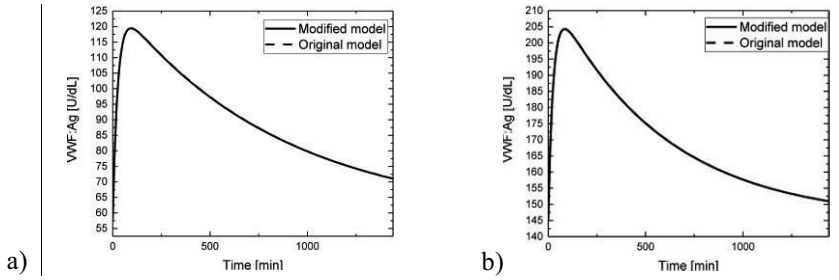


Figure 4. *Comparison between the simulated profiles of VWF:Ag response with the original PK model and modified PK model for patient 1 (a) and 19 (b).*

It can be seen that the VWF:Ag profiles generated with the original and modified PK model of VWD overlap, meaning that the predictive capability of the original PK model of VWD has been retained.

5. Conclusions

In this work the possibility to use basal clinical tests to estimate a subset of kinetic parameters of a PK model of VWD has been demonstrated. Response surface methods have been used to define regression models, which have been substituted into the original model. Other parameters can be estimated via a shorter DDAVP test, lasting only 3 hours instead of the standard 24 hours. Clinical data have allowed assessing the predictive performance of the modified model, which appears to be highly satisfactory. Future work will be carried out to optimize the time reduced DDAVP protocol among the different VWD categories and to reinforce model validation by applying the modified model and the time-reduced DDAVP to characterize new subjects.

6. References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on automatic control*, 19, 716-723.

Bard, Y. (1974). Nonlinear parameter estimation. Academic Press.

Box, G., Benhken, D. (1960). Some new 3 level designs for the study of quantitative variables. *Technometrics*, 2, 455-475.

Casonato, A., et al. (2011). Assesment of von Willebrand factor propeptide improves the diagnosis of von Willebrand disease. *Seminary Thrombosis Haemostasis*, 37, 456-463.

Castaldello, C., et al. (2017). A Model-based protocol for the diagnosis of von Willebrand disease. *Can. J. Chem. Eng.* doi: 10.1002/cjce.22964

Ferrari, M. et al. (2018). A mechanistic model to quantify VWF release, survival and proteolysis in patients with VWD. *Thrombosis and Haemostasis*. DOI: 10.1160/TH17-05-0375

Galvanin, F. et al. (2013). A model-based approach to the automatic diagnosis of von Willebrand disease. *AIChE Journal*, 60, 1718-1727.

Lillicrap, D. (2007). Von Willebrand disease-phenotype versus genotype: Deficiency versus disease. *Thrombosis Research*, 87, 57-64.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50346-3

A shortcut approach for decision-making and operational analysis of an integrated end-to-end continuous pharmaceutical process

Seyed Soheil Mansouri[a,§], Isuru A. Udugama[a,§], Jakob Kjøbsted Huusom[a], Krist V. Gernaey[a], Brahim Benyahia[b]

[a] *Process and Systems Engineering Centre (PROSYS), Department of Chemical and Biochemical Engineering, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark*
[b] *Department of Chemical Engineering, Loughborough University, Epinal Way, Loughborough, Leicestershire LE11 3TU, United Kingdom*
b.benyahia@lboro.ac.uk; [§]*The authors have contributed equally.*

Abstract

In this work, a plant-wide model-based shortcut approach for decision-making and operational analysis of integrated end-to-end continuous pharmaceutical processes is developed and tested. The proposed hierarchical approach is implemented through a computer-aided framework. Using concepts of Layer of Protection Analysis (LOPA) combined with Net Present Value (NPV) are embedded in the framework from a model-based point of view to comprehensively assess the overall benefits of a specific operational point. This in turn can be used to make an informed decision between competing operation alternatives. The application of the framework is demonstrated through a validated dynamic process model of an integrated end-to-end continuous pharmaceutical pilot plant. In this application, the implications of changing process operations from single pass operation to raw material recycling have been investigated.

Keywords: Decision-making, layer of protection analysis (LOPA), NPV

1. Introduction

The pharmaceutical industry has traditionally been dominated by batch processing which has many inherent disadvantages, such as high costs and environmental footprint, long processing time and poor flexibility. Moreover, drug development is a complex, long and costly process, entrenched with high degrees of uncertainty regarding the success of the clinical trials and scale-up (Mascia et al. 2013). To address some of these issues more effectively, the economic and regulatory drivers are more than ever pushing the pharmaceutical industry towards continuous manufacturing, which demonstrated tremendous advantages over batch processing. Integrated end-to-end continuous pharmaceutical manufacturing has particularly attracted a lot of interest over the last few years. However, the integrated continuous approach can be very challenging and requires more rigorous and robust decision making tools for design, development and supply chain management, to meet the regulatory, economic, and resource recovery and safety requirements (Mansouri et al. 2017; Udugama et al. 2017). In pharmaceutical process development, the standard is to first develop different production scales to supply clinical trials. Afterwards, feasible and viable technologies for full production scale and

operational windows are identified in a very short time during the post approval period, to be able to produce as long as possible while the API is still patent protected. These singularities make the pharmaceutical process design and development very challenging and several operational alternatives (within a fixed design space) need to be systematically evaluated during start-up, steady-state and non-routine process upsets (presence of disturbances) (Benyahia et al., 2012, Lakerveld et al., 2013).

2. Decision-making approach

In this section, the decision-making framework for operational analysis of an end-to-end integrated pharmaceutical process is described. Figure 1 shows an overview of framework embedded in a computer-aided work-flow is given.

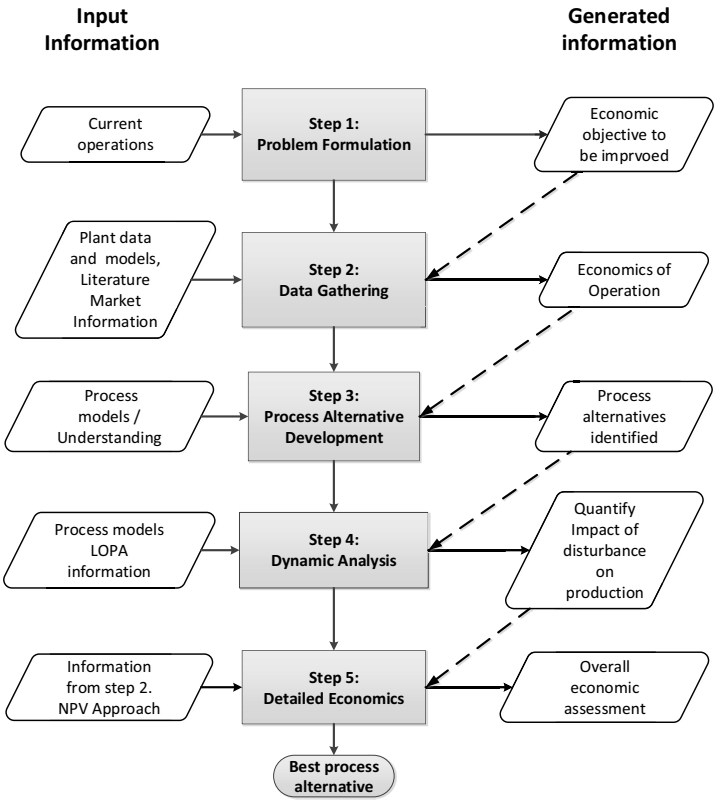


Figure 1: The decision-making workflow

Step 1 – Problem formulation: In the first step the overall objective(s) to be optimized improved is defined. The formulated objective primarily has economic elements, which in turn would influence decision-making in throughout the application of the framework. This could be also in form of identifying and evaluating various process alternatives subject to satisfying feasible economic objectives.

Step 2 – Data gathering: In this step all the relevant information that are required to carry out the analysis are systematically gathered. The information that is relevant to is categorized into three main categories, these are i) technical/process information, ii)

process and operational information related to process upsets, and iii) economic and market information. This Information related to technical/process information can be gathered from plant data or validated process models, while economic information can be obtained from literature and stakeholders.

Step 3 – Process alternative development: In context of pharmaceutical industry the development of process alternatives are constrained by regulatory agencies. Thereby, it is often necessary to develop process alternatives within the fixed approved design space. As such, operational alternatives are more worthwhile for retrofit design of an existing process. The development of process alternatives should be carried out using plant setup or validated process models taking into account the practical limitations such as operating temperatures, recycle ratios, unit mechanical constraints, solvent usage, as well as regulatory certified unit specifications.

Step 4 – Dynamic analysis: Once process alternatives are identified and developed, it is important to take into account their dynamic performance. This is because pharmaceutical processes are prone to non-routine process disturbances, which may have significant economic impact. This is due to high value of product as well as tight specifications. To this end, the disturbance probability and its economic influence on the process needs identification. The concept of layer of protection analysis (LOPA) is one method that can be adapted to calculate the probability and the consequences of process disturbances. The information generated in this step will be input to *Step 5* for detailed economic analysis.

Step 5 – Detailed economics: This step will provide a comprehensive analysis for the economics of the process alternatives identified in Step 3 taking into consideration information generated in Step 4. The final objective of this step is generation of a Net Present Value (NPV) number for a given process alternative that illustrates the economic incentives. Thereby, making decision to implement that process alternative over the current operations. To calculate the NPV value one requires i) the quantification of net profit generated by the process alternative ii) net profit/loss generated by the process alternative (based on information generated from LOPA analysis) iii) Cost of implementing the process alternatives (which will be at a minimal). As such, Eqs. 1 and 2 can be employed to carry this analysis

$$\Delta Profit_{p.a} = \Delta Profit_{cycle} \times \frac{cycles}{year} + \Delta Profit_{nro} \quad (1)$$

$$\mathrm{NPV} = \Delta\mathrm{Profit}_{\mathrm{p.a}} \times \frac{1-(1-\mathrm{r})^{\mathrm{n}}}{\mathrm{r}} - \mathrm{IC} \quad (2)$$

Where Eq. 1 calculates the annual change profit between the process alternative and current operations by taking into account the change in profit during improved operations during process cycles, while also considering the economics of non-routine operations generated in Step 4. Once the annual change in profit is calculated in can be used in a NPV analysis taking into account the time value of money by factoring in r the cash discount rate and n the number of years the analysis will be carried out while also taking into consideration the internal cost IC.

3. Application example

Step 1 – Problem formulation: In this work, a continuous end-to-end pharmaceutical process for organic synthesis of an API selected as case study. Benyahia et al. (2012) first proposed the dynamic model of the continuous pharmaceutical plant presented in Figure 2, which will be used as case study in this work. The objective in this example is to

improve the overall economics of the integrated end-to-end pharmaceutical through development of process alternatives within the given process design space.

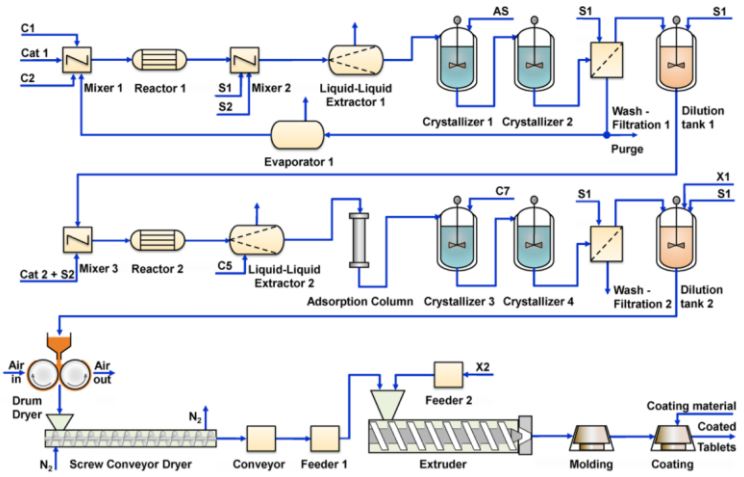


Figure 2: Flowsheet for continuous pharmaceutical pilot plant (Benyahia et al., 2012)

Step 2 – Data gathering: Here, the technical process information is collected through a validated dynamic model from Step 1. The market information is gathered from literature and open source information. The process-upset information is also obtained from Benyahia et al. (2012).

Step 3 – Process alternative development: In this example, the process alternatives are selected based on the recycle ratios (see recycle loop from wash filtration 1 in Figure 2). Four alternatives are selected i) base case – 13% purged (87% recycle) with wash factor 2 equal to 1.5, ii) case A – 0% recycle, and iii) case B – 100% recycle, iv) optimized base case – with wash rate of 2,and v) optimized case B - with wash factor 2 of 5.6. This is to evaluate the economic benefits of recycling at different rates in an integrated end-to-end pharmaceutical process.

Step 4 – Dynamic analysis: In this step, the dynamic process performance in the presence of upset scenarios is evaluated. Figure 3 shows the key plant dynamics for case B it is more sensitive to process disturbances than Case A and Optimized base case due to the employment of total recycle that amplifies process disturbances (due to the limited number of pages and for sake of brevity , the results are not fully presented here). It can be observed from Figure 3a and 3b that the in presence of a process upset the levels of the main impurities are above acceptable specification, which results in off spec tablets. Due to the tight regulations around processes in the pharmaceutical industry, these products cannot be easily recycled to be used as feedstock's, or purified to meet acceptable standards. Because of these reasons, Case B, which makes more products (due to a higher recycle rates), will also produce more off-specification products, especially during non-routine process disturbances. In Figure 3c and 3d, this effect is illustrated where Case B has a higher total production rate but a lower on-specification production, despite high API production. Therefore, when analyzing process control alternatives with a total recycle, it is important to consider the sensitivity to process disturbances, which will influence the overall NPV calculation.

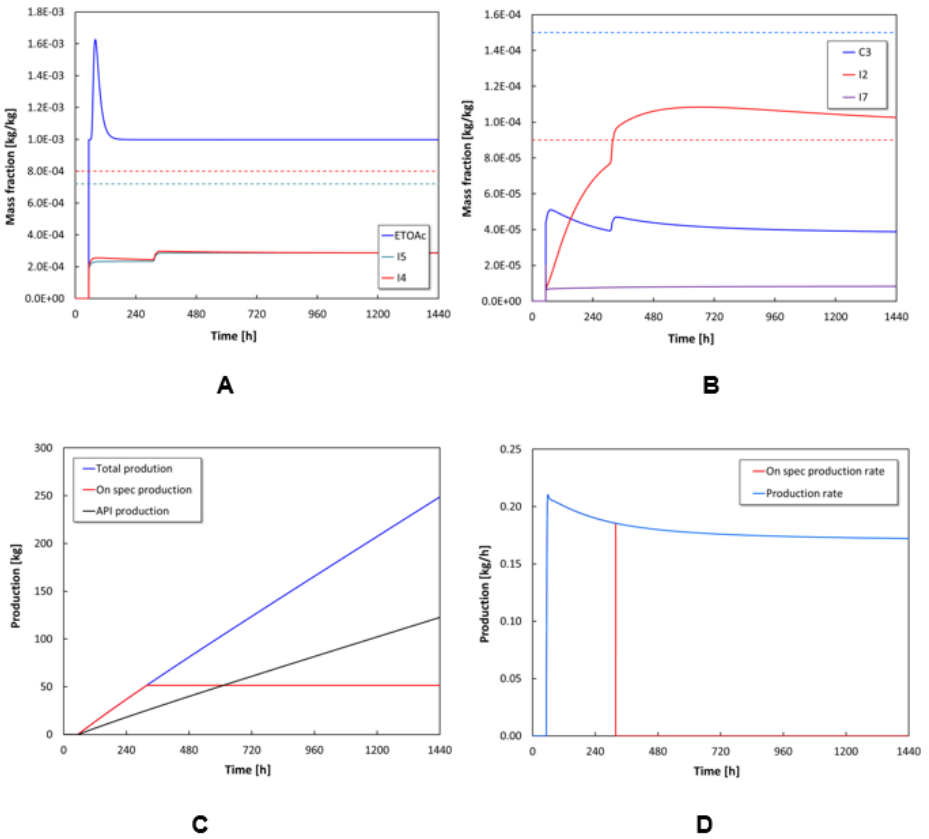


Figure 3: Plant key dynamics under total recycle (purge ratio = 0) at optimal wash factors (f_{w1} = 1.5 and f_{w2} = 5.6) with a process upset at t =300 hours, (a) I5, I4 and Ethyl acetate in the tablet (b) C3, I2 and I7 in the tablet overall impurities, (c) Total production (both on spec and off spec) and total API production, (d) on spec production rate and total production rate

Step 5 – Detailed Economics: In Table 1, the overall economics of the four process control alternatives during a standard operational cycle are listed where they have been compared against base case (originally reported in Benyahia et al. 2012). Analysis of the table illustrates all cases except for case B has better economic performance than the current base case. The best performing process alternative is the optimized base case. Analysis of optimized case B illustrates economically favourable operations in comparison to the current base case. However, concurrently analysing the optimized case B's performance during non-routine process operations illustrates that some of the economic gains during routine operations cycles will reduced due to the process alternatives increased sensitivity to non-routine process disturbances. Based on calculations performed on optimized case B, this case will lose the equivalent of 20,000-25,000 USD per production cycle due non-routine disturbances. Based on information listed in Table 1 as well as the understanding of process disturbances identified in Step 4, the optimized base case provides the best economic outcome out of the four process control alternatives examined. Since the optimized base case performed similarly to the current base case during process upsets

and because this type of operational change has minimal installation cost we can calculate the NPV of the optimized base case. Based on these values the NPV of the optimized base case would be 2.26 million USD in comparison to current operations. Note that this NPV is a differential analysis between base case and the optimized base case (case iv), which reduces the complexity and information required to carry out the analysis.

Table 1: Detailed economics of the process alternatives during routine operation cycles

	Costs ($/kg)	Case a	Case B	Optimized base case	Optimized case B
C8 (aliskiren free base - Lactone)	3,000	54,270	-31,350	-	-31,350
C9 (3-amino-2,2-dimethylpropionamide)	66	-	-	-	$ -
HA (2-Ethylhexanoic acid)	113	-	-	-	$ -
H2O	6	-	-	-	$ -
HEP (heptane)	92	-	-	-	1
HCL	194	-1,341	-2,051	-	-2,051
NaOH	161	-2,718	-4,155	-	-4,155
FA (Fumaric acid PharmaGrade)	773	-4,878	-7,320	-	-7,320
SiO2	49	-1	-2	-	-2
PEG	73	-223	-326	-4	-341
EOA	59	-577	-929	1,976	14,169
Marginal changes in Cost of Production		**$ 44,532**	**-$ 46,132**	**$ 1,972**	**-$ 31,050**
Marginal change inRevenue from Sales	**$ 7,600**	**$ 795,492**	**-$ 147,212**	**$ 841,244**	**$ 770,868**
Marginal change in Profit per cycle		**$ 750,960**	**-$ 101,080**	**$ 839,272**	**$ 801,918**

4. Conclusions

In this work, a systematic shortcut decision-making framework for operational analysis of process alternatives with an application on integrated end-to-end pharmaceutical processes was proposed. Through the application of the framework, it was demonstrated that economic considerations need to be taken into account for routine and non-routine operations when selecting an operating mode. This framework can be advantageous for both retrofit design of existing pharmaceutical processes as well as early-stage process design development activities. Current and future work is to extend the application of the framework to select economic control structures to better manage process routine and non-routine operations.

Acknowledgements

The project received financial support from Innovation Fund Denmark through the BIOPRO2 strategic research center (Grant number 4105-00020B).

References

B. Benyahia, R. Lakerveld, P.I. Barton, 2012. Industrial and Engineering Chemistry Research, 51 (47), 15393-15412.

R Lakerveld, B Benyahia, PL Heider, H Zhang, A Wolfe, C Testa, S Ogden, D.R. Hersey, S. Mascia, J.M. B. Evans, R.D. Braatz, P.I. Barton, 2015. Organic Process Research & Development 19 (9), 1088-1100

S. Mascia, P.L Heider, H. Zhang, R. Lakerveld, B. Benyahia, P.I. Barton, D.R. Braatz, R.D. Cooney, J.M.B. Evans, T.F. Jamison, K.F. Jensen, A.S. Myerson, B.L. Trout, 2013. Angewandte Chemie International Edition, 52(47), 12359-12363.

S.S. Mansouri, I.A.. Udugama, S. Cignitti, A. Mitic, X. Flores-Alsina, K.V. Gernaey, 2017.?. Current Opinion in Chemical Engineering, 18, 1-9.

I.A. Udugama, S.S. Mansouri, A. Mitic, X. Flores-Alsina, K.V. Gernaey, 2017. Processes, 5 (3), 48.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50347-5

Development and Analyses of a Database of Antibody – Antigen Complexes

Varun M. Chauhan[a], Sumaiya Islam[a], Alexis Vroom[a] and Robert Pantazes[a*]

[a]*Department of Chemical Engineering, Auburn Univeristy, Auburn, Alabama 36849, United States*
rjp0029@auburn.edu

Abstract

Antibodies are medically and experimentally important immune system proteins. By binding to foreign molecules with high specificity and affinity, they act as flags that guide the immune response. The high-quality binding characteristics of antibodies have led to their widespread use as both experimental reagents and therapeutic molecules. Due to their importance, the interactions between antibodies and antigens has been an ongoing area of research. Here, a non-redundant database of 492 antibody – protein complexes was studied. Using the CHARMM and Amber force fields, the energies of the complexes were minimized and the interaction energies of individual residues were determined. This revealed that the percent contributions to binding energy of amino acids follows an exponential decay, with only a few residues being responsible for most of the energy.

Keywords: Antibody, Antigen, Epitope.

1. Introduction

Antibodies are vital to the immune system as a line of defense against infectious diseases. The body continues to make antibodies even after an infection has been cured because the immunological memory lives long after infection or vaccination; and the antibodies undergo clonal expansion and differentiation after an immune response (Janeway et al. 2001). Antibodies are often used in research and are powerful tools in the study of proteins (Edwards et al. 2011). Antibodies are also a $100 billion global market as therapeutic agents (Ecker et al. 2015), including for autoimmune diseases and cancer (Zhou and Zhou 2014)(Buss et al. 2012). Identification of the binding sites of antibodies and their target molecules (i.e. antigens) is crucial for development in biomedical research applications such as pharmaceutical molecules and immuno-therapeutics (Chevalier et al. 2017).

The key performance feature of antibodies is the high specificity, high affinity manner in which they bind antigens. The portions of antigens that antibodies recognize are known as epitopes and antibodies bind only at these locations. This is achieved with six complementarity-determining regions (CDRs), which form the paratope of the antibody. Epitope residues must be close to one another in space and may also be located within a short stretch of continuous sequence. If so they are classified as linear epitopes and if not as structural epitopes.

For over two decades, features of antibody-antigen binding interfaces have been studied in detail. Initially, Mian et al. (1991) studied six antibody structures which included only one antigen. Work by Lo Conte et al. (1999) and Sundberg and Mariuzza (2002) studied 19 and 30 complexes respectively. The former analyses studied paratope features in more

detail when compared with epitope properties. W. Chen et al. (2009) studied interaction methods and molecular surface properties of both paratopes and epitopes in 155 protein-antibody complexes, with 53 unique structures. The work of Fellouse et al. (2004) constructed paratope libraries with a diversity of only four amino acids and studied the features of two complexes for the generation of synthetic antibody libraries. Ramaraj et al. (2012) studied features of 53 non-redundant complexes using change in solvent accessible surface area and distance cut offs. Kuroda and Gray (2016) studied the features of hydrogen bonds and shape complementarity in 6637 non-redundant protein-protein interfaces which included 191 antibody-antigen complexes. Nguyen et al. (2017) analyzed 403 antibody-antigen complexes using computational methods to find hydrogen bonds and van der Waals, hydrophobic and ionic interactions.

These previous studies have identified numerous features of paratopes and epitopes, including which amino acids are most commonly used, the distribution of linear versus structural epitopes, and the types of interactions that can occur. However, these studies have done so using binary metrics (e.g. residues are in contact or not, an epitope is linear or structural). Here, we use the CHARMM (Brooks et al. 2009) and Amber (Case et al. 2017) molecular mechanics force fields to quantify the percent contribution to binding of every residue in a database of 492 antibody-protein complexes. This permits a more nuanced assessment of interactions. To our knowledge, a similar study has not previously been conducted.

2. Methodology

Initially, 2498 antibody structures were gathered from the International Immunogenetics Information System 3D Structure Database (Ehrenmann et al. 2009). 1344 antibodies with at least five CDR mutations from one another were identified, of which 492 were in complex with protein antigens. Those structures were selected for this study.

Molecular mechanics force fields allow for the computational study of protein structures and interactions. Rosetta is the most commonly used force field for protein engineering. However, it is not well-suited for studying this database because its electrostatics potential is parameterized for the hydrophobic interiors of proteins rather than the polar exteriors (Alford et al. 2017). Instead, the CHARMM and Amber force fields were selected. Missing atoms were added and the overall energies of each of the 492 complexes were minimized in each force field. Next, the energies of interaction between every antibody residue and every antigen residue were tabulated. Due to the pairwise additive nature of the energy functions, the total energy can vary significantly from complex to complex and generally increases in magnitude as the total number of amino acids increases. Therefore, the energies were converted to the percent of total energy for each complex to facilitate comparisons between complexes.

In CHARMM, the calculated energies included van der Waals and electrostatic forces, while Amber accounted for van der Waals, electrostatics and solvation (columbic and non-columbic) energies. The usage of solvation energy in Amber can be explained by the differences in the electrostatic equations in CHARMM and Amber. The CHARMM equation includes an additional distance-dependent scaling factor that reduces energies to zero at distances of 12 Å or greater and accounts for the electrostatic shielding due to the solvent. Amber does not include this term but instead calculates solvation energy using the modified implicit Gibbs-Born model by A. Onufriev, D. Bashford and D.A. Case and

the Linear Combination of Pairwise Overlaps (LCPO) method. The van der Waals energy was calculated using the 12-6 Lennard Jones potential in both CHARMM and Amber.

3. Results and Discussion

3.1. Percent contributions to binding

The percent contribution to binding energy of residues was determined to follow an exponential decay, as shown in Figure 1. This was true for both antibody and antigen residues and with both force fields. The average R^2 of the exponential fits for antibodies were 0.957 ± 0.023 and 0.965 ± 0.021 with the Amber and CHARMM force fields, respectively. The corresponding antigen values were 0.958 ± 0.025 and 0.963 ± 0.023. For both force fields, the first several antigen residues typically contribute more to binding energy than an equal number of antibody residues. This suggests that on average some antigen residues have strong interactions with multiple antibody residues. Also, the percent contributions to binding energy of the first several residues are greater using CHARMM than with Amber. This is a result of the differences in the force fields and how they handle electrostatics shielding from solvent. Without experimental testing, it is not clear which is a more accurate prediction, but both force fields show the same exponential decay trend.

3.2. Antigen

Experimental studies have found that in most complexes five antigen residues are sufficient for specific and high affinity binding (Sykes et al. 2013; Pellois et al. 2002). To remove effects of small antigens biasing results, only the 385 complexes with protein antigens of 50 or more amino acids were considered in the antigen calculations. The five most attractive antigenic residues contribute an average of 62 % and 74 % of the total Amber and CHARMM binding energies, respectively. Notably, as can be seen in Figure 1, on average the sixth residue contributes less than 6 % of total binding energy for both force fields, which is less than 10 % of these values. Thus, residues that are "significant to binding" were defined as those that contribute at least these percentage thresholds to binding energy. It is important to note that these values are likely an overestimate of critical binding requirements, as five residues are enough to establish binding for the majority of antigens rather than half.

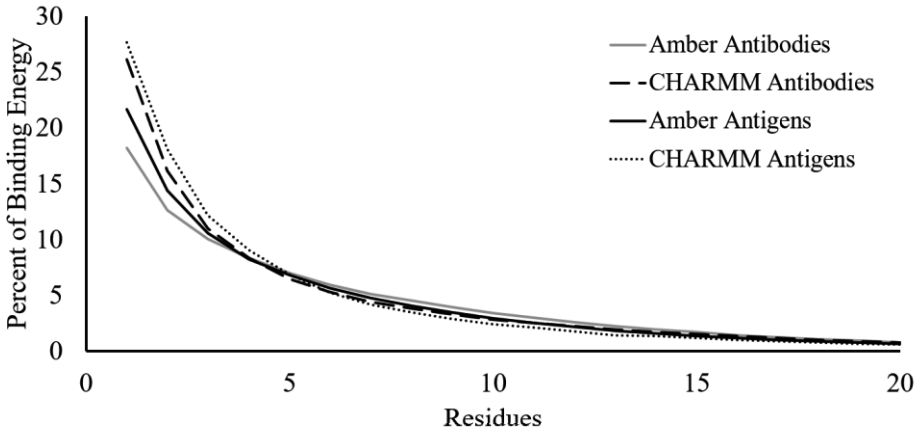


Figure 1: Average percentage of binding energy contributed by antibody and antigen residues

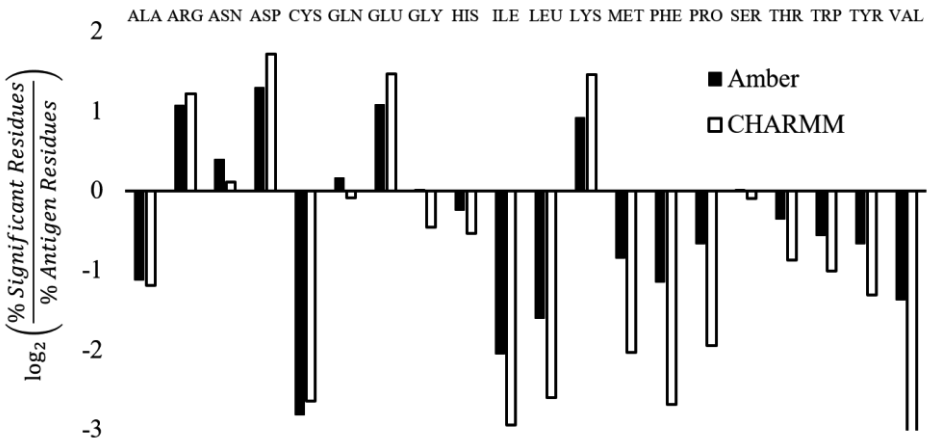


Figure 2: Log_2 change in amino acid usage in significant residues versus the antigens as a whole

Figure 2 depicts the Log_2 change in the relative usages of amino acids in the significant residues versus the antigens as a whole. The trends are consistent with expected behaviours: charged and acidic amino acids are overrepresented in the significant residues, hydrophobic amino acids (which are expected to be in the interior of antigens) are underrepresented, and most polar amino acids are significant at a rate that is comparable to their usage in antigens. In nearly all cases, the CHARMM values show larger magnitude over / underrepresentation than the Amber values, which arises from differing treatment of electrostatics shielding.

An important question in classifying epitopes is whether they are structural or linear. With known binding energy contributions, it is now possible to quantify the degree of linearity of epitopes instead of using a binary definition. Figure 3 shows the fraction of the threshold binding energy reached by the best consecutive 12 amino acid sequence in the antigens. 77 % of epitopes were at least 50 % linear, 27 % were at least 80 % linear and 11 % were at least 100 % linear in the Amber results. The corresponding CHARMM values were 70 %, 21 % and 7.5 %, respectively. Using alternative peptide lengths did not significantly change these results.

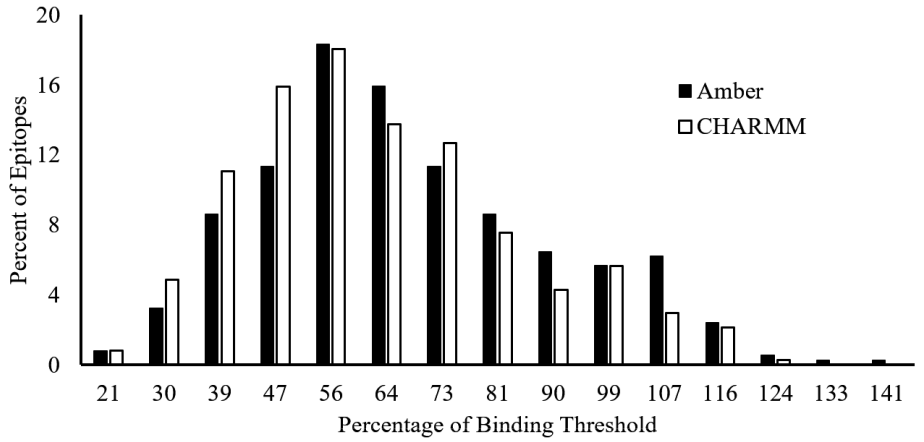


Figure 3: Percentage of binding threshold in the best 12 consecutive antigen amino acid sequence

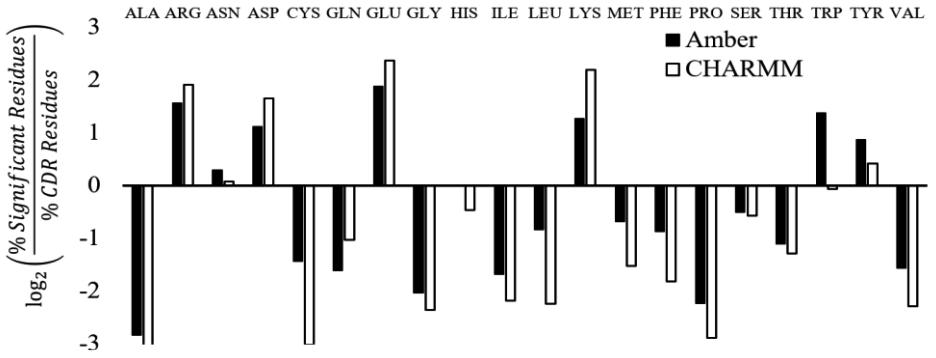


Figure 4: Log2 change in amino acid usage in significant antibody residues versus in the CDRs

3.3. Antibody

For the antibodies, significant residues were defined using the same percentage binding energies as required for the antigens. For both force fields, six amino acids were required on average to reach the thresholds in antibodies instead of five as in antigens. Figure 4 depicts the Log_2 change in amino acid usage in the significant residues versus the CDRs as a whole. The most notable difference between these values and the antigen ones are the increased importance of tyrosine and tryptophan (only in the Amber data). The importance of tyrosine in antigen binding is well-established (Ramaraj et al. 2012).
The distribution of significant amino acids among the CDRs followed expected trends. Using both force fields, the H3 CDR had about 35 % of the significant residues and H2 had ~22 %. The H1, L1 and L3 CDRs each had between 12 and 14 % of the significant residues and the L2 CDR contained around 4 %. This is consistent with the belief that H3 is typically considered the most important CDR for antigen binding and corresponds with previously observed mutation rates during affinity maturation (Pantazes and Maranas 2013).

4. Conclusions

We developed a non-redundant database of 492 antibody-protein complexes and used the CHARMM and Amber force fields to energetically minimize and analyze their binding interactions. To our knowledge, these data are the first quantification of the contributions to binding of single residues in antibodies and antigens. The primary finding is that the contributions follow an exponential decay. Considering the biomedical importance of antibodies, the knowledge that only a few antibody residues contribute most of the binding energy should aid in the creation of novel computational algorithms to design next generation therapeutics, vaccines and diagnostics.

References

Alford, Rebecca F., Andrew Leaver-Fay, Jeliazko R. Jeliazkov, Matthew J. O'Meara, Frank P. DiMaio, Hahnbeom Park, Maxim V. Shapovalov, et al. 2017. "The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design." *Journal of Chemical Theory and Computation* 13 (6): 3031–48. doi:10.1021/acs.jctc.7b00125.

Brooks, B. R., C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, et al. 2009. "CHARMM: The Biomolecular Simulation Program." *Journal of Computational Chemistry* 30 (10): 1545–1614. doi:10.1002/jcc.21287.

Buss, Nicholas A.P.S., Simon J. Henderson, Mary McFarlane, Jacintha M. Shenton, and Lolke De

Haan. 2012. “Monoclonal Antibody Therapeutics: History and Future.” *Current Opinion in Pharmacology* 12 (5). Elsevier Ltd: 615–22. doi:10.1016/j.coph.2012.08.001.

Case, D.A., D.S. Cerutti, T.E. Cheatham III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, et al. 2017. “AMBER 2017.” University of California, San Francisco.

Chevalier, Aaron, Daniel-Adriano Silva, Gabriel J. Rocklin, Derrick R. Hicks, Renan Vergara, Patience Murapa, Steffen M. Bernard, et al. 2017. “Massively Parallel de Novo Protein Design for Targeted Therapeutics.” *Nature* 550 (7674). Nature Publishing Group: 74–79. doi:10.1038/nature23912.

Conte, L Lo, C Chothia, and Joel Janin. 1999. “The Atomic Structure of Protein– Protein Recognition Sites.” *J Mol Biol* 285: 2177–2198. doi:10.1006/jmbi.1998.2439.

Ecker, Dawn M., Susan Dana Jones, and Howard L. Levine. 2015. “The Therapeutic Monoclonal Antibody Market.” *mAbs*. doi:10.4161/19420862.2015.989042.

Edwards, Aled M., Ruth Isserlin, Gary D. Bader, Stephen V. Frye, Timothy M. Willson, and Frank H. Yu. 2011. “Too Many Roads Not Taken.” *Nature* 470 (7333): 163–65. doi:10.1038/470163a.

Ehrenmann, François, Quentin Kaas, and Marie Paule Lefranc. 2009. “IMGT/3dstructure-DB and IMGT/domaingapalign: A Database and a Tool for Immunoglobulins or Antibodies, T Cell Receptors, MHC, IgSF and MHcSF.” *Nucleic Acids Research* 38 (SUPPL.1). doi:10.1093/nar/gkp946.

Fellouse, Frederic A, Christian Wiesmann, and Sachdev S Sidhu. 2004. “Synthetic Antibodies from a Four-Amino-Acid Code: A Dominant Role for Tyrosine in Antigen Recognition.” *Proceedings of the National Academy of Sciences of the United States of America* 101 (34): 12467–72. doi:10.1073/pnas.0401786101.

Janeway, Charles A., Paul Travers, Mark Walport, and Mark Shlomchik. 2001. “Immunobiology: The Immune System In Health And Disease.” *Immuno Biology 5*, 892. doi:10.1111/j.1467-2494.1995.tb00120.x.

Kuroda, Daisuke, and Jeffrey J. Gray. 2016. “Shape Complementarity and Hydrogen Bond Preferences in Protein-Protein Interfaces: Implications for Antibody Modeling and Protein-Protein Docking.” *Bioinformatics* 32 (16): 2451–56. doi:10.1093/bioinformatics/btw197.

Mian, I S, A R Bradwell, and A J Olson. 1991. “Structure, Function and Properties of Antibody Binding Sites.” *Journal of Molecular Biology* 217 (1): 133–51. doi:10.1016/0022-2836(91)90617-F.

Nguyen, Minh N., Mohan R. Pradhan, Chandra Verma, and Pingyu Zhong. 2017. “The Interfacial Character of Antibody Paratopes: Analysis of Antibody-Antigen Structures.” *Bioinformatics* 33 (19): 2971–76. doi:10.1093/bioinformatics/btx389.

Pantazes, Robert J, and Costas D Maranas. 2013. “MAPs: A Database of Modular Antibody Parts for Predicting Tertiary Structures and Designing Affinity Matured Antibodies.” *BMC Bioinformatics* 14 (1): 168. doi:10.1186/1471-2105-14-168.

Pellois, Jean Philippe, Xiaochuan Zhou, Onnop Srivannavit, Tiecheng Zhou, Erdogan Gulari, and Xiaolian Gao. 2002. “Individually Addressable Parallel Peptide Synthesis on Microchips.” *Nature Biotechnology* 20 (9): 922–26. doi:10.1038/nbt723.

Ramaraj, Thiruvarangan, Thomas Angel, Edward A. Dratz, Algirdas J. Jesaitis, and Brendan Mumey. 2012. “Antigen-Antibody Interface Properties: Composition, Residue Interactions, and Features of 53 Non-Redundant Structures.” *Biochimica et Biophysica Acta - Proteins and Proteomics* 1824 (3). Elsevier B.V.: 520–32. doi:10.1016/j.bbapap.2011.12.007.

Sundberg, EJ, and RA Mariuzza. 2002. “Molecular Recognition in Antibody-Antigen Complexes.” *Advances in Protein Chemistry* 61: 119–60.

Sykes, Kathryn F., Joseph B. Legutki, and Phillip Stafford. 2013. “Immunosignaturing: A Critical Review.” *Trends in Biotechnology* 31 (1): 45–51. doi:10.1016/j.tibtech.2012.10.012.

W. Chen, Shu-wen, Marc H.V. Van Regenmortel, and Jean-Luc Pellequer. 2009. “Structure-Activity Relationships in Peptide-Antibody Complexes: Implications for Epitope Prediction and Development of Synthetic Peptide Vaccines.” *Current Medicinal Chemistry* 16 (8): 953–64.

Zhou, Juhua, and Juhua Zhou. 2014. “Advances and Prospects in Cancer Immunotherapy.” *New Journal of Science* 2014: 1–13. doi:10.1155/2014/745808.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50348-7

Quantitative Analysis on the Relevance of Regulatory Constraints in Designing Pharmaceutical Manufacturing Processes

Hirokazu Sugiyama[a*], Yusuke Morikawa[a], Mai Matsuura[a], Menghe Xu[a]

[a]*Department of Chemical System Engineering, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*
sugiyama@chemsys.t.u-tokyo.ac.jp

Abstract

This work presents the relevance of regulatory constraints on the outcome of process design in pharmaceutical manufacturing with a case study on waste solvent recovery. The role of the investigated process was to recover tetrahydrofuran from an azeotropic mixture with water and methanol. As the technologies to overcome the distillation boundary, different alternatives were modeled and evaluated in a multiobjective manner. The target concentration of recovered solvent, which cannot be altered because of regulations, was imagined to be modifiable, and two design problems, initial and extended, were formulated. A type of process that was found to be optimal in the extended problem was better than any of the optimal alternatives in the initial design problem. These results confirmed quantitatively that the way in which regulatory constraints are considered makes a difference in the outcome and that the appropriate formulation of a design problem is critical for pharmaceutical manufacturing processes.

Keywords: Multiobjective optimization, Active pharmaceutical ingredients, Waste solvent recovery, Azeotrope, Decision-making

1. Introduction

In the research & development of pharmaceuticals, time is regarded as the critical factor. Limited time is available for the process design, which makes it difficult to avoid the launch of sub-optimal processes for commercial production. After the launch, compliance to the regulations known as Good Manufacturing Practice (GMP) becomes the most important factor, and GMP itself functions as the hard constraint. The environment hinders companies from substantial changes, such as changes in process technologies, even when the change itself is beneficial.

Many process systems engineering (PSE) researchers are entering the field of pharmaceutical manufacturing in order to design/redesign inherently superior processes. García-Muñoz and Mercado (2013) presented a method for selecting raw materials in active pharmaceutical ingredient (API) manufacturing; Jolliffe and Gerogiorgis presented process models of continuous manufacturing of ibuprofen and artemisinin (2016); Sugiyama, et al. (2017) presented applications of PSE in the design of tablet and injectable manufacturing processes. However, the questions of (i) to what extent the obtained solutions are constrained by the regulations, and (ii) whether different conclusions would be drawn if the regulatory conditions had been treated differently, have not been addressed in previous studies.

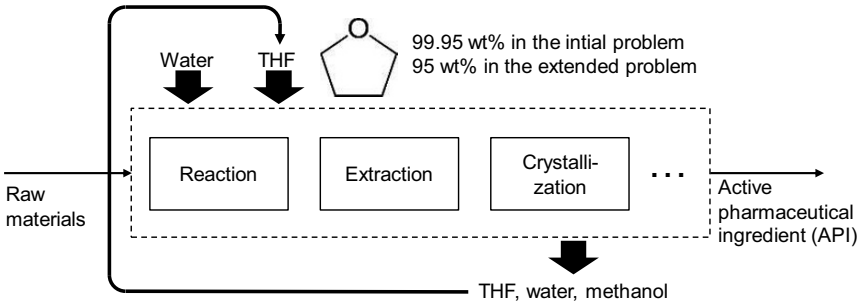


Figure 1. Overview of the main process and the THF recovery process

We present the relevance of regulatory constraints in the design of pharmaceutical manufacturing processes. Case study is performed on solvent recovery process for an actual API manufacturing, which discharges an azeotropic mixture of tetrahydrofuran (THF), water and methanol. Multiple alternatives are generated for recovering THF, and are evaluated considering economy, environmental impacts, and environmental, health and safety (EHS) hazard. Here, the design problem is formulated in two different ways, either considering the regulatory constraint as given (*initial design problem*) or as modifiable (*extended design problem*). The obtained results in both design problems are compared to quantify the impact of the way to consider the regulatory constraint differently. The full results are being submitted elsewhere (Sugiyama, et al, 2018).

2. Definition of design problem

2.1. Process overview

An overview of the investigated process is shown in Figure 1. The API production process (main process) is operated batch-wise with unit operations such as reaction, extraction, and crystallization. The process uses water and THF as solvents, where the concentration of the inlet THF is specified as higher than 99.95 wt%. After each batch operation, an azeotropic mixture of THF, water and methanol is discharged, with the molar ratio of 64.3%, 35.5% and 0.2%, respectively. Methanol is a byproduct of the reaction, i.e., methanol has to be removed constantly in the recovery process in order to avoid accumulation in the main process. In the current status (as-is process), the mixture is sold to an external company at the price of USD 0.5 to 1.0 per kg-mixture. The inlet THF is purchased at a higher price, e.g., USD 4.0 per kg, which triggered the investigation whether and how to separate the solvent mixture to recover THF onsite.

2.2. Problem formulation

Equation (1) presents the general mathematical formulation of the design problem.

$$\begin{aligned} & \max_{x} \left(Y_{\mathrm{NPV}}(x), Y_{\mathrm{CO_2}}(x), Y_{\mathrm{E}}(x), Y_{\mathrm{H}}(x), Y_{\mathrm{S}}(x)\right) \\ \text{s.t.} \quad & \boldsymbol{C}(x) \geq \boldsymbol{C}_{Problem} \\ & y_{\mathrm{NPV}}(x) \geq y_{\mathrm{NPV}}(\mathrm{as-is}) \\ & x \in P_{Problem} \\ & Problem \in \{\mathrm{Initial}, \mathrm{Extended}\} \end{aligned} \tag{1}$$

where x represents a process alternative, Y is a normalized objective function regarding net present value (NPV), CO_2 reduction, and EHS hazard, y is the original objective function before normalization, P is a set of alternatives depending on *Problem*, which is further an element of initial/extended design problems. The first constraint indicates that the concentration vector of the recovered solvent, $\boldsymbol{C}$, needs to be larger than the target $\boldsymbol{C}$ depending on *Problem*; the second indicates that the NPV of an alternative needs to be higher than that of the as-is process. The target THF concentrations were set as 99.95 wt% and 95 wt% in the initial and extended design problems, respectively.

The working assumption of this work is presented in Eq. (2).

$$\boldsymbol{Y}^{\mathrm{opt}}|_{Problem=\mathrm{Extended}} - \boldsymbol{Y}^{\mathrm{opt}}|_{Problem=\mathrm{Initial}} \geq 0 \tag{2}$$

where the vector $\boldsymbol{Y}^{\mathrm{opt}}$ represents the evaluation result of an optimal alternative. If a valid relation in Eq. (2) is found, it indicates that the optimal solution in the extended problem is better than any of the optima in the initial problem, i.e., the way in which the regulatory constraint is considered makes a difference in the final outcome.

2.3. Evaluation indicators

Equations (3) to (7) presents the evaluation indicators: y_{NPV} [USD] for NPV, y_{CO2} [kg-CO_2/kg-inlet] for CO_2 reduction, $y_{\mathrm{E,\,H,\,S}}$ [kg/kg-inlet] for EHS hazard. Eqs. (5) to (7) are adopted from Sugiyama et al. (2008), which is further based on the hazard assessment method of Koller et al. (2001) that translates the physical property of chemicals—e.g., flash point—to a nondimensional index value I. The index I has the range from 0 (least-hazardous) to 1 (most-hazardous) and can be calculated in different categories. In this work, the following 10 categories were considered: mobility, fire/explosion, reaction/decomposition, and acute toxicity for safety; irritation and chronic toxicity for health; and water-, air-mediated effects, and persistency for environment.

$$y_{\mathrm{NPV}} = -INV + \sum_{i=1}^{N} \frac{CF_i}{(1+r)^i} \tag{3}$$

$$\begin{aligned} y_{\mathrm{CO_2}} = {} & \text{Decreased } \mathrm{CO_2} \text{ in THF production} \\ & - \text{Increased } \mathrm{CO_2} \text{ in THF recovery} \end{aligned} \tag{4}$$

$$y_{\mathrm{E}} = \sum_{cE} \left(\sum_{j} (z \cdot \max_{F}(m_j{}^{F}) \cdot I_j{}^{cE}) + \sum_{OUT} \sum_{j} (m_j{}^{OUT} \cdot I_j{}^{cE}) \right) \tag{5}$$

$$y_{\mathrm{H}} = \sum_{cH} \max_{j}(m_j{}^{UN} \cdot I_j{}^{cH}) \tag{6}$$

$$y_{\mathrm{S}} = \sum_{cS} \max_{F} \left(\sum_{j} m_j{}^{F} \cdot I_j{}^{cS} \right) \tag{7}$$

Here, INV [USD], CF_i [USD], r [–], and N [year] represent investment cost, annual cash flow in year i, interest rate, and the planned years of operation, respectively. The

parameter m_j^F [kg/kg-inlet] represents the mass flow of substance j relative to the amount of the inlet solvent mixture in flow F, and m_k^{OUT} [kg/kg-inlet] represents that in waste effluent flow OUT, and m_j^{UN} [kg/kg-inlet] is the unit mass (UN), i.e., 1 kg of substance j in the process. The parameters I_j^{cE} [–], I_j^{cH} [–], and I_j^{cS} [–] are the hazard index values of substance j in category cE, cH, and cS, respectively. The parameter z [–] is the fraction of mass emitted to the environment in case of an accident from maximum mass present in the overall process, and it was set as 0.1 in the calculation.

To facilitate the multiobjective evaluation, the normalized evaluation indicator Y [–] was defined for each design objective, as shown in Eq (8), where x^* is an alternative considered in the initial design problem. The indicators y_{NPV} and y_{CO2} are to maximize; y_{E}, y_{H} and y_{S} are to minimize. The Ys are all to maximize.

$$\begin{cases} Y(x)_{\mathrm{NPV,CO_2}} = \dfrac{y(x)_{\mathrm{NPV,CO_2}}}{\max\limits_{x^*}\left(y(x^*)_{\mathrm{NPV,CO_2}}\right)} \\ Y(x)_{\mathrm{E,H,S}} = \dfrac{y(x)_{\mathrm{E,H,S}} - \min\limits_{x^*}\left(y(x^*)_{\mathrm{E,H,S}}\right)}{\max\limits_{x^*}\left(y(x^*)_{\mathrm{E,H,S}}\right)} \end{cases} \tag{8}$$

3. Generation and analysis of process alternatives

3.1. Generation of process alternatives

Figure 2 (a) shows the ternary phase diagram of the THF/water/methanol system that was calculated using the UNIQUAC property estimation method in Aspen Plus®. The point "inlet" represents the concentration of the waste solvent mixture (64.3%, 35.5%, 0.2%). The constraint lines I and II indicate the required concentration of THF in the initial and extended design problems, respectively; the line III indicates the required methanol concentration, which is common to both problems. The target area of the initial design problem is almost on the vertex of THF; the target area of the extended design problem contains that of the initial problem and is stretched on the top left side of the triangle. For overcoming the distillation boundary, the alternatives of zeolite membrane, pressure swing, azeotropic distillation, and entrainer processes using ethylene glycol or glycerol were generated, and modelled using Aspen Plus®. The main paths of the first three processes are indicated in Figure 2 (a).

3.2. Analysis of process alternatives

The flow rate of the "inlet" was set at 100 kmol/hr, which was fast enough to recover THF within the cycle time of the main process. The annual operating time of this continuous recovery process was estimated based on the annual production amount of the main process. Figure 2 (b) shows the flowsheet of the pressure swing process, where P_1 [atm] indicates the pressure of the first column. The standard pressure P_1 was set as 0.1 atm. Notably, in the extended design problem, the processes inside the dotted line were sufficient to reach the relaxed target concentration, when $P_1 \leq 0.07$ atm. This simplicity lead to several advantages in the evaluation results shown in the next chapter.

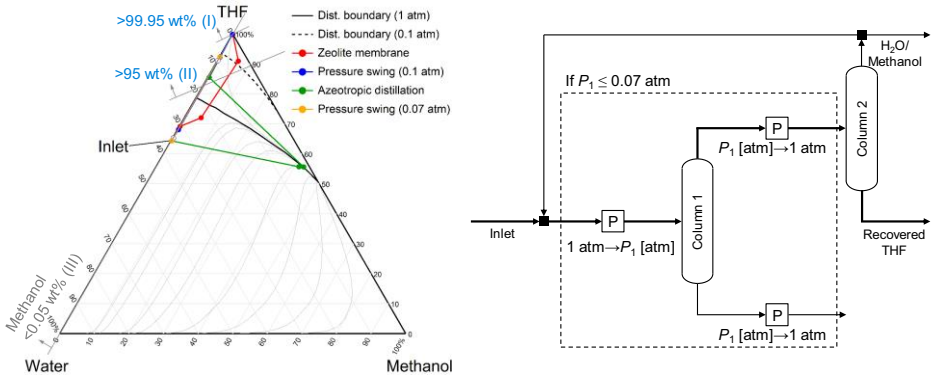


Figure 2. (a) ternary phase diagram of THF/water/methanol (b) pressure swing process

4. Multiobjective evaluation of process alternatives

4.1. Multiobjective evaluation result

Figure 3 (a) shows the result for y_{NPV}. All the alternatives except for the azeotropic distillation process showed a larger y_{NPV} than the as-is process; i.e., economically it makes sense to build and operate a recovery process. The azeotropic distillation process showed a negative value with violation of the constraint on y_{NPV} (see Eq. (1)), because of the poor THF yield and the heavy investment requirements. In the initial design problem, the zeolite and the pressure swing processes resulted in large values of y_{NPV}, which are followed by the two entrainer processes. The main contributors to this result are the THF yield, as well as the investment cost that is affected by the number of columns. In the extended design problem, the value of y_{NPV} increased in all comparable alternatives, and the largest value was given by the pressure swing process (P_1 = 0.07 atm) that is presented in the graph.

Figure 3 (b) shows the result of y_S. As defined in Eq. (7), the indicator highlights the hazard of the flows with the maximum flow rate, which in this case were the flows after recycling. This setup resulted in the large indicator values for the zeolite membrane and the pressure swing processes that showed large maximum flow rates. Similar to y_{NPV}, an improvement was observed between the initial and extended design problems, and the smallest value was given by the pressure swing process (P_1 = 0.07 atm).

4.2. Implications for process design

Figure 3 (c) shows the multiobjective evaluation results using the normalized indicators for the zeolite membrane, pressure swing, and entrainer processes in the initial problem, and for the pressure swing process (P_1 = 0.07 atm) in the extended problem. The validity of Eq. (2) can be seen by investigating the difference in Y between the pressure swing process with P_1 = 0.07 atm and the four alternatives in the initial design problem, all of which were Pareto-optimal. Remarkably, the maximum Y_{NPV} increased by about 17% from the initial to the extended design problem. These results quantitatively confirmed our working assumption that the way in which the regulatory constraint is considered makes a difference regarding the process performance and the alternative to be chosen.

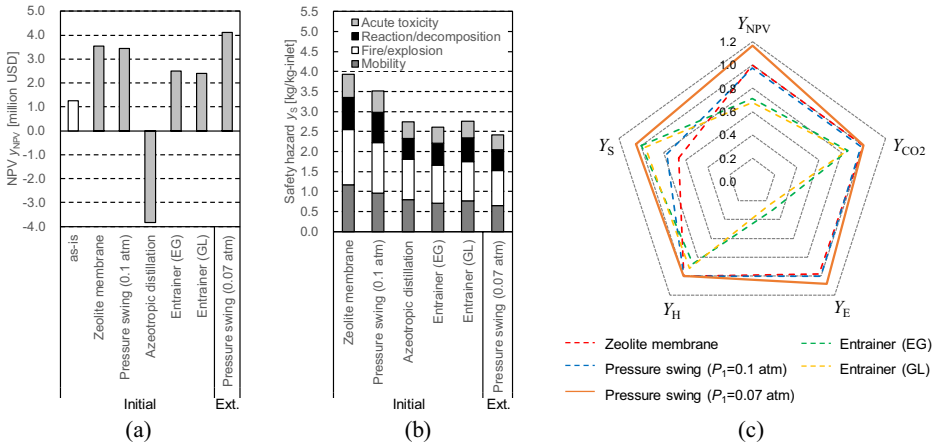


Figure 3. Evaluation results of (a) NPV, (b) safety hazard, and (c) normalized indicators

5. Conclusions

We presented a quantitative analysis of the relevance of regulatory constraints in designing pharmaceutical manufacturing processes. In the case study on waste solvent recovery, an optimal alternative found in the extended design problem showed performance equal to or better than any of the optimal alternatives in the initial design problem. These results confirmed quantitatively that the way to consider the regulatory constraint made a difference in the assessment results and had an impact on the decision.

Acknowledgements

The authors would like to thank Mr. S. Yoshikawa, Mr. A. Oda, Mr. Y. Miyako, Mr. K. Baba, and Mr. H. Nakanishi from Shionogi & Co. Ltd. for fruitful discussions. Financial support from the Japan Society for the Promotion of Science (No. 17H04964) as well as from the Nagai Foundation Tokyo (Research Grant 2017) are gratefully acknowledged.

References

S. García-Muñoz, J. Mercado, 2013, Optimal selection of raw materials for pharmaceutical drug product design and manufacture using mixed integer nonlinear programming and multivariate latent variable regression models. *Ind. Eng. Chem. Res.*, 52, 5934–5942.

H. G. Jolliffe, D.I. Gerogiorgis, 2016, Plantwide design and economic evaluation of two Continuous Pharmaceutical Manufacturing (CPM) cases: Ibuprofen and artemisinin. *Comput. Chem. Eng.*, 92, 269–288.

G. Koller, et al., 2000, Assessing safety, health, and environmental impact early during process development. *Ind. Eng. Chem. Res.*, 39, 960–972.

H. Sugiyama, et al., 2008, Decision-making framework for chemical process design including different stages of environmental, health and safety (EHS) assessment, *AIChE J.*, 54, 1037–1053.

H. Sugiyama, et al., 2017, Process systems engineering approaches for drug product manufacturing: from tablets to injectables, *Comput. Aided. Chem. Eng.*, 40, 2785–2890.

H. Sugiyama, et al., 2018, Relevance of regulatory constraints in designing pharmaceutical manufacturing processes: a case study on waste solvent recovery, *Chem. Eng. Res. Des.*, in submission.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50349-9

A Simulation-based Optimization Approach to Develop Personalized Colorectal Cancer Screening Strategies

David Young, Selen Cremaschi*

Department of Chemical Engineering, Auburn University, Auburn, AL 36849, USA
selen-cremaschi@auburn.edu

Abstract

Colorectal cancer (CRC) is the third leading cause of death due to cancer, with survival closely linked to the stage of cancer at diagnosis. This paper presents a simulation-based optimization approach to generate an optimal screening strategy for CRC. We applied our approach to male and female cohorts, and studied the impact of identified strategies on CRC incidence and death rates for the 1975 US population (a population that had little to no screening). The best strategy identified for men was a colonoscopy every 40 years starting at age 32 and ending at age 96 and for women was a FIT every 17 years starting at age 49 and ending at age 100. If those strategies were implemented for the US population of 1975, it would have resulted in a 45 % reduction in CRC cases and a 53 % reduction in CRC related deaths compared to no screening.

Keywords: Colorectal Cancer, Microsimulation, Cancer Screening, Genetic Algorithm

1. Introduction

Colorectal cancer survival is closely linked to stage of cancer at diagnosis, starting from 92% for early stage diagnosis, stage 1, decreasing to 11% for advanced stage, stage 4. Currently, there are three main methods for CRC screening, endoscopy, stool testing, or computed tomography (CT) scans, and numerous guidelines for CRC screening, e.g., by American College of Gastroenterology (Rex et al., 2009) and by American Cancer Society (Smith et al., 2016). These guidelines, or strategies, are summarized in Fig. 1(A). A screening strategy includes: (1) recommended screening method(s), (2) frequency of screening for each method, and (3) start and potentially end of screening window. The ideal outcome from screening is to disrupt the natural progression pathway to CRC.

An estimated 85% of all CRC cases develop through what is referred to as an adenoma-carcinoma sequence (Gopalappa et al., 2010): A healthy individual develops a precancerous lesion, known as an adenoma within the colon. Not all adenomas will become cancerous, but of the ones that do become cancerous, their probability of transitioning to a cancerous state increases with the size of the adenoma. The remaining 15% of CRC cases occur through a non-visible, or flat, adenoma with no immediate visual signs of an adenoma. The adenoma-carcinoma pathways are demonstrated in Fig. 1(B).

Due to the complexity of CRC and the ethical issues that arise from direct studies of CRC progression, researchers have turned to the use of microsimulations (e.g., Loeve et al., 2000, Frazier et al., 2000, Rutter and Savarino, 2010) that contain a CRC natural development and progression, i.e. natural history model, to conduct epidemiological studies. These microsimulation models have been used primarily to study the cost

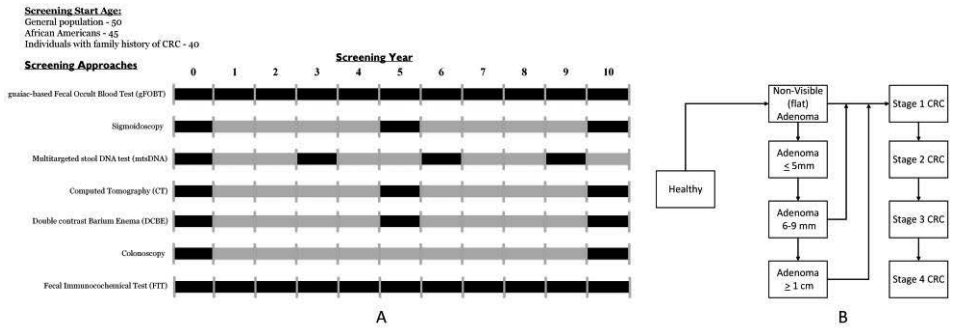


Figure 1. A) Recommended CRC screening strategies. B) Adenoma-carcinoma pathways.

effectiveness of specific screening strategies using what-if analysis. For example, they specify a set of screening strategies, e.g., colonoscopy every 10 years starting at the age of 50 and a FIT every 2 years starting at the age of 45 (Haug et al., 2015), and compare these strategies in terms of their cost-effectiveness and health outcomes. However, the optimum screening strategy may be different from the considered strategies, and it is difficult to efficiently explore all potential screening strategies using what-if-analysis. In this paper, we introduce a simulation-based optimization (SIMOPT) framework to explore the screening strategy space using a CRC microsimulation for efficiently determining the optimum screening strategy for a given cohort.

2. Model Framework

The overall framework can be separated into a CRC natural history model and a screening section, which are implemented in a microsimulation, and an optimization component, which generates various screening strategies that are passed to the microsimulation. For each strategy, the microsimulation constructs a cohort of individuals and their life histories without screening using the CRC natural history model. Then, each individual is screened according to the strategy altering their life history. The microsimulation is repeated multiple times for statistical significance. At completion, the microsimulation computes the cohort's expected gain in life span and expected costs associated with screening and cancer, which are passed to the optimization component for evaluation.

2.1. Natural History

The CRC natural history model assigns each individual a death year and CRC related parameters randomly drawn from distributions constructed by Rutter and Savarino (2010). An individual's death year is stochastically assigned based on the Centers for Disease Control's 2000 life tables (Arias, 2002). The individual's CRC related parameters are used to calculate the development of adenomas, preclinical cancer, clinical cancer, and death due to cancer within their lifetime. These calculations utilize CRC-SPIN's continuous time model (Rutter and Savarino, 2010). The CRC-SPIN's model only incorporates age and gender as CRC related risk factors, and the adenoma-carcinoma pathway, where the probability of transition to cancer is proportional to the adenoma's size. A full description of CRC-SPIN model can be found in Rutter and Savarino (2010).

2.2. Screening

The screening component of the microsimulation takes a screening strategy and an individual's CRC natural history as inputs and outputs an updated life history, where that

individual undergoes screening according to the strategy. The simulation timeline is discretized into one-year increments. The simulation determines when an individual is due for screening, preforms the screening, and takes appropriate action given the screening result. Our model currently assumes a 100 % compliance rate for each individual for every screening strategy and follow-up procedure.

The simulation currently includes three possible screening tests: FIT, sigmoidoscopy, and colonoscopy. For each individual, the screening section determines if an individual has any adenomas or preclinical cancer, as well as location and size of each adenoma at the time of screening. An individual is defined to be at a preclinical cancerous state if the individual has CRC that has not yet been detected or become symptomatic. Detection rate of adenomas and cancers are dependent on the location of the adenoma/cancer, size of the adenoma, how far the testing method reaches within the colon, and the test's sensitivity and specificity (whose values are taken from Aronsson et al., 2017). Each screening test is assumed to have no risk of complications associated with them. Upon positive FIT results, the simulation schedules a diagnostic colonoscopy. Once an adenoma is detected via colonoscopy or sigmoidoscopy it is removed, and the corresponding adenoma progression pathway to CRC is also removed. If the sigmoidoscopy detects any adenomas, a diagnostic colonoscopy is also scheduled. If a sigmoidoscopy or colonoscopy detects preclinical cancer, that individual then transitions into a clinically detected cancer state, where the stage of cancer at diagnosis is potentially less advanced than that of the natural history. Once an individual transitions into a clinically detected CRC state, they are no longer eligible for screening, and instead receive treatment for CRC. The treatment methods are not currently modelled in our simulation but the associated average annual cost of treatment is considered.

2.3. Optimization

The optimization component utilizes genetic algorithm as implemented in the package Distributed Evolutionary Algorithms in Python outlined by Fortin et al (2012). Genetic algorithm was chosen because it is a derivative-free optimization approach and is easy to implement as a wrapper around a microsimulation. In theory, the identification of optimum screening problem can be modelled as a stochastic program, yielding a mixed integer linear program for its deterministic equivalent. However, the sheer size of scenarios (which is the size of microsimulation cohort – 1 million individuals) necessary to represent the population adequately make this approach computationally prohibitive.

The objective is to minimize the total incurred cost due to CRC, which includes screening, diagnostic procedures, and treatment costs, per total quality adjusted life years (QALY) gained due to screening. The decision variables are screening type (w), frequency (x), start age (y), and end age (z). For each decision variable set, the total incurred cost and QALY gained are calculated using the microsimulation.

The objective function is given in Eq. (1), where C_i is the total incurred cost due to CRC and $QYG_{i,n}$ is the QALY gained for individual $i \in I = \{0,1,\dots,|I|\}$ of the simulated cohort for the n^{th} iteration of the microsimulation. Eq. (2) calculates $QYG_{i,n}$ using QALY per cancer stage ($s \in S = \{0,1,2,3,4\}$), $QY_{s,i,t,n}$, based on time of death without screening, $Td_{i,n}$, and time of death with screening, $Td'_{i,n}$. In set S, $s = 0$ represents a cancer free state. In Eq. (3), $C_{i,n}$ is defined as the sum of screening costs, $Cs_{i,t,n}$, and costs due to cancer treatment, $Cc_{s,i,t,n}$, over the individual's lifetime. The values of $Cs_{i,t,n}$, $Cc_{s,i,t,n}$, and $QY_{s,i,t,n}$ are determined by the microsimulation, and these relationships are represented by Eq. (4).

$$min \ \frac{1}{|N|}\sum_{n=1}^{|N|}\frac{\sum_{i=1}^{|I|} C_{i,n}}{\sum_{i=1}^{|I|} QYG_{i,n}} \tag{1}$$

$$C_i = \sum_{t=0}^{Td'} Cs_{i,t} + Cc_{s'_{i,t},i,t} \qquad s'_{i,t} \in S, n \in N \tag{2}$$

$$QYG_i = \sum_{t=0}^{Td'} QY_{s'_{i,t},i,t} - \sum_{t=0}^{Td} QY_{s_{i,t},i,t} \qquad s'_{i,t} \in S, s_{i,t} \in S \tag{3}$$

$$Cs_{i,t,n} = f(w,x,y,z), Cc_{s_{i,t,i,t,n}} = g(w,x,y,z), QY_{s_{i,t,i,t,n}} = h(w,x,y,z) \tag{4}$$

$$w \in W = \{\text{Colonoscopy, Sigmoidoscopy, FIT}\} \quad x \in X = \{1,2,\dots,40\}$$
$$y \in Y = \{20,21,\dots,80\} \qquad z \in Z = \{50,51,\dots,100\} \tag{5}$$

Each gene in the genetic algorithm corresponds to a decision variable, and a unique gene set defines a screening strategy. The algorithm randomly constructs a population of gene sets, evaluates the fitness value of each set as defined by Eq. (1), and induces random changes amongst the initial population to create a new generation of child gene sets. The changes are the random mutations of genes and crossover. The algorithm keeps track of the best preforming set of genes, and continues the generation of child gene sets for a predetermined number of iterations. At termination, it outputs the best three and worst three sets of genes, or screening strategies.

3. Results

The framework was used to determine the optimum screening strategies for both a male cohort and a female cohort. The initial population of gene sets included current screening guidelines for each screening type (three gene sets) and seven randomly constructed screening strategies. The gene sets were then allowed to evolve for ten generations. The probabilities of gene crossover and mutation were 20 % and 50 %, respectively. For each set of genes, the microsimulation's cohort size was one million individuals to coincide with previous screening studies (e.g. Roberts et al., 2007, Loeve et al., 2000). Each individual in the cohort is assumed to be born at time zero. The simulation was run for one hundred iterations, where each iteration had its random number generator seeded by the iteration number to maintain consistency over the comparison of screening strategies. The associated costs were discounted at a rate of 3 % each year after the start of the simulation to account for the time value of money.

Table 1 summarizes the three best and worst screening strategies identified by the framework. Strategies BF and CM (Table 1) closely resemble current recommended screening strategies (Fig. 1(A)). Identification of the accepted screening strategies can be interpreted as a validation of the framework. The results suggest that starting screening later in life (65+ years) yields the highest costs per QALY gained as adenomas developed early in life might have already progressed to CRC. The best strategies favor the use of colonoscopy, which visually inspects the entire colon for adenomas and CRC.

The best strategy identified for the male cohort (AM – Table 1) suggests screening with colonoscopy once every 40 years, at ages 32 and 72. This result demonstrates the importance of early screening. When comparing strategy AM with a similar screening strategy, only starting later in life, it is clear that the early screening has a large effect (e.g. a colonoscopy every 40 years starting at age 64 and ending at 96 yields a cost per QALY gained of $ 12,226.10). A plot comparing the number of clinical CRC cases with and

without screening for the male cohort reveals a small reduction in total number CRC cases due to the initial screening at age 32. However, screening at age 32 deceases the maximum total adenomas in the cohort by 18 %.

Table 1. The three best and worst gender specific screening strategies suggested by the framework

	Female			Male		
	Strategy Name	Screening type, Frequency (y), Start age (y), End age (y)	Cost per QALY gained ($/y)	Strategy Name	Screening type, Frequency (y), Start age (y), End age (y)	Cost per QALY gained ($/y)
Best 3 strategies	AF)	FIT, 17, 49, 100	$ 2,167.41	AM)	Colonoscopy, 40, 32, 96	$ 1,396.25
	BF)	Colonoscopy, 10, 50, 100	$ 2,201.99	BM)	Colonoscopy, 27, 38, 100	$ 1,498.53
	CF)	Colonoscopy, 22, 38, 82	$ 2,626.69	CM)	Colonoscopy, 13, 44, 100	$ 2,042.36
Worst 3 strategies	DF)	Sigmoidoscopy, 1, 80, 92	$ 64,136.28	DM)	Colonoscopy, 13, 80, 100	$ 367,638.92
	EF)	Sigmoidoscopy, 17, 80, 95	$ 74,976.68	EM)	Sigmoidoscopy, 11, 80, 88	$ 433,578.40
	FF)	Sigmoidoscopy, 17, 80, 92	$ 74,976.68	FM)	Sigmoidoscopy, 13, 80, 100	$ 453,126.84

The best strategy identified for the female cohort (AF- Table 1) suggests screening with FIT once every 17 years starting at age 47. The large deviation from the recommended FIT screening strategy may be an artifact of employing diagnostic colonoscopy with perfect compliance upon a positive FIT result. The best screening strategies identified for the female cohort have higher costs per QALY gained than the ones for the male cohort. Males have higher risk of developing CRC than females (Rex et al., 2009), which is also reflected in our microsimulation. Data used to determine the model parameters had 33 % more cases of CRC for males than females (Rutter and Savarino, 2010). Hence, screening has a higher impact on QALY gained for a male cohort.

We applied strategies AF and AM (Table 1) to the 1975 US population, a population in which screening for CRC was not prevalent. We assumed 100 % compliance rate and no risk of complications from screening to investigate the impact screening would have had on the population's CRC incidence and death rate. Figure 2(A) reveals that the screening strategies reduced CRC related deaths by an average of 53 % and decreased the clinical CRC incidence rate by an average of 45 %. Figure 2(B) demonstrates an average of 66 % reduction in the total number of adenomas and an average of 51 % reduction in preclinical CRC incidence. The reduction of the adenomas is responsible for the overall reduction in CRC incidence both clinical and preclinical as well as the reduction of death due to CRC by removing the progression pathway to CRC. These drastic reductions were expected because the original 1975 population received no screening. It should be noted that our 100 % compliance rate assumption artificially inflates the reductions observed in CRC incidence and death rates due to screening. The current compliance rate to colonoscopies is 38 % (Aronsson et al., 2017). Assuming a linear relationship (as a back of the envelope estimation), the observed reductions of CRC related deaths and CRC incidence would have been 20 % and 17 % respectively, which are still significant.

4. Conclusion and Future Work

In this paper, we present a framework to identify a potentially optimal screening strategy for CRC. The framework employs a microsimulation to model the natural progression of CRC within an individual, and to alter the CRC progression using a screening strategy. A genetic algorithm was chosen to determine the optimal screening strategies for male and female cohorts. We applied the best screening strategies identified to the 1975 US population to investigate the effect screening would have had on that population. It was shown that the framework identified screening strategies that are similar to the current guidelines. It should be noted that the current guidelines only consider gender and CRC

history as their risk factors, and hence the similarity of our results to current guidelines increases the confidence in our framework. Future work will focus on incorporating imperfect compliance rates for screening and a diverse set of CRC risk factors within our natural history section, to further describe our investigated cohort creating a more personalized screening strategy.

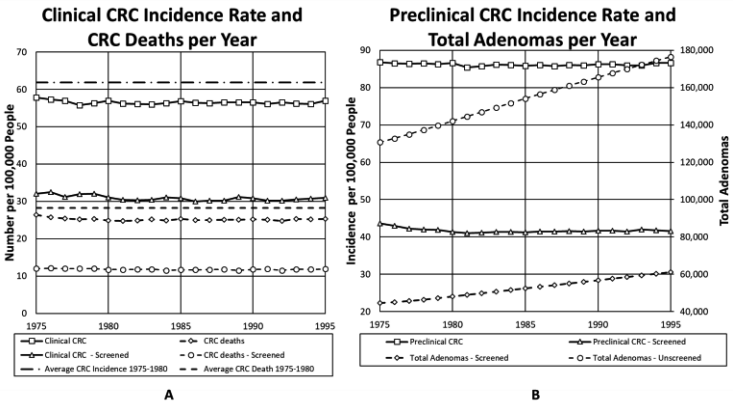


Figure 2. A) Predicted clinical CRC incidence and deaths B) Predicted preclinical CRC incidence and total adenomas.

Acknowledgements: The authors would like to acknowledge financial support from Department of Education GAANN Fellowship (P200A150074).

References

E. Arias, E., 2002, United States life tables, 2000. National Vital Statistics. 51, 1–63.

M. Aronsson, P. Carlsson, L. Levin, J. Hager, and R. Hultcrantz, 2017, Cost-effectiveness of high-sensitivity faecal immunochemical test and colonoscopy screening for colorectal cancer. Br. J. Surg. 104, 1078–1086.

F. Fortin, F. De Rainville, M. Gardner, M. Parizeau, and C. Gagné, 2012, "DEAP: Evolutionary Algorithms Made Easy", Journal of Machine Learning Research, 13, pp. 2171-2175.

A. Frazier, G. Colditz, C. Fuchs, & K. Kuntz, 2000, Cost-effectiveness of Screening for Colorectal Cancer in the General Population. . *JAMA* 284, 1954–1961

C. Gopalappa, S. Aydogan-Cremaschi, T. K. Das. and S. Orcun, 2011, Probability model for estimating colorectal polyp progression rates. Health Care Manag. Sci. 14, 1–21

U. Haug, A. Knudsen, I. Lansdorp-Vogelaar, K. Kuntz, 2015, Development of new non-invasive tests for colorectal cancer screening: The relevance of information on adenoma detection. International journal of cancer Journal international du cancer. 2015;136(12):2864-2874.

F. Loeve, M. Brown, R. Boer, M. van Ballegooijen, G. van Oortmarssen, and J. Habbema, 2000, Endoscopic colorectal cancer screening: a cost-saving analysis. J. Natl. Cancer Inst. 92, 557–563.

D. K. Rex, et al., 2009, American College of Gastroenterology guidelines for colorectal cancer screening 2009 [corrected]. *Am. J. Gastroenterol.* 104, 739–50.

S. Roberts, L. Wang, R. Klein, R. Ness, and R. Dittus, 2007, Development of a simulation model of colorectal cancer. ACM Trans. Model. Comput. Simul. 18, 1–30.

C. Rutter and J. Savarino, 2010, An evidence-based microsimulation model for colorectal cancer: Validation and application. *Cancer Epidemiol. Biomarkers Prev.* 19, 1992–2002.

R. Smith, K. Andrews et al., 2016, R. C. W. Cancer Screening in the United States , 2016 : A Review of Current American Cancer Society Guidelines and Current Issues in Cancer Screening. *A Cancer J. Clin.* 66, 95–114

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50350-5

Economic Evaluation of Batch and Continuous Manufacturing Technologies for Solid Drug Products during Clinical Development

Kensaku Matsunami[a*], Shuichi Tanabe[b], Hiroshi Nakagawa[b], Masahiko Hirao[a], Hirokazu Sugiyama[a]

[a]*Department of Chemical System Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan*
[b]*Pharmaceutical Technology Division, Daiichi Sankyo Co., Ltd, 1-12-1, Shinomiya, Hiratsuka, Kanagawa. 254-0014, Japan*
k-matsunami@pse.t.u-tokyo.ac.jp

Abstract

This work presents a model-based economic evaluation of batch and continuous manufacturing technologies for the solid drug products. The intended phase of applications is during clinical development where process/technology needs to be specified along with the development of formulation. The research was conducted in the three steps: (I) generation of alternatives and recognition of uncertainty, (II) creation of the evaluation models, and (III) interpretation of the evaluation results. As a case study, net present values of the process alternatives using either batch or continuous technology were calculated using industrial data, which were compared considering process uncertainty. Monte Carlo simulation was applied to obtain the difference of net present value between the two alternatives as a probability distribution, which could support more informed decision-making.

Keywords: Continuous manufacturing, Monte Carlo simulation, Economic evaluation, Uncertainty analysis, Decision-making.

1. Introduction

In the pharmaceutical industry, processes and products are designed in conjunction with clinical development. Generally, the results of clinical trials determine the necessary amount of active pharmaceutical ingredients (APIs) per product, and the manufacturing processes are designed accordingly. Continuous manufacturing is a novel technology that is being intensively developed both in the academia and in the industry, especially for the manufacturing of solid drug products (Ierapetritou et al., 2016). The choice between batch and continuous technologies affects the clinical development as well as the commercial production largely regarding the economy, quality, safety, and other technical aspects.

Various kinds of studies have been conducted on continuous manufacturing. Case studies were presented which compared two technologies regarding economic performance (Schaber et al., 2011), and product quality (Järvinen et al., 2015). Sundaramoorthy et al (2012) proposed a framework of capacity planning in drug product manufacturing by considering the uncertainty of clinical trials using continuous manufacturing; Walther et al (2015) compared the economic performance of batch and

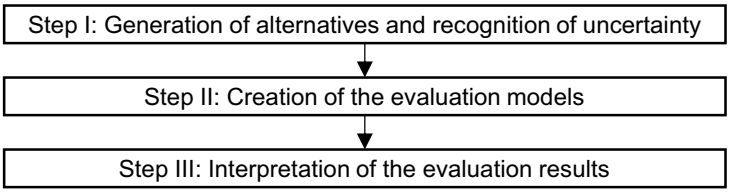


Figure 1 Research strategy for economic evaluation of solid drug products

continuous technologies in biopharmaceutical manufacturing considering uncertainty in technical transfer, product failure during development, and product demand profile. However, general evaluation tools are still to develop, which can support the choice of the batch and continuous technologies with incorporating uncertainty at both clinical development and production stages. To this end, we present an uncertainty-conscious economic evaluation of the two technologies for solid drug product manufacturing, using net present value (NPV) from clinical development to production as the indicator.

2. Research strategy for economic evaluation

Figure 1 shows the three steps of this study: (I) generation of alternatives and recognition of uncertainty, (II) creation of the evaluation models, and (III) interpretation of the evaluation results.

2.1. Generation of alternatives and recognition of uncertainty

In the first step, design alternatives and possible uncertainties at the stages of clinical development and commercial production are listed in order to identify the input parameters to be considered. Table 1 shows the overview of the design alternatives regarding tablet products and the manufacturing processes. Regarding the products, formulation needs to be considered because different types of tablets containing different amount of API are often required for clinical trials. In the industry, two strategies are widely recognized, i.e., “same tablet weight” with different API content and “same composition ratio” with different tablet weight. Given the product, different combinations of unit operations are possible, which is summarized in Figure 2 as the superstructure. Generally, there exist three manufacturing methods: wet granulation, dry granulation, and direct compaction methods. Milling and coating processes are optional, which means that the necessity of these processes is judged mainly by the type of the

Table 1 Design alternatives of pharmaceutical tablets and the manufacturing processes

	Alternatives			
Formulation	A.	Same tablet weight	B.	Same composition ratio
Manufacturing methods	A.	Wet granulation method	B.	Dry granulation method
	C.	Direct compaction method		
Milling	A.	With	B.	Without
Coating	A.	With	B.	Without
Technologies	A.	Batch technology	B.	Continuous technology
Granulator	A.	Fluidized bed granulator	B.	High shear granulator

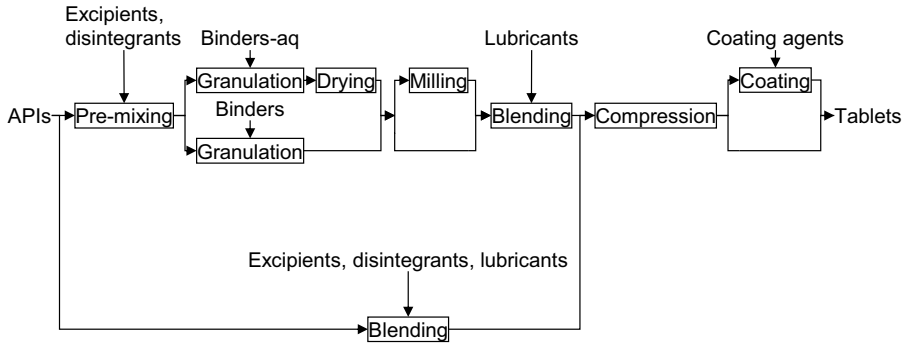


Figure 2 Superstructure of pharmaceutical tablet manufacturing processes

product. After specifying the manufacturing process, the details of unit operations are determined including the choice of batch and continuous technologies and machine types such as fluidized bed and high shear granulation machine. Given the unit operation, there is a range of the process parameters such as blade rotation speed and manufacturing rate. The above mentioned discrete and continuous options need to be specified along with the clinical development towards commercial production, even when some necessary information is unknown. To deal with such decision-making under uncertainty, one powerful approach is Monte Carlo simulation, as Walther et al. (2015) demonstrated in the work of biopharmaceutical process design. We also adopted this simulation technic to incorporate parameter uncertainty in the evaluation result in order to assist more informed decision-making.

2.2. Creation of evaluation models

In the second step, economic evaluation models are developed that can reflect the characteristics of each process and the effect of parameter uncertainty into the result. The objective function is determined as NPV [\$] where the costs for both clinical development and commercial production are considered. In this study, NPV is defined by Eq.(1):

$$NPV = -\sum_{i=0}^{\tau_{\mathrm{dev}}} \frac{C_{\mathrm{dev}}(i)}{(1+r)^i} - \frac{C_{\mathrm{invest}}}{(1+r)^{\tau_{\mathrm{dev}}}} + \sum_{i=\tau_{\mathrm{dev}}}^{\tau_{\mathrm{dev}}+\tau_{\mathrm{sales}}} \frac{C_{\mathrm{sales}}(i) - C_{\mathrm{op}}(i)}{(1+r)^i} \tag{1}$$

where τ_{dev} [y] and τ_{sales} [y] represent the estimated time required for launch and the selling period from launch, respectively. The parameters $C_{\mathrm{dev}}(i)$ [\$ y^{-1}], C_{invest} [\$], $C_{\mathrm{sales}}(i)$ [\$ y^{-1}] and $C_{\mathrm{op}}(i)$ [\$ y^{-1}] represent the annual clinical development cost in the ith year, investment cost, annual sales, and annual operating cost, respectively. The parameter r [–] represents the interest rate.

There are two types of uncertainty in the above mentioned NPV, i.e., one at the clinical development stage, and the other at the commercial production stage. The former contains success/failure of the clinical trials or strategy of formulation development, whereas the latter includes the choice of process technologies and parameters. By combining discrete scenarios and parameter distributions, NPV can be obtained as probability distribution using Monte Carlo simulation.

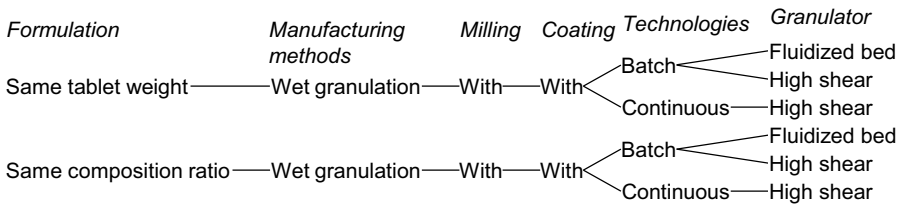


Figure 3 Alternatives of manufacturing processes considered in the case study (details of alternatives are also mentioned in Table 1)

2.3. Interpretation of evaluation results

The evaluation results that are produced as probability distribution are interpreted in order to make decisions on the alternatives (see Table 1). In case of comparing batch and continuous technologies, one approach would be to calculate the difference of NPV between the two technologies in each iteration of Monte Carlo simulation. Sensitivity analysis can be performed between the input and the output distributions in order to identify most contributing parameters.

3. Case study

A case study was conducted along with the steps mentioned above. Based on the actual drug development in the industry, the following situation was assumed. A drug producer has a new tablet product in clinical development. The producer wants to decide which technology to apply for the commercial manufacturing, batch or continuous, both of which are present in the plant. The decision is to be made considering the net present value from the time point of the development until commercial production. Evaluation models were created on the assumption that the products obtained by all alternatives are pharmacologically equivalent for the patient needs.

3.1. Generation of alternatives and recognition of uncertainty

An overview of the product and process alternative was created as shown in Figure 3. The property of the new drug was set as follows: mass fraction of API is high, and the tablet needs to be coated due to the color of API. These product properties reduced the number of alternatives shown in the superstructure in Figure 2 down to the six combinations in Figure 3. More specifically, direct compaction method had to be avoided; milling, granulation, and coating processes were needed.

In this paper, we focus on the choice between batch and continuous for the manufacturing technology with treating other alternatives on, e.g., formulation, as given. Uncertainty for the parameters such as manufacturing rate in continuous manufacturing or numbers of campaign manufacturing lots was considered as relevant, which are to be incorporated in the evaluation.

3.2. Creation of evaluation models

For simplification, the forth cost item of Eq.(1), i.e., operating cost, was selected as the relevant factor, and the economic evaluation model was defined as shown in Eq.(2). Operating cost consists of material, disposal, labor, and utility cost, and the detail of Eq.(2) was presented in our previous research (Sugiyama et al., 2017).

$$NPV = -\sum_{i=0}^{\tau_{\text{sales}}} \frac{C_{\text{op}}(i)}{(1+r)^i} \tag{2}$$

For around 80 parameters contained Eq.(2), probability distributions were determined; for a demonstration purpose, triangular distribution was used, as shown in Eq.(3):

$$f(x) = \begin{cases} \dfrac{2(x - x_{\min})}{(x_{\max} - x_{\min})(x_{\text{std}} - x_{\min})} \,|\, x_{\min} \le x \le x_{\text{std}} \\ \dfrac{2(x_{\max} - x)}{(x_{\max} - x_{\min})(x_{\max} - x_{\text{std}})} \,|\, x_{\text{std}} \le x \le x_{\max} \\ 0 \,|\, x < x_{\min}, x > x_{\max} \end{cases} \tag{3}$$

where $f(x)$ stands for the probability density function of the triangular distribution, x_{std} [–] for the estimated standard value, $x_{\min}$ [–] and $x_{\max}$ [–] for the minimum and the maximum values, respectively. The values of x were provided by the industrial co-authors, where typical manufacturing data was set as x_{std}. For $x_{\min}$ and $x_{\max}$, relatively large margin was assumed to avoid underestimate the effects of parameters. The expected ranges of NPV in batch (NPV^{batch}) and continuous (NPV^{cont}), and the difference ($NPV^{\text{cont}} - NPV^{\text{batch}}$) were calculated by Monte Carlo simulation (10,000 iterations). Figure 4 (a), (b), and (c) shows simulation results of NPV^{batch}, NPV^{cont}, and $NPV^{\text{cont}} - NPV^{\text{batch}}$, respectively, where the last was calculated at each iteration.

3.3. Interpretation of evaluation results

The result in Figure 4 suggests that the probability that NPV^{cont} is smaller than NPV^{batch} is 67 %. If all values of the input parameters are set to $x_{\text{std.}}$, $NPV^{\text{cont}} -$

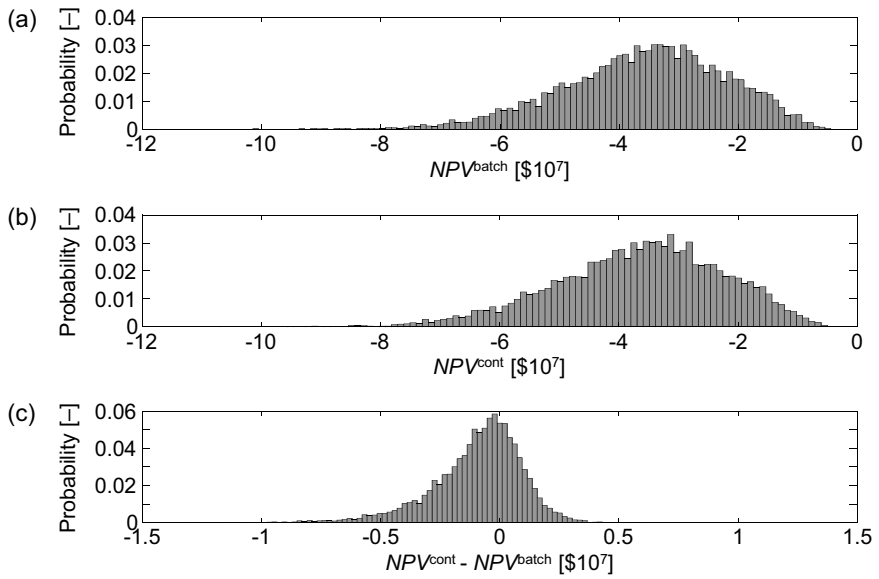


Figure 4 Simulation results of NPV. (a) NPV of batch manufacturing technology, (b) NPV of continuous manufacturing technology, (c) Difference amount of NPV

NPV^{batch} was calculated to be \$6.5×10^5, i.e., if everything is as expected, continuous manufacturing technology is economically beneficial. However, there is 67 % probability that batch manufacturing technology becomes economically more beneficial, i.e., the conclusion could differ when uncertainty was considered.

The ranges of the obtained probability distributions are strongly influenced by the input distributions set, more specifically, by the distribution type as well as the parameters x_{min} and x_{max}. In this case study, a large margin was set for x_{min} and x_{max}; the values should be revised in case more accurate estimation is needed. Sensitivity analysis would quantify the contribution of the input parameters on the output result, which would support the parameter updates and finally decision-makings.

4. Conclusions and outlook

We presented an economic evaluation of batch and continuous technologies for solid drug products as the decision support information during clinical development. An overview of product and process alternatives was shown that are possible to be chosen and are under uncertainty during the stage of development. A case study was conducted using industrial information, where the uncertainty of the process parameters was incorporated in the calculation using Monte Carlo simulation. The NPV using the standard input values was larger for continuous, however, there is 67 % probability that batch becomes more beneficial, i.e., a different conclusion was obtained by considering uncertainty. These findings will be incorporated in our future development of the general tools, which can support the choice of the batch and continuous technologies comprehensively.

Acknowledgement

Financial supports by Grant-in-Aid for Young Scientists (A) No. 17H04964 from Japan Society for the Promotion of Science are acknowledged.

References

M. Ierapetritou, F. Muzzio, G. Reklaitis, 2016. Perspectives on the continuous manufacturing of powder-based pharmaceutical processes, *AIChE J.*, 62, 1846–1862

M. A. Järvinen, M. Paavola, S. Poutiainen, P. Itkonen, V. Pasanen, K. Uljas, K. Leiviskä, M. Juuti, J. Ketolainen, K. Järvinen, 2015. Comparison of a continuous ring layer wet granulation process with batch high shear and fluidized bed granulation processes, *Powder Technol.*, 275, 113–120

S. D. Schaber, D. I. Gerogiorgis, R. Ramachandran, J. M. B. Evans, P. I. Barton, B. L. Trout, 2011. Economic analysis of integrated continuous and batch pharmaceutical manufacturing: a case study, *Ind. Eng. Chem. Res.*, 50, 10083–10092

H. Sugiyama, K. Matsunami, K. Yabuta, 2017. Process systems engineering approaches for drug product manufacturing: from tablets to injectables, *Comput. Aided Chem. Eng.*, 40, 2785–2790

A. Sundaramoorthy, J. M. B. Evans, P. I. Barton, 2012. Capacity planning under clinical trials uncertainty in continuous pharmaceutical manufacturing, 1: mathematical framework, *Ind. Eng. Chem. Res.*, 51, 13692–13702

J. Walther, R. Godawat, C. Hwang, Y. Abe, A. Sinclair, K. Konstantinov, 2015. The business impact of an integrated continuous biomanufacturing platform for recombinant protein production, *J. Biotechnol.*, 213, 3–12

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50351-7

Precision healthcare supply chain design through multi-objective stochastic programming

Xiaonan Wang[a,*], Qingyuan Kong[b], Maria M. Papathanasiou[b], Nilay Shah[b]

[a]*Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore 117585 Singapore*
[b]*Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, London SW7 2AZ, UK*
chewxia@nus.edu.sg

Abstract

Following the FDA's historic approval of the first cell-based, autologous, cancer therapy in 2017, there has been an increasing growth in the personalized cell therapy market. Both the personalized character as well as the sensitive nature of these therapies, has increased the complexity of their supply chain design and optimisation. In this work, we have addressed key issues in the cyclic supply chain for simultaneous design of the supply chain and the manufacturing plan. A comprehensive optimisation based methodology through both deterministic and stochastic programming is presented and applied to study the Chimeric Antigen Receptor (CAR) T cell therapies. Multiple objectives including maximisation of the overall net present value (NPV) and minimisation of the average response time of all patients are evaluated, while accounting the uncertainties in patients' demand distribution. Results indicate that the total benefits from the optimized supply chain management are significant compared with the current global market.

Keywords: Precision healthcare, supply chain, optimisation, economic.

1. Introduction

Precision healthcare and medicines, especially cell and gene therapies, are progressing rapidly in clinical trials (Wang and Shah, 2017). As a breakthrough, Chimeric Antigen Receptor (CAR) T-cell therapies have proved to be an effective approach to the treatment of hematological malignancies, such as B-cell leukemia and lymphoma. In August 2017, the U.S. Food and Drug Administration (FDA) made the historic approval of Novartis' Kymriah; an autologous CAR T-cell therapy for B-cell acute lymphoblastic leukemia (ALL), opening the road to novel, cell-based, autologous cancer therapies. Currently, Kymriah's cost is estimated to be $475,000 per patient (Bach et al., 2017), making autologous therapies not only very expensive for patients to afford, but also very challenging to be reimbursed by health insurance schemes (Malik and Durdy, 2016). Due to their sensitive nature, such products are coupled with stringent handling requirements (e.g. temperature hold times, storage times) during transportation and manufacturing. As a result, mapping and then optimizing the supply chain of such products is a challenging task, as one needs to consider multiple factors, ranging from the availability of raw materials to transportation costs that usually, highly affect the total cost (Rafiq et al., 2017). Given their immaturity in the market, such therapies are currently using suboptimal manufacturing and delivery routes. Due to market growth, the latter will soon prove to be insufficient, driving the pharmaceutical industry towards novel solutions that will assist a smooth scale out in process development. The optimisation of the supply chain in cell-based, autologous therapies, is of key importance as it can effectively decrease therapy cost and delivery time and bring such therapies to a much larger group of patients. An

innovative supply chain from laboratory inventions to full-scale commercialization is desired to maximize healthcare and economic benefits (Shah, 2004).

Multi-objective stochastic programming has been effectively adopted to optimize supply chains of various products (Hugo and Pistikopoulos, 2005; Liu and Papageorgiou, 2012; Nikolopoulou and Ierapetritou, 2012; You and Grossmann, 2008). To address the challenges in precision healthcare and medicine supply, here, we establish a stochastic model that can capture the uncertainties in patients' treatment needs, while evaluates objectives including maximizing the overall net present value (NPV) and minimizing the average response time of all patients. Compared with classical supply chain problems that mainly address costs or NPV, the additional objectives of patients' treatment delivery and quality control are equally or even more important for the biomedical industry. Following that, the decision variables such as number of manufacturing modules and distribution centres to be constructed and locations of modification for each sample can be determined under the supply chain constraints (e.g. processing capacities of hospitals and manufacturing sites).

This work first provides a qualitative analysis of the current supply chain in CAR T-cell therapies in Section 2, touching upon main bottlenecks/challenges that are reported in the open literature and/or indicated by the pharmaceutical industry. A mathematical model is formulated in Section 3 and solutions using a deterministic scenario and scenarios in which the demand uncertainty evaluated are presented in Section 4. Conclusions and future work are discussed in Section 5.

2. Qualitative analysis of the CAR T-cell supply chain

We first focus on a qualitative analysis of the current supply chain of CAR T-cells, collecting information both from the open literature, as well as from interviews with industrial project partners.

2.1. Current state-of-the-art

Figure 1 presents the current supply chain of CAR T-cells that begins from the collection of the cells from the patient at a specialized clinical site (*leukapheresis & cell separation/enrichment*). At this point (point (1), Figure 1), the patient sample can be frozen and stored, until a larger batch of patient samples is created. Optimal shipment schedules are defined by the capacity of the manufacturing site, as current facilities can handle a finite number of samples.

During manufacturing, samples undergo a series of modifications (*activation, genetic modification, expansion, formulation, and cryopreservation*) until the final product is ready to be shipped back to the patient. The majority of the manufacturing steps are based on the supply of common raw materials, such as medium, cell washing accessory sets, selection reagents etc. Nevertheless, during *genetic modification* (point (2), Figure 2), patient cells are transduced by viral vectors carrying the genetic material required for the CAR expression. Unlike other materials used in the manufacturing process, viral vector products are highly sensitive, with relatively short shelf-life and strict stability windows, with respect to temperature. Moreover, they are characterized by a complex, lengthy manufacturing process. Therefore, there is an evident interdependence between the two supply chains (CAR T-cells & viral vectors), that stresses the need for the design of optimal planning and coordination, in order to avoid product losses and/or shortages. In particular, the demand in the CAR T-cell space automatically indicates the respective demand of viral vectors.

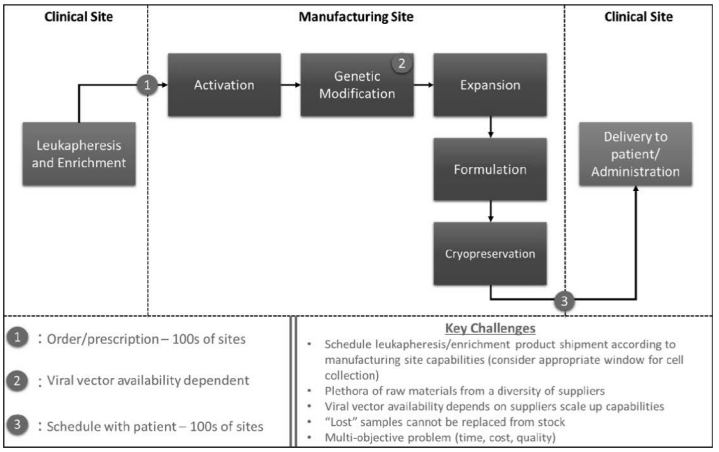


Figure 1. Current CAR T-cell process/distribution steps along with key bottlenecks and challenges (Levine et al., 2017).

Following their manufacturing, CAR T-cells need to be transferred back to the clinical site where they will be administered to the patient. That can often be translated into airfreight shipping, as sites might be located not only in different countries, but also in different continents (Griffiths, 2016). Therefore there are significant transport costs involved, but also challenging cross-border regulations to consider. Moreover, treatment administration must be scheduled based on patient condition; thus posing additional challenges in defining the optimal task scheduling profile.

2.2. Process bottlenecks/challenges

It is observed that commercialisation of CAR T-cell therapies is a challenging, multi-factorial task. Firstly, such therapies introduce 1-to-1 business models, making the patient part of the supply chain. Moreover, due to their autologous nature and specificity, they restrict the possibility of product losses to be replaced from stock, thus stressing the need for minimisation of failure rates. Reflecting the sensitive nature of the studied therapies, *temperature and time excursions* are reported to be amongst the most important challenges. Meanwhile, *sample identification* is at high risk at all times as in this case the patient becomes his/her own donor. Therefore it is essential that samples are not only tracked from the collection point to the manufacturing center, but also and most importantly during their delivery back to the patient, ensuring that the right therapy is administered to the right patient. Other barriers/challenges to commercialisation may arise from: manufacturing (e.g. insufficient facilities), current business models (e.g. one-to-many), reimbursement & clinical adoption procedures.

3. Problem formulation and methodology

3.1. Problem statement

Given a set of hospitals j and the number of patients i that go through cell-therapy cancer treatment at each hospital in a given time period T, the objective is to establish a number of cell manufacturing facilities k and distribution (storage) centres m among the potential available sites with an optimal economic goal meeting the treatment demand while keeping the responsive time of each patient below a certain value and maximising utilisation of existing assets. Different transportation modes l are considered depending

on feasible connections between two sites.The clinical treatment data obtained from ClinicalTrials.gov is used as the foundation to estimate global demand. Time series analysis and regression is first conducted as the most basic methodology for future demand forecasting. The predicted treatment needs follow a certain type of curves for each individual region, mainly in the U.S. (nearly linear growth) and China (exponential growth).

3.2. Supply optimisation

Decision making in the network modelling is formulated as a mixed-integer linear programming (MILP) model, as the objective functions and part of constraints listed in Table 1. The interactive method is used to evaluate the preferences of multiple objectives, which is adapted from the weighting and ε-constraint methods and solved using CPLEX.

Table 1. Mathematical model formulation as designed for the optimisation problem

Index	Mathematical Formulation	Description
Objective functions		
(1)	$Z_1\,(NPV) = Revenue - C_{invest} - C_{operate} - C_{manufacture} - C_{transport} - C_{store} - C_{R\&D}$	Total NPV
(2)	$Z_2\,(Response\ Time) = T_{transport} + T_{manufacturing} + T_{storage}$	Maximisation of patients' satisfaction
Constraints		
(3)	$MCP \times N_k^M - \sum_i S_{i,k}^{MT} \geq 0$ $SCP \times N_m^D - \sum_i MS_{i,m}^{DT} \geq 0$	Manufacturing & distribution capacity constraints
(4)	$\sum_k Y_k^M \leq \alpha\,,\ \ \sum_m Y_m^D \leq \beta$	Total number of manufacturing sites and distribution centres
(5)	$n_i Y_{i,j}^H = \sum_{k,l} S_{i,j,k,l}^H\,,\ \ S_{i,k}^{MT} = \sum_{j,l} S_{i,j,k,l}^M\,,\ \ MS_{i,m}^{DTS} = \sum_{j,l} MS_{i,j,m,l}^{DH}$	Sample balance at each node (hospital, manufacturing and distribution centre)

The multi-objective programing aims to maximize the total NPV (Eq. (1), Table 1). The economic cost structure of cell therapies supply chain includes capital investments, nominal operating costs, manufacturing, transportation and storage costs, as well as potential research and development costs. While the maximisation of patients' satisfaction is represented by the response time of the total supply chain (Eq. (2), Table 1). The two objectives are evaluated in a weighed sum manner, or by converting one to a constraint. In this paper, we will pose an upper bound to the response time while mainly aiming to maximize the NPV of the value chain.

The optimisation constraints mainly include capacity constraints, material and sample balance, and other logistics constraints (Eq. (3)-(5), Table 1). As discussed in Section 2, a key feature of the precision healthcare is that each patient is treated and manufactured individually, while tracked with a label through the whole life cycle and back to the patient in a loop. The sample balance and cell tracking are incorporated into the optimization program as a set of binary constraints.

4. Case study and results

Following the framework proposed, a case study of the CAR-T supply chain is presented in this section. The output provides a whole seamless supply chain with distribution, logistics, and storage capacity for cell and gene therapies in order to accelerate advanced therapy development and commercialisation.

4.1. Deterministic case

In the base case, we consider four types of hospitals are accepting patients at fixed cohort size. The real distance and transportation time are used as input to characterize multiple

routes through each manner (e.g. air and road). The deterministic case is characterized by the absence of uncertain variables, which treats the total patients distributed to be treated worldwide are aggregated to groups. The problem is formulated as a multi-objective MILP optimisation model with the assumption that each batch of patients pass through the same supply chain route as Figure 1. An estimated size of the total patients group is 9,364, which is consistent with the clinical statistics. The 11 considered hospitals that accept patients are located in London, Paris, California, New York, Texas, Florida, Washington (Seattle), China North (Beijing), China South (Shanghai), Singapore and Australia. Another 6 candidate locations can build cGMP manufacturing centres.

The results of the selected sites to for CAR-T manufacturing are shown in Figure 2. The desired responsive time plays a key role in deciding the optimal supply chain strategies. When relaxing the time constraint, the NPV can increase by 64.7% compared with the base case (normal demand level) shown in Table 2.

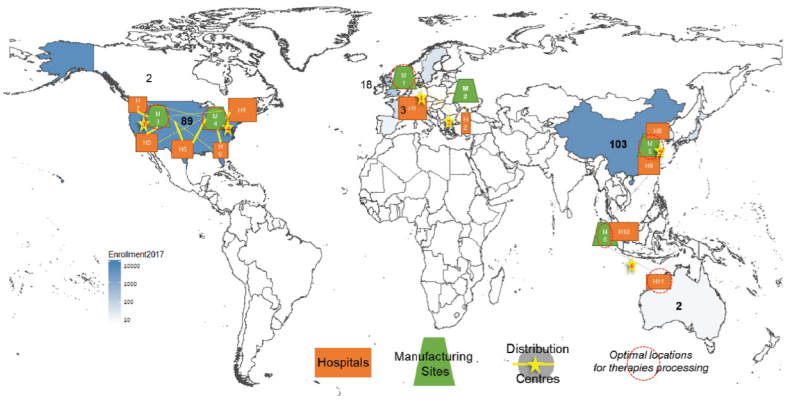


Figure 2. Graphical representation of the results from the deterministic scenario.

4.2. Stochastic case

The operation problem becomes more complex when probabilities are evaluated through stochastic programing approaches. In this stochastic case, the model can be adjusted to accommodate operational issues such as uncertainties in end-point treatment needs, manufactory inventory, and quality of samples. Here we only demonstrate the simplified results from the scenario-based uncertainty analysis, by evaluating each demand stage, e.g., low, medium and high levels. There is a trade-off between total responsive time and economies of scale. Large-scale centralised manufacturing centres are optimal under deterministic conditions, while distributed small-scale flexible plan are more cost-effective under stochastic conditions.

Table 2. Optimisation results of scenarios considering demand uncertainty.

Demand of treatment globally	NPV (billion USD)	Response time (days)	Total manufacturing modules
Low level (121 each month)	1.103	15.31	45 (at 6 manufacture centres)
Normal (242 each month)	1.636	15.31	85 (at 1 hospital and 5 manufacture centres)
High level (484 each month)	0.738 if no time constraint, otherwise infeasible	15.46	167 (at 6 manufacture centres)

5. Conclusions and future work

An illustrative example has been presented to describe the global supply chain optimisation problem for cell therapies for precision healthcare. A commercial cryogenic storage, distribution and logistics solution consisting of configurable, modular components is demonstrated. Through the proposed supply chain optimisation, the NPV increases while the response time is much shortened, compared with the current situation where all therapies are manufactured on site in a completely decentralised manner. In the current discussion, we only consider a single-stage planning using a "snap-shot" description of the supply chain. Future work will also discuss dynamic supply chain operational problems with the development of manufacturing capability through robust, repeatable and scalable routes. Machine learning techniques and data-driven approaches will also be applied to forecast the number of patients and expansion of treatment capacities, associated with multiple factors such as the morbidity, costs, technology advances and clinician provisions, to obtain effective supply chain management solutions.

Acknowledgments

Funding from the UK Engineering & Physical Sciences Research Council (EPSRC) for the Future Targeted Healthcare Manufacturing Hub hosted at University College London with UK university partners is gratefully acknowledged (Grant Reference: EP/P006485/1). Financial and in-kind support from the consortium of industrial users and sector organisations is also acknowledged.

References

Bach, P.B., Giralt, S.A., Saltz, L.B., 2017. FDA Approval of Tisagenlecleucel. JAMA. https://doi.org/10.1001/jama.2017.15218

Griffiths, R., 2016. Successfully managing the unique demands of cell therapy supply chains.

Hugo, A., Pistikopoulos, E.N., 2005. Environmentally conscious long-range planning and design of supply chain networks. J. Clean. Prod. 13, 1471–1491. https://doi.org/10.1016/J.JCLEPRO.2005.04.011

Levine, B.L., Miskin, J., Wonnacott, K., Keir, C., 2017. Global Manufacturing of CAR T Cell Therapy. Mol. Ther. Methods Clin. Dev. 4, 92–101. https://doi.org/10.1016/j.omtm.2016.12.006

Liu, S., Papageorgiou, L.G., 2012. Multiobjective optimisation of production, distribution and capacity planning of global supply chains in the process industry. https://doi.org/10.1016/j.omega.2012.03.007

Malik, N.N., Durdy, M.B., 2016. CAR T-cell therapies: formulation–product–proposition framework for commercialisation. Drug Discov. Today 21, 1731–1734. https://doi.org/10.1016/j.drudis.2016.09.008

Nikolopoulou, A., Ierapetritou, M.G., 2012. Optimal design of sustainable chemical processes and supply chains: A review. Comput. Chem. Eng. 44, 94–103. https://doi.org/10.1016/J.COMPCHEMENG.2012.05.006

Rafiq, Q.A., Nienow, A.W., Hewitt, C.J., Coopman, K., 2017. The role of biopreservation in cell and gene therapy bioprocessing. Cell Gene Ther. Insights. https://doi.org/DOI: 10.18609/cgti.2017.037

Shah, N., 2004. Pharmaceutical supply chains: key issues and strategies for optimisation. Comput. Chem. Eng. 28, 929–941. https://doi.org/10.1016/j.compchemeng.2003.09.022

Wang, X., Shah, N., 2017. Supply Chain Logistics for Personalized Medicines, in: AIChE Annual Meeting.

You, F., Grossmann, E., 2008. Design of responsive supply chain under demand uncertainty. Comput. Chem. Eng 32, 3090–3111.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50352-9

Stochastic Modelling of CTB-receptor Binding Kinetics

Dongheon Lee[a,b], Alec Mohr[a], Joseph Sang-Il Kwon[a,b*], Hung-Jen Wu[a], Akshi Singla[a]

[a]*Artie Mcferrin Department of Chemical Engineering, Texas A&M University, 3122 TAMU, College Station, 77843, USA*
[b]*Texas A&M Energy Institute, 3372 TAMU, College Station, 77843, USA*
kwonx075@tamu.edu

Abstract

How cholera toxin subunit B (CTB) binds to its receptor in host cell membrane is still not fully understood due to its complex nature. In this study, a kinetic Monte Carlo (kMC) model is proposed to describe the binding kinetics between CTB and its receptor on lipid membrane. The proposed kMC model considers receptor migration, CTB attachment and detachment, and surface forward and backward reactions as discrete microscopic events. At every moment, an event to be executed is selected randomly based on the rates of all possible events, and this procedure continues till the end of simulation. By accounting for these events, the predictions from the proposed kMC model qualitatively agree with the experimental measurements.

Keywords: kinetic Monte Carlo, multivalent protein-glycan interaction, cholera toxin

1. Introduction

In the intestinal lumen, *Vibrio Cholerae*, the main culprit of cholera, secrets cholera toxin (CTx), which is responsible for acute symptoms of cholera once CTx enters cytoplasm. In order for a CTx to move into cytoplasm, its pentameric CTx subunit B (CTB) must bind with its receptor, ganglioside, on host cell membrane that initiates the CTx internalization process. Without the CTB-ganglioside binding, a CTx is unable to cross the host cell membrane as a cell membrane is impermeable to a CTx. Consequently, an effective inhibition of the CTB-ganglioside binding can prevent the cholera infection even in the presence of the bacteria.

Role and structure of CTB have been extensively studied, but the binding dynamics between CTB and ganglioside still remains elusive, particularly at the microscopic level, since its nature is intrinsically complex. First, a CTB can bind up to five ganglioside receptors in host cell membrane; second, multiple types of receptors present in cell membrane can bind with CTB with different affinities (Krishnan et al., 2017); last, the low number of ganglioside receptors in host cell membrane renders the binding process stochastic (Wilkinson, 2009).

Motived by these considerations, this work aims to implement kinetic Monte Carlo (kMC) methodology (Nayhouse et al., 2013) to model the binding kinetics between CTB and GM1 receptors on the surface, which will help us enhance our understanding of the underlying physics of the CTB-ganglioside binding mechanisms at the microscopic level. The kMC is chosen since it is a powerful tool to simulate microscopic reactions on a surface and can track the temporal evolution of the surface

configuration (Nayhouse et al., 2013). Specifically, the dynamic model incorporates four microscopic phenomena of the CTB-receptor binding mechanisms: 1) diffusion of a CTB to surface and attachment to a receptor in host cell membrane, 2) dissociation of a CTB-receptor complex, 3) association between a CTB-receptor complex and a free receptor, 4) migration of free receptors. By accounting for these events, the predictions from the proposed kMC model qualitatively agree with the experimental measurements. In the future, the proposed methodology can be extended to study the binding kinetics between CTB and lipid membrane containing more than one type of receptors to investigate the roles of different receptors.

2. Methodology

This study aims to model the binding kinetics between CTB proteins and their receptors in a cell membrane mimicking environment, where CTB proteins are in the solution and receptors are localized on surfaces of spherical lipid bilayers. The solid-to-solid model and kMC algorithm are implemented to simulate the CTB-receptor binding kinetics (Nayhouse et al. 2013; Fallahi-Sichani and Linderman, 200). In this model, the surface of lipid bilayer is modeled as a square lattice with a periodic boundary condition, and receptors are randomly distributed on the surface. Description of microscopic phenomena

Here, the CTB-receptor binding kinetics are described by five microscopic phenomena as follows

- Receptors and CTB-receptor complexes migration on a lipid bilayer surface.
- CTB proteins diffuse within the solution.
- If a CTB is sufficiently close to a receptor on the surface, CTB can attach to the surface by binding to the receptor.
- A membrane-bound CTB binds to an additional receptor if the CTB has an unfilled binding spot and a receptor is sufficiently close to the CTB.
- A receptor can dissociate from a membrane-bound CTB, and the CTB will detach from the lipid bilayer after all receptors dissociate from it.

For the purpose of this study, the above descriptions are simplified with the following two assumptions. First, the simulation domain is restricted to two-dimensional as our main interest is the dynamics on the membrane. By assuming the solution is well-mixed, the transport of CTB proteins to and from the surface via diffusion is considered implicitly by introducing effective association and dissociation rate constants, which are defined as follows (Lauer et al., 2002):

$$k_f = \frac{k_0}{1 + \frac{5k_0R}{4\pi Da}}, \qquad k_r = \frac{k_{-1}}{1 + \frac{5k_0R}{4\pi Da}} \tag{1}$$

where k_f and k_r are effective association and dissociation rate constants, k_0 and k_{-1} are intrinsic association and dissociation rate constants, R is the number of receptors available for binding, D is a diffusivity coefficient of CTB in the solution, and a is the radius of lipid bilayers. Please note this simplification is derived based on the steady-state diffusion equation and has been experimentally verified (Erickson et al. 1987).

Second, we assume CTB-receptor complexes on the surface is immobile. With these simplification, only the following five events on the surface are considered during the kMC simulation.

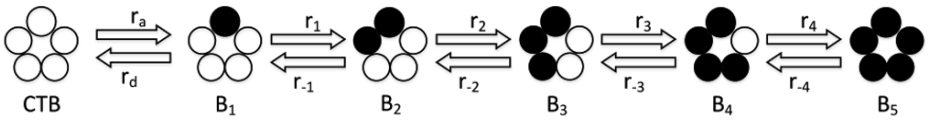


Figure 1. The schematic diagram for the CTB-receptor binding kinetics. B_i is a CTB bound to i number of receptors. Empty and filled spheres represent CTB binding spots without and with bound receptors, respectively.

2.1. Surface kinetics

Description of the CTB-receptor binding kinetics is adapted from that of Lauer et al. (2002) as illustrated in Fig. 1. As a CTB has five binding spots, a CTB will bind to or dissociate from receptors in a step-wise manner, which results in ten reactions to be considered in the kMC (Fig. 1). It should be noted that the steps from CTB to B_1 and B_1 to CTB should be treated separately since the effective rate constants k_f and k_r need to be used to account for the transport of CTB to the membrane implicitly as described above. First, the attachment and detachment rates are defined as

$$r_a = 5k_f CR, \qquad r_d = k_r B_1 \tag{2}$$

where r_a is the attachment rate from solution to the surface, C is the CTB concentration in solution, r_d is the CTB detachment rate from the membrane to solution, and B_1 is the number of CTB-receptor complex with one bound receptor. Here, a factor of five is multiplied because a CTB protein is symmetric molecule with five identical binding spots for receptors (Lauer et al., 2002).

Since a CTB can take up to five receptors, a membrane-bound CTB can associate with or dissociate from additional receptors, which are termed as forward and backward reactions, respectively, hereafter (Fig. 1). The reaction rates of these two events on the membrane are computed as follows

$$r_i = (5-i)k_1 B_{i-1} R, \qquad r_{-i} = (i+1)k_{-1} B_{i+1}, \qquad \forall\, i = 1, \ldots, 4 \tag{3}$$

where r_i is the forward reaction rate from B_i to B_{i+1}, B_i is the number of CTB bound to i number of receptors, k_1 is the forward reaction rate, and r_{-i} is the backward reaction rate from B_{i+1} to B_i.

Finally, the receptor migration rate is defined as (Nayhouse et al., 2013; Fallahi-Sichani and Linderman, 2009)

$$r_m = R\frac{k_m}{l^2}, \tag{4}$$

where k_m is the receptor migration rate constant, and l is the distance between two lattice sites. Nominal parameters for the kinetic parameters and diffusivity are taken from the literature and listed in Table. 1 (Lauer et al. 2002; Krishan et al. 2017).

2.2. Kinetic Monte Carlo Implementation

An event is selected based on a random number and the total reaction rate, r_t, which is defined as

$$r_t = r_a + r_d + \sum_{i=1}^{4} (r_i + r_{-i}) + r_m \tag{5}$$

In order to execute an event, a uniform random number, $\xi_1 \in [0,1)$, is sampled. If $\xi_1 \leq r_a/r_t$, the attachment event is selected. If $r_a/r_t < \xi_1 \leq (r_a + r_d)/r_t$, the detachment event is selected. If $(r_a + r_d + \sum_{i=1}^{j-1} r_i)/r_t < \xi_1 \leq (r_a + r_d + \sum_{i=1}^{j} r_i)/r_t$, the forward reaction from B_j to B_{j+1} is selected. If $(r_a + r_d + \sum_{i=1}^{4} r_i + \sum_{i=1}^{j-1} r_{-i})/r_t < \xi_1 \leq (r_a + r_d + \sum_{i=1}^{4} r_i + \sum_{i=1}^{j} r_{-i})$, backward reaction from B_{j+1} to B_j is selected to occur; if $\xi_1 > (r_a + r_d + \sum_{i=1}^{4} r_i + \sum_{i=1}^{4} r_{-i})/r_t$, the migration event is selected.

When an attachment event is selected, a free receptor is randomly selected to associate with an incoming CTB protein. After a receptor for binding is selected, it is required to check whether there is enough free space around the selected receptor for a CTB without overlapping with other CTB molecules that are already bound to the host cell membrane. If there is no enough space for an incoming CTB, the attachment will be rejected. As the membrane becomes more crowded with the increasing number of membrane-bound CTB molecules, the available space for an additional CTB to attach to the host cell membrane decreases significantly; hence, the rejection rate will increase accordingly.

When a detachment event is selected, one CTB bound to one receptor, which is denoted as B_1 (Fig. 1), is randomly selected and dissociates with the receptor, leaving it on the host cell membrane. Whenever the attachment or detachment event occurs, the concentration of CTB in solution is updated via mass balance by counting the number of proteins undergoing attachment and detachment process

When a forward reaction event from B_i to B_{i+1} is selected, whose rate is denoted as r_i, one CTB bound to i number receptors is randomly selected, and another free receptor will be bound to it. Here, it is required to check whether there are any free receptors sufficiently close to the selected binding spot, which is determined by the distance between the binding spot to free receptors on the host cell membrane. If the distance is smaller than the threshold distance (l_c), the corresponding receptor is classified as a free receptor that can bind with the CTB. If there is no receptor close to the selected binding spot, the forward reaction will not occur; if there is at least one such receptor, one receptor is randomly selected and bind to the binding spot.

Similarly, when a backward reaction event from B_{i+1} to B_i is selected, one CTB bound to $i+1$ number receptors is randomly selected, and one of its bound receptors is randomly selected for dissociation.

Table 1. Nominal parameter values for the CTB-GM1 binding kinetics

$k_0\ (M \cdot s^{-1})$	$k_{-1}\ (s^{-1})$	$k_1\ (cm^2 s^{-1})$	$D\ (m^2/s)$	$k_m\ (m^2/s)$
2.8×10^4	3.2×10^{-3}	3.5×10^{-15}	9.8×10^{-11}	8.25×10^{-12}

When a migration event is selected, one free receptor is randomly selected and move to one of its neighbouring empty site.

After one event is selected and proceeds as described above, the time increment for the selected event is calculated by generating a new random number $\xi_2 \in [0,1)$, and the time increment is computed as follows (Nayhouse et al., 2015)

$$\tau = -\ln \xi_2 / r_t \tag{6}$$

And the simulation will proceed by $t + \tau$ seconds. The kMC simulation is written in C#, and a few realizations are computed to calculate the average kinetics.

3. Results & Discussion

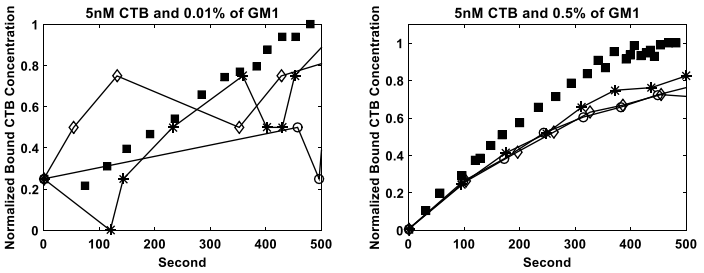


Figure 2. Simulated profiles of total number of bound CTB on membrane with different surface densities of GM1. On each figure, each line represents a different realization of the kMC simulation while filled square points are experimental measurements from Lauer et al. (2002).

The proposed kMC model was applied to describe the number of CTB bound to receptors on lipid membrane. Specifically, the initial CTB concentration was 5nM, and lipid membrane contained two different amounts of GM1 receptors. Here, 0.01% and 0.5% GM1 lipid membranes were considered for the kMC simulations. In Fig. 2, three different realizations of the kMC simulation for each GM1 surface density were presented. Predicted kinetics qualitatively agreed with the experimental measurements (Lauer et al., 2002), which validated the predictive capability of the proposed kMC simulation.

At low GM1 concentration such as 0.01% GM1 membrane, three realizations showed distinct dynamic behaviors due to the dominance of stochasticity in the kinetics caused by low copy number of receptors (Wilkinson, 2009). On the other hand, three realizations of the binding kinetics at 0.5% GM1 surface concentration showed a very consistent behavior with each other, which indicated that stochasticity was minimal under the higher GM1 surface concentration. Two contrasting kMC simulation results under two different receptor surface densities indicated that one of the main sources of stochasticity in the ligand-receptor binding kinetics is the initial amount of receptor available.

The other important feature of this kMC simulation is that it can take into account surface configurations such as hindrance effect of neighboring bound proteins and receptor aggregations. As Fig. 3 demonstrates, the surface containing 2% GM1 became more crowded, which increases the likelihood to reject the attachment event. The consideration of the surface configuration is a desirable feature of the proposed kMC simulation framework since there should be a maximum amount of CTB proteins that

can be bound to the membrane in reality. Previous deterministic models to predict the CTB-GM1 binding kinetics (Lauer et al., 2002) did not account for this issue, and this model would erroneously predict the CTB-receptor binding would occur even when there was no space to accommodate an additional CTB on the membrane. Moreover, it would be difficult to incorporate the surface configuration in a deterministic model. Hence, the kMC modelling approach is a more natural tool to simulate the CTB-receptor binding kinetics.

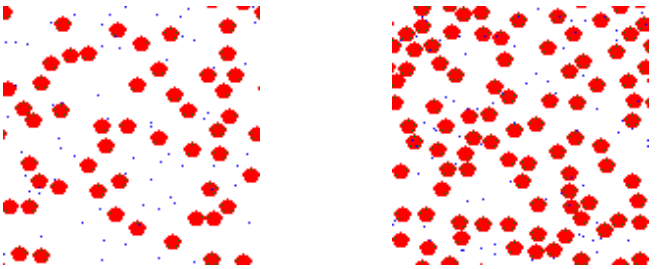


Figure 3. Schematic diagram of CTB surface coverages after 2000 seconds of binding with 1% (left) and 2% (right) GM1-containing membranes under 50 nM CTB. Pentagon-like shapes are membrane-bound CTB, and small dots are GM1 receptors

4. Conclusions

In this study, the kMC simulation framework was applied to model the binding kinetics between CTB and lipid membrane containing one type of ganglioside receptor by considering several important microscopic events as well as the surface configuration. Specifically, the proposed kMC model considered receptor migration, CTB attachment and detachment, and surface forward and backward reactions. The predicted CTB binding kinetics were qualitatively consistent with the experimental observations in the literature. In future, after rigorous parameter estimation of the proposed kMC model against available experimental measurements, it can be applied to study the CTB binding kinetics in the presence of multiple types of ganglioside receptors.

References

J. Erickson, B. Goldstein, D. Holowka, B. Baird, 1987, The effect of receptor density on the forward rate constant for binding of ligands to cell surface receptors. Biophyical Journal, 52, 657-662.

M. Fallahi-Sichani, J. J. Linderman, 2009, Lipid Raft-Mediated Regulation of G-Protein Coupled Receptor Signaling by Ligands which Influence Receptor Dimerization: A Computational Study, PLoS ONE, 4, 8, e6604.

P. Krishnan, A. Singla, C.-A. Lee, J. D. Weatherston, N. C. Worstell, and H.-J. Wu, 2017, Hetero-multivalent binding of cholera toxin subunit B with glycoplipid mixtures, Colloids and Surface B: Biointerfaces, 160, 281-288.

S. Lauer, B. Goldstein, R. L. Nolan, and J. P. Nolan, 2002, Analysis of cholera toxin – ganglioside interaction by flow cytometry, Biochemistry, 41, 1742-1751.

M. Nayhouse, J. S. Kwon, P. D. Christofides, and G. Orkoulas, 2013, Crystal shape modeling and control in protein crystal growth, Chemical Engineering Science, 87, 216-223.

D. J. Wilkinson, 2009, Stochastic modelling for quantiative description of heterogeneous biological systems, Nature Review Genetics, 10, 122-133.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50353-0

Sensor Network for Continuous Tablet Manufacturing

Sudarshan Ganesh[a], Mariana Moreno[a], Jianfeng Liu[a], Marcial Gonzalez[b], Zoltan Nagy[a], Gintaras Reklaitis[a,*]

[a]*School of Chemical Engineering, Purdue University, West Lafayette, IN 47907, USA*
[b]*School of Mechanical Engineering, Purdue University, West Lafayette, IN 47907, USA*
**reklaiti@purdue.edu*

Abstract

The progress in the mechanistic understanding of the unit operations and the availability of multiple sensor technologies enable the inline implementation of data reconciliation and gross error detection methods in continuous pharmaceutical manufacturing. In this work, we demonstrate the benefits of accurate real-time monitoring of the process state in a continuous tableting process, with case studies representative of common situations in pilot-plant or manufacturing line implementation.

Keywords: Pharmaceutical Tableting, Continuous Manufacturing, Process Analytical Technology, Sensor Network, Data Reconciliation.

1. Introduction

Developing robust process operations for continuous production and real-time release testing in pharmaceutical manufacturing is progressing in earnest with the encouragement from regulatory agencies and prominent companies. The Process Analytical Technology (PAT) guidance set the foundation for modernizing pharmaceutical manufacturing by proposing an integration of subject matter expertise in pharmaceutics, powder technology and real-time process management. Accurate real-time measurements of Critical Quality Attributes (CQAs) and Critical Process Parameters (CPPs) are crucial. Some of these variables can be measured directly in real-time, while others require soft sensors, mainly using spectroscopic methods for inline measurements. Ierapetritou et al. (2016) discuss the process and technical challenges for continuous tableting.

In this work, we describe our progress in developing a reliable monitoring system for continuous pharmaceutical tableting**.** We limit this discussion to the application of data reconciliation (DR) and gross error detection (GED) for accurate inline measurements of the process state by using data from individual sensors. The paper illustrates the application of DR and GED in a subsystem of a continuous tableting line which uses dry granulation.

2. Continuous Tableting Process Description

Loss-in-weight (LIW) feeders are used for feeding the active pharmaceutical ingredient, excipients, inert additives, etc. into the continuous blenders. Then the blend is compacted into tablets. Depending on material properties, the blends may require granulation, milling, etc. before tableting. The sensor network is comprised of built-in equipment sensors for CPPs and external sensors for CQAs. The communication framework for

connecting equipment and sensors to a distributed control system (DCS) is crucial to effective network utilization (Singh et al. 2015). Our pilot-plant (Fig. 1) includes an Alexanderwerk WP120 roller compaction unit, two Schenck Accurate AP-300 LIW feeders for metered feed of API (acetaminophen) and excipient (microcrystalline cellulose PH-102), and a Gericke GCM 250 continuous blender. The blend is granulated and blended with a lubricant (0.5 wt% Magnesium Stearate) or glidant (0.2% Silicon Dioxide) using a Schenck DP-4 LIW feeder in another Gericke blender. The final blend is compacted in a Natoli BLP-16 tablet press. As shown in Fig. 1, the sensor network consists of load cell measurements from all the LIW feeders, NIR sensors at locations 2, 3 and 5, microwave sensor (Gupta et al., 2015) at 3; an InnopharmaLabs Eyecon camera at 4, x-ray sensor (Ganesh et al., 2017a) at location 3 and a Sotax AT4 automatic tablet tester at location 6. Additional details of our system are available in Moreno et al. (2017).

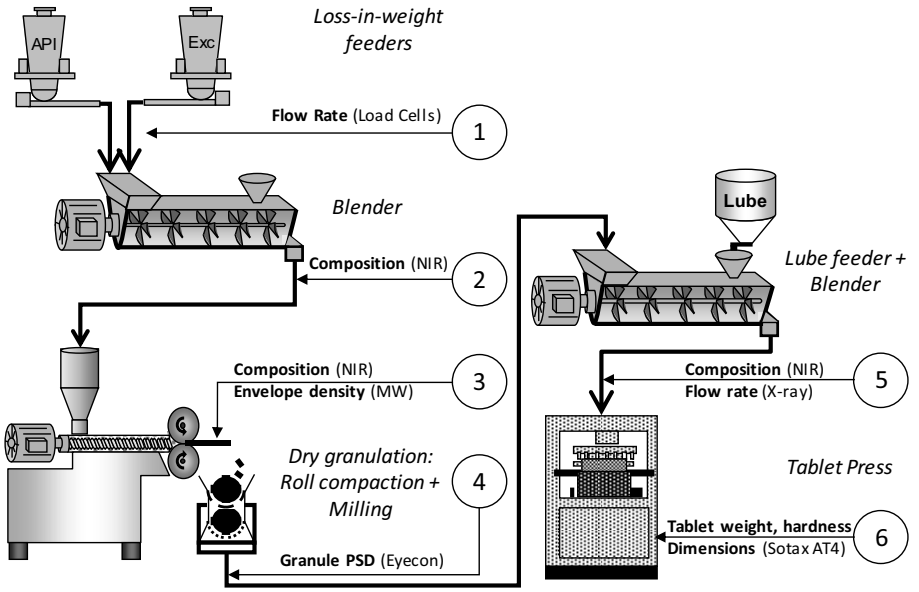


Figure 1. Schematic of pilot plant and available sensor network at Purdue University

3. Systematic approach to measurement accuracy

Real-time release testing in continuous tableting requires accurate and reliable inline measurements. However, there exist several operational challenges. Spectroscopic sensors, such as NIR and microwave, require data pre-processing and analysis before recording in the DCS. Calibration of such sensors is material and sensor location specific. Due to a lack of well-established communication protocols in the industry, tasks such as data acquisition, filtering and processing are performed in separate software. Software issues can result in possible communication failures, rendering the measurement unavailable for specific periods during the process. Handling particulate streams can be subject to frequent fouling of sensor interfaces, leading to biased measurements. Further, tablet properties such as hardness, weight and dimension are measured at set time intervals minutes apart by destructive testing of the samples drawn. Moreover, measurements are always subject to random errors arising from sources such as power supply fluctuations, network transmission delays, changes in ambient conditions, etc.

Data reconciliation (DR), gross error detection (GED) and sensor network design (SND) have been demonstrated to address such challenges for improved measurement accuracy (Narasimhan and Jordache, 1999). DR and GED require direct simultaneous measurement of a number of variables which is larger than the process degrees of freedom to permit the estimation of all variables in the sensor network. Multiple measurements of the same process variable can improve the reliability of that measured variable; however, it does not affect the observability of the unmeasured variables. Besides, sensors using the same technology can miss certain process features which can only be seen via measurements utilizing a palette of alternative technologies. Thus, though the blend uniformity is the primary concern for a continuous tableting process, integrating available mechanistic understanding with sufficient measurements to maintain network redundancy and observability of unmeasured variables is essential for reliable continuous operations.

4. Case Study

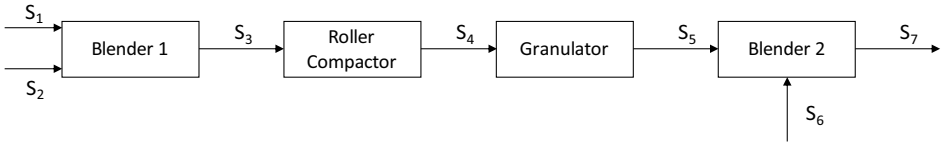


Figure 2. Block diagram for subsystem case study

In this study, the improvements in the accuracy of estimating the process state achieved using DR and GED for a subsystem of the continuous tableting line are demonstrated using three case studies representing situations commonly encountered in experimental implementation. We only consider the feeding, blending and granulating operations in the continuous manufacturing process, leading to the block diagram shown in Figure 2. There are seven material streams, and we assume there is no loss of material. The process variables are summarized in Table 1.

The measurement technologies, if applicable, are indicated in the 'Tool' column. The corresponding expected true values, and relative standard deviations ('RSD' column) of measurements obtained from experimental data under steady-state operations are as indicated. All the unit operations are represented using overall material balances and

Table 1. Process variables with measurements, NOC mean and standard deviations

	Variable	Tool	True Value (units)	RSD
F_1	API Flow	Load Cell	1.00 kg/h	2%
F_2	Excipient Flow	Load Cell	9.00 kg/h	2%
F_3	Blender1 Flow	-	-	-
x_3	Blender1 CU	NIR	10.00 wt%	6%
F_4	Ribbon Flow	-	-	-
S	Roll Gap	RC equipment	1.960×10^{-3} m	3%
ρ_R	Ribbon Density	NIR	0.963 kg/m^3	6%
x_4	Ribbon CU	NIR	10.00 wt%	6%
F_5	Granule Flow	-	-	-
x_5	Granule CU	-	-	-
F_6	Lubricant/Glidant Flow	LIW Load Cell	0.055 kg/h	15%
F_7	TP Inlet Flow	X-ray	10.00 kg/h	3%
x_7	TP Inlet CU	NIR	10.00 wt%	6%

component balance for the API across each node. Thus, there are eight material balance equations. Further, the flow rate from the roller compactor can be calculated using measured values for ribbon density and roll gap as shown in the equation (1). Also, mechanistic understanding of the roll compaction as shown in Equation (2) can be integrated in to the framework. Details of the model, notations and parametric values can be referred to Ganesh et al. (2017b). The roll speed (N_R) and hydraulic pressure (P_H) are treated as fixed parameters, given the low standard deviations of their measurements.

$$F_4 = \pi \rho_R N_R W D S \tag{1}$$

$$\rho_R = \rho_0 P_0^{1/K} \tag{2}$$

Where $1/2\ P_0(1 + \sin \delta_E) W D F\ =\ P_H * A$

and $F = \int_{\theta=0}^{\theta=\alpha} \left[1 + \frac{1-f_0}{f_0}\left(\frac{\theta}{\alpha}\right)^n\right]^K \left[\frac{S/D}{(1+S/D-\cos\theta)\cos\theta}\right]^K \cos\theta\, d\theta$

The set of equations has 3 degrees of freedom for the process, which means the process has a minimum requirement of four gross error free measurements for data reconciliation. Table 1 shows the availability of nine inline measurements, leading to six degrees of redundancy (DoR) in the system for the GED tests. However, given the frequency of fouling, communication errors, requirement of feeder refilling, etc., consistent availability of gross error free measurements from these sensors is challenging. Moreover, LIW tuning parameters are material bulk density specific, which can vary for the raw materials. Also, calibration models for all the spectroscopic sensors are material, location and probe position specific. Hence, the redundancy is crucial to achieving robustness of this system.

The model based DR and GED problems are solved in MATLAB using the approach reported in Moreno et al. (2017). The GED involves solving the global test (GT) and measurement test (MT). MT requires linearization of constraints. The bilinear component balances and Equation (1) are linearized using Taylor series expansion. Equation (2) is highly nonlinear and is linearized as a linear function of roll gap at the corresponding operating conditions. A total of 1000 random normal measurements using the mean and standard deviations given in Table 1 are simulated. Average values of these noisy measurements are compared with the corresponding reconciled values to demonstrate the improvement in measurement accuracy and are presented in Table 2, with units same as those given in Table 1.

Table 2. Measured variable estimates after data reconciliation and gross error detection

Var	NOC		Case 1		Case 2		Case 3	
	Mean	RSD	Mean	RSD	Mean	RSD	Mean	RSD
F_1	1.000	1.79	1.001	1.92	1.003	1.87	1.001	1.80
F_2	8.997	1.57	9.005	1.55	8.997	1.95	9.004	1.55
x_3	10.002	2.04	10.003	2.13	10.032	2.38	10.004	2.03
S	1.929	1.68	1.931	1.66	1.929	2.06	1.930	1.66
ρ_R	0.955	0.24	0.955	0.24	0.955	0.30	0.955	0.24
x_4	10.002	2.04	10.003	2.13	10.032	2.38	10.004	2.03
F_6	0.055	15.20	0.055	15.04	0.055	15.14	0.055	14.77
F_7	10.051	1.43	10.060	1.41	10.056	1.76	10.060	1.42
x_7	9.948	2.04	9.949	2.12	9.977	2.38	9.949	2.03

4.1 Normal operating conditions (NOC)

At NOC, all the nine measurements are expected to be active with true values and RSD as shown in Table 1. DR and GED for the system of equations result in improved accuracy for most process variables, particularly for the CU at all locations, as shown in Table 2 ('NOC' column), while ensuring material balance closure across all units and the process.

4.2 Case 1: Biased measurement from the NIR sensor

Suppose the NIR sensor reports the CU for x_3 as 15 wt%, while the measurements from the rest of the sensors are normal. In such situation, a bias resulting from fouling might be expected. By performing GED and DR, this faulty measurement can be rectified. The reconciled estimate for the CU at Location 3 is closer to the NOC conditions. The reconciled measurement for a shorter duration of gross errors in x_3 is shown in Fig.3 (left). The results of the case study are in Table 2 ('Case 1' column).

4.3 Case 2: Biased measurement from X-ray sensor

During steady-state plant operations, the X-ray sensor at Location 3 (F_7) reports a reduced flow rate of 8 kg/h. With a granulation process, material losses or ratholing in hoppers are always a possibility. However, if both cases are dismissed, the measurement must simply be biased. For the X-ray sensor, faulty measurements arising from communication failure or calibration error resulting from corrective actions for mitigating fouling are possible. In this case, the GED and DR use the existing redundancy in the sensor network to confirm that the process state remains within bounds and thus, a shutdown of the process is avoided. The reconciled measurement for a shorter duration of gross errors in F_7 is shown in Fig. 3 (right) and the results are presented in Table 2 ('Case 2' column).

4.4 Case 3: Unavailability of ribbon density sensor

Calibration of sensors for providing inline measurements of physical properties such as ribbon density is challenging. Moreover, installation of such sensors within the compactor at the ribbon location may require modifications to the equipment and plant setup. The decision to avoid these complexities will result in the unavailability of a direct density measurement. In this case study, we assume ρ_R is an unmeasured variable, reducing the DoR to five. However, an estimate of the ribbon density is important for downstream tableting. DR can accommodate sensor unavailability for estimation of CQAs using process models and measurement redundancy. DR can provide accurate state estimates as shown in Table 2 ('Case 3' column), confirming that the process is within bounds.

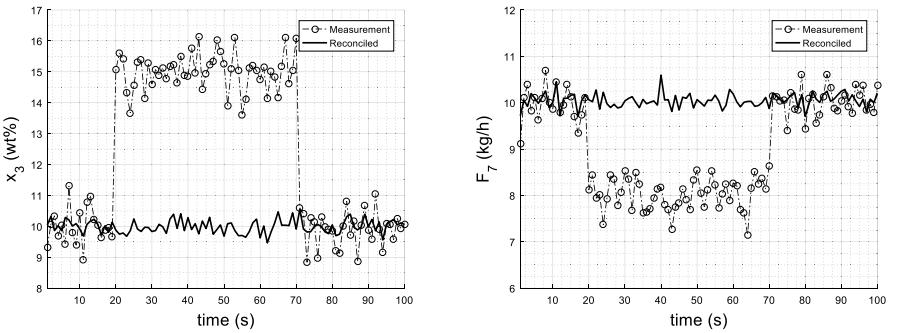


Figure 3. The reconciled values for measurements with gross errors. Bias in x_3 (Case 1, Left figure) and F_7 (Case 2, Right figure)

5. Discussion

For pharmaceutical processes, a measurement of the blend CU inline after every unit operation is essential and is typically achieved using NIR. The measurement RSD for the sensor depends on the material, location, spectra averaging, smoothing, etc. The challenge is to maintain the RSD within regulatory limits. Newer technologies for direct measurement of process variables with simplified calibration are very much desired. However, maintenance action to correct fouling of a sensor could result in bias for the measured variable which is beyond acceptable limits. It would be infeasible to pause a continuous process frequently to check for such errors. Also, ensuring material balance closure is crucial to maintain robust and profitable operations.

6. Conclusions

The case studies in this paper illustrate some of the practical challenges in the implementation of robust inline sensing in continuous pharmaceutical tableting. Specifically, we demonstrate the application of DR and GED to the system of unit operations and sensors using mechanistic models and material balance to obtain reliable and accurate estimates of the process state. Expanding this framework to add the models of other unit operations and measurement technologies is a part of our current research. Moreover, in ongoing work, we show that the application of a DR framework facilitates effective implementation of process control systems (Su et al., 2017). Robust process operations using systematic sensing and control systems are essential for reliable function of a material-tracking framework, leading to real-time release testing in pharmaceutical manufacturing.

Acknowledgements

The authors acknowledge the support of all collaborators, FDA Grant U01FD005535 and NSF-ERC Grant EEC-0540855 for this work.

References

Ganesh, S., Troscinski, R., Schmall, N., Lim, J., Nagy, Z., Reklaitis, G. *J. Pharm. Sci.* (2017a), Vol 106 (12): pp 3591-3603.

Ganesh, S., et al. *AIChE Annual Meeting* (2017b), Talk 438e

Gupta ,A., Austin, J., Davis, S., Harris, M., Reklaitis, G. *J. Pharm. Sci.* (2015) Vol 104 (5): pp 1787-1794

Ierapetritou, M., Muzzio, F. and Reklaitis, G. *AIChE J.* (2016), Vol 62: pp 1846–1862.

Moreno, M., Liu, J., Su, Q., Leach, C., Giridhar, A., Yazdanpanah, N., O'Connor, T., Nagy, Z., Reklaitis, G. *(In preparation)*.

Narasimhan, S., Jordache, C. *Data Reconciliation and Gross Error Detection* (1999), pp 1-31.

Singh, R., Muzzio, F., Ierapetritou, M., Ramachnadran, R. Comput. Aided Chem. Eng. (2016), Vol 38: pp 1473-1478.

Su, Q., Moreno, M., Giridhar, A., Reklaitis, G., Nagy, Z. *J. Pharm. Innov.* (2017) Vol (12): pp 337-346

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50354-2

Evaluating Hospital Performance Using Process Systems Engineering Tools

Jangwon Lee[a], Q. Peter He[a*]

[a]*Department of Chemical Engineering, Auburn University, Auburn AL 36849, USA*
qhe@auburn.edu

Abstract

Healthcare spending in the United States accounts for about 17% of US GDP, making the US the highest healthcare expenditure per capita in the world. There are areas for improvement in the healthcare system and hospital care is one of them, which represents the single largest national health expenditure by the type of services. Because hospital operations and manufacturing processes share some similarities at systems level, we believe that some of the theories and techniques developed for manufacturing processes are also applicable to hospital operations. In this work, we examine whether the so called focused factory theory (i.e., factories that concentrate on narrow range of services or operations produce better products at low costs) is applicable to hospital operations. Specifically, we examine whether the hospitals that are specialized in certain diseases achieve better results in terms of costs and patient outcomes using a national healthcare cost and utilization project (HCUP) dataset. Data analytics tools including multiple linear regression (MLR), principal component regression (PCR), Fisher discriminant analysis (FDA), ordinal regression and logistic regression are used to investigate effects of hospital specialization on the total charge and death of patient during hospitalization.

Keywords: Focused factory theory, hospital performance, specialization,

1. Introduction

The healthcare spending in the United States accounts for about 17% of US GDP, making the US the highest healthcare expenditure per capita in the world. However, the United States has not seen an increase in life expectancy to match its huge outlay on healthcare. Therefore, there are areas for improvement in the healthcare system and hospital care is one of them. This is because hospital care represents the single largest national health expenditure by the type of services, accounting for approximately 31% of total healthcare costs (Hartman et al., 2009). One of many potential routes is through healthcare specialization based on the so called focused factory theory, which states that factories that concentrate on narrow range of services or operations produce better products at low costs. There have been many debates over the subject. Consequently, several studies have been conducted to investigate the effect of specialization on the hospital costs and patient outcomes. However, most of these studies were based on few individual cases, or a limited number of hospitals (Eastaugh, 1992; Herzlinger, 2004). There were studies that have investigated factors associated with high-quality/low-cost hospital performance using national databases such as healthcare cost and utilization project (HCUP). However, specialization was not considered in those studies (Jiang et al., 2006). Because hospital operations and manufacturing processes share some similarities at systems level, we believe that some of the theories and techniques developed for manufacturing processes are also applicable to hospital operations. In this

work, we examine whether the hospitals that are specialized in certain diseases achieve better results in terms of costs and patient outcomes using a national healthcare cost and utilization project (HCUP) dataset. Specifically, multiple linear regression (MLR), principal component regression (PCR), Fisher discriminant analysis (FDA), ordinal and logistic regressions are used to investigate the effects of hospital specialization on the total charge and death of patient during hospitalization. It is worth noting that the effects of specialization on other measures of hospital performance such as efficiencies quantified by data envelopment analysis (DEA)(Ozcan, 2014) require further investigation. The rest of the paper is organized as the follows. Section 2 introduces the dataset used in this study. Section 3 introduces the index used in quantifying hospital specialization and analysis methods to investigate the effect of hospital specialization on costs and patient outcomes. Section 4 presents the results and Section 5 draws the conclusions.

2. The National Inpatient Sample (NIS) dataset

The Healthcare Cost and Utilization Project (HCUP) is a family of healthcare databases and related software tools and products developed through a Federal-State-Industry partnership and sponsored by the Agency for Healthcare Research and Quality (AHRQ). HCUP databases bring together the data collection efforts of State data organizations, hospital associations, private data organizations, and the Federal government to create a national information resource of encounter-level healthcare data (HCUP Partners). HCUP includes the largest collection of longitudinal hospital care data in the United States, with all-payer, encounter-level information beginning in 1988. These databases enable research on a broad range of health policy issues, including cost and quality of health services, medical practice patterns, access to healthcare programs, and outcomes of treatments at the national, State, and local market levels (Steiner et al., 2001).

Table 1. The five most expensive DRG's with at least 5,000 cases in the 2012 NIS-HCUP dataset

DRG	No. cases	Mean TOTCHG	Median TOTCHG	Std. Deviation	Short DRG Description
233	5,506	$207,098	$169,938	$139,804	CORONARY BYPASS W CARDIAC CATH W MCC
25	6,536	$133,707	$101,442	$116,105	CRANIOTOMY & ENDOVASCULAR INTRACRANIAL PROCEDURES W MCC
238	6,847	$99,567	$86,949	$54,958	MAJOR CARDIOVASC PROCEDURES W/O MCC
246	12,337	$97,447	$81,086	$66,060	PERC CARDIOVASC PROC W MCC
470	67,626	$54,210	$47,836	$27,551	MAJOR JOINT REPLACEMENT

MCC: multiple chronic conditions

The National Inpatient Sample (NIS) used in this study is part of HCUP, which covers all patients, including individuals covered by Medicare, Medicaid, or private insurance, as well as those who are uninsured. Overall, NIS covers more than 95 percent of the U.S. population and includes more than 94 percent of discharges from U.S. community hospitals. The 2102 NIS dataset used in this study includes 7,296,968 cases (*i.e.*, inpatient stays). Each case has 481 variables, which include de-identified (*i.e.*, all personally identifiable information has been removed to protect individual identities and privacy) patient data such as age, gender, ethnicity, etc.; disease, diagnosis and procedure data such as disease severity, length of stay (LOS), diagnosis related group

(DRG), cost, total charge (TOTCHG), whether the patient died during the hospitalization, etc. In addition, each case is also linked to hospital related data/information such as bedsize, location, ownership, and teaching status of the hospital, which enable us to evaluate the performance of different hospitals. In this study, we focus on the five most expensive DRG's based on the national average TOTCHG with at least 5,000 cases. Based on this criterion, the five selected DRG's are listed in Table 1.

3. Methods

3.1. Specialization quantification

Farley and Hogan (1990) have proposed an index of specialization based on Information Theory. It has been shown that the index provided intuitively reasonable results in characterizing patterns of specialization across hospitals. In Farley and Hogan (1990), let Φ_i represent the baseline proportion of cases in DRG category i, and let p_{ih} denote proportion of cases in the h^{th} hospital observed in DRG category i. The Information Theory index of specialization (ITI) for hospital h collapses information about differences between the Φ_i's and p_{ih}'s as follows (Farley and Hogan, 1990):

$$ITI_h = \sum_{i=1}^{I}\{p_{ih} \cdot ln(p_{ih}/\Phi_i)\} \quad (1)$$

In our study, because we only focus on the five DRG's listed in Table 1, the specialization index (I_s) of h^{th} hospital is modified based on Eqn. (1) as follows:

$$I_{sh} = \sum_{i=1}^{5}\{w_{ih} \cdot ln(p_{ih}/\Phi_i)\} \quad (2)$$

where $w_{ih} = p_{ih}/\sum_{i=1}^{5} p_{ih}$ such that $\sum_{i=1}^{5} w_{ih} = 1$.

3.2. Multiple linear regression (MLR)

Multiple linear regression (MLR) is performed to investigate the effect of each independent variable on total charge (TOTCHG). Because TOTCHG is strongly right skewed, in order to satisfy the assumption of linearity and constant variance of residual for MLR, the transformed TOTCHG is used. It was found that the 4^{th} root of TOTCHG provide satisfactory performance, which is therefore used as the dependent variable. The independent variables include the following continuous variables: specialization index (I_s), wage index, length of stay (LOS), age, number of chronical conditions, number of diagnosis, number of procedures; as well as the following categorical variables: bed size (small/medium/large), location (rural/urban), teaching status (non-teaching/teaching), ownership (public/private not-for-profit/private for profit), region (northeast/Midwest /south/west), payment (private/Medicare/Medicaid/ others), DRG type (470/233/24/238/ 246), race (white/black/Hispanic/others) and sex (male/female). Dummy variables were introduced for categorical variables and the first variable in each category was used as the reference. The MLR in R was used for the analysis.

3.3. Principal Component Regression (PCR) with mixed variables

The same dependent and independent variables used in MLR were used in principal component regression (PCR). The standard principal component analysis (PCA) cannot be directly applied for this case because of the categorical variables. Therefore, the R package of PCAmixdata was used, which handles a mixture of qualitative and quantitative variables. Because of the large number of independent variables and potential multicollinearity among them, we hope that PCR could address the potential multicollinearity while reducing the impact of noise through dimensionality reduction.

3.4. Principal Component Regression (PCR) with continuous variables only

We also performed standard PCR using the following continuous variables only: specialization index (I_s), wage index, length of stay (LOS), age, number of chronical conditions, number of diagnosis, number of procedures, and total discharge (to represent hospital bed size). The dependent variable is still the 4th root of TOTCHG.

3.5. Fisher discriminant analysis (FDA)

In this analysis, the total charges (TOTCHG's) of all cases were divided into two groups: the lowest TOTCHG group (*i.e.*, the bottom 25%) and the highest TOTCHG group (*i.e.*, the top 25%). Fisher discriminant analysis (FDA) is performed to find which variables highly contribute to the separation of these two groups, and whether specialization index is one of them. The same independent variables used in MLR were used in FDA, which was performed in SPSS.

3.6. Ordinal regression

To reduce the noise in TOTCHG, we categorize TOTCHG for all cases into four groups and use ordinal regression to investigate the effect of specialization. In this study, TOTCHG is divided into four groups based on the quartiles of cases (i.e., dividing all cases into four TOTCHG groups containing an approximately equal number of cases in each group based on TOTCHG). The same independent variables used in MLR were used in the ordinal regression, which was performed in SPSS.

3.7. Logistic regression

Logistic regression is performed to investigate the effect of each independent variable on the death of patients during hospitalization (i.e., one of the patient outcome measures). The dependent variable is the death of patients during the hospitalization (alive/death). For the independent variables, beside the ones used in MLR, the following two variables were included: severity of illness (minor/moderate/major/extreme), risk of mortality (minor/moderate/major/extreme).

3.8. Adjusted R^2

In this study, the adjusted R^2 is used to compare the explanatory power of regression models. The adjusted R^2 is a modified version of R^2 that has been adjusted for the number of predictors in the model. The adjusted R^2, which is always lower than R^2, increases only if the new independent variable improves the model more than by chance. Similarly, the adjusted McFadden's R^2 was used for logistic regression.

4. Results

To visually inspect the effect of specialization on the total charge, in Figure 1 (a), we plot TOTCHG vs. I_s for all cases of the five expensive DRG's listed in Table 1. Because TOTCHG is strongly right skewed, it can be seen from Figure 1 (a) that some of the positive residual of a linear fit line are rather large. In contrast, we also plot $(TOTCHG)^{1/4}$ vs. I_s as shown in Figure 1 (b), which shows more balanced residual of a linear fitting between $(TOTCHG)^{1/4}$ and I_s, supporting our choice of $(TOTCHG)^{1/4}$ as the dependent variable in all regression analyses rather than the original TOTCHG. We have also performed PCA on the combination of dependent variable $(TOTCHG)^{1/4}$ and all independent variables excluding I_s. In the score plot of the first and the third principal components, we use different colors for different ranges of I_s as shown in Figure 2. Although the scores were generated without the inclusion of I_s, they are

clustered for the same range of I_s, indicating the underlying correlation between I_s and other variables.

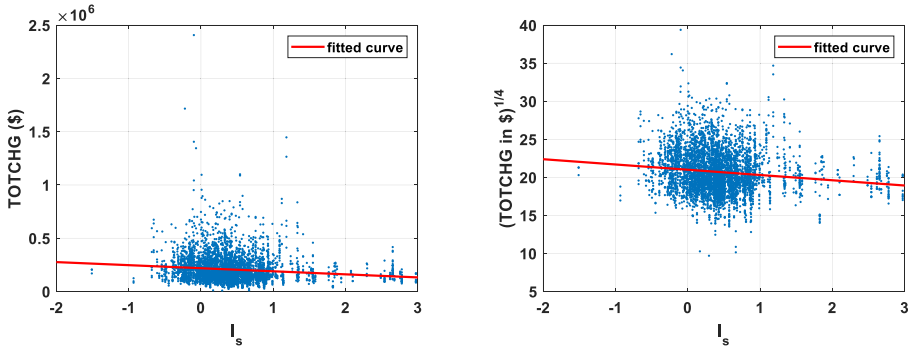


Figure 1. (left) Scatter plot of TOTCHG vs. I_s (dots) with linear fitting (solid line); (right) Scatter plot of $(TOTCHG)^{1/4}$ vs. I_s (dots) with linear fitting (solid line).

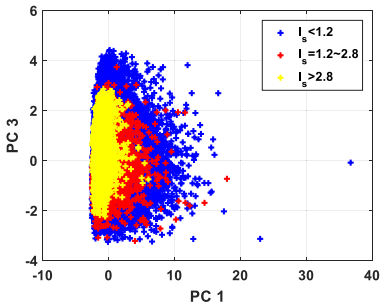


Figure 2. Score plot of the 1st and 3rd principal components of PCA on all variables other than I_s

Next, different regression and discriminant analysis methods discussed in Section 3 are used to model the relationship between $(TOTCHG)^{1/4}$ and various independent variables and the results are listed in Table 2. The second column of Table 2 shows that I_s has a negative effect on $(TOTCHG)^{1/4}$ in all analyses, indicating that higher hospital specialization leads to lower TOTCHG and the p-values of I_s coefficients listed in the last column indicate that the effects are statistically significant in all regression models.

Table 2. The results of regression and discriminant analyses on TOTCHG

Method	**Effect of I_s on $(TOTCHG)^{1/4}$**	**Adj. R^2**	**p-value of I_s coefficient**
MLR	(-)	0.5871	<0.0001
PCR (mixed variables)	(-)	0.5151	<0.0001
PCR (continuous variables)	(-)	0.4009	<0.0001
FDA	(-)	-	-
Ordinal regression	(-)	-	<0.0001

Finally we perform logistic regression to investigate the effect of various factors, including I_s, on the death of patients during hospitalization as one of the patient outcome measures. In the original NIS data, the proportion of patient is severely imbalanced between alive (98.0%) and death (2.0%) cases during hospitalization. One

approach to improve the classification performance for imbalanced data is to either oversample the minority class and/or under-sample the majority class. In this study, balanced dataset is obtained using synthetic minority over-sampling technique (SMOTE) (Chawla et al., 2002). As a result, the fraction of dead patients increases to 17.65% from 2% and prediction ability of model is improved based on the adjusted McFadden's R^2 as shown in Table 3. Table 3 also shows that I_s has a negative effect on the death of patient during hospitalization, indicating higher hospital specialization leads to fewer patient deaths during hospitalization, or better patient outcomes.

Table 3. The logistic regression on the death of patient during hospitalization

Data	The effect of I_s on death during hospitalization	Adj. McFadden's R^2	p-value of I_s coefficient
Orig. imbalanced data	(-)	0.403	<0.0001
Balanced data	(-)	0.575	<0.0001

5. Conclusions

We investigate the effect of hospital specialization on healthcare cost (total charge) and patient outcome (death during hospitalization). This is the first time that a nation-wide dataset of over seven million cases is used for such a study. Various regression methods (including MLR, PCR and ordinal regression) and a linear discriminant analysis method (*i.e.*, FDA) are used to study the effect of hospital specialization on the total charge and the findings are consistent among different methods: higher hospital specialization leads to lower total charge. In addition, logistic regression is used to investigate the effect of I_s on the death of patients during hospitalization as one of the patient outcome measures. Logistic regressions are performed on the original imbalanced data and the balanced data through synthetic minority over-sampling technique (SMOTE). Both results indicate that higher hospital specialization leads to fewer patient deaths during hospitalization or better patient outcomes. This work provides evidence for supporting hospital specialization to achieve low-cost and high-quality hospital care.

References

Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. 2002. SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**:321–357.

Eastaugh SR. 1992. Hospital specialization and cost efficiency: benefits of trimming product lines. *J. Healthc. Manag.* **37**.

Farley DE, Hogan C. 1990. Case-mix specialization in the market for hospital services. *Health Serv. Res.* **25**.

Hartman M, Martin A, McDonnell P, Catlin A, Team NHEA. 2009. National health spending in 2007: slower drug spending contributes to lowest rate of overall growth since 1998. *Health Aff.* **28**:246–261.

Herzlinger RE. 2004. Specialization and Its Discontents. *Circulation* **109**:2376–2378.

Jiang HJ, Friedman B, Begun JW. 2006. Factors Associated with High-Ouality/Low-Cost Hospital Performance. *J. Health care Financ.* **32**:39–52.

Ozcan YA. 2014. Performance measurement using data envelopment analysis (DEA). *Health Care Benchmarking Perform. Eval.* Springer.

Steiner C, Elixhauser A, Schnaier J. 2001. The healthcare cost and utilization project: an overview. *Eff. Clin. Pract. ECP* **5**:143–151.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50355-4

Characterization of Electroencephalography of strokes based on Time-frequency analysis

Ning Wang[a,&], Yuehua Pu[b,&], Ying Li[a], Jia Tian[c], Dacheng Liu[b], Zhe Zhang[b], Qingrong Shen[a], Liping Liu[b,*], Wei Sun[a,*]

[a]*Beijing Key Lab of Membrane Science and Technology, College of Chemical Engineering, Beijing University of Chemical Technology, Beijing 100029, China*
[b]*Beijing Tiantan Hospital of Capital Medical University, Beijing 10050, China*
[c]*The Second Hospital of Hebei Medical University, Shijiazhuang, 050000, Hebei, China*
[&]*Both authors contributed equally to this work and should be considered as co-first authors*
sunwei@mail.buct.edu.cn

Abstract

Stroke is a medical condition, under which poor blood flow to the brain results in cell death. Electroencephalography, EEG, is an electrophysiological monitoring method to record electrical activity of the brain. If EEG signal is processed properly, the feature of each individual stroke case can be analysed to indicate location and severity, which could be used for the diagnosis and continuous monitoring of the treatment process in clinic. In this work, EEG signals from 11 patients with ischemic stroke and 1 healthy person are analysed by Short-time Fourier Transformation (STFT). All patients have malignant middle cerebral artery infarction or received reperfusion therapy. EEG data from ten of them are investigated, which indicates that the delta frequency band energy percentage (DFBEP) in the ipsilateral hemisphere is always higher than the contralateral side, while the alpha frequency band energy percentage (AFBEP) and alpha/delta (AD) are the opposite. Based on this feature, one EEG signal collected along 15 hours from the 11th patient is further analysed by STFT. Compared to the medical record of this patient, the energy of each frequency content changes with the progress of treatment and recovery of patient accordingly, which could be used to monitor the patient condition during the treatment.

Keywords: STFT, Continuous monitoring, Non-invasive diagnosis, Feature extraction

1. Introduction

Stroke is an acute cerebrovascular disease caused by a sudden rupture of the blood vessels in the brain or the inability of blood to flow into the brain due to the occlusion of blood vessels, including ischemic and hemorrhagic stroke.

There are many detection methods available currently, including transcranial Doppler (TCD), magnetic resonance imaging (MRI), computed tomography (CT) and digital subtraction angiography (DSA), etc. Most of them are structural examination, snapshot, and cannot be performed bedside.

As a wide applied and non-invasive detection method, Electroencephalography (EEG) is the general reflection of the electrophysiological activity of brain cells on the surface of

the cerebral cortex or scalp at real time. EEG contains a large number of physiological and functional information.

Delta/alpha power ratio (DAR) was proposed for discriminating between acute ischaemic Stroke (IS) patients and control groups, as DAR from IS patients is significantly higher than that from control groups (Finnigan et al., 2016). Their work reveals that DAR can be an index to recognize the IS symptom, which inspires people that the lesion location could be identified also if sufficient number of montages along scalp are placed, while their analysis is based on Fourier Transform(FFT), and no frequency information is extracted along time. The short-time Fourier transform (STFT) and wavelet transform (WT) was applied to analyse EEG signals for the identification of a child having an epileptic seizure (Kıymık et al., 2015). Both STFT and WT can provide time-frequency information, and STFT is more straightforward to give frequency information along time. Putten and Tavy (2004) introduced brain symmetry index (BSI) for the characterize EEG signals from both hemispheres. However, BSI is calculated within the frequency range up to 25HZ, which fails to differentiate the EEG contribution corresponding to various cerebral conditions, as it is reported that dominant frequency will change for different brain activities.

The aim of this work is to extract more corresponding information from EEG signal regarding to the stroke symptom, not only the frequency content and also the time-frequency content, so as to facilitate the diagnosis and treatment in clinic practice. The rest of this paper is organized as follows: The data is described in Section 2; Methods are shown in Section 3; Result and discussion are included in the Section 4; Conclusion is given in Section 5.

2. Data

2.1. Data acquisition

EEG data acquisition were performed in Beijing Tiantan Hospital of Capital Medical University. We included patients with malignant middle cerebral artery infarction, Tomographic evidence of early hypodensity involving cortico-subcortical topography > 50% of the MCA territory, National Institutes of Health Stroke Scale (NIHSS) > 15, and age ≥ 18 years, or patients who received reperfusion therapy. Patients who received decompressive craniotomy therapy before EEG monitoring were excluded. The EEG examination was performed within 72 hours of symptom onset and more than 30min. The placement of electrodes were based on the international 10-20 system. The electrocardiogram signals collected synchronously. Acquisition frequency was 500Hz. The montage of EEG data is Fp1-C3, C3-O1, Fp1-T3, T3-O1, Fp2-C4, C4-O2, Fp2-T4, T4-O2. All the data saved as ASCII format.

2.2. Details of data

Eleven patients (Patient 1-11) who have infarcts in the hemisphere and a healthy person are analyzed in this paper. Ten short pieces of data taken from the healthy person with total monitoring time of 35.38min are used as references. The details are shown in Table 1.

Table 1. Details of the data

Patient Number	Time	Lesion Location	Reference episode	Time
1	5.33min	Left	1	69s
3	9min	Left	2	45s
4	5min	Left	3	56s
6	5.28min	Left	4	119s
7	3.15min	Left	5	85s
9	1.15min	Left	6	146s
2	5min	Right	7	84s
5	2.77min	Right	8	161s
8	1.2min	Right	9	81s
10	8.63min	Right	10	45s
11	15h	Left		

3. Methodology

3.1. Fourier transform and Short-time Fourier transform

The Fourier transform (FT) decomposes a function of time into the frequencies. Discrete Fourier transform (DFT) is actual used in the processing, can be conducted by Eq.(1).

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-i \cdot 2\pi k n / N} \tag{1}$$

where $X[k]$ is the frequency domain , $x[n]$ is the time domain and N is the number of the sampling points.However, the Fourier analysis is performed based on the whole signal across the time domain, which only gives the frequency component for the overall signal, and fails to reveal the local information of a signal in time domain. For this purpose, STFT is proposed as in Eq.(2), which involves representing the signal in the time-frequency domain to show the changes in its frequency components with respect to time.

$$STFT\{x[n]\}(m, \omega) \equiv X(m, \omega) = \sum_{-\infty}^{\infty} x[n] \omega[n - m] e^{-jwn} \tag{2}$$

In Eq.(2), $\omega[n]$ is the window function which is a Hamming window and $x[n]$ is the signal to be transformed.

3.2. Frequency band energy

Frequency changes over time can be obtained through STFT, and then the energy is calculated through Eq.(3).

$$\varepsilon = |X(m, \omega)|^2 \tag{3}$$

The frequency band in clinic practice is defined in Table 2.

Table 2. The frequency band adopted in clinic practice

Frequency band	Name
0.5-4Hz	Delta
4-8Hz	Theta
8-13Hz	Alpha
13-30Hz	Beta
30-35Hz[#]	Gamma

[#]30-35Hz: The frequency range which is considered is 0-35Hz in this paper.

The definition of frequency band refers to the guideline (Deuschl and Eisen, 1999). The energy of each frequency band is calculated by integration in Eq.(4).

$$E_i = \int_{f_1}^{f_2} \varepsilon \, df \tag{4}$$

where ε is the transform results of window whose length is N. Each frequency band energy percentage is a more important indicator as defined in Eq.(5).

$$E_{norm} = \frac{E_i}{\sum E_i} \tag{5}$$

where E_{norm} is the normalized frequency band energy.In this paper, delta frequency band energy percentage (DFBEP), alpha frequency band energy percentage (AFBEP) and alpha/delta (AD) are used as main indicators. AD is the ratio of the AFBEP to the DFBEP.

3.3. Processing of data

The procedure of data process is shown in Figure 1.

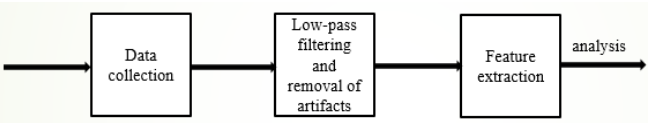


Figure 1. Procedures for data processing.

Frequency range of the low-pass filtering is 0-35Hz. STFT is used in feature extraction in every step of analysis. The energy percentage of each frequency band is calculated.

4. Results and Analysis

4.1. STFT analysis result relative to the lesion location

A comparison of Patient 1 and the reference is shown in Figure 2. Columns with odd numbers correspond to left cerebral hemisphere, otherwise to the right side.

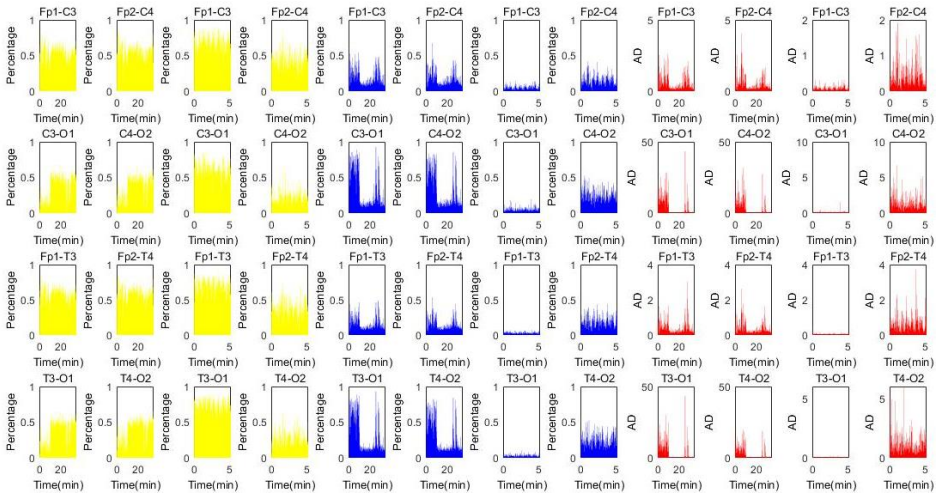


Figure 2. Frequency band energy percentage of the data of patient 1 and the reference. The first and the second columns are DFBEP, the fifth and the sixth columns are AFBEP, and the ninth and the tenth columns are AD. The two columns following each of them are the reference.

As it can be seen from Figure 2, the DFBEP, the AFBEP and the AD of the healthy person are generally symmetrical between left and right cerebral hemisphere. However, the DFBEP of left cerebral hemisphere is higher than that of the right side, while the AFBEP and the AD are opposite to it with respect to the patient 1. AD is a better indicator than AFBEP when compared the frequency band energy percentage of left and right cerebral hemisphere.

Then the average values of the DFBEP, the AFBEP and the AD of left and right cerebral hemisphere of patient 1-10 are calculated respectively. The results are shown in Table 3. Figure 3 depicting symmetry between left and right cerebral hemisphere are drawn according to the data of Table 3

Table 3. The average values of the DFBEP, the AFBEP and the AD of left and right cerebral hemisphere

No	DFBEP		AFBEP		AD		Re	DFBEP		AFBEP		AD	
	L	R	L	R	L	R		L	R	L	R	L	R
1	0.70	0.35	0.05	0.22	0.08	1.01	1	0.27	0.28	0.43	0.35	4.93	4.13
2	0.85	0.71	0.03	0.09	0.05	0.16	2	0.50	0.48	0.11	0.11	0.36	0.33
3	0.62	0.37	0.07	0.18	0.14	0.75	3	0.54	0.54	0.10	0.10	0.23	0.25
4	0.62	0.36	0.07	0.24	0.15	1.03	4	0.43	0.44	0.19	0.17	0.85	0.64
5	0.58	0.24	0.09	0.33	0.20	2.36	5	0.37	0.39	0.32	0.26	4.21	2.87
6	0.75	0.56	0.04	0.11	0.06	0.26	6	0.50	0.50	0.12	0.13	0.31	0.32
7	0.61	0.79	0.09	0.04	0.23	0.05	7	0.49	0.51	0.10	0.10	0.27	0.27
8	0.25	0.58	0.36	0.13	2.80	0.32	8	0.50	0.50	0.11	0.12	0.30	0.32
9	0.59	0.80	0.15	0.04	0.40	0.06	9	0.38	0.39	0.29	0.25	2.66	2.03
10	0.57	0.67	0.13	0.07	0.31	0.12	10	0.49	0.53	0.13	0.11	0.33	0.27

No: Patient number; L(R): Left (Right) cerebral hemisphere; Re: Refence episode.

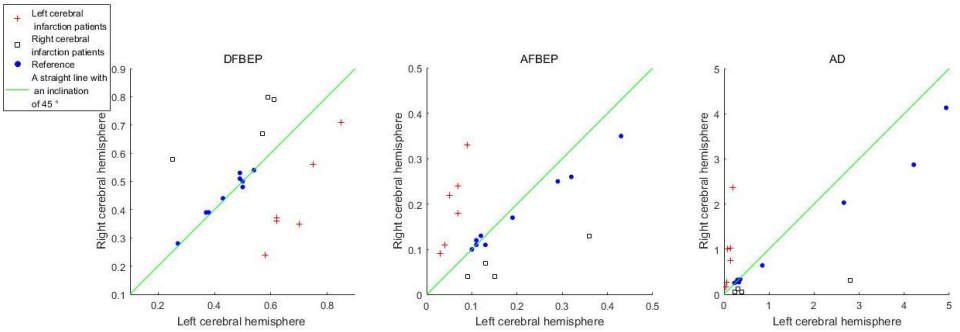


Figure 3. The symmetry between left and right cerebral hemisphere

As it can be seen from Figure 3, the average of the DFBEP, the AFBEP and the AD of the healthy person are generally symmetrical between left and right cerebral hemisphere. However, with respect to the left cerebral infarction patients, the average of the DFBEP of left cerebral hemisphere is higher than that of the right side, while average of the AFBEP and the AD are the other way around. The same results are observed with respect to the right cerebral infarction patients.

To summarize the results obtained from first ten patients, it can be seen that DFBEP is relatively higher, AFBEP and AD are relatively lower when the lesion exists in corresponding cerebral area, thus, DFBEP, AFBEP and AD can be used as indicators to monitor patient condition.

4.2. Monitoring the patient condition

Patient 11 suffered the left-sided cerebral infarction, and was getting better during clinic treatment. His EEG data related to the left cerebral area within the first 15 hours was analysed. DFBEP, AFBEP and AD are illustrated in Figure 4.

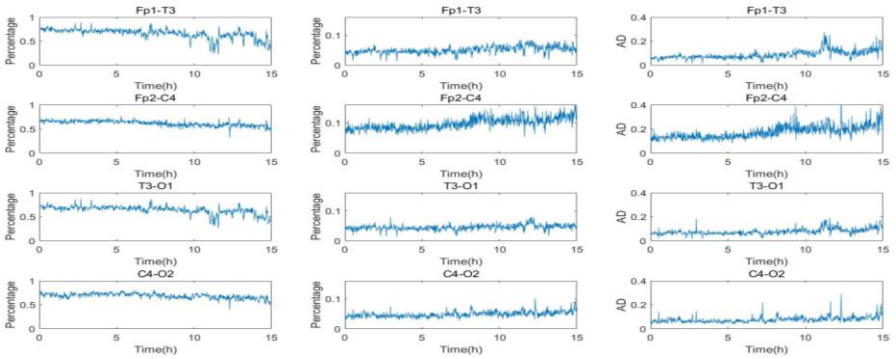


Figure 4. Frequency band energy percentage of patient 11 in the left cerebral hemisphere. The first to the third columns are DFBEP, AFBEP and AD in turn.

It can be seen that that the DFBEP is decreasing, the AFBEP and the AD are increasing along time, indicating that the patient's condition is getting better, which is highly consistent with the clinic record.

5. **Conclusions**

In this paper, EEG signal is analysed by STFT, from which the frequency band energy is obtained. The lesion location of patients with ischemic stroke can be identified by comparing DFBEP, AFBEP and AD among each pair of montages. The preliminary results indicate that DFBEP, AFBEP and AD can be used as indicators for stroke evolution and patient recovery. With further analysis on EEGs it can be employed for more sophisticate clinical applications, e.g. early diagnosis, treatment decision, outcome prediction, and clinical monitoring.

References

G. Deuschl, A. Eisen, 1999, Recommendations for the practice of clinical neurophysiology: guidelines of the international federation of clinical neurophysiology. Electroencephalography & Clinical Neurophysiology Supplement, 52, 1

M. J. van Putten, D. L. Tavy, 2004, Continuous quantitative eeg monitoring in hemispheric stroke patients using the brain symmetry index, Stroke, 35(11), 2489

M Kemal Kıymık, İnan Güler, A Dizibüyük, Mehmet Akın, 2015, Comparison of STFT and wavelet transform methods in determining epileptic seizure activity in EEG signals for real-time application, Computers in Biology & Medicine, 35(7), 603

S Finnigan, A Wong, S Read, 2016, Defining abnormal slow EEG activity in acute ischaemic stroke: delta/alpha ratio as an optimal QEEG index, Clinical Neurophysiology Official Journal of the International Federation of Clinical Neurophysiology, 127(2), 1452

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50356-6

Integrated Process and Ionic Liquid Design by Combining Flowsheet Simulation with Quantum-Chemical Solvent Screening

Simon Bechtel[a,*], Zhen Song[a], Teng Zhou[a], Tanja Vidakovic-Koch[a], Kai Sundmacher[a,b]

[a]*Max Planck Institute for Dynamics of Complex Technical Systems, Department Process Systems Engineering, Sandtorstr.1, D-39106 Magdeburg, Germany*
[b]*Otto-von-Guericke- University Magdeburg, Department Process Systems Engineering, Universitätsplatz 2, D-39106 Magdeburg, Germany*
bechtel@mpi-magdeburg.mpg.de

Abstract

Motivated by the need for fast design approaches for sustainable chemical processes, this work focuses on the combination of a quantum chemistry based screening method for the selection of suitable ionic liquids (ILs) as separating agents in chemical gas phase processes and the analysis of the overall process with the help of process simulations. By close integration of the systematic screening procedure for suitable ionic liquids based on constraints defined by the overall process and the flowsheet simulation, the proposed methodology allows for an environmentally and economically feasible process development. The approach was applied to an industrial example process, namely the electrochemical oxidation of hydrogen chloride to chlorine, which is a crucial part of the polyurethane and polycarbonate production. The most efficient industrially realized electrochemical process was developed by Bayer and UHDENORA. It uses aqueous hydrochloric acid as feed stock. As an alternative, our group has shown recently that the electrolysis of gaseous hydrogen chloride is much more efficient and leads to exergy savings of 36% on the reactor level alone. In the present work the separation of unreacted hydrogen chloride from chlorine in the reactor effluent stream by means of selective HCl absorption in ILs is investigated in order to determine whether the exergetic advantage of the novel gas phase reactor can be maintained on the overall process level when the necessary separation steps are included. For this purpose, an extensive screening procedure of ILs available in the COSMOtherm IL database is carried out in order to find the most suitable candidates. Hereby, key properties like the chemical stability in the harsh chlorine environment as well as important physical and thermodynamic properties of the ILs are tailored to the demand of this specific gas phase process. Subsequently, the overall process is simulated with Aspen Plus and a detailed exergy analysis of all process units is performed. Additionally, the results are compared to simulations of the Bayer UHDENORA process. The analysis shows that not only the electrochemical reactor but also the separation steps in the proposed gas phase process display a significantly lower exergy demand than the industrial Bayer UHDENORA liquid phase variant. This demonstrates that the proposed integrated design methodology efficiently supports the development of sustainable, yet profitable chemical processes.

Keywords: Ionic Liquids, Process Design, Process Simulation, Process Integration

1. Introduction

In order to reduce the impact of chemical processes on the environment and society while still preserving economic competitiveness, a more sustainable process design in the chemical industry is imperative. This can be understood as the development of a process that leads to lower values of a set of targeted performance criteria, like costs or environmental impact, in comparison to a reference process (Segovia-Hernandez and Bonilla-Petriciolet 2016). In most chemical processes a complete conversion of the reactant is not achievable so that the feasibility of purifying the reaction product is oftentimes decisive for the overall energetic efficiency and profitability of the chemical process. In the last two decades ionic liquids have gained significant attention in various fields, such as separating agents for the absorption of gases like CO_2 (Zhang 2012). Due to their green and designer properties, they are often regarded as environmental friendly, less hazardous and sustainable alternatives to organic solvents (Rogers and Seddon 2012). However, owing to the vast variety of possible cation-anion combinations, a systematic screening approach for finding the best suited candidate for each specific chemical process is absolutely essential (Song 2017). In addition, the screening process must consider constraints imposed by the particular chemical process, such as the inertness and stability of ILs under the given conditions, the overall process performance in terms of energy consumption and many more. Therefore, it is critical to combine a quantum chemical based IL screening method with a process flowsheet simulation in order to find a the best suited IL for the product purification and at the same optimal process conditions that lead to a more energy efficient and sustainable overall process. In the present work, this combined approach is exemplified by an important process in the chemical industry, the electrolysis of hydrogen chloride to chlorine. Up to now, the most efficient industrially employed process for the HCl electrolysis is the Bayer UHDENORA process, in which aqueous hydrochloric acid is fed into an electrochemical reactor, the emerging chlorine is separated and the depleted hydrochloric acid stream is recycled. More detailed information about this process can be found in our recent publication (Bechtel 2017). This process has the major disadvantage of a significant energy demand on the reactor level as well as in the separation step. Recently, our group has investigated an electrochemical reactor in which gaseous hydrogen chloride is oxidized to chlorine so that the reactor outlet stream consists only of anhydrous hydrogen chloride and chlorine. This reactor concept leads to significant exergy saving on the reactor level (Bechtel 2017). However, in this previous work we have not investigated the use and the advantages of ILs as separating agents to obtain the pure chlorine.

Based on the methodology described above, the goal of this contribution is to determine, whether the use of ionic liquids as a solvent allows for an energy efficient purification of the product and a more economically feasible and sustainable overall process in comparison to the Bayer UHDENORA reference method. For this reason a suitable IL is determined based on the screening procedure explained in chapter 2 and a flow sheeting simulation of the overall process is carried out. Contrary to many other absorption processes reported in literature (Farahipour 2016), the gas mixture consisting of chlorine and hydrogen chloride is very reactive which strongly influences and restricts the selection process of possible ionic liquids. Finally, based on the process simulation a detailed exergy analysis was carried out and the overall exergy demand of the novel proposed IL process was compared to the reference process.

2. Screening procedure and set up of the Process simulation

Although the computer-aided molecular design (CAMD) method has been extended to guide the selection of IL solvents for different separation problems, the state-of-the-art IL screening process mainly concentrates on the estimation of thermodynamic and physical properties due to the unavailability of predictive models for other IL properties (Austin 2016). Considering the highly reactive nature of HCl and Cl_2, a modified IL screening background is proposed for this particular case as depicted in Figure 1.

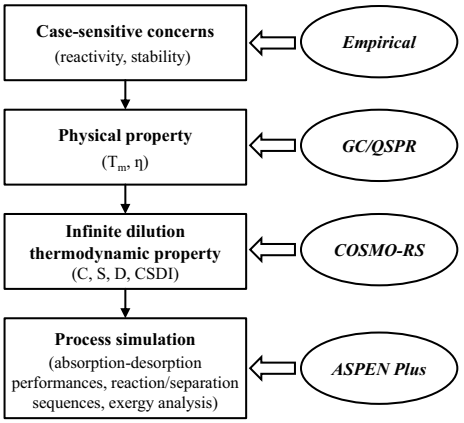


Figure 1, Process design methodology consisting of a three step IL screening procedure and the flowsheet simulation.

In the first step, ILs with functional groups that are prone to react with either chlorine, hydrogen chloride or traces of water that can be present in the gas stream due to crossover in the membrane of the electrochemical reactor are excluded. Therefore, amines and ammonium cations have to be excluded in order to avoid the formation of chloramines. Furthermore, no aromatic systems or allyl groups are allowed, since they are prone to electrophilic substitution with chlorine. This is especially critical, if the aromatic system or the allyl group is substituted with functional groups that increase the electron density in the π-bond. Additionally, no carbonyl, carboxyl, imine or amide group should be present since they can be halogenated in alpha position (Hauptmann 1991). Moreover, metal salts that are prone to hydrolysis in the presence of water like tetrachloroaluminiate have to be discarded in this first screening step (Huheey 2006). In addition cyanides or related groups are also prone to react with halogens (Weyl 2005). Finally, highly fluorinated molecules were disregarded for environmental and economic reasons.

In the second step the melting temperatures T_m of the remaining ILs are estimated using the GC method developed by Lazzús *et al.* (2012) and the viscosities η are calculated with the GC feed-forward artificial neural network (FFANN) method developed by Paduszyń and Domańska (2014) by group contribution (GC) or quantitative structure-property relationship (QSPR) methods. Only ILs with a melting temperature below 298.15 K and a viscosity smaller than 150 mPa s are considered for the following screening step in which the infinite dilution thermodynamic properties are investigated. The higher the viscosity the more difficult and energy intensive is the circulation of the ionic liquid within the process and its units, which is the reason for the imposed maximum value. Here, the absorption capacity *C*, the absorption selectivity towards

hydrogen chloride *S*, the desorption ability *D* and the overall *CSDI* value are calculated with the help of COSMO-RS (Farahipour 2016, Zhao 2017):

$$C = \frac{1}{H_{HCl}^{25}} \times \frac{MW_{HCl}}{MW_{IL}} \tag{1}$$

$$S = \frac{H_{Cl2}^{25}}{H_{HCl}^{25}} \times \frac{MW_{HCl}}{MW_{Cl2}} \tag{2}$$

$$D = H_{HCl}^{55} / H_{HCl}^{25} \tag{3}$$

$$CSDI = C \times S \times D = \frac{H_{HCl}^{25} \times H_{HCl}^{55}}{\left(H_{HCl}^{25}\right)^3} \times \frac{\left(MW_{HCl}\right)}{MW_{Cl2} \times MW_{IL}} \tag{4}$$

In these equations H_i stands for the Henry constant of the substance *i* in the IL and MW_i for the molar mass of the according substance. The most promising candidate of this screening procedure was then chosen for a process simulation in Aspen Plus 8.8. For the process simulation, the IL was treated as a user-defined component for which the molecular structure as well as the boiling point, the density, the critical properties and the acentric factor was manually entered. These properties were determined using the fragment contribution-corresponding states method proposed by Huang *et al.* (Huang 2013). The temperature dependence of the Henry constants of chlorine and hydrogen chloride in the IL was fitted with the method of least squares based on values determined with COSMO RS in a temperature range from 298.15 K – 483.15 K with an interval of 20 K. As a property method COSMO-SAC was chosen and the screening charge density profile as well as the molecular volume were calculated with COSMO-RS and implemented into the flowsheet simulation. The corresponding properties for chlorine and HCl are already implemented in the Aspen Plus databases. In order to allow for a comparison with the simulation of the Bayer UHDENORA process in (Bechtel 2017) the conversion in the electrochemical reactor was set to be 80 % and the remaining 20 % of not converted hydrogen chloride was absorbed in water forming 37 wt% hydrochloric acid, which leaves the process as an additional product. For more information please see (Bechtel 2017). The process flowsheet is shown in Figure 2.

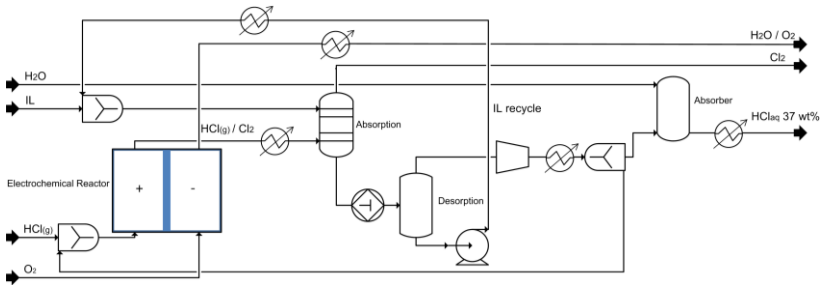


Figure 2, Flow chart of the IL process. The hydrogen chloride in the anode outlet stream in the reactor is absorbed in the IL leaving pure chlorine in the gas stream leaving the absorption unit. Afterwards, the absorbed hydrogen chloride is desorbed at a pressure of 50 mbar and a temperature of 483.15 K. Subsequently the IL is compressed to atmospheric pressure and recycled into the process again. The hydrogen chloride leaving the desorption unit is compressed in 4 steps with intermediate cooling towards atmospheric pressure. Subsequently, the hydrogen chloride can be recycled back to the electrochemical reactor or absorbed in water forming concentrated hydrochloric acid.

It should be mentioned that the proposed multi-step IL screening essentially follows the main idea of the state-of-the-art of IL design method (Song 2017, Farahipour 2016). However, it is interesting to see that only very few ILs out of a large number of possible candidates remain after the first three steps, which makes it possible to perform the detailed process simulation and evaluation in the fourth step. Due to this, and the complex considerations regarding the reactivity, such an approach appears to be very purposeful in this specific case and comparable gas phase processes. Finally, the exergy analysis of the overall process was carried out based on the equations described in (Bechtel 2017).

3. Results

As a result of the second screening step, only 39 ILs remained for which the C, S, D and the CSDI values were calculated and depicted in Figure 3:

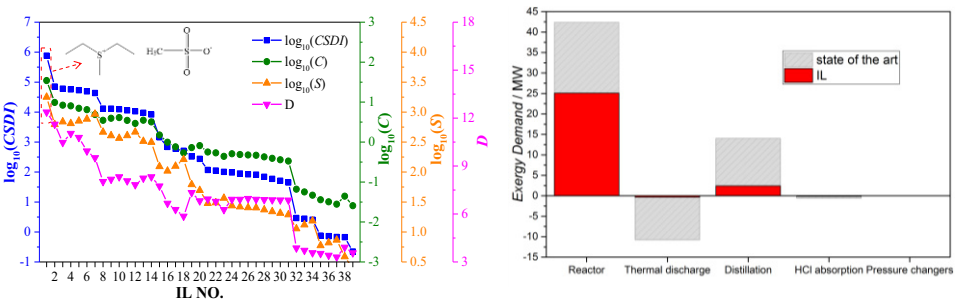


Figure 3, (a) Infinite dilution thermodynamic properties calculated according to Eq. (1-4) for the remaining 39 ILs. (b) Contributions of the different unit operations to the exergy demand of the state-of-the-art process (Bechtel 2017) (shaded) and the novel IL process (filled).

The IL with the highest values in all four categories is diethyl-methylsulfonium methanesulfonate ([S221][MeSO3]). The chemical structure is included in Figure 3. Since it furthermore has a comparatively low heat capacity of only 310 J mol^{-1} K^{-1} at room temperature, it was chosen for the process simulation in Aspen Plus. It is worth noting that the heat capacity of the IL turned out to also significantly influence the exergetic efficiency of the process and should therefore be directly included into the screening process in future investigations. In Figure 3b the exergy demand of the IL-based process is compared with the state of the art process from Bayer UHDENORA. The overall exergy demand of the newly proposed process is 36 % lower compared to the Bayer UHDENORA state of the art process. Additionally, less apparatus are needed for the IL based process. Furthermore, the size of the apparatus can be reduced, since the large water containing recycle streams of the Bayer UHDENORA process (see Bechtel 2017) can be avoided. Finally, it can also be expected that the material costs will be lower in the IL process, since the hydrogen chloride only comes into contact with water at the very end of the process when parts of it are facultatively absorbed in water to generate concentrated hydrogen chloride as a byproduct. Contrary to that, in the Bayer UHDENORA process the hydrogen chloride is absorbed in water right at the beginning of the process which leads to the need of more corrosion resistant and therefore more expensive material throughout the whole process. In future work, a comparison of the capital and production cost between both processes should be carried out to also further quantify the financial profitability. Additionally, the VLE data used for the process simulation could be experimentally validated in future investigations.

4. Conclusions

In the present work, a combination of quantum chemical screening methods for ILs was combined with flowsheet simulations in order to develop an energy efficient and sustainable alternative to a state-of-the-art reference process. This methodology was tested on an industrially important example process, the electrochemical recycling of hydrogen chloride to chlorine. During the screening process, not only physical and thermodynamic properties of the ILs were taken into account but also restrictions arising from the specific process like the chemical stability of the IL. Based on a detailed exergy analysis, it was possible to show that the newly proposed process based on the above explained methodology leads to exergy savings of 36 % compared to the Bayer UHDENORA reference process. Furthermore, significant savings in material costs can be expected which leads to an overall more sustainable and profitable process.

References

J. G. Segovia-Hernandez, A. Bonilla-Petriciolet, 2016, Process Intensification in Chemical Engineering, Springer International Publishing Switzerland

X. Zhang, X. Zhang, H. Dong, Z. Zhao, S. Zhang, Y. Huang, 2012, Carbon capture with ionic liquids: overview and progress. Energy & Environmental Science, 5(5), 6668-6681

R. D. Rogers, K. R. Seddon, 2003, Ionic liquids--solvents of the future? Science, 302(5646), 792-793

Z. Song, T. Zhou, Z. Qi, K. Sundmacher, 2017, Systematic Method for Screening Ionic Liquids as Extraction Solvents Exemplified by an Extractive Desulfurization Process. ACS Sustainable Chemistry & Engineering, 5(4), 3382-3389

S. Bechtel, T. Vidakovic-Koch, K. Sundmacher, submitted 12.30.2017, Novel Process for the exergetically efficient recycling of chlorine by gas phase electrolysis of hydrogen chloride, Chemical Engineering Journal

R. Farahipour, A. Mehrkesh, A. T. Karunanithi, 2016, A systematic screening methodology towards exploration of ionic liquids for CO 2 capture processes. Chemical Engineering Science, 145, 126-132

S. Hauptmann, 1991, Reaktion und Mechanismus in der organischen Chemie, Teubner Publishing Stuttgart

J. E. Huheey, E. A. Keiter, R. L. Keiter, O. K. Medhi, 2006, Inorganic Chemistry: Principles of Structure and Reactivity, Pearson Education Publishing

T. Weyl, J. Houben, J. G. Knight, 2005, Science of Synthesis, Thieme Publisher

J. A. Lazzús, 2012, A group contribution method to predict the melting point of ionic liquids. Fluid Phase Equilibria, 313, 1-6

K. Paduszyński, U. Domańska, 2014. Viscosity of ionic liquids: an extensive database and a new group contribution model based on a feed-forward artificial neural network. Journal of chemical information and modeling, 54(5), 1311-1324

Y. Huang, H. Dong, X. Zhang, C. Li, S. Zhang, 2013, A new fragment contribution-corresponding states method for physicochemical properties prediction of ionic liquids. AIChE Journal, 59(4), 1348-1359

N. D. Austin, N. V. Sahinidis, D. W. Trahan, 2016, Computer-aided molecular design: an introduction and review of tools, applications, and solution techniques, Chemical Engineering Research and Design, 116, 2-26

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50357-8

Multiscale scheme for the optimal use of residues for the production of biogas

Manuel R. Taifouris, Mariano Martín[a*]

[a]*Departamento de Ingeniería Química. Universidad de Salamanca, Salamanca, 37008, Spain*

Abstract

This work evaluates the use of four different types of waste such as sludge, manure, municipal solid wastes (MSW) and lignocellulosic residues for the production of biogas and fertilizer using a feedback multiscale analysis. Waste yield to both products is determined and the digestion kinetics is computed using a parameter estimation approach with experimental data. Hydrolysis is the limiting stage. Next, a techno-economic evaluation of a processing facility is carried out. Finally, a supply network model is developed to evaluate the optimal use of the residues over a region in Spain. The network is formulated as a mixed integer linear programming problem selecting the type of residue, the number and size of the digesters and the location of the plants across 59 shires of a particular region in Spain for the available budget. MSW is the selected waste due to its wide availability and large yield to fertilizer. Lignocellulosic residues are the second best option. Only intrashire transport is suggested. The shires selected for its processing are rather scattered as a function of the resources available and the already installed facilities. However, the amount of residues available can provide more than three times the region demand for natural gas, but the investment required adds up to more than 4.4 B€.

Keywords: Waste, Anaerobic digestion, Biogas, Power, Multiscale analysis.

1. Introduction

Developed societies generate large amounts of waste. In the European Union, 5 t of waste are produced per person and year (ec.europa.eu). Not only the volume but also the composition becomes a challenge. Processing the waste requires large and costly facilities but it is required to avoid dangerous effects on the environment. One single technology that is capable of both, processing and generating added value out of waste, is anaerobic digestion. It produces biogas, a mixture of methane and CO_2, and digestate, that can be used as fertilizer. This kind of facilities can be considered as part of the so-called circular economy aiming at the reuse of waste to bring value again to residues extending their lifespan (Lieder and Rashid, 2016). However, the evaluation of the potential of the use of waste requires a multiscale evaluation scheme, from the digestor analysis, kinetics and yield (Weinrich and Nelles, 2015), to the supply chain (Santibañez-Aguilar, 2013), including the techno-economic analysis of the process.

2. Methodology

The design and operation of waste to power systems based on anaerobic digestion have to be evaluated in a multiscale framework, from the process to the supply chain. Due to

the strong link between the various scales, a retrofit analysis is required. High level decisions are made at macroscopic scale, but detail information on the lower levels, process and unit performance and economics, are essential for informed decisions. Thus, we first evaluate the operation of the plant for each of the residues, namely, sludge, lignocellulosic residues, MSW and manure. Detail kinetics for each one is required because the residence time at the reactor is a direct function of the residues physical structure. Therefore, for the same reactor size, typically standardized, the yield is different as a function of the kinetics. Next, a techno-economic evaluation of the production is carried out to determine the investment and production costs. The processing plants allocation depends on the residue availability and its yield to valuable products. A supply chain study is developed to evaluate the use of waste and the size and number of facilities to be installed, considering residue availability per region and type, presence of other transformation plants and transportation costs. A two-level supply chain is formulated considering residues production and transformation layers. Economic or sustainability metrics can be added as objective function.

2.1. Process description

The process consists of the preparation of the feed to the digester to operate at 37ºC and with up to 10% by weight in solids. The reaction is endothermic; therefore an external heating system is in place. The yield depends on the kinetics. We assume a residence time corresponding with 95% of the maximum methane production per residue. Biogas is cleaned from NH_3, using a scrubber, H_2S, using an iron based fixed bed, and CO_2, followed by a pressure swing adsorption (PSA) system. The digestate is filtered. The cake is stored to be sold, and the water, after treatment, reused. The number of standard digesters depends on the processed waste. The economic evaluation is carried out using the factorial method (Sinnot, 1999) based on ‘equipment sizing and costing, Figure 1.

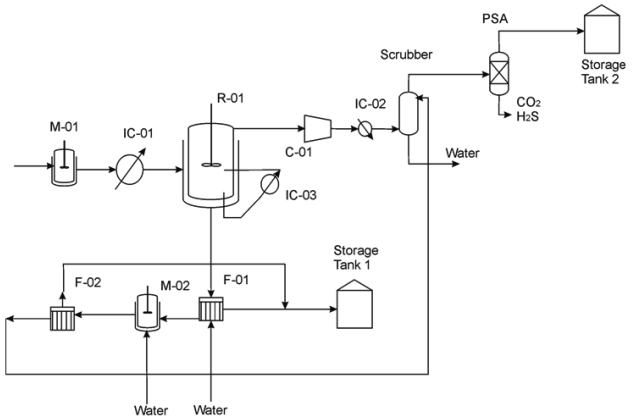


Figure 1.- Scheme of a one digestor process for waste to power.

2.2. Reactor Kinetics

Anaerobic digestion consists of four stages, hydrolysis, acetogenesis, metanogenesis. Detailed models allow simulation but they are too complex to evaluate the effect of waste type on the kinetics. We consider the fact that, for the residues used and the operating conditions, the limiting stage is the hydrolysis. Therefore, we use a hydrolytic kinetics that assumes first order with respect to lipids, carbohydrates and protein based on the work by Weinrich and Nelles (2015) to fit the experimental results in the

literature for four different waste types: manure, lignocellulosic, sludge and MSW. To come up with the model, eq. (1), we consider an average composition for the lipids, carbohydrates and protein within the biomass. However, the use of an average composition results in the fact that a correction parameter, α, is needed to match experimental data. The model requires three input values, X_{CH0}, X_{Pr0}, X_{Li0}. However, only the biodegradable part of the residue will produce biogas (Montañés et al., 2015). This value is experimental for each waste according to its structure. Thus, biodegradability is averaged from different sources for each type of waste to come up with the values used in our mode. Thus, the biodegradability for manure is 46% (Kalamaras and Kotsopoulos, 2014), for MSW 50% (Rao and Singh, 2004), for sludge 55% (Park et al., 2016) and for lignocellulosic residues 74% (Bouallagui et al., 2004).

$$\begin{aligned}
\frac{dX_{Ch}}{dt} &= -k_{Ch}\cdot X_{Ch} + 0.18\cdot k_{DEC}\cdot X_{BAT}\\
\frac{dX_{\Pr}}{dt} &= -k_{Pr}\cdot X_{Pr} + 0.82\cdot k_{DEC}\cdot X_{BAT}\\
\frac{dX_{Li}}{dt} &= -k_{Li}\cdot X_{Li}\\
\frac{dX_{BAT}}{dt} &= (1+\alpha)\cdot(0.1509\cdot k_{Ch}\cdot X_{Ch} + 0.1241\cdot k_{Pr}\cdot X_{Pr} + 0.1857 k_{Li}\cdot X_{Li})\\
\frac{dS_{CH_4}}{dt} &= (1+\alpha)\cdot(0.2433\cdot k_{Ch}\cdot X_{Ch} + 0.2383\cdot k_{Pr}\cdot X_{Pr} + 0.6588 k_{Li}\cdot X_{Li})\\
\frac{dS_{CO_2}}{dt} &= (1+\alpha)\cdot(0.6675\cdot k_{Ch}\cdot X_{Ch} + 0.6822\cdot k_{Pr}\cdot X_{Pr} + 0.6644\cdot k_{Li}\cdot X_{Li})\\
\frac{dS_{H_2O}}{dt} &= (1+\alpha)\cdot(-0.039\cdot k_{CH}\cdot X_{ch} - 0.2047\cdot k_{Pr}\cdot X_{Pr} - 0.4810\cdot k_{Li}\cdot X_{Li})\\
\frac{dS_{H_2S}}{dt} &= (1+\alpha)\cdot(0.0012\cdot k_{Pr}\cdot X_{Pr})\\
\frac{dS_{NH_3}}{dt} &= (1+\alpha)\cdot(-0.0227\cdot k_{Ch}\cdot X_{Ch} + 0.1645\cdot k_{Pr}\cdot X_{Pr} - 0.0280\cdot k_{Li}\cdot X_{Li})
\end{aligned} \tag{1}$$

2.3. Supply chain

This work focuses its attention on a particular agricultural area in the South of Europe. We divide Castile and Leon into the official 59 shires. Each of the shires is characterized by its position, computed using the latitude and longitude of the centre, and the availability of the four waste types considered. (Medioambiente JCYL, 2016). The presence of waste processing facilities is considered by including a penalty that reduces the market of digestate and the profit obtained from it as well as the waste availability in the region and its souroundings. The objective function is given by eq. (2)

$$OF = \sum_i \sum_j TInc(i,j)) - \sum_i \sum_j TotalC(i,j)) \tag{2}$$

The total cost, eq. (3), (TotalC) involves transportation (TC), amortization (FC), raw materials (RM), labour (LB) and utilities (Util). The reference for all computations is the anaerobic digesters used per plant. Three standard sizes are considered, 1,000, 3,000 and 6,000 m^3. The maximum size recommended is based on the difficulties in mixing https://mediathek.fnr.de/. Any plant at a shire consists of a number of digesters of each size. For the transportation a two level supply chain is developed, one for the waste production another one for processing. Only the waste in the processing level will be used in the plants. Transportation takes place at the production layer and from there it is sent to the processing plants. Raw materials only consider water, since waste is obtained at zero cost. Utilities accounts for 15% of the total annual cost. Labour is computed as

per plant size, small, medium or large plant. Finally, amortization is computed using as reference the digester size, the number of them used per plant and its lifetime. A correction factor based on the contribution of the reactor to the total cost of the plant is computed from the detailed process design. Raw material availability is limited in regions where waste to energy plants are already available.

$$\mathrm{TotalC}(i,j) = \mathrm{FC}(i,j) + \mathrm{VC}(i,j) = \frac{\left(\mathrm{FC}(i,j) + \mathrm{TC}(i,j) + \mathrm{RMs}\ (i,j) + \mathrm{Lb}(i,j)\right)}{Util} \tag{3}$$

The benefit, eq. (4) (Tinc), from the operation comes from the production of biogas (Ibio) and digestate (Id). Each digester size generates a certain amount of biogas and digestate as a function of the waste processed and its size. The yield is computed from the reactor modelling stage. Only a fraction of the digestate can be sold in the market the rest generates no benefit.

$$TInc = \sum_i \sum_j (\mathrm{Ibio}(i,j) + \mathrm{Id}(i,j)) \tag{4}$$

The model is formualted as and MILP consisting of 16,500 continuous variables, 708 binary variables and 2,373 equations written in GAMS® and solved in a Windows 10, intel core i7 machine with CPLEX 12.6.

3. Results

3.1. Yields and reactor operation.

It turns out that k_{Ch}, k_{Pr} and k_{Li} are similar for each waste, k_i in Table 1, since there are no recording of the generation of CO_2 , biomass or H_2S that will allow full validation of the model. Next, we fix the residence time for the time that it takes to reach 95% of the maximum methane yield. Thus, the residence time is 15 days for sludge, 30 days for MSW, 24 days for manure and 32 days for lignocellulosic residues. Next, the yield per unit volume of the reactor and day is computed, see Table 2.

Table 1.- Kinetic parameters for various waste types (37 ºC)

Residue	$k_i(d^{-1})$	$X_{ch}(g_{Ch}/g_{vs})$	X_{Pr} (g_{Pr}/g_{vs})	X_{li} (g_{Li}/g_{vs})	Degradability (%)
Sludge	0.21	0.44	0.182	0.309	55
MSW	0.07	0.367	0.268	0.013	50
Manure	0.11	0.83	0.1	0.03	46
Ligno	0.1	0.672	0.125	0.05	74

Table 2.- Yield per residue type.

	CH_4 $(kg/m^3 \cdot d)$	CO_2 $(kg/m^3 \cdot d)$	H_2S $(kg/m^3 \cdot d)$	Biogas $(kg/m^3 \cdot d)$	Biomass $(kg/m^3 \cdot d)$	Digestate $(kg/m^3 \cdot d)$
Sludge	1.057	1.862	0.000653	2.919	0.437	58.721
MSW	0.241	0.650	0.000317	0.891	0136	71.182
Manure	0.359	0.940	0.000176	1.300	0.210	65.070
Ligno	0.502	1.258	0.000219	1.760	0.280	40.149

3.2. Techno-economic analysis of the process.

The use of MSW as reference is based on a previous run of the supply chain model to see which is the most promising residue. This is another feedback required within the multiscale analysis. We consider the flowsheet presented in Figure 1 to identify the units involved. For a facility of 1 single reactor and a production capacity of a reactor of

6,000 m^3, the total cost of the units adds up to 1.5 M€. The cost of IC-03 is included in the reactor cost. Based on the factorial method, factors of 3.4 and 1.45 for a plant dealing mainly with fluids (Sinnot, 1999) and considering that the working capital is 10% of the fixed capital we compute the investment cost of the facility. However, for the supply chain model we need to compute the ratio between digester cost and plant investment, γ. We assume that we can scale the system using a digester as reference unit. Thus, the ratio between the total investment cost and the reactor cost, γ turned out to be equal to 0.10.

3.3. Infrastructure for waste processing.

The regional government has a budget for infrastructure development that accounts for 80% of the total budget for the environmental department. Using a fraction of that budget, from 25% to 100% (88 M€ to 351 M€, JCYL, (2017)), we evaluate the network that can be operated, see Figure 2. Only intrashire transportation is recommended. Thus, sources and sinks coincide. We see that with the entire annual budget, we reach the exploitation of 8.95 % of the biogas potential in the residues. Even if there is a large difference in the fraction processed, the mass of lignocellulosic residue processed is the same as that of MSW when investing 351 M€. That means that 50% of the budget goes to MSW and when it is not economical to continue exploiting MSW, lignocellulosic residues are used due to its larger yield to biogas. No manure is recommended to be used, due to the scattered availability and low yield to digestate, see Table 3.

Table 3.- Waste processing capabilities as a function of the budget

Investment (M€)	88	175	263	351
% Total Waste processing	2.14	4.32	6.71	8.95
% waste processing	22% MSW	43% MSW	48% MSW 2.62% Ligno	70.83% MSW 2.29% Ligno
Biogas cost (€/MMBTU)	59	57	56	51

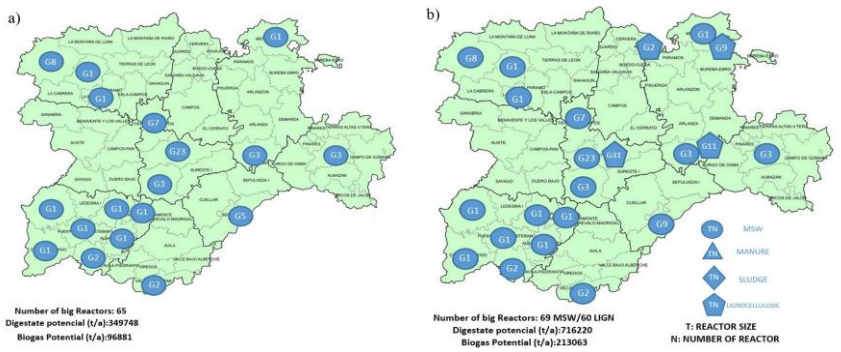


Figure 2.- Waste selection and plant size for a) 175M€ and b), 351 M€

However, this is a one year program. A multiperiod investment plan may allow increasing the processing capacity of residues and the production of biogas. The biogas potential out of the residues in the region reaches $2.09 \cdot 10^9$ m^3 (at 60% methane). The total budget to process this much residue adds up to 4,426 M€, requiring the installation of 1,389 big reactors, 44 medium size and 14 small size reactors. The methane potential from residues can provide 3 times the natural gas demand in Castile and Leon (http://www.cescyl.es) and it can provide 5.1% of the total natural gas consumption in Spain (http://www.cores.es). The production cost of the gas is 10 times higher than

current prices, and 5 times the cost of methane produced from solar and wind (Davis and Martín, 2014) see Table 3. Spain, does not have major sources of natural gas, therefore, the use of residues is an interesting alternative to evaluate.

4. Conclusions

A multiscale approach, from digester analysis to supply chain design, has been developed to evaluate the use of four different types of waste for the production of biogas and digestate. The reactor is characterized using a parameter estimation scheme assuming first order kinetics to evaluate the yield to biogas and digestate of the various waste fractions. Based on this, a techno-economic study of a plant is developed. The macroeconomic scale determines the potential and the location of the processing plants. A supply chain network is developed to evaluate the potential of residues, the type of residue, reactor size and the best location to install the processing facilities.

MSW is the selected waste due to its wide availability and large yield to fertilizer. Lignocellulosic residues are the second best option. Only intrashire transport is suggested. The shires selected for its processing are rather scattered as a function of the resources available and the already installed facilities. However, the amount of residues available can provide more than three times the region demand for natural gas, but the investment required adds up to more than 4.4 B€. Therefore, a multiyear investment program is to be put in place.

Ackowledgement

The authors would like to acknowledge Salamanca Research for optimization software licenses and MINECO grant DPI2015-67341-C2-1-R

5. References

H. Bouallagui, O., Haouari, Y., Touhami, R.B., Cheikh, L. Marouani, M., Hamdi, 2004. Effect of temperature on the performance of an anaerobic tubular reactor treating fruit and vegetable waste. Proces. Biochem. 39 (12), 2143-2148.

W. Davis, M. Martín (2014) Optimal year-round operation for methane production from CO_2 and Water using wind and/or Solar energy. J. Cleaner Prod. 80, 252-261

JCYL 2017 http://www.jcyl.es/web/jcyl/Hacienda/es/Plantilla100/1173961932821/_/_/_

S.D. Kalamaras, T.A. Kotsopoulos, 2014. Anaerobic co-digestion of cattle manure and alternative crops for the substitution of maize in South Europe. Biores. Technol. 172, 68-75

Lieder, M.;, Rashid, A., 2016. Towards circular economy implementation: A comprehensive review in context of manufacturing industry. J. Clean. Prod. 115, 36-51.

Medicomabiente JCYL, 2016b. http://www.medioambiente.jcyl.es

K.Y. Park, H.M. Jang, M-R Park, K. Lee, D. Kim, Y.M. Kim, 2016. Combination of different substrates to improve anaerobic digestion of sewage sludge in a wastewater treatment plant. Int. Biodet. Biodeg. 109, 73-77.

M.A. Rao, S.P. Singh, 2004. Bioenergy conversion studies of organic fraction of MSW: Kinetic studies and gas yield organic loading relationships for process optimisation. Biores Technol. 95, 173-185

J.E. Santibañez-Aguilar, J.M. Ponce-Ortega, J.B. González-Campos, M.M. Serna-González, M. M., M. El-Halwagi, 2013. Optimal planning for the sustainable utilization of municipal solid waste. Waste Manage., 33(12), 2607-2622.

Sinnot, R.K., 1999. Chemical Engineering Design. 4th Ed. Elsevier Oxford

S. Weinrich, M. Nelles, 2015. Critical comparison of different model structures for the applied simulation of the anaerobic digestion of agricultural energy crops Biores. Technol. 178, 306–312

http://www.cores.es/sites/default/files/archivos/estadisticas/est-gas-consumo-2016.pdf

http://www.cescyl.es/

https://mediathek.fnr.de/media/downloadable/files/samples/l/e/leitfadenbiogas-es-2013.pdf

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50358-X

Direct Air Capture of CO_2 in Enclosed Environments: Design under Uncertainty and Techno-Economic Analysis

Anshuman Sinha[a], Harshul Thakkar[b], Fateme Rezaei[b], Yoshiaki Kawajiri[a,c], Matthew J. Realff[a*]

[a]*Georgia Institute of Technology, 311 Ferst Drive, Atlanta, GA 30332, USA*
[b]*Missouri S&T, 1101 North State Street, Rolla, MO 65409, USA*
[c]*Nagoya University, Materials Process Engineering, Graduate School of Engineering, Nagoya 466-8603, Japan*
matthew.realff@chbe.gatech.edu

Abstract

CO_2 capture from enclosed environments such as commercial buildings can result in reduced ventilation, thereby leading to decreased energy loads on the HVAC systems. We propose a model which regulates the air quality inside the room by adsorption of CO_2 on monoliths coated with zeolite 13X, and water adsorption on packed beds with silica gel. We perform a modeling study and energy assessment by simulating a multi-component, multi-bed system which controls carbon dioxide level inside the room on a 24 hr basis. The results obtained indicate a 75% reduction in energy load with the CO_2 capture system as compared to the conventional ventilation system. We have identified areas of uncertainty in the model and compared non-intrusive polynomial chaos method with Monte Carlo simulations to quantify the model uncertainties. The results show improvement in computational efficiency with non-intrusive methods as compared to Monte Carlo simulations. We conclude that the CO_2 capture system can lead to improvement in building energy performance and application of polynomial chaos methods can result in reduction of computational time.

Keywords: CO_2 capture, Concentration Swing Adsorption, Zeolite 13X, Monoliths

1. Introduction

In enclosed environments such as commercial buildings, temperature and air quality are two important factors which are controlled to maintain a comfortable and workable environment. Air conditioning controls temperature, often at the expense of humidity, and particulate filters reduce suspended matter, but other quality factors, particularly CO_2 concentration, have remained unaddressed. There are health risks associated with high CO_2 concentration (>0.5%) (Fisk, 2000). Occupants may suffer headaches, nausea, fatigue and listlessness when exposed to high CO_2 concentration for prolonged durations. Hence, removal of CO_2 from enclosed environment is essential for improving human productivity.

There have been recent studies on minimizing energy inside buildings (Prívara et al., 2011; Salsbury et al., 2013). However most of them focus on controlling the thermal standards inside buildings (Aktacir et al., 2010; Videla and Lie, 2006). The conventional method of regulating the air quality is through ventilation (Seppänen et al., 1999).

However, such techniques are not energy efficient in case of large and enclosed space because air has to be conditioned to lower concentrations of air pollutants such as CO_2 or air-borne volatile organic compounds (VOCs). An alternate route to lower CO_2 level is to capture the CO_2 from indoor air and recirculate the CO_2 depleted air. The major challenge to implement such a strategy is the small driving force for CO_2 removal due to the low concentration. Recently a 3D monolith coated with zeolite was developed (Thakkar et al., 2016) which shows a promising pathway for dilute CO_2 capture.

In this study, we design and model a modified air conditioning system which regulates temperature as well as air quality, as quantified by CO_2 and H_2O concentrations, inside an enclosed environment. The model includes human occupancy of the room which is time dependent. Human respiration and perspiration dynamics have been taken into consideration while modeling the complex multi-bed system. 13X zeolite adsorbent is considered to capture the CO_2 adsorption dynamics within the 3D monoliths and silica gel packed beds are used to regulate the humidity within the system. Finally, the polynomial chaos expansion (PCE) method (Sanzida and Nagy, 2014; Najm, 2009) is used to quantify uncertainty in the model.

2. Method

2.1. System Design

In the current study, we propose a CO_2 capture system which removes CO_2 from the inside air through concentration swing adsorption cyclic process. The adsorbent used for this study is 3D monoliths coated with zeolite 13X which have shown a significantly large equilibrium capacity (~1 mol/kg) at the CO_2 concentration of 5000 ppm (Thakkar et al., 2016). However, zeolite 13X would show reduction in CO_2 uptake on exposure to humid conditions (Ahn and Lee, 2004) since water molecules compete for the active adsorption sites. To avoid this problem, the indoor air is passed through a dehumidification unit prior to passage through monoliths.

Figure 1 depicts the schematic of the air capture process. The CO_2 rich humid air is directed to the dehumidifier which consists of two packed beds silica gels. One of the packed beds adsorbs water molecules from the indoor air while the other packed bed is regenerated by heating it and re-humidifies the air inside the enclosed space. This ensures the humidity inside the room is within a certain interval (30 – 60 %) as suggested by A.S.H.R.A.E standards (A.S.H.R.A.E, 2004). The dry air rich in CO_2 then passes through the CO_2 capture system which consists of parallel honeycomb like 3D zeolite monoliths (Figure 1). Most recently, 3D printing technique has been applied to manufacture adsorbent monoliths with high adsorbent loading (ca. 90 wt%) and good mechanical strength with comparable adsorption performance to their powder analogues (Thakkar et al., 2016).

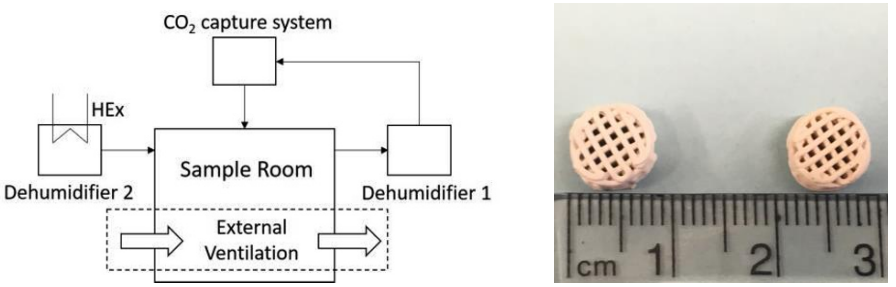


Figure 1: Schematic of the air capture process (left) and 3D zeolite 13X monolith (right)

Coupled heat and mass transfer partial algebraic differential equations are developed for monoliths (Sinha et al., 2017) and packed bed (Bollini et al., 2012). A sample room measuring 5×5×2.5 m^3 is considered to model the gas dynamics inside an enclosed space. It is assumed that the air inside the room is well mixed. It is also assumed that the room is occupied with five healthy adults during regular office hours (9 AM to 5 PM) and the respiration and perspiration dynamics is the same for each of these adults. The gas concentration dynamics inside the room is modelled by the following equation:

$$V_1 \frac{dC_{i,R}}{dt} = k_i + F_R C_{i,in} - F_R C_{i,R} + F_M C_{i,\mathrm{M}} - F_M C_{i,D} \tag{1}$$

where, V_1 is the room volume (m^3), k is the perspiration/respiration rate of healthy adult (mol/s), C is the concentration of i^{th} gas component (mol/m^3) and F is the gas flow rate (m^3/s). The subscripts R, M and D represent the room, monolith and dehumidifier respectively. The multi-bed model is implemented in gPROMS, which is a commercial dynamic process modeling and optimization software.

2.2. Uncertainty Quantification

The assumption regarding the fixed number of people inside the room with constant respiration and perspiration dynamics may not hold true in real life situations. Hence, it is essential is quantify the uncertainty caused due to such variable parameters. In this study, the PCE method is used to quantify model uncertainty in the number of room occupants. In particular, we have used non-intrusive polynomial chaos (NIPC) as described by Eq. 2.

$$\arg\min \sum_{i=0}^{n} \left(y^{(i)} - \sum_{k=0}^{P} y_k \psi_k(\xi^{(i)}) \right)^2 = 0 \tag{2}$$

where, y_k are output PCE coefficient, ξ is standard normal vector, ψ_k are basis functions and y^i are outputs of input data sampled using the Latin Hypercube Sampling scheme. Further details on implementation of non-intrusive polynomial chaos can be found in Najm (2009).

3. Results and Discussion

The multi bed model is simulated for a cycle of 24 hr. Table 1 lists the physical parameters used in the model. The number of monolith channels is selected according to channel adsorption capacity, sample room size and number of people such that CO_2 concentration

inside the room remain below 0.5% at all times. The size of dehumidifier packed bed is determined such that its capacity is sufficient to adsorb water during an entire day of operation of the CO_2 capture system; after which it is replaced by the second dehumidifier for the following day.

Table 1: Physical parameters of the Enclosed Air Capture system

Parameter	Value	Unit
Monolith cell density	400	cpsi
Number of channels	30,000	-
Monolith dimension	0.60×0.25	m^2
Packed bed dimension	0.60×0.30	m^2

The system is designed to constrain the CO_2 concentration inside room below 5000 ppm at all times. Figure 2 shows the CO_2 concentration inside room (Figure 2a) and at monolith exit (Figure 2b) on a 24 hr time basis. It can be seen from Figure 2a that CO_2 concentration rises from 9 AM – 5 PM. This is because it is assumed that the room is occupied by people during that time. However, the rise of CO_2 concentration during that time is not linear and is accompanied by periodic drops in the concentration. This happens because CO_2 is removed from the room by adsorption on monoliths at periodic intervals from 9 AM – 5 PM as can be seen from Figure 2b.

The cyclic nature of this system can be exploited to minimize the energy cost and deal with uncertainty. From 5 PM – 3 AM, the room is assumed empty. However, the CO_2 capture system remains active to lower the room CO_2 concentration back to the initial level. This can be seen from Figure 2a, where the CO_2 concentration inside the room can be seen to drop until 2 AM. The CO_2 capture system during that time can be seen from Figure 2b, where the cyclic adsorption – desorption continues until 2 AM. From 3 AM – 9 AM, the CO_2 concentration in the room stabilizes at the lower bound (Figure 2a) and we stop the CO_2 capture system, as seen from Figure 2b. The current design provides flexibility during this period (3 AM – 9 AM) should there be need for uncertainty quantification due to variations in input parameters such as variable respiration or perspiration rates, variation in number of people inside the room, etc.

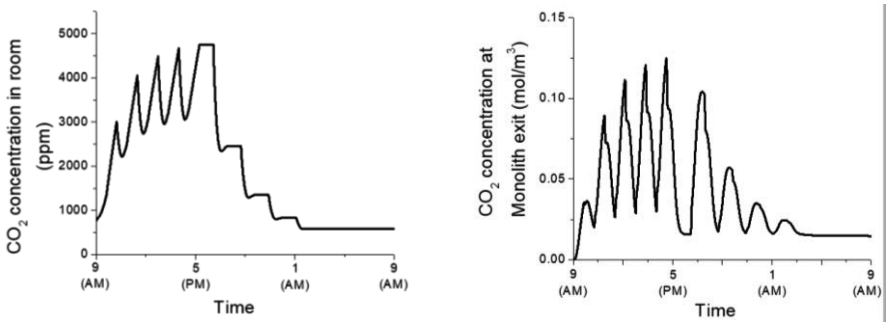


Figure 2: CO_2 concentration dynamics a) inside room and b) at monolith exit

The CO_2 capture system results in reduction of external ventilation rate since the external air flow is now governed by O_2 depletion rate inside the room instead of increase in CO_2 concentration. This results in improvement in energy performance of the overall system

as compared with the conventional ventilation methods. Figure 3 compares the energy performance of both the systems on an hourly basis. It can be seen from the figure that the CO_2 capture system consumes 75% less energy during working hours as compared to the conventional ventilation methods. After working hours (after 5 PM), the energy consumption is negligible for both systems. This is because after 5 PM, it is assumed that the external ventilation is turned off. The only source of energy consumption for CO_2 capture system after 5 PM is because of 3D monolith adsorption–desorption operation which turns out to be negligible as compared to the energy consumption for external ventilation.

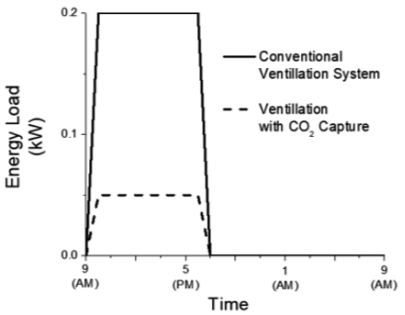


Figure 3: Comparison of energy performance on an hourly basis

Using the mathematical model, we look into quantification of model uncertainty. In order to quantify the uncertainty in respiration dynamics and number of people, we have assumed that the respiration rate varies on a daily basis with a normal distribution, $N(2.64\times10^{-4}$ mol/h, 0.25×10^{-4} $mol^2/h^2)$. Figure 4 shows the performance of NIPC as compared to Monte Carlo simulations. It can be seen from the Figure 4a, that the output CO_2 concentration inside the room converges with the results obtained through Monte Carlo Simulations (MCS). However, NIPC simulations required only 4 input sampling data as compared to 200 sample points required for MCS. This has direct implication on computational performance. Figure 4b provides the computational time incurred due to NIPC as well as Monte Carlo simulation. It can be seen from the figure that NIPC takes less computation time as compared to Monte Carlo simulation. Thus using NIPC, similar convergence in model output can be realized at lower computation cost.

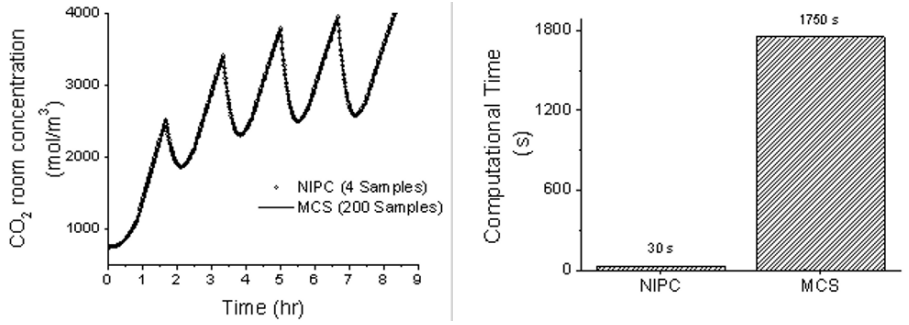


Figure 4: a) Output convergence with NIPC and Monte Carlo, b) Computational time analysis

4. Conclusion

This study proposes a novel strategy to capture CO_2 from enclosed environments. A numerical model is developed with CO_2 adsorption on zeolite 13X 3D monoliths. The optimized results control the CO_2 level inside the room within 5000 ppm at all times. Further, enclosed CO_2 capture system shows improved energy performance compared to conventional ventilation methods. The uncertainty due to variability in respiration and perspiration dynamics is quantified through non-intrusive polynomial chaos and compared to Monte Carlo simulations. It is found that non-intrusive polynomial chaos shows improvement in computational time compared to Monte Carlo simulation. This study establishes direct air capture of CO_2 as feasible strategy for reduction of energy requirements from enclosed environments. Better adsorbents with higher equilibrium capacity for CO_2 will result in better performance. Further, quantification of additional uncertainty such as room size, number of people etc. and design under uncertainty can result in risk minimization and better prediction of system output performance.

References

A.S.H.R.A.E. Standard 62.1. (2004). Ventilation for acceptable indoor air quality, 15.

Ahn, H. and Lee, C.H. (2004). Effects of capillary condensation on adsorption and thermal desorption dynamics of water in zeolite 13X and layered beds. Chemical Engineering Science, 59 (13), 2727-2743.

Aktacir, M. A., Büyükalaca, O., & Yılmaz, T. (2010). A case study for influence of building thermal insulation on cooling load and air-conditioning system in the hot and humid regions. Applied Energy, 87(2), 599-607.

Bollini, P., Brunelli, N. A., Didas, S. A., & Jones, C. W. (2012). Dynamics of CO_2 adsorption on amine adsorbents. 1. Impact of heat effects. Industrial & Engineering Chemistry Research, 51(46), 15145-15152.

Fisk, W. J. (2000). Health and productivity gains from better indoor environments and their relationship with building energy efficiency. Annual Review of Energy and the Environment, 25(1), 537-566.

Najm, H. N. (2009). Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics. Annual review of fluid mechanics, 41, 35-52.

Prívara, S., Vana, Z., Cigler, J., Oldewurtel, F., & Komárek, J. (2011). Role of MPC in building climate control. Computer Aided Chemical Engineering, 29, 728-732.

Salsbury, T., Mhaskar, P., & Qin, S. J. (2013). Predictive control methods to improve energy efficiency and reduce demand in buildings. Computers & Chemical Engineering, 51, 77-85.

Sanzida, N., & Nagy, Z. K. (2014). Polynomial Chaos Expansion (PCE) Based Surrogate Modeling and Optimization for Batch Crystallization Processes. Computer Aided Chemical Engineering, 33, 565-570.

Seppänen, O. A., Fisk, W. J., & Mendell, M. J. (1999). Association of ventilation rates and CO_2 concentrations with health andother responses in commercial and institutional buildings. Indoor air, 9(4), 226-252.

Sinha, A., Darunte, L. A., Jones, C. W., Realff, M. J., & Kawajiri, Y. (2017). Systems design and economic analysis of direct air capture of CO2 through temperature vacuum swing adsorption using MIL-101 (Cr)-PEI-800 and mmen-Mg_2 (dobpdc) MOF adsorbents. Industrial & Engineering Chemistry Research, 56(3), 750-764.

Thakkar, H., Eastman, S., Hajari, A., Rownaghi, A. A., Knox, J. C., & Rezaei, F. (2016). 3D-printed zeolite monoliths for CO_2 removal from enclosed environments. ACS applied materials & interfaces, 8(41), 27753-27761.

Videla, J. I., & Lie, B. (2006). Library for modeling and simulating the thermal dynamics of buildings. Computer Aided Chemical Engineering, 21, 1777-1782.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50359-1

Modelling and Simulation for Regional Ozone Impact by Flaring Destruction and Removal Efficiency of Oil & Gas Industries

Sijie Ge[a], Sujing Wang[b], Jian Zhang[a], Qiang Xu[a*], Thomas Ho[a]

[a] *Department of Chemical Engineering, Lamar University, Beaumont, TX 77710, USA*
[b] *Department of Computer Science, Lamar University, Beaumont, TX 77710, USA*
qiang.xu@lamar.edu

Abstract

Flaring is a necessary measure for plant safety in oil & gas industries. However, the industrial flaring generates large amounts of VOC and NO_x, which could transiently aggravate regional ozone concentrations that could affect on human health and the environment. Currently, the national standard value of 98% or 99% for destruction and removal efficiency (DRE) is regulated and presumed for industrial flaring operations. However, the DRE value could be lower than the national standard under various atmospheric and operating conditions, such as the high cross-wind speed and flare jet velocity. Thus, it is important to know the quantity and sensitivity of the regional ozone impact due to lower DREs in oil & gas industries. In this paper, a systematic study on industrial ozone pollution impacts due to the lower flaring DREs caused by cross-wind speed and flare jet velocity has been performed. First, the DRE formulas considering crosswind speed and jet velocity were derived based on available references. Second, the emission inventory assisted by the DRE formulas was updated in a certified episode of air-quality model from Texas Commission on Environmental Quality (TCEQ), where compressive information of meteorology (e.g., solar radiation intensity, wind speed, temperature, pressure and humidity), geography (e.g., population densities and land use) and emission inventory (e.g., point source emissions, area emissions, on-road emissions and biological emissions) were modeled. After that, comprehensive air quality model with extensions (CAMx) based air-quality modelling was conducted to simulate the regional ozone formation and transformation, as well as to study the ozone increments due to effects on DRE changes. This study could provide valuable and quantitative support for all relevant stakeholders, including environmental agencies, regional plants, and local communities.

Keywords: Flare emissions; Air quality; DRE; Cross-wind speed; Jet velocity

1. Introduction

Flaring is the practice of burning unusable or unwanted flammable gas which might otherwise be vented directly to the atmosphere. The oil & gas industry processes millions of cubic feet of hydrocarbon gases every day so that a slight decrease in flaring performance means a release of tens of thousands of cubic feet of emissions into the atmosphere (Baukal and Schwartz, 2001; Ge et al., 2018). It should be noted that flaring is crucial to plant safety in oil & gas industries. However, flaring generates emissions such as CO_2, CO, NO_x (nitrogen oxides), VOCs (volatile organic compounds) due to the incomplete flaring combustion. Given the 98 % flaring efficiency (TCEQ, 2015), the

resultant air emissions include at least 15.4 million pounds of CO_2, 40.0 Klbs CO, 7.4 Klbs NO_x, 15.1 Klbs hydrocarbons, and 100.0 Klbs HRVOC (Xu et.al, 2009). These emissions may cause highly localized and transient air pollution events and negative societal impacts (Ge et al., 2016; Ge et al., 2017a; Ge et al., 2017b). It should be noted that under certain conditions (e.g., strong crosswind, jet velocity, low combustion heating value and stoichiometric ratio), the flare destruction and removal efficiency (DRE) can be reduced and the portion of unburned species can increase (Singh et al., 2014). The main concern is the potential release of unburned VOCs, which is potent to transiently aggravate ground-level ozone concentration locally. The reason is that ozone is formed by the photochemical reaction between NO_x and VOCs (i.e., ozone precursors) in the presence of sunlight.

Ozone is detrimental to humans and many other living organisms (Kampa, 2008). Ozone is one of six common pollutants regulated by the Federal Clean Air Act. U.S. EPA (Environmental Protection Agency) set the National Ambient Air Quality Standard (NAAQS) for ground-level ozone since July 1997. Starting from Oct 1st, 2015, a more stringent ozone standard (70 ppb) has been issued to further enhance protection for public health (EPA, 2016). To improve air quality in nonattainment areas and meet the requirements of, each state was required to issue the State Implementation Plan (SIP) for their future air quality controls. According to EPA regulations, a 98% DRE or higher can be achieved if the flare is operated according to 40 CFR Section 60.18 (McDaniel, 1983). However, flare operations outside of these parameters DRE may be lower than 98% due to the crosswind, jet velocity and others. Very recently, Ge et al. (2016) has studied the ozone impact due to lower DREs via the optimal scheduling of multiple ethylene plant start-ups through virtual case studies, where the 8-h ozone increment under DREs of 95%, 96%, 97% and 98% are 10.1, 8.5, 6.7 and 3.2 ppb, respectively. In general, plant start-up operations generate larger amounts of flare emissions than the normal operating conditions. The study on air quality impact under normal flare situations with lower DREs in oil & gas industries is still lacking.

Thus, it is environmentally important to know the regional ozone impact due to the excess VOCs and NO_x released from oil & gas industries when the DRE value is lower than the national standard. In this paper, a systematic study on ozone pollution impacts due to the lower flaring DREs caused by cross-wind speed and flare jet velocity has been performed. The case study indicates that the maximum ozone impact from the crosswind speed and jet velocity is 0.18 ppb. This study could provide valuable and quantitative support for all relevant stakeholders, including environmental agencies, regional plants, and local communities.

2. Problem statement

This study first utilizes results from computational fluid dynamics (CFD) simulations to obtain DRE profile changed with the cross-wind speed and jet velocity. Next, the emission inventory files for flaring emissions are modified based on the adjusted DRE. For each flare source of oil & gas industries in the studied domain, adjusted DRE is calculated based on its cross-wind speed and jet velocity. After that, emissions of VOC and NO_x are adjusted. Finally, the new emission inventory can be generated then used for the air quality model simulation to ozone impacts due to the cross-wind speed and jet velocity. In this study, the comprehensive air quality model with extensions (CAMx) is employed to simulate the spatial and temporal distribution of ozone concentrations (CAMx User's Guide, 2015). Detailed ozone episodes including emission inventories

and meteorological data used as the base for the air-quality modelling can be obtained from the website of Texas Commission on Environmental Quality (TCEQ).

3. General methodology

3.1 Methodology framework

Figure 1 summarizes the general methodology framework of this study. The computational fluid dynamics (CFD) simulations are firstly employed to obtain DRE profile changed with the cross-wind speed and jet velocity. After that, flaring emissions (i.e., VOC and NO_x) are adjusted by DREs for each elevated point source from the ozone episode. Finally, the new emission inventory can be generated then used for the air quality model simulation to obtain ozone impacts due to flare DREs changed with the cross-wind speed and jet velocity. In the following Sections of 3.2 and 3.3, major technical methods involved in this methodology framework will be elaborated.

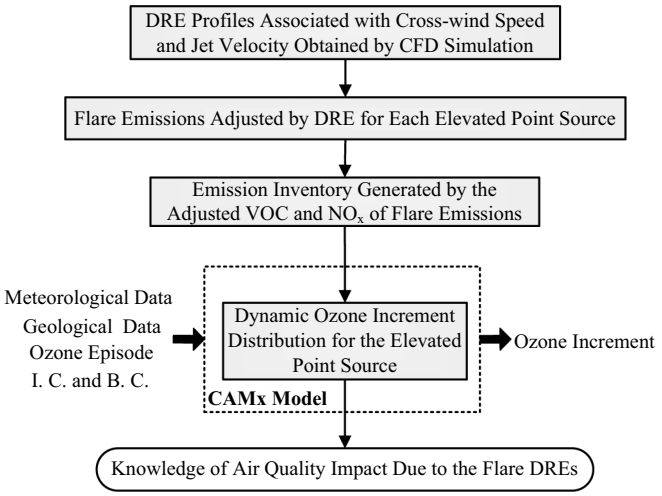


Figure 1. General methodology framework.

3.2 Model descriptions

CAMx is employed in this study to simulate the spatial and temporal distribution of ozone concentration. CAMx is a Eulerian photochemical dispersion model that allows for integrated "one-atmosphere" assessments of tropospheric air pollutants (i.e. ozone, particulates, air toxics, and mercury) over spatial scales ranging from neighborhoods to continents (CAMx User's Guide, 2014). An ozone episode with the completed meteorological and emission inventories data from May 31, 2006 to June 15, 2006 established by TCEQ is selected as the base case model simulation. All the CAMx simulations were run on a Dell computer with four 3.6 GHz CPUs and 8 GB memories. The used CAMx model is a nested regional-to-urban scale with four different grid resolutions (Ge et al., 2016).

3.3 Calculation methods

The amount of flare emissions is calculated based on both the flow rate and chemical species of the vent gases and the appropriate DRE value. The emission inventory files for elevated point sources (i.e., oil & gas industrial flare emissions) were modified

based on the adjusted DREs. For each flare source, the adjusted DRE value is calculated based on its jet velocity and crosswind speed. Next, emissions of VOC and NO_x can be calculated. After all flare emissions are obtained, new elevated point emission files for CAMx will be generated. Given that the original flare emissions data are based on 98 % DRE, the adjusted emissions of VOC and NO_x are as follows:

$$f_{VOC} = f_{0,VOC} \frac{1-\eta(U,V)}{1-0.98} \tag{1}$$

$$f_{NO_x} = f_{0,NO_x} \frac{\eta(U,V)}{0.98} \tag{2}$$

where f_{0,NO_x} and $f_{0,VOC}$ are original emissions of VOC and NO_x from the elevated point sources in the studied domain; the DRE value η is a function of the cross-wind speed of U and the jet velocity of V; f_{VOC} and f_{NO_X} are adjusted VOC and NO_x emissions by the crosswind speed and jet velocity. The units of all emissions are mol/h.

To quantitatively study the ozone concentration impacted from the adjusted DREs, the amount of ozone difference between adjusted DRE case and the base case is defined by the following equation.

$$\Delta C_n^{O_3}(d,h,\mathbf{x}) = C_n^{O_3}(d,h,\mathbf{x}) - C_0^{O_3}(d,h,\mathbf{x}), \quad h = 0, 1, \ldots, 23 \tag{3}$$

where $\Delta C_n^{O_3}(d,h,\mathbf{x})$ represents the hourly ozone difference due to the adjusted $\eta(U,V)$ at hour h on day d in grid x, ppb. $C_n^{O_3}(d,h,\mathbf{x})$ represents the hourly ozone concentration of the n^{th} simulation case at hour h on day d in grid **x.** $C_0^{O_3}(d,h,\mathbf{x})$ represents the background ozone concentration of the base case at hour h on day d in grid x. Note that $C_0^{O_3}(d,h,\mathbf{x})$ are obtained by the CAMx simulation based on the standard 98 % DRE value of flare emissions. Based on the results of $\Delta C_n^{O_3}(d,h,\mathbf{x})$, the significance of ozone impact from adjusted flare DREs due to the cross-wind speed and jet velocity will be obtained to provide quantitative technical support for relevant decision makers.

4. Case studies

4.1 DRE Formula Obtained based on CFD Simulations

CFD simulations show that four variables affect the DRE of industrial flares: crosswind speed, jet velocity of flare vent gas, combustion zone heating value (CZHV), and stoichiometric ratio. Due to limited data sources, only the crosswind speed and jet velocity are considered to adjust DREs in this study. The crosswind speed data can be obtained from meteorological data of the studied ozone episode, while the jet velocity can be obtained from the emission inventory. Based on CFD simulation results from Singh et al. (2014) with practical amendment and mathematic interpolations, the complete relation of DRE with respect to U and V can be generated as shown in Figure 2. It is a 3-D plot of DRE with changes of crosswind speed and jet velocity. It can be seen that the DRE keep the standard value of 98 % when jet velocity is less than 15 m/s. When the jet velocity is larger than 30 m/s and the crosswind speed is larger than 5 m/s, DREs are dropped lower than 80 %.

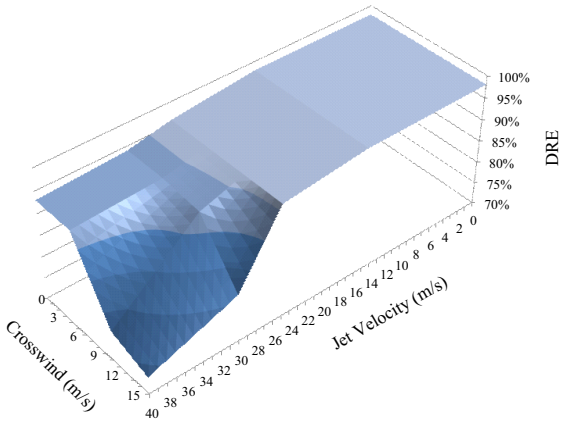


Figure 2. 3D plot of DREs with changes of crosswind speed and jet velocity.

4.2 Simulation Results based on CAMx Simulations

The modified emission inventory files are imported to CAMx model for air quality simulations. The episode of Houston-Galveston-Brazoria May 31-June 15, 2006 (base case) is chosen for the air quality modelling. First, the modelling was run once as the base case to get the background ozone concentrations. Second, the modified emission inventory files were employed, and the air quality model was run again to calculate ozone concentrations under the adjustment of crosswind speed and jet velocity. After that, results from two runs were compared to estimate the effect of cross-wind speed and jet velocity. It can be seen that the maximum hourly ozone increment fluctuates between 0.002 and 0.18 ppb. The maximum hourly ozone decrement fluctuates between 0 and 0.03 ppb. The maximum ozone increment occurred on June 11, 2006.

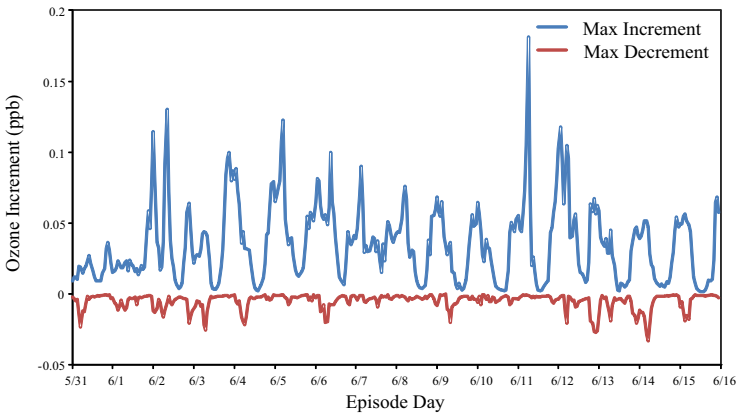


Figure 3. Maximum ozone increment and decrement after the DRE adjustment.

Based on the case studies, it shows that although the individual DREs could be significantly changed by the crosswind speed and jet velocity (e.g., the adjusted DREs can be as low as 74 % at the jet velocity of 40 m/s and the crosswind speed of 15 m/s.), the regional ozone impact will not be affected much (the maximum hourly ozone increment is only 0.18 ppb). It should also be highlighted that the DRE impact factors of combustion zone heating value (CZHV) and stoichiometric ratio are set as fixed values

as normal values in our study. Thus, their impacts on DREs as well as the regional air quality are neglected. Meanwhile, this is the initial study coupling process DRE estimation and regional air impacts. All the results listed in the paper should be treated as preliminary and subject to a variety of risks, uncertainties, and assumptions.

Conclusions

By coupling flare DREs from oil & gas industries with the air-quality model based CAMx simulation together, the ozone impacts generated from the lower DREs due to the meteorological condition (i.e., cross-wind speed) and the process operating condition (i.e., jet velocity) has been investigated in this paper. The CFD based DRE correlations are employed to adjust DREs. Through case studies, it shows that although the individual DREs could be significantly changed by the crosswind speed and jet velocity (e.g., the adjusted DREs can be as low as 74 % at a jet velocity of 40 m/s and a crosswind speed of 15 m/s.), the regional ozone impact will not be affected much (the maximum hourly ozone increment is only 0.18 ppb). This study could provide valuable and quantitative support for all relevant stakeholders, including environmental agencies, regional plants, and local communities.

References

Baukal CE, Schwartz RE. (2001). The John Zink Combustion Handbook. New York, CRC Press.

CAMx User's Guide, 2014.

Kampa, M; Castanas, E (2008). Human health effects of air pollution. Environmental pollution,151(2):362-367.

EPA. National Ambient Air Quality Standards (2016). Available online at https://www.epa.gov/criteria-air-pollutants/naaqs-table. Accessed March 29, 2016.

Ge, S., Wang, S., Xu, Q., Ho, T. C (2016). An Air-quality Considered Study for Multiple Olefin Plant Start-ups. Ind. Eng. Chem. Res., 55: 9698-9710.

Ge, S., Wang, S., Xu, Q., Ho, T. C (2017a). Air-quality conscious coordination for multiple olefin plant startup and shutdown operations, Proceedings of FOCAPO/CPC, Tucson, Arizona, USA.

Ge, S., Wang, S., Xu, Q., Ho, T.C (2017 b). Impact of chemical plant start-up emissions on ambient ozone concentration. Atmospheric Environment, 164:20-30.

Ge, S., Wang, S., Xu, Q., Ho, T.C (2018). Ozone impact minimization through coordinated scheduling of turnaround operations from multiple olefin plants in an ozone nonattainment area. Atmospheric Environment, 176:47-53.

McDaniel, M. Flare Efficiency Study; United States Environmental Protection Agency: Washington, DC, Report No. 600/2-83-052, July 1983.

Singh, K.D., Gangadharan, P., Dabade, T., Shinde, V., Chen, D., Lou, H.H., Richmond, P.C. and Li, X (2014). Parametric Study of Ethylene Flare Operations Using Numerical Simulation. Engineering Applications of Computational Fluid Mechanics, 8(2): 211-228.

TCEQ (2015). 2013 Emission Inventory Guidelines, Appendix A: Technical Supplements. Available online at http://www.tceq.state.tx.us/publication s /rg/rg-360-13/index.html. Accessed June 5, 2015.

Xu, Q; Yang, X.T; Liu, C.W; Li, K; Lou, H.H; Gossage, J.L (2009). Chemical plant flare minimization via plant-wide dynamic simulation. Ind. Eng. Chem. Res, 48(7): 3505-3512.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50360-8

Smart adaptive sampling for developing surrogate approximations of physicochemical systems

Sushant Garud[a], Iftekhar Karimi[a*], Markus Kraft[b,c]

[a]*Department of Chemical and Biomolecular Engineering, 4 Engineering Drive 4, National University of Singapore, Singapore-117585*
[b]*Department of Chemical Engineering & Biotechnology, University of Cambridge, West Cambridge Site, Philippa Fawcett Drive, Cambridge, CB3 0AS, UK*
[c]*School of Chemical and Biomedical Engineering, Nanyang Technological University, 62 Nanyang Drive, Singapore-637459*
**cheiak@nus.edu.sg*

Abstract

In this article, we present an application of our smart sampling algorithm (SSA) for designing real complex physico-chemical experiments. SSA consists of a novel strategy to select sample points systematically in an adaptive and optimized manner. It ascertains the placement of points in the most unexplored regions of the domain with complex/nonlinear behaviour by solving a series of placement optimization problems. To this end, we employ SSA to design experiments for butanol separation using multicomponent adsorption. Our numerical evaluation shows that the experimental design of SSA requires fewer experimental runs compared to the conventional designs. Overall, SSA is a novel modern experimental design technique that reduces experimental expenses through smart placement.

Keywords: Adaptive sampling, Intelligent placement, Surrogates, Experimental design

1. Introduction

First principles-based or physics-based modelling approach is typically used to model, study, and analyse complex nonlinear physicochemical systems. However, evaluation of such models is generally computationally intensive, thus, prohibiting their repeated evaluations in an analysis procedure. Often, these models are treated as black-box, hence no system information is available to the users without evaluating an instance of their costly simulation. Furthermore, such models are typically validated by conducting real-physical experiments which are obviously expensive. Therefore, it is beneficial to convert such high-fidelity models and corresponding physical experiments into their computationally inexpensive counterparts, *viz.* surrogate models that capture the essential features with prescribed numerical accuracy. In other words, surrogate modelling approach generates a mathematical or numerical representation of the underlying complex system using some sampled input-output data. The literature is replete with a variety of surrogate modelling techniques and their applications. For a given modelling technique, the quality of the surrogate model depends on the input-output data set used for its construction. Since this input-output data set is gathered by evaluating high-fidelity models or by performing rigorous physical experiments, sampling paradigm is the most expensive task in surrogate construction.

Commonly used sampling techniques employ uniform, quasi-random, or systematic distributions. Typically, these one-shot techniques can result in under/oversampling and

thus, poor system approximation. To tackle these issues, a dynamic class of DoE called adaptive sampling (sequential sampling) has attracted attention of the research community (Eason & Cremaschi, 2014). Adaptive sampling has two major advantages over the one-shot techniques namely low computational expenses and better approximations. In this work, we present a practical application of our adaptive sampling technique namely smart sampling algorithm (SSA) for intelligently designing real physical experiments using surrogate models. This article is organized as follows. Section 2 gives the overview of SSA and section 3 presents an application of SSA for designing multi-component adsorption experiments. Finally, we draw our conclusions in section 4.

2. Overview of SSA

Let $y = f(x)$; $f: \mathbb{R}^N \rightarrow \mathbb{R}^M$ for $\mathcal{D}: x^L \leq x \leq x^U$ describe the behaviour of a unit/process/system whose experimental or computational quantification is complex and computationally expensive. We need an analytical or numerical surrogate model $S(x)$ to replace $f(x)$ so that $y \approx S(x)$. Hence, SSA solves a problem of generating K_{max} sample points that can give the best S. In other words, experiments can be conducted at these sample points to extract maximum information from the underlying system. Our thought-process behind the development of SSA considers two fundamental concepts *viz.* exploration and exploitation. To this end, we quantify the exploration of $\mathcal{D}$ using crowding distance metric (CDM). It is based on the Euclidean distance of a point from all the other sample points as given in Eq. (1).

$$\boldsymbol{CDM(x)} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{I}} \left(\left\|\boldsymbol{x} - \boldsymbol{x^{(i)}}\right\|\right)^{\boldsymbol{2}} \tag{1}$$

where $\|.\|$ is the Euclidean norm.

We quantify the exploitation using the departure function defined as follows. Let $S^{(I)}(x)$ denote a surrogate constructed using sample set $X_N^{(I)} = \{x^{(i)} | i = 1, 2, \cdots, I\}$ and its corresponding response set. Let $x^{(j)}$ be a sample point in $X_N^{(I)}$, and let $S^{(I/j)}(x)$ be the surrogate constructed using $X_N^{(I/j)} = \{x^{(i)} | i = 1, 2, \cdots, I, i \neq j\}$ and its corresponding sample set. Then, the departure function is given by Eq. (2).

$$\Delta^{(\boldsymbol{I/j})}(\boldsymbol{x}) = \boldsymbol{S^{(I)}(x) - S^{(I/j)}(x)} \;\; \boldsymbol{j = 1, 2, \cdots, I} \tag{2}$$

Qualitatively, it determines the impact of locating a sample point in the neighbourhood of $x^{(j)}$ on $S^{(I)}$. The larger the values of the departure function, the greater the placement impact; hence, the more plausible the region for sample placement. A single objective that combines the above discussed two concepts can yield the best new sample point. This is achieved by formulating a point placement optimization problem as follows. Given a sample set $X_N^{(I)}$, its corresponding response set, and the surrogate $S^{(I)}(x)$ constructed using them, we aim to place the new point as far away from existing points as possible, and the new point should have the highest impact on $S^{(I)}(x)$. To this end, we formulate a series of NLPs given in Eq. (3). These NLPs are iteratively solved to generate new samples which are in turn used to update the surrogate.

$$\boldsymbol{NLP(j)}: \max_{\boldsymbol{\mathcal{D}}} \left(\Delta^{(I/j)}(\boldsymbol{x})\right)^2 \times \boldsymbol{CDM}(\boldsymbol{x}) \quad \boldsymbol{j = 1, 2, \cdots, I} \tag{3}$$

The optimal solution to the above NLP in Eq. (3) can be a suitable candidate for the new sample point. SSA iteratively solves these NLPs using the multi-start Hooke-Jeeves algorithm for adaptive sample placement as discussed by Garud et al. (2018).

For prespecified $K(< K_{max})$ initial sample points and a chosen surrogate model type S, SSA proceeds as follows:

1. Generate a sample set $X_N^{(I)} = \{x^{(i)} | i = 1, 2, \cdots, K\}$ using any modern DoE technique like Sobol sampling (QS).
2. Compute $Y_M^{(I)} = \{y^{(i)} | i = 1, 2, \cdots, K\}$ using a sample set $X_N^{(I)}$. Note that throughout our discussion $M = 1$. Thus, it is dropped from the subscript here onwards for the sake of convenience.
3. Set $k = K$.
4. Construct $S^{(k)}(x)$ using $X_N^{(k)}$ and $Y^{(k)}$.
5. If $k = K_{max}$, then $S(x) = S^{(k)}(x)$ and STOP. Otherwise, proceed to step 6.
6. Compute $CDM^{(j)} = CDM\left(x^{(j)}\right) \forall\, x^{(j)} \in X_N^{(k)}$ and $j = 1,2, \cdots, k$ using Eq. (1). Arrange $CDM^{(j)} (j = 1, 2, \cdots, k)$ in descending order and define the order as $p = 1, 2, \cdots, k$.
    a. Set $p = 1$.
    b. Construct $S^{(k/p)}(x)$ using data from Steps 1, 2, and $i \neq p$.
    c. Construct and solve $NLP(p)$ given in Eq. (3). Let x^* be the optimal solution. If $\left\|x^* - x^{(i)}\right\| \leq \varepsilon$ for any $i = 1, 2, \cdots, k$, then set $p = p + 1$ and go to Step 6b. Otherwise, $k = k + 1$ and $x^{(k+1)} = x^*$. Evaluate $y^{(k+1)}$, go to Step 4 and update the input-output data set. Note that if p exceeds k, then perform pure exploration-based (space-filling based) sample placement and proceed to Step 4.

Note that this is a brief overview of SSA and the detailed algorithm can be found in Garud et al. (2018).

3. Case Study

In our previous works, we have successfully employed SSA for constructing surrogate approximations of analytical test functions as well as process simulations. Here, we illustrate its application for designing multicomponent adsorption experiments to separate butanol from ABE (acetone, butanol, ethanol) solution.

3.1. Multicomponent adsorption system

Biobutanol has gained significant attention over the last few years as one of the promising biofuels. Its production involves fermentation process which produces broth containing acetone, butanol, ethanol, butyric acid, acetic acid, glucose, and xylose. Moreover, concentration of butanol in the broth is relatively lower due to the presence of inhibiting bacteria. Therefore, the major challenge in biobutanol production is its separation from the broth with high recovery.

Abdehagh et al. (2016) have modelled the multicomponent adsorption (MCA) of ABE for separating butanol. Their aim is to develop an isotherm model for MCA system. For

this, they perform extensive physical experiments and use the experimental data to construct isotherms. These isotherms essentially show the amount of particular component adsorbed, say butanol as a function of component equilibrium concentration. Their study considers both Langmuir and Freundlich isotherm models and concludes that Freundlich isotherm appears to emulate the experimental data very well. However, the number of parameters of Freundlich isotherm increases with increasing number of components, thus increasing the required experimental data set size. Therefore, the classical experimental design approaches may not be efficient for such systems.

To this end, we employ SSA to design experiments for the above discussed system using surrogates. While designing the experiments with SSA, one needs to perform these experimental runs sequentially as SSA generates samples adaptively. However, in this work, we do not perform any physical experiments, rather we rely on the exhaustive isotherm built by Abdehagh et al. (2016) to emulate those experiments. Henceforth, an evaluation of this isotherm serves as performing the real physical experiment.

3.2. Evaluation plan

We build surrogate approximation of MCA system using SSA which guides the generation of new sample points as discussed earlier in Section 2. For this, we consider 7 input variables (experimental conditions or factors) given in Table 1 and equilibrium adsorption capacity (q) as the output/response variable. We generate initial sample points $(K = 20)$ via Sobol sampling and use second order polynomial response surface (PRSM) as a surrogate. We then use this sample set to evaluate experiments and generate input-output data set which is then used to construct the initial PRSM surrogate. SSA adaptively updates the initial input-output data set, in turn the initial surrogate, by generating new sample points and using them to perform new experiments. We consider the following three scenarios for SSA: $K_{max} = \{42, 50, 80\}$. Abdehagh et al. (2016), on the other hand, conduct 42 experiments for multicomponent system and 38 experiments for pure component system, amounting to total 80 experiments. This is necessary since the complete representation of Freundlich isotherm for 7 component system requires estimation of 42 competitive adsorption coefficients and 14 pure component coefficients, resulting in total of 56 parameters.

Besides, we also compute the quality of PRSM approximation using three error-based metrics given in Eqs. (4)-(6) *viz.* average absolute error (AAE) and root mean squared error (RMSE), and pooled error (PE). These metrics quantify the performance of SSA in constructing PRSM approximation.

Table 1: Experimental settings for evaluation of SSA.

Input variables	Description of variables	Lower bounds	Upper bounds
x_1	Butanol concentration (g/L)	0.09	25.00
x_2	Acetone concentration (g/L)	0.17	15.00
x_3	Ethanol concentration (g/L)	0.05	5.00
x_4	Butyric acid concentration (g/L)	0.09	11.00
x_5	Acetic acid concentration (g/L)	0.06	12.00
x_6	Glucose concentration (g/L)	0.09	15.00
x_7	Xylose concentration (g/L)	0.05	10.00

$$\mathbf{AAE} = \frac{\sum_{q=1}^{Q}\left|\boldsymbol{y}^{(q)} - \boldsymbol{S}\left(\boldsymbol{x}^{(q)}\right)\right|}{\boldsymbol{Q}} \tag{4}$$

$$\mathbf{RMSE} = \sqrt{\frac{\sum_{q=1}^{Q}\left(\boldsymbol{y}^{(q)} - \boldsymbol{S}(\boldsymbol{x}^{(q)})\right)^{2}}{\boldsymbol{Q}}} \tag{5}$$

$$\mathbf{PE} = \sqrt{\mathbf{AAE} \times \mathbf{RMSE}} \tag{6}$$

3.3. Results and discussion

We now compare the performance of our SSA-based surrogates with the Freundlich isotherms constructed by Abdehagh et al. (2016). They consider two relatively extreme scenarios of multi-component broths *viz.* low-level and high-level to ascertain that the developed model performs well over the ranges of concentrations. The interested readers may refer to Abdehagh et al. (2016) for mixture compositions for low and high-level scenarios. Figures 1 and 2 show the trends of SSA surrogates and the real isotherms for low and high-level scenarios respectively. Clearly, SSA performs very well in emulating the real isotherm with fewer experimental runs. Freundlich isotherm is constructed using 80 experimental runs while SSA-based surrogate requires only 50 runs to emulate the real isotherm with 3% average error. Overall, we can reduce the total experimental runs by 37.5% to achieve the performance with prescribed numerical accuracy.

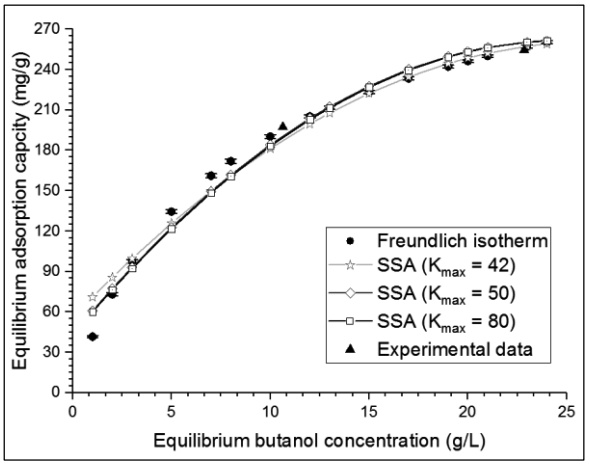


Figure 1: Comparison of SSA surrogates with Freundlich isotherms for low-level scenario.

Additionally, we assess the quality of the surrogates using error-based metrics discussed earlier. Table 2 shows these metrics for three cases. AAE quantifies the average error in surrogate prediction while RMSE captures its sense of distribution. PE combines these two metrics to give an overall performance measure. The lower the values of these metrics, the better the surrogate prediction. SSA shows significant improvement in the performance for $K_{max} = 50$ compared to $K_{max} = 42$. However, the improvement stagnates after $K_{max} = 50$.

Table 2: Performance metrics for surrogates constructed using SSA.

Legend	AAE	RMSE	PE
SSA ($K_{max} = 42$)	8.59	12.14	10.21
SSA ($K_{max} = 50$)	6.31	8.83	7.46
SSA ($K_{max} = 80$)	5.67	7.89	6.69

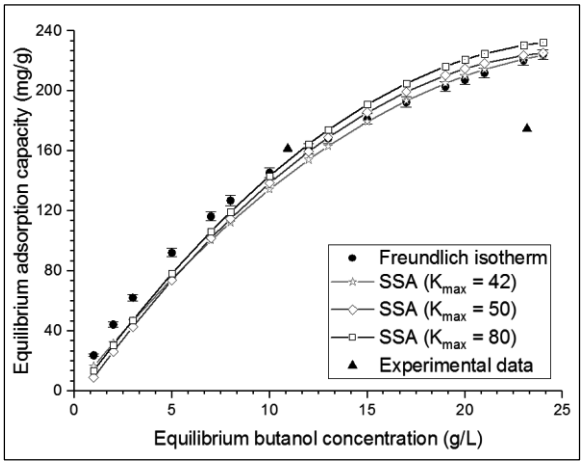


Figure 2: Comparison of SSA surrogates with Freundlich isotherms for high-level scenario.

4. Conclusions

In this work, we show the application of our novel sampling algorithm for designing physical experiments of multicomponent adsorption system. Our numerical evaluation shows that SSA reduces experimental expenses by approximately 37.5% while maintaining 3% average error. Additionally, we compare our SSA-based models with the literature data and it shows that our models perform well over ranges of component concentrations. Overall, SSA is a novel modern experimental design technique that reduces experimental expenses through smart placement.

Acknowledgement

This publication is made possible by the Singapore National Research Foundation under its Campus for Research Excellence And Technological Enterprise (CREATE) programme.

References

J. Eason, & S. Cremaschi, (2014). Adaptive sequential sampling for surrogate model generation with artificial neural networks. Computers & Chemical Engineering, 68, 220-232.

N. Abdehagh, F. H. Tezel, & J. Thibault, (2016). Multicomponent adsorption modelling: isotherms for ABE model solutions using activated carbon F-400. Adsorption, 22, 357-370.

S. S. Garud, I. A. Karimi, G. P. E. Brownbridge, & M. Kraft, (2018). Evaluating smart sampling for constructing multidimensional surrogate models. Computers & Chemical Engineering, 108, 276-288.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50361-X

Visual Modelling with Networks

Heinz A Preisig*, Arne Tobias Elve, Sigve Karolius

Chem Eng, Norwegian Uni. of Sci. & Tech., 7491 Trondheim, Norway
Heinz.Preisig@chemeng.ntnu.no

Abstract

Visual modelling serves the purpose of designing process models, discussing them on the back of an envelope, a serviette, the meeting-room board and define & edit them in the graphical user interface *ModelComposer,* which is a component of our ontology-based simulation environment. Models that spread over large range of time-scales do not connect straightforwardly as the fast time-scale usually is computationally intensive to the extent where on the large time-scale one uses surrogate models derived from the detailed models. The network-of-network approach extends readily to multi-scale system – from quantum to mechanical properties.

Keywords: Ontology-base modelling, process simulation, multi-scale.

1. Background

It is not an exaggeration if one states that process systems engineering is completely dependent on the availability of models, even being driven by models. Recognising the fact that models are so central, we put them right into the centre and make them the main object of activities. It is our goal to enable a more creative interaction with models, allowing people who model processes, which we refer to as ‘’modellers’’, to operate on a high level using primitive building blocks to construct their process models. The primitive building blocks must be fundamental, thus simple without any whistles and bells, but they must also be rich enough to construct process models with all their whistles and bells. Thus, the term primitive or simple.

At this point we have developed our ontology-based modelling environment to the level, where we feel confident to capture the majority of macroscopic processes in mostly mechanistic models; mostly because kinetics is not strictly mechanistic in our view.

The ontology is built in stages, firstly defining the main components of the model, which consecutively is used to define the mathematical representation of all model components expanding to alternative mathematical representations of the same phenomena. It is also used to define two plug-ins for the *ModelComposer* that implements “visual modelling” through a graphical user interface. One of the plug-ins is an automaton, implementing the rules for the generation of the graphs and the other is the graphical appearance of the simple and composite object being available through the graphical interface.

Here we will expand on the visual modelling, the graphical language to define and modify models. Whilst the *ModelComposer* generates mainly index files, which together with the ontology augmented with the equations are put into a model factory that generates code for different simulation environments (Elve and Preisig 2017), we shall attempt to demonstrate the power of the visual modelling approach for model design, thus as an enabling tool for discussions and model representations on a conceptional level and the natural extension to multi-scale problems and the generation of workflows.

2. The basic concepts

A good story starts at the beginning, here with the basic concepts:
Concept Systems -- can be represented as a super network of context-dependent networks of nodes and arcs, namely directed graphs, with the nodes representing the capacities and the arcs the interactions.
Concept Tokens -- live in the nodes and are transported through the arcs. Thus, nodes are capacities for the tokens and arcs transport tokens. The direction is the reference coordinate system for the transport of the tokens.
Concept Typing -- allows for refinement. Typed graphs reflect the different contexts, typed nodes are used to reflect the dynamic and distribution nature, typed arcs reflect the transport mechanisms, whilst typed tokens detail tokens all in each context.
Concept Frame -- generalised from the frame of reference used in physics. It usually includes time, and in physics extends with the spatial coordinates.
Concept Behaviour -- provides the components for the mathematical description, like *state, constant, state time derivative,* and in the context of physical/chemical/biological systems, *transport,* and *conversion.*
These elements are used to define the equations in a lower triagonal form (context-free definition space) given the states and the constants as a starting point. This ensures that any model "extruded" from the ontology via the *Model Composer* is of zero degree of freedom and with zero-order differential index.
For the graphical representation of physical systems, we typically use: 0-3 dimensional distributed systems (0D shown as circles,1-3D as ellipses), point systems (small circles) and reservoirs (half open circles).
For illustrating the concepts and their application we define two contexts, physical systems and control systems, whereby the physical systems may be detailed into the states of aggregation.

3. Illustration of generating alternatives

As an example, we choose a bioprocess consisting of a tank reactor and a tubular reactor with an immobilised culture (Figure 1).

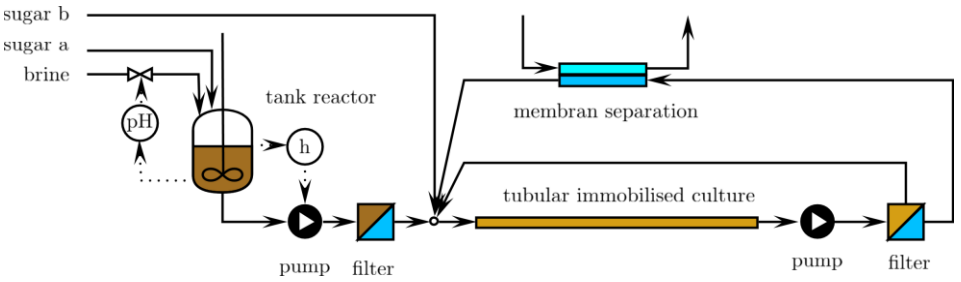


Figure 1. Example bio process

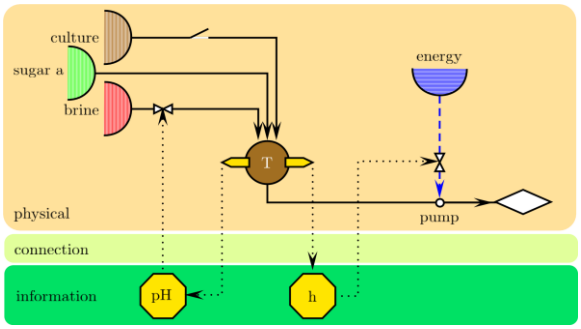


Figure 2. The controlled tank -- three connected networks

We pick a couple of elements from the bio process to show topologies of alternative representations. The first example is the stirred tank reactor and its control. It defines three networks, namely a physical network, and the control network connected with each other over the connection network. For the **physical network**, the tank is represented as a single lump, thus mixing is neglected. In the actual plant, there is a gas washing extracting some of the volatile components, which is also neglected. The collective reservoirs form the environment. Their intensive quantities are assumed constant and available as the reservoir is an infinite-size capacity and cannot be balanced. The rhomboid indicates a connection to the next section of the plant.

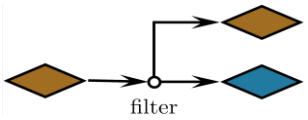


Figure 3. Very simple

Next, we have a look at a **filter**. Many possibilities: we start with a very simple model, namely just a splitter (Figure 3).

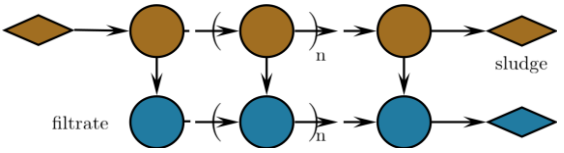


Figure 4. In slices

We can model it as a series of slices each of which has three elements, namely a lump for the sludge, a lump for the filtrate and a transfer through the filter plate, whereby the filter plate itself does not appear as a separate object. Rather it is hidden in the transfer as a resistant. Also, no build-up of the filter cake is included.

The next version considers the distribution effect along the filter from the entrance to the exit as relevant (Figure 6). The simplest version then models the two liquid bodies as each distributed in one direction, namely along the main flow coordinate.

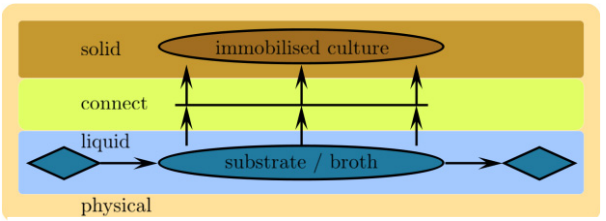


Figure 5. 1D distributed

The tubular reactor is of similar nature.

The difference is in the immobilised culture being also 1D distributed but has no inflow or outflow of the culture. It merrily exchanges material with the liquid phase.

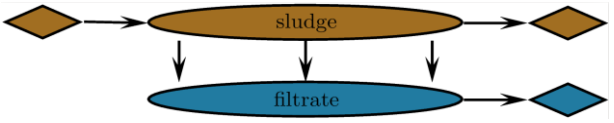


Figure 6. Simple distributed

We now have a solid phase network linked via a connection network representing the interface to the liquid phase network. In all cases, the networks have only one member node.
Considering that the culture is immobilised on fix-bed mixer elements, one can suggest a more complex model (Figure 7). Here we do not show the networks or the interfaces simply because complexity. Their presence is though apparent.
The filter and the tubular reactor case demonstrate a refinement. An interesting expansion into more detail is the culture in the broth or immobilised on the support. In this

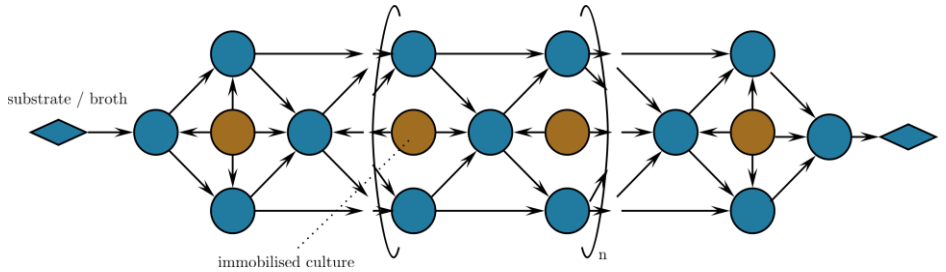


Figure 7. Fix-bed, mixing elements considered

case it is meaningful to add a new branch to the ontology (Preisig, 2016), which captures the metabolism of the cell.
The elements in the topology are linked to elements in the ontology. For example, a lumped system containing species mass is linked to an equation object representing the species mass balance, or a convective mass transport links to an equation like concentration times volumetric flow, where the latter is related to the pressure difference. In these physical systems one always handles materials, the properties of which determine the behaviour of the system. The material properties are thus an essential part of the description.
Today, most modellers would use a thermodynamic properties library to fetch the required properties for the computation of capacity and transport system's behaviours. The generation of these data has been, and still is, a costly affair and very much limited by the experimental work involved. Further, it is data points that are the basis and an interpolation is superimposed to compute the data required by the user software. If one considers multi-component systems, it is apparent that covering the experiments is infeasible given the space being spanned by the common variables temperature, pressure and concentrations. Approximate block-component type of methods like the group contribution method (Joback, 1987) offer some solutions, but are not universal applicable, though certainly useful. New alternatives are opening with the increasingly availability of particle-based computations on the molecular level (Rosso 2012). The Framework7 project MoDeNa (MoDeNa 2017) demonstrates that the integration of scales from quantum to mechanical properties is indeed feasible.

4. Down the scales and up again

With the mathematical model composed from the physical topology of the modelled system, the need is defined for the computation of material properties replacing empirical or idealised mathematical approximations. These "properties" depend on the state, primarily the intensities temperature, pressure and composition. The macroscopic system operating on a much longer time scale than the molecular simulations. The differ-

ence is several orders of magnitude -- picoseconds in contrast to minutes and hours. Taking those a macroscopic system and a molecular system as an extreme and views them from the abstract level of network modelling, the linking of two times scale may be visualised as shown in Figure 8.

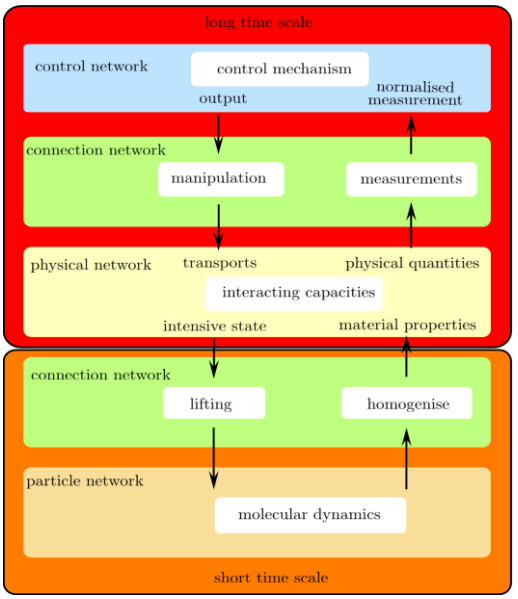


Figure 8. Coupling framework

The downwards information transfer inserts a *lifting* operation, which maps the low-dimensional state of the macroscopic system onto a large dimensional state on the molecular level, usually in the form of Boltzmann distributions for velocity of the molecules. The opposite happens on the way up, a large-dimensional state is mapped into a low dimensional one often called *homogenisation*.

4.1. Network view on molecular simulations

Considering a molecular simulation being approached by representing the molecules as a network of atoms and the material as a network of molecules, one can either view it as a hierarchical network or as a global network with different connections, where the connections represent the intra and inter molecular forces. In all cases, to get reasonable results, one must operate on very large networks, which renders any straight forward setup of the numerical problem infeasible. Methods must and have been developed to approach the super-imposed large-scale optimisation problem requiring a very large number of simulations. As a result, there exist now a zoo of highly specialised simulation environments of such systems.

4.2. Linking and Coupling Scale-Specific Software

The ontology-based modelling environment implements a super-network of context-dependent networks. Whilst the representation of the behaviour of the components can be defined in the form of a context-free, super-bi-partite graph of equations and variable one can also provide a "black box" representing an input/output representation of a complete network. The idea is to allow for the compilation of different context-dependent networks, or groups thereof, as separate computational tasks, which later are combined into a workflow management program handling the interaction and logical linking of the different tasks. Thus, in this context, a mathematical representation network like an atom/molecule & interacting forces network can be either compiled and linked with a solver for a stand-alone task or one inserts a prefabricated program. The *ModelComposer* then generates a workflow for a multi-scale system which then is passed on to a *ModelOrchestrator* for execution thereby combining the different time-scale related programs. This typically involves the on-line generation of surrogate models as explicated in Karolius (2016). The self-contained computational task can be sup-

plied with strategies for adapting the workflow depending on the quality of the surrogate models.

The *ModelOrchestrator* can perform the computational task in a distributed manner, on a high-performance cluster or a dedicated server. As long as the individual computational tasks are self-contained the simulations can be rerun; thus, facilitating a possible avenue of producing simulations that are truly reproducible. The latter is a particularly pressing issue when discussing multi-scale modelling, which typically involves multiple software tools that have a large number of version-specific dependencies. This does not ease the burden of the modeller to provide the strategies for the execution of the scale-specific models. However, the advent of web-based marketplaces for sharing of simulation models, such as nanoHUB (Klimeck and Zentner et al. 2011), can potentially remedy this as well.

5. Conclusions

Visual modelling is a powerful tool for the discussion and generation of alternative model structures. Top down, bottom up and inside outside designs are all possible. The idea of using networks of context-dependent networks provides the necessary richness to capture controlled chemical and biological processes and extends readily for to the representation of multi-scale systems and naturally to the generation of a multi-program environment supervised by a model orchestrator, a platform to handle dynamic workflows.

References

Elve, A T & Preisig, H A; 2017 A framework for multi-network modelling, PSE2018

Rosso, L. & de Baas, A. F. 2012; What makea a material function; European Commission.

MoDeNa 2017, http://modena.units.it

Karolius, S.; Preisig, H. A. & Rusche, H., 2016 Multi-scale modelling software framework facilitating simulation of interconnected scales using surrogate models; Comp. Aided Chem. Eng. 38, 463-468

Karolius, S.; Preisig, H. A. & Rusche, H.; 2017, Sequential Multi-Scale Modelling Concepts Applied to the Polyurethane Foaming Process; Comp. Aided Chem. Eng. 40, 487-492

Preisig, H. A. & Elve, A. T.; 2016 Ontology construction for multi-network models Comp. Aided Chem. Eng. 39, 2016, 1087-1092

Joback, K G & Reid, R C; 1987; Estimation of pure-component properties from group-contribution methods, Chem Eng Com, 57, 233-243.

G. Klimeck, L. Zentner, K. Madhavan, V.A. Farnsworth, and M. Lundstrom, “Network for Computational Nanotechnology – A Strategic Plan for Global Knowledge Transfer in Research and Education,” IEEE Nanotechnology Materials and Devices Conference, The Shilla Jeju, Korea, Oct. 18-21, 2011.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50362-1

Integration of Design, Scheduling, and Control of Combined Heat and Power Systems: A Multiparametric Programming Based Approach

Baris Burnak[a,b], Justin Katz[a,b], Nikolaos A. Diangelakis[a,b], Efstratios N. Pistikopoulos[a,b*]

[a]*Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station TX, 77840, USA*
[b]*Texas A&M Energy Institute, Texas A&M University, College Station TX, 77840, USA*
stratos@tamu.edu

Abstract

The operation of multiproduct/multipurpose processes involves decisions at different time scales; such as short-term for control, medium-term for scheduling, and long(er)-term for design. In this work, we present a methodology towards the unification of these decisions via multiparametric programming. Based on our previously introduced PARametric Optimization and Control (PAROC) framework, we derive (i) design dependent and scheduling aware control strategies, and (ii) scheduling strategies that are design dependent and control aware. We utilize the proposed methodology to simultaneously determine (i) the distributed control schemes for different operating policies and a range of different designs, and (ii) the coordination of operations in a network of two cogeneration of heat and power (CHP) units as a function of the constituent designs.

Keywords: process design, model based control, scheduling, multiparametric programming.

1. Introduction

Sequential assessment of process design and multi-scale operational activities often leads to suboptimal solutions as each individual problem dictates different, and sometimes conflicting, objectives. A unified formulation of the process design, scheduling, and control problems addresses the inherent interactions among these constituent decision making mechanisms. A general representation of such an integrated problem is presented in Eq. (1).

$$\begin{aligned} \min_{u,S,D,Y} \quad & J = \int_0^\tau P(x,y,u,S,De,Y,d,t)dt \\ s.t. \quad & \frac{d}{dt}x = f(x,u,S,De,Y,d,t) \\ & y_{min} \le y = g(x,u,S,De,Y,d,t) \le y_{max} \\ & u_{min} \le u = h(x,u,y,S,De,Y,d,t) \le u_{max} \\ & S_{min} \le S = m(x,y,S,De,Y,d,t) \le S_{max} \\ & Y \in \{0,1\} \\ & [x_{min}^T, De_{min}^T, d_{min}^T]^T \le [x^T, De^T, d^T]^T \le [x_{max}^T, De_{max}^T, d_{max}^T]^T \end{aligned} \tag{1}$$

where x are the states of the system, y are the system outputs, u are the manipulated variables, S are the scheduling decisions, De are the design variables, Y are the binary

decision variables, d is the system uncertainty, P is the objective function of the system accounting for the short term (control), medium term (scheduling), and longer term (design) costs, f and g are first principle equations, h and m are generic expressions that define the control and scheduling actions, respectively. Eq. (1) is a mixed integer dynamic optimization (MIDO) problem, which is usually highly nonlinear and large scale for chemical processing plants. The interactions between (i) process design and control variables, and (ii) control and scheduling variables in Eq. (1), and the efforts towards integration of these problems are reviewed and discussed in Vega et al. (2014) and Baldea and Harjunkoski (2014). Recent efforts are focusing on the integration of all constituent parts of Eq. (1) (Koller and Ricardez-Sandoval, 2017).

In this work, we extend our previous efforts towards the integration of process design, scheduling, and control (Diangelakis and Pistikopoulos, 2017) to a network of heat and power systems comprising two cogeneration of heat and power (CHP) units operating in parallel, accounting for their start-up and shut down costs. We utilize the PARametric Optimization and Control (PAROC) framework introduced by Pistikopoulos et al. (2015), and the dynamic high fidelity model of the CHP described by Diangelakis et al. (2014).

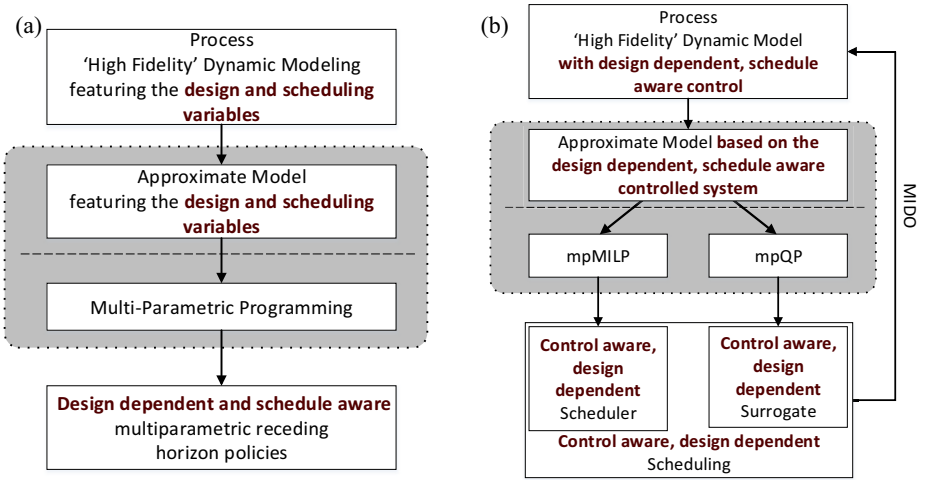


Figure 1. (a) Phase I: Methodology to build design dependent, schedule aware controller, (b) Phase II: Methodology to build design dependent, control aware scheduler. Actions within the gray area happen once and offline.

2. Integrated Design, Scheduling, and Control via the PAROC Framework

The PAROC framework allows for the design of advanced control strategies based on high fidelity models benefiting from the most recent advances in the field of multiparametric programming (Pistikopoulos et al., 2015). In this study, we utilize the PAROC framework to decompose the unified problem formulation in Eq. (1) into two phases, as depicted in Figure 1. The first phase, described in Figure 1a, focuses on the derivation of a control strategy that explicitly accounts for a range of design and scheduling options. The multiparametric solution to this problem allows for an offline optimal control strategy that can be embedded in the system of interest as described in Figure 1b. Accordingly, the Phase II focuses on building the control aware, design dependent scheduler, based on the optimal control strategies derived in the Phase I. Note

that the explicit solutions of the optimal control and scheduling problems are piecewise affine functions of the design De.

2.1. Phase I: Design dependent, schedule aware control strategy

Phase I consists of the following steps described in Figure 1a.

Step 1: "High fidelity" dynamic modelling – A first principle model for the CHP network is developed in the form of Differential Algebraic Equations (DAE) based on Diangelakis et al. (2014). The high fidelity model features the design and scheduling decisions as degrees of freedom of the overall problem. In this work, we use the gPROMS® environment to construct the model and simulate the dynamics of the system.

Step 2: Model approximation – The high fidelity model developed in the first step is highly nonlinear and nonconvex in most applications, and renders the practice of advanced control algorithms quite challenging. Therefore, we appeal to model reduction or system identification techniques to approximate the high fidelity model with a discrete time affine state space representation as presented in Eq. (2).

$$\begin{aligned} x_{t+1} &= A_c \cdot x_t + B_c \cdot u_t + C_c \cdot [De^T, S_t^T, d^T]^T \\ \hat{y}_t &= D_c \cdot x_t + E_c \cdot u_t + F_c \cdot [De^T, S_t^T, d^T]^T \end{aligned} \tag{2}$$

where index t indicates the discrete time steps, and $\hat{y}$ is the output of the system predicted by the approximate model. Note that Eq. (2) explicitly features the design and scheduling variables. In this work, model approximation is performed via the MATLAB® System Identification Toolbox™ using the N4SID algorithm, hence the states x may not preserve the physical meanings of the original states in Eq. (1).

Step 3: Design of the multiparametric model predictive controller (mpMPC) – The approximate model given in Eq. (2) is used to formulate an mpMPC based on the control scheme described in Diangelakis et al. (2016). We further treat the design and scheduling decisions (De and S) as bounded uncertain parameters. The solution of the mpMPC formulation yields an offline map of optimal control actions as affine functions of the design and scheduling decisions, as described by Diangelakis et al. (2017).

Step 4: Closed loop validation – The control strategy is validated via closed loop implementation against the high fidelity model derived in Step 1. The assessment of the closed loop performance, including set point tracking, fast adaptation to changes in the operational level, applicability on a range of different designs, constraint violation, and operational stability dictates whether a new approximate model is necessary.

2.2. Phase II: Design dependent, control aware scheduler

Production scheduling problem can be represented by a state space model as presented by Subramanian et al. (2012), and the derivation of its multiparametric counterpart is presented in Kopanos and Pistikopoulos (2014). Diangelakis and Pistikopoulos (2017) introduces a methodology to integrate the design and control into the schedule on a continuously operating CHP unit. In this work, we consider a network of two CHP units operating parallel, accounting for their start-up and shut-down costs in the schedule.

Step 1: High fidelity model with the closed loop dynamics – The design dependent, schedule aware control strategy is embedded in the high fidelity model.

Step 2: Model approximation – A discrete time affine state space model is derived based on the closed loop behavior of the high fidelity model via system

identification, the N4SID algorithm in this case. Note that the discretization step of the identified model is several orders of magnitude larger than the controller. The identified model is also resampled (via built-in MATLAB function `d2d`) in the time steps of the controller in order to formulate a surrogate model that bridges the time gap between the controller and the scheduler. The identified models in both discretization times are aware of the closed loop dynamics.

Step 3: Scheduling formulation and solution – Based on the approximate model derived in Step 2, we formulate an mpMILP problem accounting for the economic considerations and operational feasibility of the CHP network, as described in Kopanos and Pistikopoulos (2014). The scheduling formulation treats the disruptive scheduling events and the design of the CHP network as parameters. The start-up and shut-down time restrictions are taken into consideration by including lifting variables, as described by Subramanian et al. (2012). Design dependence of the approximate model imposes parametric bounds on the operation. Furthermore, in addition to our previous studies, the cost coefficients on the states of the system, and the start-up and shut-down decisions are affine functions of the design of the CHP network, which is treated as an uncertain parameter. The scheduling formulation yields explicit expressions of operational set point decisions as affine functions of the design, to be passed down to the controller. However, due to the approximation of the scheduling model and the large discretization time, there exists a plant-model mismatch. This discrepancy is addressed via implementing the surrogate model derived in Step 2 into a mpQP problem that recalculates corrective scheduling actions at the control time steps.

Step 4: Closed loop validation – The control strategy, surrogate model formulation, and the scheduler are operated simultaneously in a rolling horizon fashion against the original high fidelity model.

3. Design, scheduling, and control of a CHP network

The problem statement of this study is defined as follows:

Given: A high fidelity model of the CHP, described by Diangelakis et al. (2014), unit costs for (i) energy production, (ii) start-up and shut-down, (iii) purchasing, selling, and disposing of heat and power, and (iv) heat storage, and a residential demand profile for heat and power.

Determine: Operating levels and on/off switching times of the CHPs, quantity of heat to purchase/dispose, quantity of power to purchase/sell, and optimal control actions to achieve the target operating levels.

Objective: Minimize the total cost of operation.

For the methodology to acquire design dependent control strategies, the reader is referred to Diangelakis et al. (2017). In this work, we utilize these strategies to be embedded in the CHP network and the operating schedule.

The integrated control and scheduling strategies are cross validated against different CHP networks with different designs and operated simultaneously in a rolling horizon fashion. Figure 2 shows the Gantt chart of the schedule for a network with two CHPs with internal combustion engines (ICE) of sizes 1.5 L and 4.5 L, respectively. The schedule is given a residential heat and power demand to determine the optimal operating levels of the CHP network.

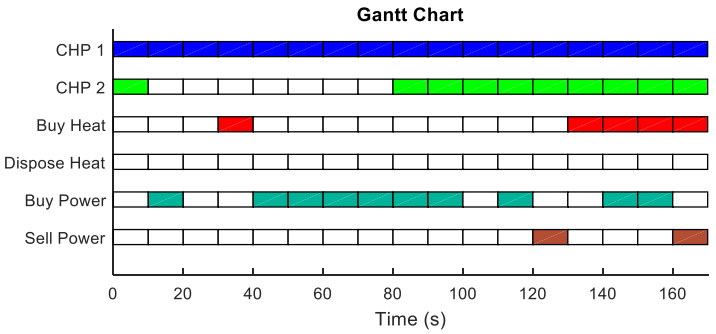


Figure 2. Gantt chart including CHP 1 ($V_{ICE} = 1.5\ L$), CHP 2 ($V_{ICE} = 4.5\ L$).

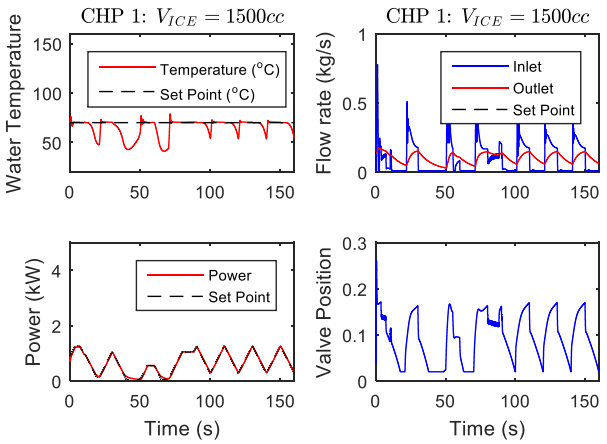


Figure 3. Closed loop response of the integrated schedule and control for CHP 1 ($V_{ICE} = 1.5\ L$).

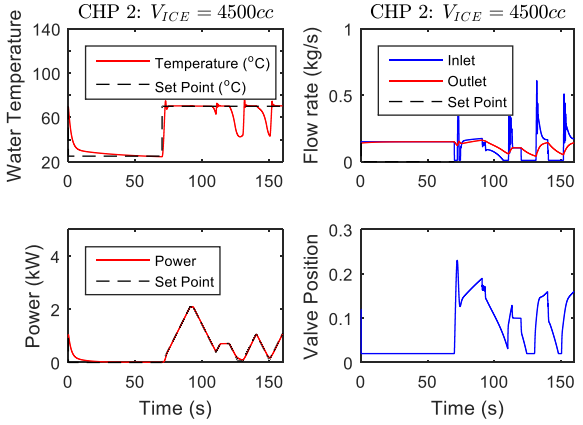


Figure 4. Closed loop response of the integrated schedule and control for CHP 1 ($V_{ICE} = 4.5\ L$).

Figures 3 and 4 show the closed loop response of the CHP units based on the scheduling decisions demonstrated in Figure 2. The responses of both CHP units show that design dependent controllers are able to track the set points provided by the scheduling level. The scheduler allocates different loads on the CHP units recognizing their closed loop behaviour, and their operating costs. For instance, the large CHP unit is switched off

during the early operation due to relatively low demand rate. However, it is switched back on and assigned a higher load than the small CHP unit as the demand increases.

4. Conclusions

In this work, we extended our previous study on simultaneous design, scheduling, and control on a single CHP unit to a network of CHP units, accounting for the start-up and shut-down times and costs. The design, scheduler, and the controller were built based on a single high fidelity model. The complete map of optimal operation in the short and medium term were derived offline via model based multiparametric programming. The frequently encountered time scale gap between the economic scheduler and the controller was bridged via a surrogate model. This work presents a framework to implement CHP network systems on an ensemble of residential units on a larger scale.

5. Acknowledgements

We acknowledge the financial support from the NSF (Grant no. 1705423), under the project “SusChEM: An integrated framework for process design, control and scheduling [PAROC]”.

References

M. Baldea, I. Harjunkoski, 2014, Integrated production scheduling and process control: A systematic review, Computers & Chemical Engineering, 71, 377-390.

N. A. Diangelakis, S. Avraamidou, E. N. Pistikopoulos, 2016, Decentralized multiparametric model predictive control for domestic combined heat and power systems, Industrial & Engineering Chemistry Research, 55, 12, 3313-3326.

N. A. Diangelakis, B. Burnak, J. Katz, E. N. Pistikopoulos, 2017, Process design and control optimization: A simultaneous approach by multi-parametric programming, AIChE Journal, 63, 11, 4827-4846.

N. A. Diangelakis, C. Panos, E. N. Pistikopoulos, 2014, Design optimization of an internal combustion engine powered CHP system for residential scale application, Computational Management Science, 11, 3 , 237-266.

N. A. Diangelakis, E. N. Pistikopoulos, 2017, Model-based multi-parametric programming strategies towards the integration of design, control and operational optimization, 27th European Symposium on Computer Aided Process Engineering, Computer Aided Chemical Engineering, 40, 1867-1872.

R. W. Koller, L. A. Ricardez-Sandoval, 2017, A dynamic optimization framework for integration of design, control and scheduling of multi-product chemical processes under disturbance and uncertainty, Computers & Chemical Engineering, 106, 147-159.

G. M. Kopanos, E. N. Pistikopoulos, 2014, Reactive scheduling by a multiparametric programming rolling horizon framework: A caseof a network of combined heat and power units, Industrial & Engineering Chemistry Research, 53, 11, 4366-4386.

E. N. Pistikopoulos, N. A. Diangelakis, R. Oberdieck, M. M. Papathanasiou, I. Nascu, M. Sun, 2015, PAROC – An integrated framework and software platform for the optimisation and advanced model-based control of process systems, Chemical Engineering Science, 136, 115-138.

K. Subramanian, C. T. Maravelias, J. B. Rawlings, 2012, A state-space for chemical production scheduling, Computers & Chemical Engineering, 47, 97-110.

P. Vega, R. Lamanna de Rocco, M. Francisco, 2014, Integrated design and control of chemical processes – Part I: Revision and classification, Computers & Chemical Engineering, 71, 602-617.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50363-3

Next Generation Multi-Scale Process Systems Engineering Framework

David C. Miller[a*], John D. Siirola[b], Deb Agarwal[d], Anthony P. Burgard[a], Andrew Lee[a], John C. Eslick[a], Bethany Nicholson[b], Carl Laird[b], Lorenz T. Biegler[c], Debangsu Bhattacharyya[e], Nikolaos V. Sahinidis[c], Ignacio E. Grossmann[c], Chrysanthos E. Gounaris[c], Dan Gunter[d]

[a]*National Energy Technology Laboratory, 626 Cochrans Mill Road, Pittsburgh, PA 15236, USA*
[b]*Sandia National Laboratories, P.O. Box 5800, Albuquerque, NM 87185-1326 USA*
[c]*Department of Chemical Engineering, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213-3890 USA*
[d]*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720 USA*
[e]*Department of Chemical and Biomedical Engineering, West Virginia University, 401 Evansdale Drive, Morgantown, West Virginia 26506-6070 USA*
david.miller@netl.doe.gov

Abstract

The IDAES PSE framework represents a new approach for the design and optimization of innovative steady state and dynamic processes by integrating an extensible, equation-oriented process model library with the Pyomo algebraic modeling language. Built specifically to enable rigorous large-scale mathematical optimization, the framework includes capabilities for conceptual design, steady state and dynamic optimization, multi-scale modeling, uncertainty quantification, and the automated development of thermodynamic, physical property, and kinetic submodels from experimental data.

Keywords: Multiscale modeling, conceptual design, process optimization, dynamic optimization, parameter estimation

1. Introduction

The U.S. Department of Energy's Institute for the Design of Advanced Energy Systems (IDAES) was formed in 2016 to develop new advanced process systems engineering (PSE) capabilities to support the design and optimization of innovative new processes that go beyond current equipment/process constraints. Engineers currently rely on state-of-the-art simulation packages to synthesize, model, and optimize process flowsheets with sufficiently detailed models. However, commercial simulation tools struggle with flowsheet optimization and typically limit consideration of nonlinear optimization problems to less than 100 degrees of freedom. As a result, process optimization with large-scale solvers is rarely performed by industry. On the other hand, many of the capabilities for large-scale optimization are available in general algebraic modeling languages (AMLs), which generally require fully open, equation-oriented (EO) models with exact first and second derivatives. Large-scale NLP algorithms routinely solve optimization problems with thousands of variables and degrees of freedom. Moreover, the extension of these optimization models for MINLP, global optimization, and optimization under uncertainty (OUU) requires a flexible and open optimization modeling environment.

The open-source IDAES process systems engineering framework addresses the capability gap between state-of-the-art simulation packages and AMLs by integrating an extensible, equation-oriented process model library with the Pyomo AML (Hart et al, 2011, 2017), which enables formulating, manipulating, and solving large, complex, structured optimization problems. The IDAES framework includes tools for: (1) process synthesis and conceptual design, including process intensification, (2) process design and optimization, including process integration, (3) process control and dynamic optimization, (4) using advanced solvers and computer architectures, (5) automated development of thermodynamic, physical property, and kinetic submodels from experimental data, (6) integration of multi-scale models, (7) comprehensive, end-to-end uncertainty quantification (UQ) and model validation, (8) maintaining complete provenance information, and (9) the ability to support multiple time and length scales, from materials to process to market.

2. IDAES PSE Framework

The IDAES framework consists of multiple components as shown in Figure 1. The core of the IDAES framework is the customizable process model library, built on the Pyomo AML. The core models can be combined to capture the hierarchical structure of sub-processes and complex unit models. In addition, the process model library can be used to support conceptual design via a superstructure formulation, process design, optimization, and integration, and dynamic optimization and control. A suite of machine learning techniques based on ALAMO (Cozad et al., 2014) enables the development of submodels for physical properties, thermodynamics and kinetics directly from experimental data, enabling more effective tracking of uncertainty. To enable the solution of large-scale parameter estimation problems, the machine learning tools can utilize a variety of high performance computing (HPC) capabilities, include GPUs and the National Energy Research Scientific Computing Center (NERSC). Finally, the framework utilizes PySP (Watson et al., 2012) to incorporate and assess uncertainty across models and scales as well as to enable simultaneous, multiscale parameter estimation. All components are developed using open-source best practices on the Github collaborative coding platform.

2.1. Pyomo Algebraic Modeling Language

The IDAES PSE Framework leverages the open source, DOE-funded platform, Pyomo (Hart et al., 2011, 2017), which addresses challenges in formulating, manipulating, and solving large complex structured optimization problems. Pyomo is an extensible algebraic

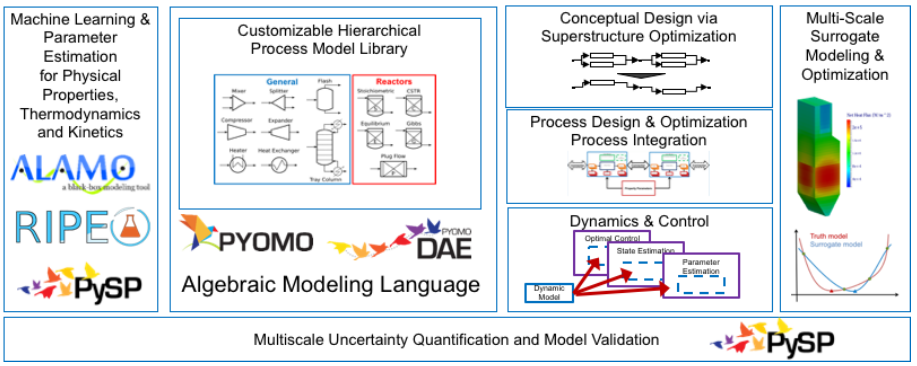


Figure 1. The major components and connection of the IDAES PSE Framework to support conceptual design, process design & optimization, process integration, dynamic optimization, and control while incorporating multi-scale surrogate models.

modeling environment written in the Python® programming language. It allows users to formulate, initialize, solve, and manipulate large-scale optimization problems with concise notation all within the Python ecosystem. Pyomo provides interfaces to dozens of optimization solvers, both open source and commercial, and supports executing solvers both locally and in distributed environments. Most importantly, Pyomo is extensible beyond "standard" algebraic optimization models. In particular, Pyomo provides extensions for block-oriented modeling and Generalized Disjunctive Programming (Friedman, et al., 2013), stochastic programming (Watson, et al., 2012), and dynamic systems containing (partial) differential-algebraic equations (Nicholson, et al., 2017). This enables modeling flexibility and the consideration of constraints and objective functions in non-standard forms not easily handled by other modeling frameworks. Further, Pyomo enables complex analyses by combining these extensions and capabilities with the ability to script custom algorithms and solution processes.

2.2. Hierarchical Equation Oriented Process Models

A major strength of commercial simulation packages is their library of unit models and thermo-physical properties. Such simulation packages, however, often have limited support for incorporating non-standard, dynamic unit models, and they do not directly support UQ. AMLs, on the other hand, are eminently flexible and readily support large-scale optimization, but considerable work is required to construct process models, which are often not easily reused. To overcome this issue, the IDAES framework extends Pyomo's block-based hierarchical modeling constructs to create a library of models for common process unit operations and a framework for rapid development of process flowsheets. Rather than developing highly detailed, but rigidly structured models, the IDAES model library consists of simpler, but extensible, models for each unit operation (Lee et al., 2018). The model library contains an internally consistent suite of interoperable models at different levels of complexity and accuracy for both steady-state and dynamic analysis. Models within the library can be extended with additional detail or rigor as needed, allowing full control of the complexity and accuracy of the models. This focus on adaptability and extensibility enables them to be used for novel processes and configurations.

2.3. Machine Learning and Parameter Estimation

Although commercial simulation tools have extensive databases and pre-regressed models for thermo-physical properties, when investigating innovative, new processes with new materials, such databases are of limited use. In addition, it is not always straightforward to assess the uncertainty inherent in these models. Thus, tools for modeling physical property and reaction kinetics have been developed based on ALAMO, which builds algebraic models that work well in an optimization framework. Building these models directly from data (either from literature or experiments) ensures they address the right range of conditions and enable tight coupling of uncertainty estimates that allow end to end UQ. In addition, through the data management framework, provenance can be more clearly established enabling tracking through the development of a new technology.

The Reaction Identification and Parameter Estimation (RIPE) tool can estimate kinetic reaction parameters from data describing the time evolution of concentrations of chemicals in a reaction system that may include reversible and irreversible reactions as well as homogeneous and heterogeneous catalytic mechanisms. ALAMO is first used to obtain algebraic models of concentration profiles in terms of polynomial, exponential, and logarithmic basis functions, then the constrained regression features of ALAMO are used for

data reconciliation by requiring the time profiles to satisfy physical bounds on concentrations – even over time domains for which concentration measurements do not exist (Wilson and Sahinidis, 2017). Finally, RIPE utilizes integer programming techniques to obtain explicit mathematical representations of the dynamics of the system.

More broadly, parameter estimation for multi-scale systems is a significant challenge since only a subset of parameters can be estimated from data at a given scale. Complex interactions among the parameters can lead to biased estimates unless they are estimated simultaneously. This can lead to a large-scale optimization problem with data that span multiple scales and contain inherent noise. Traditional modeling approaches are often not amenable to such a simultaneous approach since process models and physical property/thermodynamic models are often developed separately. In the IDAES framework, all models and sub-models are implemented in a common platform allowing access to all model parameters, enabling formulation of the complete, simultaneous multiscale problem. The problem can then be readily solved using the capabilities of the PySP module (Watson et al., 2011). The IDAES framework can also be used to develop optimal design of experiments for minimizing the variance of the parameter estimates.

2.4. Process synthesis and conceptual design

The major goal in process synthesis and conceptual design is to optimize, reliably and efficiently, superstructures of process systems to help identify novel process configurations and enable process intensification. These remain major research challenges (Chen and Grossmann, 2017). The approach relies on Generalized Disjunctive Programming (GDP) (Grossmann and Trespalacios, 2013) in which a model consisting of equations, disjunctions, and logic statements is used instead of formulating and solving a single monolithic MINLP model. Thus, only NLP subproblems corresponding to fixed flowsheet structures need to be solved, which greatly increases the robustness of the nonlinear optimization. Furthermore, the conceptual design algorithms in IDAES directly leverage the native support of Generalized Disjunctive Programming in Pyomo. Several test problems have been solved successfully with this approach, including a simplified version of the chemical looping process.

2.5. Process design and optimization

The IDAES framework enables large-scale system-wide optimization using rigorous models of process equipment and thermophysical properties. Complex interactions between multiple process systems can be explored simultaneously without the need for extensive simplifying assumptions that render the problems tractable at the risk of compromising the results. For example, both traditional and advanced combustion technologies such as oxycombustion or chemical looping combustion can be optimized together with their power islands and pollution control systems without relying on fixed steam cycle designs, the use of overly simplistic “Gibbs reactors” that cannot capture key nonlinear responses to operating conditions, or fixed design points from CFD simulations that overlook critical decision variables. Methods for enabling optimization over broad ranges of operation have also been developed that can handle non-smooth features such as vapor/liquid or sub/supercritical phase transitions as well as vanishing phases. The process design and optimization capabilities developed within IDAES are applicable not only to the energy sector but also to several other industries (e.g., chemicals, petrochemicals, pharmaceuticals) where tight integration is desired.

2.6. Dynamics and control

The dynamic process simulation and optimization capability builds on Pyomo's support for expressing ordinary and partial differential equations as large-scale systems of nonlinear equations. The IDAES process model library includes both steady-state and dynamic versions of common unit models allowing users to switch between steady-state and dynamic analysis with minimal changes to their flowsheet implementation. In addition, the IDAES framework includes capabilities for state estimation and control which enable the rapid deployment of approaches such as moving horizon estimation (MHE) and nonlinear model predictive control (NMPC) to arbitrary process flowsheets.

2.7. Multi-scale surrogate modeling and optimization

Even within a powerful equation-oriented platform for process optimization, complex, black box models that resist a direct equation-based description must often be included. These include CFD models for advanced combustors, reactors and separators as well as complex physical property models. When incorporating simplified, surrogate models of such complex models, optimizers often exploit the weaknesses of surrogate models and fail to find the true optimum. To guarantee convergence to the optimal solution, we have developed a trust region filter algorithm, which monitors and controls the errors in the reduced model and guarantees convergence to first order critical points of the complex model. The trust region filter algorithm, based on the SQP filter method along with concepts of Derivative Free Optimization, is developed, analyzed and demonstrated on challenging power plant models in Eason and Biegler (2016). Now implemented in Pyomo, it enhances a powerful optimization capability to include complex black box models.

2.8. Data management

The IDAES Data Management Framework (DMF) is responsible for managing the knowledge and provenance required for the synthesis, design, optimization, and scale up of innovative new processes. The DMF stores property data, inputs, model parameters, model codes, and results in an easily searchable form. Standard representations of provenance (e.g., journal publications and websites) support validation of results. The DMF's Python API can be used within scripts or for interactive analyses in Jupyter Notebooks. The Jupyter Notebooks are themselves stored in the DMF, providing a self-contained reproducible unit of analysis. The API also enables efficient access to large scale data sets for parameter estimation to other components of the IDAES PSE framework.

2.9. Other applications

The core approach of using data to develop submodels, which can then be utilized to formulate larger-scoped optimization problems, has also be applied to materials design (Hanselman and Gounaris, 2017) and multi-scale infrastructure planning (Lara et al., 2017). In the case of designing crystalline materials, for example, computational chemistry data is used to develop simplified structure-function relationships linking atomic neighborhood descriptors to the reactivity of atomic sites. These relationships can then be incorporated into an optimization model that seeks to identify the overall material structure that performs best in the context of the process of interest. In the case of infrastructure planning, the IDAES framework has been used to determine representative dispatch days and growth forecast for an electricity grid, which forms the basis for determining what types of power plants should be built to most economically meet all the expected demand.

3. Conclusions

In conclusion, this paper has described a comprehensive, open source framework to enable large scale process optimization of complex, innovative processes spanning conceptual design and process synthesis to dynamic optimization. The capabilities of this framework are applicable at the process scale as well as for materials design and infrastructure planning.

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. KeyLogic Systems, Inc.'s contributions to this work were funded by the National Energy Technology Laboratory under the Mission Execution and Strategic Analysis contract (DE-FE0025912) for support services. This material is based upon work supported by the U.S. Department of Energy, Office of Science, under Contract No. DEAC02-05CH11231. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

Q. Chen, I.E. Grossmann, 2017, Recent developments and challenges in optimization-based process synthesis, Annual Review of Chemical and Biomolecular Engineering, 8, 249-283.

A. Cozad, N.V. Sahinidis and D.C. Miller, 2014, Learning surrogate models for simulation-based optimization, AIChE Journal, 60, 2211-2227.

J.P. Eason, L.T. Biegler, 2016, A trust region filter method for glass box/black box optimization, AIChE Journal, 62, 3124-3136.

Z. Friedman, J. Ingalls, J.D. Siirola, and J.-P. Watson, 2013, Block-oriented modeling of superstructure optimization problems, Computers & Chemical Engineering 57, 10-23.

I.E. Grossmann, F. Trespalacios, 2013, Systematic Modeling of Discrete-Continuous Optimization Models through Generalized Disjunctive Programming, AIChE J. 59, 3276-3295.

C. L. Hanselman, C.E. Gounaris, 2016, A mathematical optimization framework for the design of nanopatterned surfaces, AIChE Journal 62, 9, 3250-3263.

W.E. Hart, J.-P. Watson, and D.L. Woodruff, 2011, Pyomo: modeling and solving mathematical programs in Python, Mathematical Programming Computation 3, 3, 219-260.

W.E. Hart, C.D. Laird, J.-P. Watson, D.L. Woodruff, G.A. Hackebeil, B.L. Nicholson, and J.D. Siirola, 2017, Pyomo – Optimization Modeling in Python. Second Edition. Vol. 67. Springer.

C.L. Lara, D. Mallapragada, D. Papageorgiou, A. Venkatesh, I.E. Grossmann, 2017, Electric Power Infrastructure Planning: Mixed-Integer Programming Model and Nested Decomposition Algorithm, Submitted for publication, 1-36.

A. Lee, J.H. Ghouse, Q. Chen, J.C. Eslick, J.D. Siirola, I.E. Grossman, D.C. Miller, 2018, A Flexible Framework and Model Library for Process Simulation, Optimization and Control, Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018, July 1-5, 2018, San Diego, California, USA.

B. L. Nicholson, J. D. Siirola, J. P. Watson, V. M. Zavala, L. T. Biegler (2017). Pyomo.DAE: A modeling and automatic discretization framework for optimization with differential and algebraic equations. Mathematical Programming Computation, 10.1007/s12532-017-0127-0

J.P. Watson, D. L. Woodruff, and WHart, 2012, PySP: modeling and solving stochastic programs in Python. Mathematical Programming Computation 4, 109-149.

Z.T. Wilson, N.V. Sahinidis, 2017, The ALAMO approach to machine learning, Computers and Chemical Engineering, 106, 785-795.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50364-5

A framework for multi-network modelling

Arne Tobias Elve*, Heinz A. Preisig

Department of Chemical Engineering; NTNU; Trondheim, Norway
arne.t.elve@ntnu.no

Abstract

We present a new modelling framework called the ModelComposer. This framework combines well-known concepts such as equation-based modelling, graph-based modelling and code generation to produce simulation models automatically. Moreover, the framework uses ontologies as the bases, which means that interoperability of models produced by the modelling framework is high. The model information is stored in JSON-files, and code generation for different programming languages is used to transform the defined models into executable simulation. A result is a modelling tool that promotes the reuse of models and interoperability between networks of models of different domains. This paper will present the overall layout of the ModelComposer and how the framework handles internal conversions within a network and how different networks communicates. This is exemplified by a case study presenting the model construction of a dynamic mixing tank and the simulation of the resulting model.

Keywords: Graph-based Modelling, Automatic Code Generation, Model-Based Simulation, Customized Modelling, Multi-Scale Modelling

1. Introduction

Complex networks are present in many chemical and biochemical systems. In addition to the complexity of networks, there is also the possibility of reaction networks, such as combustions and cell metabolism. These reaction networks are of particular interest because of the emergence of new chemistry, for example, the conversion of biomass to fuels. There exist a variety of mathematical models to describe specific parts, but the challenge is to facilitate the interaction between these networks (Bogusch et al., 2001), especially when the networks operate on different scales.

The motivation for the development of our modelling framework is founded upon the objectives of generating proper and consistent models based on the principles described in (Preisig, 2010), automatically. We also want a minimization of the modelling errors and absolutely no code implementation errors in the desired output language. This also implies that we remove the hand coding of models and focus on the model formulation and not wasting time on programming technicalities.

This approach to modelling is closely related to MOSAIC (Kraus et al., 2014) and (Fedorova, 2015), which both populate model templates. We separate from their work since we extract our model knowledge from an ontology and then unite the knowledge into code based on context instead of using templates designed for special model units. Hence, we are conceptually closer related to (Brandt et al., 2008) and (Koo et al., 2017) that uses ontologies for knowledge management. We combine the equation template approach with the knowledge from the ontology to generate secure and correct models automatically.

2. The ModelComposer

The ModelComposer is a new modelling tool implementing a visual modelling language described in (Elve and Preisig, 2017) to a multi-network modelling environment, with the concepts described in (Preisig et al. 2018). The workflow used in the ModelComposer is illustrated in figure (1).

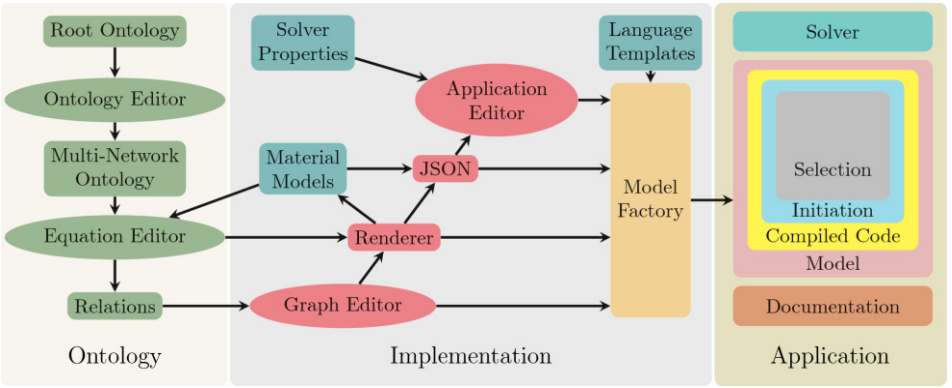


Figure 1. Illustration of the essential workflow for generating executable model code in the ModelComposer. The boxes are tasks required in the modelling procedure, the ellipses are editors that are controlled by a user and the arrows indicate a flow of information.

The framework consists of three parts that that are grouped under the terms ontology, computer implementation and the produced application containing the generated model, which is the output of the tool. We define an ontology to be a naming and definition of "things", properties and relations that exist for a particular domain. Molecules, phases, tanks and energy are examples of such "things" in the context of chemical engineering. The properties are attributes that provide the metric for the "things", and the relations describe how properties and different "things" are related. The relations are what we usually term equations and is how the relations are implemented in the produced model. For the ModelComposer, we have chosen to use a directed graph for the representation of the model topology. This means that the ontologies we construct are tailored for abstraction of a model into a topology. The construction of the ontologies was covered in (Preisig and Elve, 2016). For the ModelComposer, we start off with a root ontology, which we specialise so that we describe the domain in which the model lives. In the next step, we feed the ontology to an equation editor that aids the user in defining the relations of the ontology with structural and unit consistency checking.

The implementation part is where "things" from the ontology are used and connected such that the "things" interact with each other in a structure. This structure represents a network model, which we implement as a directed graph. The nodes of this graph represent the capacities, and the arcs represent interactions of quantities that characterise the capacities. A special graph editor is made for constructing these network structures. The resulting directed graph is then analysed, and the pieces of the graph stored as JSON files maintaining the object structure of the graph. The user then has to choose which application and what solver with properties that also should be implemented. All this information is given over to a compiler that use some language templates described in (Elve and Preisig, 2017) to transform the structure into code in the output language.

The generated model code consists of an initiation procedure, which requires that the user provide constants and initial conditions for the state description. If any alternative calculation method exists for a property, the user also has to select what method the code should use for the calculation. The compiled code are equations that describe the dynamic behaviour of the model. This model is coded into a selected application (typically simulation) that contains model documentation, the executable model code and solver packages required for solving the model. The product is a standalone application written in essentially any output language, but for now, limited to Python.

3. Ontologies and the network view

The modelling methodology involves capturing the behaviour of a set of characterizing quantities that interact with each other in a structure. For a physical system are these characterising quantities the conserved quantities. To have an understanding across different scales, we refer to these quantities as tokens. The term token comes from the analogy of Petri-nets where tokens are contained and move around within in the nets.

To capture the behaviour of multi-network models, we associate attributes of a network to specific layers in our ontology. When constructing the ontology, we start off with defining a root definition and then sequentially refine layers to form a tree of sequentially refined ontology branches. Within a leaf network in the model, all nodes and arcs are described within that refined branch in the ontology. Figure 2 illustrates the hierarchical decomposition of our ontology. Interactions between the networks are then defined on the latest shared leaf in the branch between the two networks. The illustration in figure 2 is an example of a decomposition of an ontology for chemical processes. The structure can be modified to the required specifications.

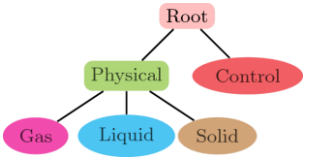


Figure 2. Network based decomposition of an ontology.

3.1. Tokens and typed tokens

We must describe the conversion of species for the modelling method to apply to chemical and biological processes. Mass, which is a conserved quantity and thus a token, consists of species, with species representing a type of mass. We consequently introduce typed tokens, which structurally can capture modelling entities such as species. The associated dynamic equations are then balances, in contrast to conservation. For species the link between them is well defined, namely the molecular-mass weighted sum of the molar species yields the total mass, described in equation (1).

$$m = \lambda^T n \tag{1}$$

The mass, m, is a scalar property in the node, while the molar mass, lambda, and species, n, are vector properties based on the amount of species in the node, which reduces in the multiplication. Nodes and arcs are already implemented as vectors, and by introducing typed tokens as a vector within a node or an arc, we get a vector of vectors structure in the mathematical implementation. This vector of vectors representation is very suitable for reaction networks (Horn et al., 1972). For example, we can use equation (2) when calculating the internal conversion of the entire network.

$$\tilde{n} = VN^T g(\mu) \tag{2}$$

$\tilde{n}$ represents the change rate of amount of moles due to chemical reaction. V is the volume, g(mu) is the extent of reaction in the node based on the chemical potential, and the N matrix captures a complete reaction network mapped into the physical network. The dimensions of the matrix are stacked reactions in nodes and stacked species in nodes. This capturing of the reaction network also applies to metabolic networks.

3.2. *Token and typed token propagation*

The objective of the modelling framework is to generate executable program code for a model. For the instantiation, we first have to calculate the propagation of tokens, which describes how the tokens are spread throughout the multi-network graph. This propagation is calculated before any compiling of code. The rules we have implemented in the ModelComposer is that tokens have to be introduced through reservoirs. Reservoirs are particular nodes where the numerical value of quantities is constant for the dynamic scale of the simulation model. From the reservoir, tokens propagate through the arcs describing the interaction of a particular token between two nodes. This procedure is recursively repeated until the graph is properly instantiated. Some rules are describing what tokens are propagated. For examples does mass also carry energy, whilst energy transport does not imply mass transport. The typed tokens follow the token. For the chemical species, a species can appear in a node if a chemical reaction is defined and all the reactants are present in the node. This node is now the origin of that species, and a new recursive token propagation procedure is initiated from this node. From the species distribution, we can also determine possible reaction networks.

When a token propagates from a network to another, the token has to be translated from token in a network to a token that is defined for the network on the other side of the boundary. We do this by selecting an already defined translation from one side to the other. This now generates an algebraic link that describes the translated token in the scales of the present network. The newly propagated token on the other side then initiates a new propagation routine.

3.3. *Scale coupling networks: Property transaction via ports*

The tokens determine the network definitions and tokens can only be present if they are defined for a particular network. Sometimes there are in some properties of a different network that determine how a token behaves in the other network. An example is temperature measurements, which are transmitted to a control system. A temperature is not described in a signal processing network. Thus it has to transacted as a measurement from the physical system over to a signal in the control system. We, therefore, introduce ports that are the connection point to other networks. The ports with the transactions between them are responsible for handling the broadcasting of variables from one network to the other. For example, the current example illustrated in figure (3). The measurement of a property is to be transacted from a network, denoted by α, of a physical nature defined in a continuous time frame. The transaction translates the measurement over to a signal, which is valid in the connected

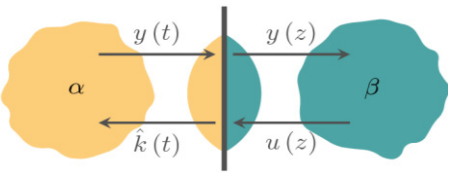


Figure 3. Illustration of transaction of a property between two networks. They are connected with a transaction system containing a algebraic link of a property between two

control network, denoted by β, described in discrete time. The control output from the discrete control network is transacted over to the physical network to a number that represents an updated value for a valve position. For a typed token, we use the same approach, but the link between them is defined in the tokens propagation. This also allows the user to inhibit transaction of some species, which reduce the mathematical space and then simplifies the model. For a scale connection between a physical network on a macroscale, state information can be broadcasted over to a network described on microscale doing molecular calculations. All variables are then available for that network due to the heritance.

4. Case study: Dynamic model of a mixing tank

The following case study highlights the concepts of the modelling framework. The simulation results are from the automatically generated code from the ModelComposer.

4.1. Model implementation

The tank consists of a liquid phase and a gas phase open to the surroundings. The feeds into the tank, illustrated in figure 4, are one acidic with pH 4 and one basic of pH 10, where the compositions of the solutions were set into the reservoirs denoted by A (acid) and B (basic). The reactions are assumed to be equilibrium reactions, which takes place in the liquid (L). The level in the tank is controlled by a level controller, that measures the level and calculates the new position of the valve opening of the outflow, following the same methodology presented in figure 3. We assume perfect mixing in the liquid (L), so we can use a lumped node for the representation. The two inflows are determined by the user, and the flow out from the tank is pressure driven and controlled.

The topology was constructed in the graph editor and the token propagation was carried

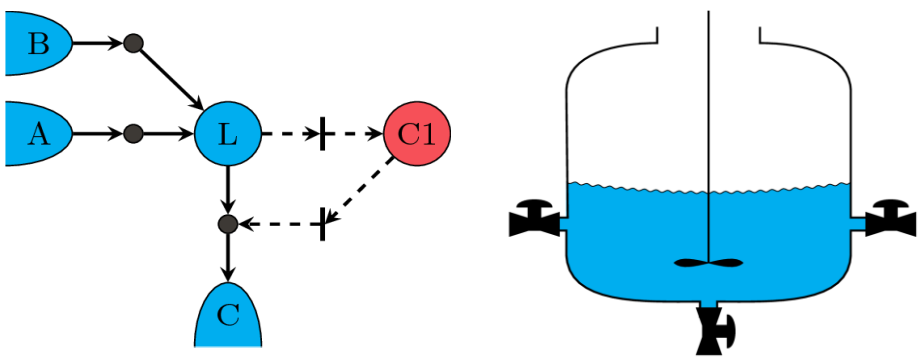


Figure 4. Illustration of the modelled mixing tank and the corresponding topology generated.

out and based on this the reaction systems were selected. Initial conditions were given to the states and constants. The initial volume of the tank was set to 0.5 m^3.

We used the same automatically generated model with three different initial conditions. For the first simulation, we set the initial composition of the tank to be of pH = 10 and only have an inflow of acid. For the second simulation, we set the initial composition of the tank to be of pH = 4 and just add basic solution. For the final simulation we kept the initial composition of the tank to be of pH = 4, but now we added equal amounts of the two solutions. We simulated the tank for 100 seconds and used the Runge-Kutta method as integrator and a root solver for obtaining the equilibrium conditions in the reactions. The results from these simulations are displayed in figure 5. Where the pH in the tank is plotted over time for the three different initial conditions.

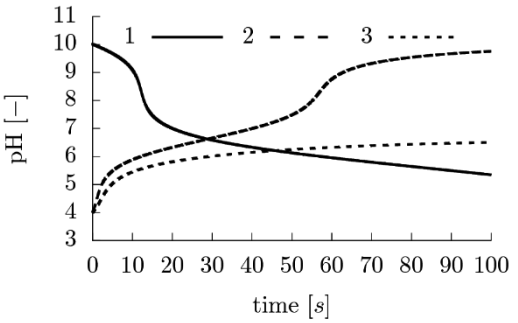


Figure 5. Results from the three simulations.

5. Conclusions

This paper presents the basic concepts in which a modelling tool named the ModelComposer is built upon. The modelling tool utilize an ontology that captures the system knowledge for formulation of models. Interactions between networks are handled in transaction systems containing an algebraic link that defines the transacted property with scaling on both sides in the network boundary. These concepts were used in the study of reactive mixing tank. The code for this system was automatically generated with the ModelComposer.

References

R. Bogusch, B. Lohmann, W. Marquardt, 2001, Computer-aided process modeling with ModKit, Computers and Chemical Engineering 25, 963–995

S. C. Brandt, J. Morbach, M. Miatidis, M. Theißen, M. Jarke, W. Marquardt, 2008, An ontology-based approach to knowledge management in design processes. Computers & Chemical Engineering 32, 320–342

A. T. Elve, H. A. Preisig, 2017, From Ontology to Executable Program Code. In: Computer Aided Chemical Engineering. Vol. 40. pp. 2317–2322

M. Fedorova, G Sin, R. Gani, 2015, Computer-aided modelling template: Concept and Application. Computers & Chemical Engineering 83, 232–247

F. Horn, R. Jackson, R. Ares, 1972, General mass action kinetics. Archive for Rational Mechanics and Analysis. Vol. 47. (2), pp. 81-116

L. Koo, N. Trokanas, F. Cecelja, 2017, A semantic framework for enabling model integration for biorefining. Computers and Chemical Engineering 100, 219–231

R. Kraus, S. Fillinger, G. Tolksdorf, D. H. Minh, V. A. Merchan-Restrepo, G. Wozny, 2014, Improving Model and Data Integration Using MOSAIC as Central Data Management Platform. Chemie Ingenieur Technik 86 (7), 1130–1136

H. A. Preisig, 2010. Constructing and maintaining proper process models. Computers & Chemical Engineering 34 (9), 1543–1555

H. A. Preisig, A. T. Elve, 2016. Ontology construction for multi-network models. In: Computer Aided Chemical Engineering. Vol. 38. pp. 1087–1092

H. A. Preisig, A. T. Elve, S. Karolius, 2018, Visual Modelling with Networks, PSE 2018

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50365-7

Approaches to Multi-Scale Modeling from Systems Engineering

Sigve Karolius*, Heinz A. Preisig

Norwegian University of Science and Technology, Sem Sælandsveg 4, Trondheim 7491, Norway
sigve.karolius@ntnu.no

Abstract

The concept of multi-scale modeling has been embraced for different purposes in various branches of science and engineering. In this context, the importance of process-systems engineering is self-evident considering that the overall goals remain to improve or invent novel products and processes through optimization and simulation-guided studies. This paper aims at presenting an overview of past and current approaches to categorize multi-scale models according to scale-integration strategies.

Keywords: multi-scale modelling, surrogate modelling.

1. Introduction

The classic approach to modeling, particularly in engineering, focuses on developing deterministic models based on conservation principles for the purpose of design and optimization of equipment and entire processes. The emphasis in Process-Systems Engineering (PSE) on establishing a diverse and solid foundation has enabled control- and systems-engineers to cover the entire range from model development to computer implementation and optimization.

However, in the domain of multi-scale modeling, this approach quickly becomes impractical. The reason is twofold: firstly, the inherent complexity of the physics and modeling paradigms themselves; secondly, the modeling frameworks required to perform efficient simulations are naturally developed for domain-specific use in a variety of programming languages and are computationally expensive. Therefore, approaches to integrating multi-scale models into optimization frameworks is based on replacing the detailed models with simplified surrogates as presented in (Biegler and Lang et al. 2014).

The purpose of this work is to present current approaches to multi-scale modeling and simulations, as well as reviewing the conceptual work that the PSE community has contributed to the domain. This is relevant for PSE as the material science community is currently focusing major efforts into formalizing documentation, reproducibility of simulation and defining an ontology for materials-modeling as discussed in (Rosso and de Baas 2012).

2. Multi-Scale Perspectives

The multi-scale nature of the physical world has been recognized and utilized as solution strategies in science and engineering. The fundamental idea is that the dynamics for a model on scale "i" are slow compared to the scale below and fast

compared to the one above. Consequently, the apparent dynamics of the lower level is that of an "event", and the scale above is a constant reservoir.

Figure 1: Illustration of models that are scale-separated.

The multi-scale modeling paradigm attempts either to utilize this scale separation or remedy the lack thereof. The work of formalizing categories for multi-scale models has been approached by the PSE community, such as in Pantelides (2001) and later by Ingram and Cameron (2004). Moreover, the review paper of Vlachos (2005) presents concepts and examples from Chemical Engineering. The later also emphasizes the need for systems engineers to facilitate optimization and control for effective utilization of these models in making business and design decisions.

2.1. Coupling and Linking Scales

All of the aforementioned references emphasize that multi-scale modeling poses some additional challenges and that the refinement strategy depends on not only the perspective and desired level of detail, but how models on different scales are intertwined. This will guide the choice of what Pantelides called the "scale-integration" strategy employed.

Ingram states the coupling can be performed from the following perspectives: top-down, bottom-up, concurrent and middle-out. Regardless of perspective, the relationship between two adjacent scale-specific models can be illustrated as shown in Figure 2. The scale coupling from the top-down involves homogenizing the state of the macroscopic scale and subsequently lifting the output from the microscale model.

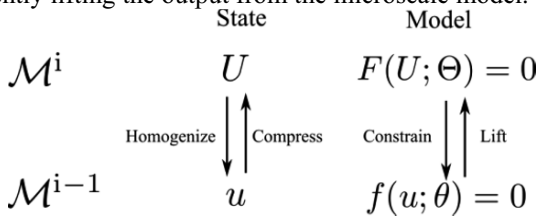


Figure 2: Illustration of the relationships between two coupled scale-specific models.

From a simulation perspective, the homogenization/lifting could be the result of sequential simulations, or the result of processing happening on-the-fly between concurrent simulations. In (Rosso and de Baas 2012), sequentially coupled models are referred to as "linked", whereas the term "coupled" is reserved for concurrently coupled models.

2.2. Model Construction

Constructing a multi-scale model requires at least one person that is familiar with both the overall process and the sub-processes that make up the inter-scale connections. However, there are methodologies that aim at formalizing the construction of different types of scale-connections. For engineering applications, the details of the microscale models are often not of interest and a "top-down" approach can be used. The heterogeneous multi-scale method (HMM) from (Weinan and Engquist et al. 2003), is a methodology for coupling macro- and microscopic models by homogenizing the macroscopic state and lifting the result of the microscopic model at regular intervals.

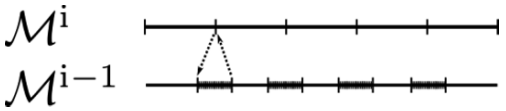


Figure 3: Illustration of the hierarchical multi-scale method (HMM). The model on the macroscopic scale employs the microscopic model, re-instantiating it at every time-step.

In HMM, Weinan and Engquist differentiate between models where the multi-scale nature is due to localized events, e.g. cracks, and those where constitutive information is missing globally and supplied by microscale models. Either way, the microscale model is initialized and averaged around each sampling point. Alternatively, the "top-down" connection can be constructed by scaling the time-step of the adjacent models in order for the simulations to run concurrently. This is the approach of the "seamless" multi-scale method presented in (Weinan and Ren et al. 2009).

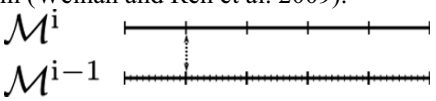


Figure 4: Illustration of the "seamless" multi-scale method. Simulations run concurrent using different time-steps allowing for information exchange at regular points throughout the interval.

At the other end of the spectrum there is the "bottom-up", or simultanious coupling, perspective, which aims at reconstructing the macroscopic model in its entirety using microscale simulations. The so-called "equation-free" approach from (Kevrekidis and Gear et al. 2003) is a formal way of achieving such a coupling by only using the macroscale model to supply information that in unobtainable from the lower scale. The bottom-up approach has the advantage of retaining details of the microscopic scale that would be lost in the HMM. There are also specialized forms of "bottom-up" approaches focusing on course-graining, e.g. modeling a solvent as a bulk-phase rather than molecules.

2.3. Model Formalization

The approach of categorizing scale connections is a useful start for constructing a multi-scale model, but a general approach requires further abstraction. The work of (Yang and Marquardt 2009) takes an even more generic approach, aiming at a formal framework for conceptualizing all multi-scale models. Yang and Marquardt address scale-integration rigorously and is recommended for anyone entering the domain.

Transistivity implies that the properties of a thing "x" contains the set of its constituent parts, here: "y" and "z". Alternatively, the ontology can incorporate the hierarchical relationship "direct parthood", meaning that a thing "y" can be a direct part of "x" without implying that "z" is a part of "x".

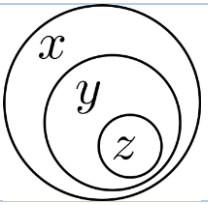


Figure 5: Illustration of the notion of transistivity and direct parthood.

However, there is a potential for conflict between the approach of Yang (2009) and that of Rossi (2012). It arises with the composition of a "thing". In material science the assumption of transistivity, illustrated in Figure 5, is not necessarily valid. In the terminology of Yang, this means that the composition of a thing "x" does not contain all the constituent parts of another thing "y", even though the latter is a subset of the former. For example, if a model based on molecular dynamics (MD) is coupled with a

model based on quantum mechanics (QM) this does not imply that the MD model contains the notion of electrons.Since the model formalization has a direct implication on software design it is imperative that the definitions are coherent between interdisciplinary groups.

3. Multi-Scale Simulation Types

Developing generalized software for the purpose of multi-scale modeling boils down to the question of generalizing different types of scale-integration.

3.1. Concurrent Simulations

The "seamless" coupling of models is a typical example of model construction requiring simulations to run concurrently. This is challenging to address in a generic way but there are examples, such as (Borg and Lockerby 2015), that achieve this by inter linking the low-level simulation models directly into the simulation tool for the macro-scale, e.g. OpenFOAM. The main reason why this is necessary is that there is no standard for passing data between scientific software. Recently there has been approaches, such as semantic interoperability (Hagelien and Chesnokov et al. 2017), that are aimed specifically at providing a generic framework within which meta-data can be defined and subsequently extracted using simple software-specific adaptors.

3.2. Sequential Simulations

Multi-scale models generally require adaptive workflows and distributed computing resources. Consequently, the workflow "orchestrator" is an integral part of any sequential simulation framework. However, the framework can be designed such that any orchestrator in principle can be used. The requirement is that model definitions can contain pre-defined strategies that the framework can employ to resolve errors occurring during a simulation. The added benefit of a sequential framework is that the workflow can be re-run; thus, providing a potential method of documenting reproducible simulations.

4. The MoDeNa Simulation Framework

The software framework Morphology Development of Micro and Nanostructures (MoDeNa), outlined in (Karolius and Preisig 2016), was designed specifically to interlink models sequentially as parts of a larger workflow. It uses adaptive surrogate models and on-the-fly parameter estimation, performing micro-scale simulations as necessary in order to satisfy a given error criteria.

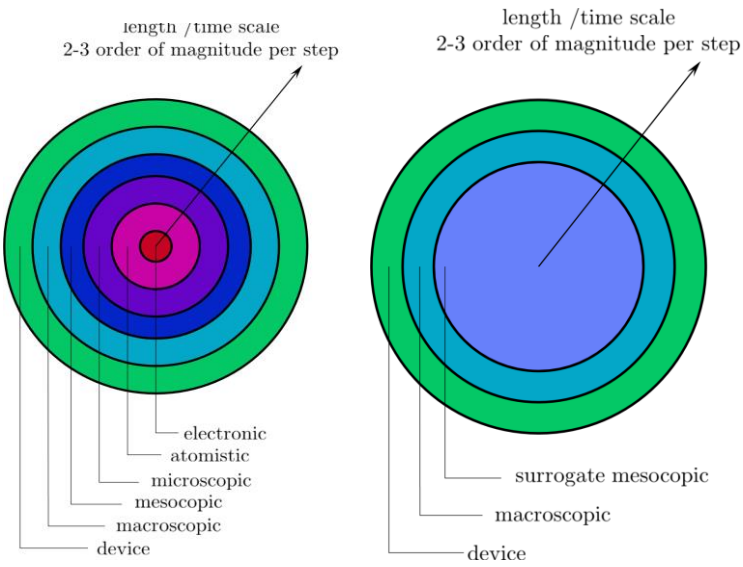


Figure 6: Onion-type multi-scale model employing a surrogate model for the mesoscopic scale.

This was a particularly challenging project as the scale-specific detailed models were developed in parallel with the software framework. It quickly became clear that it is desirable for the framework: 1) attain loose coupling between the models and 2) enforce high cohesion within each model. This is due to the practical need of maintaining the code-base of a large project where each member is familiar with only a small fraction of the complete project. The implication of loose coupling between models is that the API must be well designed such that it can handle all special cases without employing hacks. High cohesion, on the other hand, was enforced by defining that every detailed model was an atomic computational task performing homogenization, simulation and lifting. Consequently, all implementation details related to a scale-specific model is confined is the definition of a single computational task.The design on which the MoDeNa framework was based will obviously not apply to all types of multi-scale models as there is an inherent assumption of scale-separation. However, it does represent one approach for abstracting the scale-integration such that it can be defined outside of the implementation of the detailed models. The challenge of writing an ontology-based software framework capable of handling multiple types of scale-integration has been tackled by (Zhao and Jiang 2012). By extending such a tool to be capable of translating between multiple ontologies one would provide a significant contribution to the domain of multi-scale modelling.

5. Conclusions

The multi-scale modeling paradigm is continuously maturing and there is a need for development of formal conceptualization and generic software solutions. Part of the challenge for the PSE community is to become familiar with software and design concepts that are readily available. This paper presented an overview of current and past approaches to the challenge of tackling coupling of models across scales with a particular focus on the top-down strategy known as hierarchical multi-scale modeling (HMM). It was highlighted that supporting information exchange between simulations running concurrently and sequentially poses significantly different requirements on

software frameworks. The MoDeNa software framework was used as an example of how design principles from software engineering can be enforced in the development process whilst complementing the logical structure of the models that are intertwined.

References

L. Biegler, Y. Lang and L. Weijie. 2014. "Multi-Scale Optimization for Process Systems Engineering." *Computers & Chemical Engineering* 60, 17–30.

M. Borg, D. Lockerby and J. Reese 2015. "A hybrid molecular-continuum method for unsteady comp ressible multiscale flows." *Journal of Fluid Mechanics (768)*, 388-414.

J. Caballero and I. Grossmann. 2008. "An Algorithm for the Use of Surrogate Models in Modular Flowsheet Optimization." *AIChE Journal* 54 (10), 2633–2650.

P. Constantinos C. 2001. "New Challenges and Opportunities for Process Modelling." In *European Symposium on Computer Aided Process Engineering - 11* (9), 15–26.

T. Hagelien A. Chesnokov, S. Johansen, E. Meese and B.T. Løvfall. 2017. "SOFT: A Framework for Semantic Interoperability of Scientific Software." In *Proceedings of the 12 Th Int. Conf. on Comp. Fluid Dynamics in the Oil & Gas, Metallurgical and Process Industries*, 273–278.

G.D. Ingram, I.T. Cameron and K.M. Hangos. 2004. "Classification and Analysis of Integrating Frameworks in Multiscale Modelling." *Chemical Engineering Science* 59 (11): 2171–2187.

S. Karolius, H.A. Preisig, H. Rusche, 2016. "Multi-scale modelling software framework facilitating simulation of interconnected scales using surrogate models", *Comp. Aided Chem. Eng. (38)*, 463-468.

I.G. Kevrekidis, C.W. Gear, J.M. Hyman, P.G Kevrekidid, O. Runborg, and C. Theodoropoulos. 2003. "Equation-Free, Coarse-Grained Multiscale Computation: Enabling Mocroscopic Simulators to Perform System-Level Analysis." *Commun. Math. Sci.* 1 (4), 715–762.

L. Rosso and A. F. de Baas 2012; "What makea a material function"; *European Commission*.

D.G. Vlachos 2005. "A Review of Multiscale Analysis: Examples from Systems Biology, Materials Engineering, and Other Fluid–Surface Interacting Systems." *Advances in Chemical Engineering* (30), 1–61.

E. Weinan, W. Ren, and E. Vanden-Eijnden. 2009. "A General Strategy for Designing Seamless Multiscale Methods." Journal of Computational Physics 228 (15): 5437–5453.

E. Weinan, B. Engquist and Z. Huang. 2003. "Heterogeneous Multiscale Method: A General Methodology for Multiscale Modeling." *Phys. Rev. B* 67 (9). American Physical Society: 092101.

A. Yang, W. Marquardt, 2009. "An ontological conceptualization of multiscale models"; *Comp. & Chem. Eng. 33*, 822-837.

Y. Zhao, C. Jiang and A. Yang 2012, "Towards computer-aided multiscale modelling: An overarching methodology and support of conceptual modelling", *Computers & Chemical Engineering (36)*, 10-21.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50366-9

Study of the Phenomenology of Dispersion of Hydrocarbon Spillage into Freshwater Bodies

Jeffrey Leon-Pulido[a*], Cristhian C. Henao-Sanchez[b], Paola M. Parra-Herrera[b], Angel D. Gonzalez-Delgado[c] and Alvaro E. Villamizar[d]

[a]*EAN University, Bogota -Colombia. El Nogal, Cra. 11 No 78 -47*
[b]*America University Foundation, Bogota - Colombia. Calle 106 No.19-18*
[c]*University of Cartagena, Cartagena -Colombia. Av. del Consulado, Calle 30 No. 48-52*
[d]*University of Pamplona, Pamplona - Colombia. Km 1 Vía Bucaramanga*
jleonp@universidadean.edu.co

Abstract

The spillage of hydrocarbons associated with the fuel production chain, have generated irremediable environmental damages to the Colombian geography. The present study presents the analysis of the phenomenology of dispersion of a hydrocarbon in a body of fresh water, using as basis the mathematical modeling presented in the literature. The growth of the stain is studied from experimental tests established by a design of experiments to a determined time. The established criterion was developed for a controlled environment based on the model of dispersion of Fay, comparing the results of the model with the experimental data.

Keywords: Spillage, hydrocarbons, fresh water, dispersion.

1. Introduction

Due to the quantity of products and by-products derived from the processing of oil, this hydrocarbon has been one of the most demanded fuels in the world. The places of exploitation and production of crude oil are diverse in relation to the geography that contains them. This fact generates conditions of variable exploitation and transport (Parra & Haenao 2015, Alexopoulos *et al*, 2013). The effects of spillage are not exclusively associated with the extraction of crude oil (drilling), also, the terrestrial transport operations are (pipelines and tankers) and maritime (ships and offshore stations), which are vulnerable to technical failures and attacks which cause spillages and generate immense affects on natural resources.

Colombia has presented a gradual increase in the production of hydrocarbons, a precipitous increase in production that does not go to the same rhythm of the increase in the structures of fast transportation of the oil (pipelines). In the Colombian hydrocarbons production chain, ground transportation is one of the activities that has been most positioned as an alternative over time, moving crude oil and its by-products through national geography to refineries, processing units and Ports for exportation. With the increase of these transport activities, also, the risk of spillage that significantly impact the natural resources and communities close to the transport routes, increase. In 2014, the Department of Putumayo was affected by large spillages caused by actions of illegal groups, which resulted in the spillage of 5,100 barrels of crude that were transported from the Department of Putumayo to Ecuador. The attack affected 106

families, and the ravine La Canoa which flows into the Putumayo and Guamuez rivers, which supply the municipalities of Puerto Asís and Puerto Leguízamo. The environmental impact caused by hydrocarbon spillage mostly leaves irreversible damages.

This fact makes the study of phenomenology, once the spillage occurs and the provision of tools allows to predict the behavior of occurrences for timely attention, it is of great importance to the world. The present study finds results related to the dispersion of crude oil in freshwater bodies, generating knowledge for the development of spillage control techniques and, in this way, minimizing the negative effects on the ecosystems that are affected in this type of emergencies.

2. Methodology

The criteria established for the selection of the mathematical model (for the dispersion phenomenology) took into account that the variables considered were directly or indirectly measurable by means of an experimental approach, in addition to systems where the changes in the densities were not taken into account and viscosities of fluids due to changes in temperature over time, will not consider the influence of wind currents or mechanical disturbances. The experimental development that was carried out evaluated the growth of the oil stains generated by spillage of four volumes: 20 mL, 25 mL, 30 mL and 40 mL. The volumes used for experimental development were based on the experiment carried out by Cohen et al. (1980), Charbeneau et al., (1995), Qiao et al., (2014) the monitoring times were established through a pre-experimental process, by means of which it was determined that they would take records of the stains at the zero time of the spill, which corresponds to the moment at which the spillage is produced, from then on, we study the growth of the stain hourly. The hydrocarbon spillage present different phenomenologies, which are needed to be understood if the intention is to control a spillage in a timely manner, and to minimize the environmental impacts generated by it.

It is possible to see that the phenomena of dissolution and dispersion occur in all directions, for the purposes of the present study, the complexity of the movements in a two-dimensional sense will be reduced. Figure 2 shows the factors that affect the behavior of the stain over time. This way, it is possible to observe that environmental variables and boundary conditions can accelerate the phenomenologies of dispersion of hydrocarbon spillage in freshwater bodies.

3. Analysis of spillage factors and phenomenologies

The phenomena that are developed in a spillage of crude oil in bodies of water are physical, chemical, biological, and their evolution is related to the time that elapses between the beginning of the spillage and the attention of the same. The focus of study of this work centers on the behavior of the system from the dispersion phenomenon.

Natural dispersion

If a certain amount of hydrocarbon is spilled on a body of calm water, in which there is no interaction with the wind, nor streams of water, nor tides, this volume of oil would be dispersed evenly by the equilibrium of the forces associated with its weight initially, of immediately viscosity and finally by the surface tension. This phenomenon of occurrence in ideal conditions under any of the following names: runoff, mechanical dispersion, expansion or horizontal dispersion Betancourt (2014), Calderon (2007).

Vertical dispersión

The turbulence in large bodies of water is mainly caused by the wind, whereas in rivers it is mainly caused by the banks and the bottom of the channel in greater proportion, this arrangement causes part or all of the rupture of the stain, dividing it into fragments and drops of different sizes. These small droplets can become mixed at different levels of the water column. The small droplets have a greater tendency to remain dispersed in suspension when their diameter is less than 70 μm, whereas for those of larger diameter, the cohesive forces rejoin them, forming a new stain that is spread in thin films (NOAA - National Oceanic and Atmospheric Administration, 1995). The dispersion increases the surface area of the stain, favoring dissolution, biodegradation and sedimentation.

Characteristics and implications of water-oil spill systems

The different characteristics and implications that occur in hydrocarbon spillage are described in Table 1, which describes the time of action and the description of the phenomenon described in the spillage system.

Table 1. Characteristics of the processes involved in an oil spillage

PROCESS	DESCRIPTION	MASS LOSS	ACTION TIME
Mechanical dispersion	Mechanical dispersion determines the growth of the petroleum stain by the balance of gravitational, viscous and surface tension forces.	None	It depends on the initial volume shed, varying from approximately one day to three weeks
Vertical dispersion	The vertical dispersion takes place by the action of the agitation of the body in water, which separates the oil in small drops that are suspended in the water column. The permanence of the hydrocarbon drops in the water column depends on its turbulence	The percentage of mass lost may reach 20% or more depending on the environmental conditions. In the vicinity of a cliff the vertical dispersion can become total.	Vertical dispersion occurs a few hours after the spill has occurred.
Dissolution	The solution is presented by the solubilization of the hydrocarbon fractions present in the petroleum. Generally the same soluble fractions are those that first evaporate.	The amount of dissolved oil is about 1% which depends on the characteristics of the oil and the environmental conditions.	The dissolution as well as the vertical dispersion occurs a few hours after the spill has occurred.
Emulsification	Emulsification occurs by suspending water droplets in the oil, increasing its viscosity, density and environmental impact. The stability of an emulsion depends directly on the content of asphaltenes.	The water content in the oil can be up to 80% in turbulent waters. However a stable emulsion is generally 55 to 70%	The formation of emulsions takes place after a few hours, but begins to be of interest after the first 8 or 15 hours.
Sedimentation	Sedimentation occurs when the oil droplets suspended in the water column are adhered organic material or particles in suspension, causing precipitation.	The amount of settled oil does not reach large proportions and is very variable.	Sedimentation occurs after the first week of the spill.

Mathematical Model: Fay's Model (1971)

It performs an analysis of the forces acting on the interfaces formed when a spillage occurs (water-air, oil-water and oil-air) which accelerate or delay the dispersion of the oil stain once the spillage occurs. In equations 1, 2, 3, 4 and 5 are presented the four forces acting on a spillage:

a. Gravity force:

$$F_P = \Delta\rho g h\left(\frac{dh}{dX}\right) \quad (1)$$

b. Surface tension: Determines the coefficient of dispersion as the net difference between the surface tension of the water-air interface and the sum of the surface tensions of the oil-water and oil-air interfaces

$$\sigma = \sigma_{wa} - (\sigma_{oa} + \sigma_{ow}) \quad (2)$$

c. Inertia:

$$F_I = \rho h \frac{du}{dt} \quad (3)$$

d. Friction:

$$\mu \frac{u}{\delta} = \rho \delta \frac{du}{dt} \quad (4)$$

$$\delta = \sqrt{vt} \quad (5)$$

Where:

ρ = Density of the fluid
h = Height of the stain
g = gravity, X = distance
σ_{wa} = Surface tension
σ_{oa} = Surface tension raw air
σ_{ow} = Surface tension
μ = viscosity
δ = Losses by friction

From the four forces analyzed, Fay (1971) establishes three important regimes to measure the dispersion of the stain (linear or radial) as a function of time. These regimes are known as the Fay dispersion laws and were designed to measure one-dimensional or two-dimensional (axial symmetry) spills.

4. Experiment and discussions

According to the selective criteria presented above, the base model used was Fay's (1971). It performs an analysis of the forces acting on the interfaces formed when a spillage occurs (water-air, oil-water and oil-air). The properties of the crude oil and water used are shown in Tables 2 and 3 below.

Table 2. Properties of water used

Characteristics	Units	Value
Kinematic viscosity at 15 ° C	m^2/s	1.007.10-6-6
Density at 15 ° C	kg/m^3	999.1
Surface tension (water-air)	N/m	$72.08.10^{-3}$

Table 3. Characteristics of crude sales

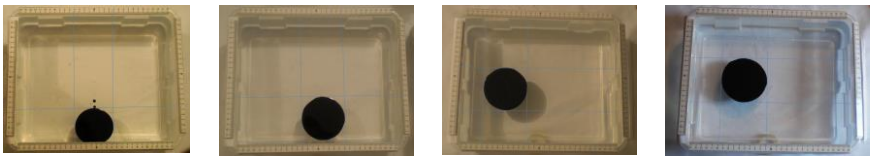
Characteristic	Units	Value
Gravity API @ 60 °F	°API	33,3
Sulphides	%W	0,144
Dissolved solids	%W	0,02
Nickel	Ppm	6
Salt content	PTB	4
Viscosity at 100 ° C	cP	9,1
Paraffins	%W	0,05
BSW	%	0,1019

Source: *Parex Resources Company*

Table 4. Dissolution of the oil stains in 24 hours

20 mL ≈ 17,6 g		
Run No.	**Dissolution (g / h) 24 hours**	**Dissolution %**
Run 1	4,507E-05	2,626E-04
Run 2	7,571E-05	4,412E-04

The analysis compared the amount of spilled crude mass with dissolved mass during each of the time periods monitored in the experimental development.



Time: 0 hours **Time:** 1 hour **Time:** 10 hours **Time:** 15 hours

Figure 1. Experiments

The pre-experimental and experimental development was carried out in a closed room, with artificial light, at temperatures between 10 ° C and 19 ° C. The spillage was manually performed at water level, the oil was measured in precipitated vessels and manually poured into the water, without controlling the speed of the spillage. To determine the growth of the stain, two diameters of the stain were measured, one called the H-diameter, which corresponds to the diameter of the stain on the horizontal axis of the container, and a diameter V corresponding to the vertical axis. The results obtained are presented in Table 5, these were established for a time of 20 mL.

Table 5. Experiment results

Picture No.	Elapsed time (h)	Diameter H (D_1) (mm)	Diameter V (D_2) (mm)
1	0	74	55
2	1	82	85
3	5	82	85
4	10	83	89

They were carried out with the same volumes of experimentation to be able to compare the behavior of the size of the oil stain spilled at different time intervals. These intervals were randomly selected in order to generate a prediction of spillage behavior from a time t = 0.5 hours to t = 48 hours, with intermediate values associated with the experimentation to compare the experimental and theoretical data..

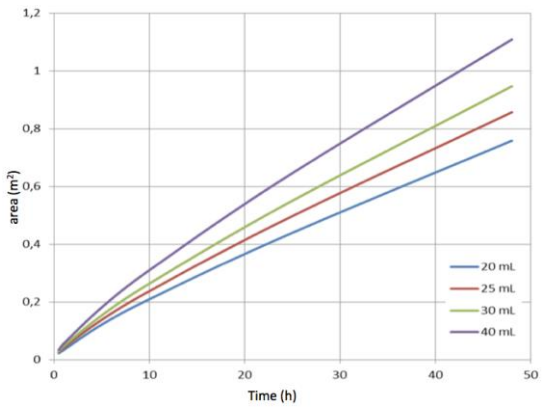


Figure 2. Increase in area over time - Numerical model

As it can be seen in Figure 3 the change of area with respect to time, the growth of the stain is gradual and constant, in conclusion the relationship or trend of the data thrown

is that the greater the volume, the greater the radius of the oil stain in wáter, and the longer the elapsed time the longer the radius. Then, it is concluded that the results generated by the simulator are in line with what was expected and could represent the growth of a crude stain at a steady state, at defined times and volumes.

5. Conclusions

The Fay Dispersion Model evidences a low correlation with the experimental behavior of spillage at laboratory scale, this is due to the sensitivity with the data field and environmental conditions going from a steady state to a dynamic behavior. It is possible to identify that as time increases by the nature of the equilibrium between fluids, the hydrocarbon area on the body of fresh water increases proportionally and without any tendency to equilibrium, this way, the dispersion of the stain is directly proportional to the time and amount of fluid spilled. The generation of more robust models that allow to have external considerations such as wind and mechanical factors are fundamental for the representation of the real phenomenology, the impact of these are significant in the speed of surface dispersion of the stain, as well as the speed in which it produces the spillage. The impact of the wind on the spillage was evidenced by the experimental results of runs one and two of the corresponding spillage to the volume of 30 mL

References

ALEXOPOULOS A.H., MILENKOVIC J, KIPARISSIDES. C. (2013) An Integrated Computational Model of Powder Release, Dispersion, Breakage, and Deposition in a Dry Powder Inhaler. Computer Aided Chemical Engineering. Vol 32, Pages 139-144.

BETANCOURT QUIROGA, Fabián. Modelado numérico de derrames de hidrocarburos en cuerpos de agua. México D.F, 2001. Tesis (Maestría en Ingeniería Área Ambiental). Universidad Autónoma de México. Facultad de Ingeniería. División de estudios de posgrado.

CALDERÓN G., Yenny C. Modelación bidimensional de derrames de crudo en cauces colombianos, aplicación Río Magdalena. Universidad de Los Andes. Bogotá D.C. Diciembre de 2007.

CHARBENEAU, Randall J.;*WEAVER, James W.;*LIEN, Bob K. The Hydrocarbon Spill Screening Model (HSSM) Vol. 1 & 2. Center of Research in Water Resources. University of Texas at Austin, Austin, Texas 78712. * Robert S. Kerr Environmental Research Laboratory. United States Environmental Protection Agency. Ada, Oklahoma 74820. Abril 1995, EPA/600/R-94/039b.

COHEN, Yoram; MACKAY, Donald; SHIU, Wan Ying. Mass Transfer Rates Between Oil Slicks and Water (1980). Institute for Environmental Studies, University of Toronto, Toronto, Ontario, Canada M5S 1A4.

FAY, James A. Physical Processes in the Spread of Oil on a Water Surface. En: Proc. Joint Conference on Prevention and Control of Oil Spills, Washington, D.C., June. Am. Petrol. Inst., pp. 463-467.

PARRA, PAOLA., HENAO, CRISTIAN, Estudio de los fenómenos de dispersión y disolución de derrames de hidrocarburos en cuerpos de agua dulce. Colombia, 2015. Tesis (Profesional en Ingeniería Química), Fundación Universitaria de América. Facultad de Ingeniería.

QIAO Z., JIAN G., FENGQI Y. (2014). Optimal Superstructure-Based Design and Synthesis of Hydrocarbon Biorefinery via Fast Pyrolysis, Hydrogen Production and Hydroprocessing Pathway. Computer Aided Chemical Engineering. Vol 33, Pages 175-180.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50367-0

Defect Data Modeling and Analysis for Improving Product Quality and Productivity in Steel Industry

Xinmin Zhang[a], Manabu Kano[a*], Masahiro Tani[b], Junichi Mori[c], Kohhei Harada[c]

[a]*Department of Systems Science, Kyoto University, Kyoto 606-8501, Japan*
[b]*Nippon Steel & Sumitomo Metal Corp., Kimitsu Works, Quality Management Div., 1 Kimitsu, Kimitsu, Chiba 299-1141, Japan*
[c]*Nippon Steel & Sumitomo Metal Corp., Kimitsu Works, Production & Technical Control Div., 1 Kimitsu, Kimitsu, Chiba 299-1141, Japan*
manabu@human.sys.i.kyoto-u.ac.jp

Abstract

The steel industry is committed to improving product quality and productivity with low production cost. To achieve these objectives, reducing product defects is of utmost importance. There is an increasing interest in developing a model to predict the occurrence of defects online. However, traditional statistical models such as multiple linear regression and Poisson model are not adequate enough to describe the observed defect count data due to the unique characteristics of non-negative integers, over-dispersion, high skewed distribution, and excess zeros in the data. This study develops an online quality monitoring system based on the zero-inflated regression modeling. Zero-inflated models are two-component mixture models that combine a count component and a point mass at zero. Intuitively, a mass of zeros observed in defect data can be attributed to two states: a safe (perfect) state, where no defect occurs, and a non-safe (imperfect) state, in which defects are possible but not inevitable. Zero-inflated models are suitable for modeling discrete and non-negative integers data and can handle both over-dispersion and excess zeros. The effectiveness of the zero-inflated models was verified through their application to the real defect data of a steelmaking plant. The application results demonstrated that the prediction accuracy of the zero-inflated models is superior to the PLS, Poisson, and negative binomial models.

Keywords: Defect data; count data; casting-rolling process; quality prediction; zero-inflated models.

1. Introduction

To keep the advantage in the fierce market competition, the steel industry is committed to improving product quality and yield with low production cost. However, manufacturing processes always suffer from various variations, including equipment malfunctions, process perturbations, improper operations, and so on. These variations may result in different types of defects, which not only degrade product quality but also increase production cost, energy consumption, and lead time. During the continuous casting-rolling process, defects are generally inspected after cold rolling. Because defects cannot be detected during the continuous casting-rolling process, many defective products will be formed. Therefore, developing an online quality monitoring system that can predict the occurrence of defects is vital to improve product quality and yield.

Virtual sensing or soft-sensor is a key technology for predicting product quality or other key variables in real time, and has been successfully applied to many industrial processes (Kano and Nakagawa, 2008; Kadlec et al., 2009). The basic idea of soft-sensor is to construct an inference model that relates product quality (response) to process operating conditions (predictors). Multiple linear regression (MLR) and partial least square (PLS) regression are the most commonly used methods (Kano and Fujiwara, 2013). However, they may not be suitable for modeling defect count data. This is because MLR and PLS depend on assumptions of normality and homoscedasticity of the dependent variable, but the observed defect count data exhibits highly skewed distribution and overdispersion. Furthermore, MLR and PLS could produce negative values for the dependent variable, although the defect count data is nonnegative integers.

Poisson regression is a basic tool for modeling count data (Cameron and Trivedi, 2013). However, it has limitations in its application to the overdispersed data due to its assumption that the variance of the dependent variable is equal to its mean, i.e., "equidispersion". This assumption is often violated by most of the data such as the defect data observed in this study. A remedy for this problem is to implement a negative binomial (NB) regression model, which could be viewed as a Poisson-gamma mixture model (Hilbe, 2011). The NB regression model has a free dispersion parameter, thus it is more flexible than the Poisson regression model in handling overdispersed data. However, the NB model is not adequate enough to account for the excess of zeros.

To account for the unique characteristics of non-negative integers, overdispersion, high skewed distribution, and excess zeros in the observed defect count data, the present work develops an online quality monitoring system based on the zero-inflated model, which is a two-component mixture model that combines a count component and a point mass at zero. Compared to the standard Poisson and NB models, zero-inflated models are not only suitable for modeling discrete and non-negative integers data, but also sufficient for handling both overdispersion and excess zeros. This research is motivated to verify the potential application of zero-inflated modeling technique to the defect data in a steel manufacturing process.

The remainder of this paper is organized as follows. In section 2, the basic theory of the Poisson regression and the NB regression is briefly introduced. Then the zero-inflated modeling technique is explained in section 3. In section 4, the performance of the zero-inflated model is evaluated through its application to the real defect data of a steelmaking plant. Section 5 gives the conclusions of this work.

2. Preliminaries

2.1. Poisson regression

Poisson regression is a basic tool for modeling count data. Given the input-output pairs $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$, where $\boldsymbol{x}_i^T \in \Re^K$ is an observation of the predictor variables, and y_i is the response, which is a non-negative integer. The Poisson probability density is

$$f(y_i|\boldsymbol{x}_i) = \frac{e^{-u_i} u_i^{y_i}}{y_i!} \tag{1}$$

where $y_i!$ is a factorial, e is the base of the natural logarithm, and u_i is the average number of events. The conditional mean and the conditional variance are given by

$$\mathrm{E}(y_i|\boldsymbol{x}_i) = \mathrm{Var}(y_i|\boldsymbol{x}_i) = u_i. \tag{2}$$

The Poisson regression model can be obtained through parameterizing the conditional mean of y_i as an exponential function of the predictor variables $\boldsymbol{x}_i$:

$$u_i = \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}) \tag{3}$$

where $\boldsymbol{\beta}$ denotes a vector of regression coefficients. Using exponential parameterization ensures the non-negativity of u_i. In general, the parameter estimates of Poisson regression can be obtained by using the maximum likelihood technique (Fox, 2015).

Since Poisson regression assumes that the conditional mean equals the conditional variance, it cannot describe overdispersed data. One way to relax this assumption is to implement the NB model.

2.2. Negative binomial regression

The NB model is a generalization of the Poisson model by allowing the Poisson parameter u_i to vary randomly following a gamma distribution (Hilbe, 2011). The NB probability density takes the form of

$$f(y_i|\boldsymbol{x}_i) = \frac{\Gamma(y_i + \tau)}{y_i!\,\Gamma(\tau)}\left(\frac{\tau}{\tau + u_i}\right)^{\tau}\left(\frac{u_i}{\tau + u_i}\right)^{y_i} \tag{4}$$

where $\tau > 0$ is a shape parameter or dispersion parameter, and $\Gamma(\cdot)$ is a gamma function. The conditional mean and the conditional variance are given by

$$\mathrm{E}(y_i|\boldsymbol{x}_i) = u_i = \exp(\boldsymbol{x}_i^T \boldsymbol{\beta}),\ \mathrm{Var}(y_i|\boldsymbol{x}_i) = u_i\left(1 + \frac{1}{\tau}u_i\right). \tag{5}$$

It is clear that the variance is larger than the mean, so the NB model allows for overdispersion. Furthermore, as $\tau \to \infty$, the NB distribution converges to the Poisson distribution. The maximum likelihood technique is used to estimate the parameters of NB regression (Hilbe, 2011).

3. Zero-inflated modeling for defect data with excess zeros

To monitor the defects online in the casting-rolling process, a quality monitoring system based on the zero-inflated modeling was proposed. Compared to the standard Poisson and NB models, zero-inflated regression models are more powerful in dealing with both the excess zeros and overdispersion in the observed defect data.

Zero-inflated regression models assume there are two underling data generating processes (Cheung, 2002; Zuur et al., 2009). For an observation, the probability that it is generated from process 1 is p_i, and the probability that it is generated from process 2 is $1 - p_i$. A binary model is used to determine which process is used. Process 1 is usually referred to as a point mass at zero, which generates only zero counts, whereas process 2 generates counts from either a Poisson or a NB model. Thus, zero counts may come from two processes. The zero-inflated probability density is given by

$$f(y_i|\boldsymbol{x}_i) = \begin{cases} p_i + \{1-p_i\}g(0|\boldsymbol{x}_i), & \text{if } y_i = 0 \\ \{1-p_i\}g(y_i|\boldsymbol{x}_i), & \text{if } y_i > 0 \end{cases} \tag{6}$$

where $g(\cdot)$ denotes a Poisson or NB probability density function.

In the zero-inflated regression, the parameters u_i and p_i are modelled as the function of the predictor variables $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$:

$$u_i = \exp(\boldsymbol{x}_i^T\boldsymbol{\beta}), \quad \text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \boldsymbol{z}_i^T\boldsymbol{\gamma} \tag{7}$$

where $\boldsymbol{z}_i$ is the vector of zero-inflated predictor variables, $\boldsymbol{\gamma}$ denotes the vector of zero-inflated coefficients, and $\text{logit}(\cdot)$ denotes the logistic regression. Generally, the regressors $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ could be different.

More formally, when $g(\cdot)$ is specified as the Poisson distribution, zero-inflated Poisson (ZIP) probability density is obtained as

$$f_{ZIP}(y_i|\boldsymbol{x}_i, \boldsymbol{z}_i) = \begin{cases} p_i + (1-p_i)e^{-u_i}, & \text{if } y_i = 0 \\ (1-p_i)\dfrac{e^{-u_i}u_i^{y_i}}{y_i!}, & \text{if } y_i > 0 \end{cases} \tag{8}$$

When $g(\cdot)$ is specified as the NB distribution, the zero-inflated NB (ZINB) probability density is obtained as

$$f_{ZINB}(y_i|\boldsymbol{x}_i, \boldsymbol{z}_i) = \begin{cases} p_i + (1-p_i)\left(1+\dfrac{u_i}{\tau}\right)^{-\tau}, & \text{if } y_i = 0 \\ (1-p_i)\dfrac{\Gamma(y_i+\tau)}{y_i!\,\Gamma(\tau)}\left(1+\dfrac{u_i}{\tau}\right)^{-\tau}\left(1+\dfrac{\tau}{u_i}\right)^{-y_i}, & \text{if } y_i > 0 \end{cases} \tag{9}$$

The conditional mean and the conditional variance of the ZIP model are

$$\text{E}_{ZIP}(y_i|\boldsymbol{x}_i, \boldsymbol{z}_i) = u_i(1-p_i) \tag{10}$$

$$\text{Var}_{ZIP}(y_i|\boldsymbol{x}_i, \boldsymbol{z}_i) = u_i(1-p_i)(1+u_ip_i). \tag{11}$$

The conditional mean and the conditional variance of the ZINB model are

$$\text{E}_{ZINB}(y_i|\boldsymbol{x}_i, \boldsymbol{z}_i) = u_i(1-p_i) \tag{12}$$

$$\text{Var}_{ZINB}(y_i|\boldsymbol{x}_i, \boldsymbol{z}_i) = u_i(1-p_i)\big(1+u_i(p_i+\alpha)\big). \tag{13}$$

Note that the variance is larger than its corresponding mean in both ZIP and ZINB models. This indicates that they are capable of accounting for overdispersion. The parameter estimates of ZIP and ZINB models can be obtained by using maximum likelihood estimation (Cheung, 2002; Zuur et al., 2009).

4. Case study

In this section, the effectiveness of ZIP and ZINB models is validated through their application to the real defect data of a steelmaking plant. The prediction performance is compared with that of PLS, Poisson, and NB models.

The defect dataset was recorded from a continuous casting-rolling process of the steelmaking plant. It consists of 3600 samples and 71 input variables. The objective is to predict the number of defects online. To construct the predictive model, the entire dataset was partitioned into two subsets: a training set with 3000 samples and a testing set with 600 samples. Figure 1 shows the frequency distribution of the defects, which is characterized by a large number of zeros and high skewness. The percentage of no defect is around 70%. Furthermore, its variance is much larger than its mean; that is, overdispersion exists in the data.

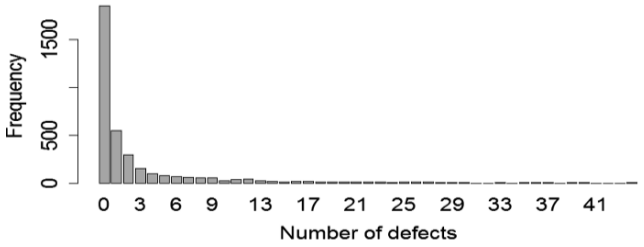


Figure 1. Distribution of the number of defects.

The prediction results of Poisson, NB, ZIP, and ZINB models for the testing data were evaluated in terms of the root mean squared error (RMSE) and correlation coefficient (R^2) criteria. In addition, the PLS model was built as a reference model. 15 variables were selected as the input variables on the basis of the nonlinear feature selection technique of ℓ_1−LSMI (Jitkrittum, 2013). The number of latent variables used in PLS was 5, which was determined by cross-validation. Table 1 presents the RMSEP (RMSE of Prediction) and R^2 of five methods. The PLS model did not perform well and provided a large RMSEP and a small correlation coefficient. As compared to PLS, the Poisson and NB models produced better performances. In contrast with the standard Poisson and NB models, the ZIP and ZINB models achieved significantly better results. The ZIP model yielded the best prediction performance among the five methods. Figure 2 shows the detailed comparison of the PLS and ZIP models. Note that some predicted values of the PLS model are below zero, although the number of defects is nonnegative. In comparison, the ZIP model ensures nonnegative predicted values. All of the above results have demonstrated that the ZIP model is preferred for modeling and analyzing the defect data.

Table 1. Prediction results of five models for the defect data.

Methods	RMSEP	R^2
PLS	4.2065	0.3843
Poisson	4.1723	0.3959
NB	4.1615	0.3775
ZIP	3.9858	0.4307
ZINB	4.0000	0.4258

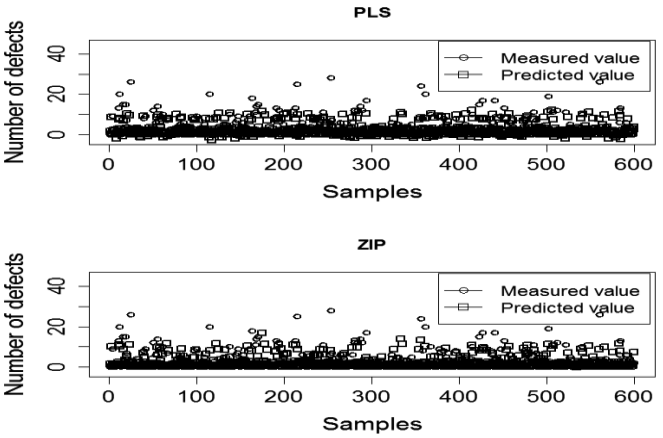


Figure 2. Prediction results of PLS and ZIP models for the defect data.

5. Conclusions

In this research, we focused on the defect data in the steel manufacturing process to predict the number of defects. To account for the unique characteristics of non-negative integers, high skewed distribution, overdispersion, and excess zeros in the observed defect data, an online quality monitoring system based on the zero-inflated model was proposed. The effectiveness of the zero-inflated Poisson (ZIP) and zero-inflated NB (ZINB) models was verified through their application to the real defect data of a steelmaking plant. The results have demonstrated that the ZIP and ZINB models are superior to the PLS, Poisson, and NB models in the prediction performance.

References

Kano, M. and Nakagawa, Y., 2008, Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry, Computers & Chemical Engineering, 32(1), 12-24.

Kadlec, P., Gabrys, B. and Strandt, S., 2009, Data-driven soft sensors in the process industry, Computers & Chemical Engineering, 33(4), 795-814.

Kano, M. and Fujiwara, K., 2013, Virtual sensing technology in process industries: trends and challenges revealed by recent industrial applications, Journal of chemical engineering of Japan, 46(1), 1-17.

Hilbe, J.M., 2011, Negative binomial regression, Cambridge University Press.

Cameron, A.C. and Trivedi, P.K., 2013, Regression analysis of count data, Vol. 53, Cambridge university press.

Fox, J., 2015, Applied regression analysis and generalized linear models, Sage Publications.

Cheung, Y.B., 2002, Zero-inflated models for regression analysis of count data: a study of growth and development, Statistics in medicine, 21(10), 1461-1469.

Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A. and Smith, G.M., 2009, Zero-truncated and zero-inflated models for count data, In Mixed effects models and extensions in ecology with R, Springer New York, 261-293.

Jitkrittum, W., Hachiya, H. and Sugiyama, M., 2013, Feature Selection via ℓ_1-Penalized Squared-Loss Mutual Information, IEICE Transactions on Information and Systems, 96(7), 1513-1524.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50368-2

States Identification of Complex Chemical Process Based on Unsupervised Learning

Shaodong Zheng[a], Jinsong Zhao[a,b*]

[a] *Department of Chemical Engineering, Tsinghua University, Beijing, China, 100084*
[b] *Beijing Key Laboratory of Industrial Big Data System and Application, Tsinghua University, Beijing, China, 100084*
jinsongzhao@tsinghua.edu.cn

Abstract

States identification is an important step of process safety management, and it can be an individual part or a precursor step of fault detection and diagnosis (FDD). Research on FDD have made good progress, but most of them are carried out on labelled data, which is hard to collect in real situation. In this paper, an unsupervised learning based state identification model is proposed to deal with unlabelled data.

Feature extraction or dimensionality reduction, and clustering are two main steps of the model. t-SNE is an outstanding tool to visualize high-dimensional data, and a deep auto-encoder network is used to reduce the dimensionality to a reasonable amount before t-SNE is implemented. Different clustering algorithm are tested on features to determine which one is better. The unlabeled data are divided into different clusters by this model. In the final part of this paper, the benchmarked Tennessee Eastman process is utilized to illustrate the performance of this unsupervised learning based model.

Keywords: states identification, TE process, deep auto-encoder network, t-SNE, cluster

1. Introduction

With rapid development of modern chemical industry, people are paying more attention to process safety management. In this big data era, data driven process monitoring is used extensively to control risk of chemical process and prevent it from accident.

As a central step of process monitoring, process fault detection and diagnosis (FDD) can be classified into three categories among which process history based methods are more promising when applying in chemical process (Venkatasubramanian et al., 2003). One set of data driven methods are multivariate statistical methods such as principal component analysis (Cho et al. 2005) and partial least squares. Methods based on artificial neural network are also studied widely. Zhang conducted a model based on deep belief network (DBN) for FDD and reached an average 82.1% fault diagnosis rate of all 20 faults of the Tennessee Eastman (TE) process (Zhang et al., 2017).

The research mentioned above have achieved good results, but most of them are carried out on labelled data, which is hard to get in real situation. Methods based on unsupervised learning can work on unlabeled data, identify different states and accordingly divide data into several clusters. This will enable experts to label the data effectively.

The rest of this paper is organized as follows: Section 2 introduces the unsupervised learning model, which consists of t-SNE, deep auto-encoder network based on restricted

Boltzmann machine and clustering algorithms. The model is applied to benchmarked TE process and the result is shown in Section 3. Finally, conclusions are drawn in Section 4.

2. Unsupervised learning based state identification model

With unlabelled data, the amount of clusters or the number of different states is also unknown. t-SNE can be used to visualize high-dimensional data to show how many clusters the data should be distributed in, and meanwhile export 2-dimensional or 3-dimensional features. It is highly recommended to use another dimensionality reduction method, which is deep auto-encoder network in this study, to reduce the number of dimensionality to a reasonable amount (e.g. 50), if the number of features is very high. This will suppress noise. Clustering algorithms are then used to divide extracted features into different groups and lead to the states identification result.

2.1. t-SNE algorithm

t-SNE is a variation of Stochastic Neighbor Embedding (Hinton and Roweis 2002). The algorithm converts the distance between two points into conditional probability and uses Kullback-Leibler divergence to describe similarity of two probability distribution, which is shown as Eq. (1) (Van Der Maaten and Hinton 2008):

$$C = \sum_i KL(P_i||Q_i) = \sum_i \sum_j p_{j|i} log \frac{p_{j|i}}{q_{j|i}} \tag{1}$$

where P_i represents the conditional probability distribution of point *i* in high-dimensional space and Q_i represents the one in low-dimensional space. The algorithm aims at minimization of cost function *C*, which means probability distribution in low-dimensional space fits the real distribution well.

2.2. RBM based deep auto-encoder network

Deep belief network (DBN) is a network consists of several middle layers of Restricted Boltzmann machine (RBM) and the last layer as a classifier. In unsupervised dimensionality reduction, the classifier is removed and a deep auto-encoder network only consisting of RBMs is used. When running the deep auto-encoder network, two steps including pre-training and fine-tuning is executed.

In pre-training stage, each layer with its previous layer is considered an RBM and trained. Low-dimensional features are extracted from input data by pre-training without losing much significant information. Wake-sleep algorithm is introduced for fine-tuning of weights to avoid training process falling into gradient diffusion and accelerate the convergence of feature extraction (Hinton 2006). In fine-tuning stage, the encoder is unrolled to a decoder, and the weights of decoder are transposed from encoder. In wake state, the weights of decoder are adjusted by errors between input data and reconstructed data. In sleep state, the weights of encoder are adjusted by errors between features extracted from input data and reconstructed data respectively. Reconstruction error (RE) shows how well the feature can represent original data.

2.3. Clustering

2.3.1. Typical clustering algorithms

Clustering aims to excavate the internal nature of unlabelled data and divide data into different groups according to their similarity. Clustering algorithms can be broadly

classified as partitioning-based, hierarchical-based, density-based, grid-based and model-based methods (Fahad et al. 2014). BIRCH is a kind of hierarchical-based algorithm, and Affinity Propagation and K-Means belong to partitioning-based algorithms. Mini-batch K-Means is a variation of K-Means. It randomly extracts a small batch of samples to do K-Means algorithm in each iteration and can accelerate the convergence.

2.3.2. Cluster validity index

Cluster validity index is the judging criteria of clustering result, and it can be classified as external index and internal index. External index is calculated by comparing the result with true division, which cannot be obtained in unsupervised learning. And for internal index, true division is not necessary. Silhouette coefficient (SC) is a kind of internal index and it is calculated as follows (Rousseeuw 1987):

$$\mathrm{SC(i)} = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{2}$$

where $a(i)$ represents the average distance between sample i and other samples in its cluster, and $b(i)$ represents the minimum average distance between sample i and samples in other clusters. The value of SC ranges from -1 to 1, and SC getting closer to 1 indicates that the cluster algorithm works better.

2.4. States identification model

With the knowledge introduced above, the flow diagram of states identification model is shown in Figure 1.

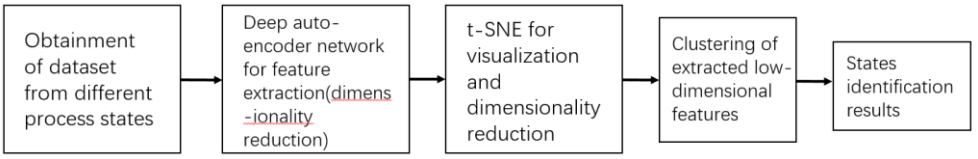


Figure 1. Flow diagram of state identification model

3. Application in the TE process

The unsupervised learning based model introduced in Section 2 is applied in Tennessee Eastman (TE) process in this section.

3.1. TE process and obtainment of dataset

TE process was proposed in 1993 with 53 process variables (Downs and Vogel 1993), but 3 of them never change value and thus can be seen as constants. There are 20 kinds of faults in TE process, and the study in this paper focuses on states identification of step-type faults, so data of Fault 1,2,3,4,5,6,7 and Normal state are sampled.

In each simulation, a kind of fault disturbance is introduced after process running normally for 10 hours, and the fault will run for 65 hours. The data is sampled from the 25th hour to 75th hour of running process with 3-minute sampling period.

TE process is a complex dynamic process, so data sampled at single time point cannot represent the process state precisely. Therefore, data sampled during certain time period are spliced together chronologically. The time window is set by 15,30,45 and 60 minutes, and one-hour-long time window appears to the best option according to cluster results.

There are 20 pieces of data sampled in one hour and they are spliced into a 1,000-dimensional sample, and 490 such samples are obtained for each state.

In conclusion, the dataset used for further study contains 3,920 1,000-dimensional samples.

3.2. Dimensionality reduction by deep auto-encoder network

The structure of deep auto-encoder network including the layer counts and the neurons counts in each layer has great influence on training result. It is assumed that numbers of neurons in hidden layers are determined by the number of layer and the number of neurons in input layer and output layer, At first, 30 different network structures are tested with the number of layer set by 3,4 and 5 and number of neurons in output layer set by 10, 20, … ,100 respectively. The structure is adjusted and optimization range is narrowed gradually on reference to REs. The test results of last optimization are shown in Table 1, and the best structure is a network with 6 layers and 1000, 346, 135, 58, 27, 12 neurons in Layer-1, Layer-2, Layer-3, Layer-4, Layer-5 and Layer-6 respectively.

Table 1. The network structures and REs in last optimization

Structure	Layer counts	Neuron counts in each layer	RE
1	5	1000-278-84-28-7	0.2360
2	5	1000-298-99-37-12	0.1857
3	6	1000-306-105-40-17-7	0.2295
4	6	1000-346-135-58-27-12	0.1734

There are some parameters that influence training progress of single RBM, such as momentum and weight-decay. The momentum is usually valued from 0.5 to 0.9, and the weight-decay is valued near 0.0001 (G. Hinton, 2010). The model is run with different values of momentum and weight-decay as shown in Table 2 and the last combination of values is chosen.

Table 2. Different values of momentum and weight-decay and corresponding REs

Momentum	Weight-decay	RE
0.7	0.0001	0.1803
0.7	0.0002	0.1766
0.8	0.0001	0.1735
0.8	0.0002	0.1698

3.3. Visualization and 2-dimensional features extraction by t-SNE

The deep auto-encoder network as optimized above is run to reduce data dimensionality to 12 for implement of t-SNE. t-SNE extracts 2-dimensional features and makes the visualization with them, which is shown as Fig 2.

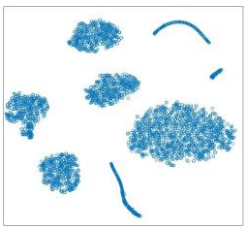


Figure 2. Visualization of the data

3.4. Clustering

The 2-dimensional features are then clustered by Affinity Propagation, BIRCH, K-Means and Mini-batch K-Means algorithm with clusters number set by 8, if needed, which can be seen from Fig 2. The results are shown in Table 3. K-Means and Mini-batch K-Means algorithm work better than the other two, and Mini-batch K-means algorithm is chosen for shorter running time.

Table 3. Comparison of different cluster algorithms

Algorithm	SC
Affinity Propagation	0.6229
BIRCH	0.6362
K-Means	0.6805
Mini-batch K-Means	0.6796

Mini-batch K-Means algorithm is run on the 2-dimensional features and the result is shown as Table 4 and Fig 3.

Table 4. Final clustering result

Cluster	1	2	3	4	5	6	7	8	total
Normal	0	253	0	0	0	0	237	0	490
Fault-1	0	0	0	490	0	0	0	0	490
Fault-2	0	0	0	0	0	490	0	0	490
Fault-3	0	270	0	0	0	0	220	0	490
Fault-4	0	0	490	0	0	0	0	0	490
Fault-5	0	207	0	0	0	0	283	0	490
Fault-6	0	0	0	0	266	0	0	224	490
Fault-7	490	0	0	0	0	0	0	0	490

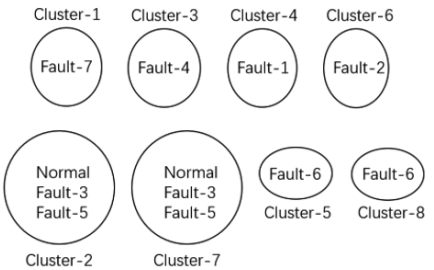


Figure 3. Schematic of clustering result

3.5. Discussion of the result

It can be seen from the result that Cluster-1,3,4,5,6,8 have pure components, which means they consist of data of a single state respectively, and data of Fault-1,2,4,6,7 only appear in these clusters. This cluster result enables experts to label all data of Fault-1,2,4,6,7 precisely and effectively with their expertise, because they only need to identify the state of cluster centres.

Data sampled from Fault-3,5 and Normal state are not separated and they gather to Cluster-2,7. Take Fault-3 for example, the process disturbance of Fault-3 is feed

temperature of Stream 2, and the variable influenced most is temperature of reactor cold water. But it only deviates from Normal state slightly and thus cannot be distinguished from Normal state. Many previous FDD research on TE process based on supervised learning also indicate that some faults are difficult to diagnose correctly since the data do not have obvious features, such as Fault-3,5,8,9,12,13,14.

4. Conclusions and outlooks

In this paper, an unsupervised learning based states identification model is proposed. Deep auto-encoder network and t-SNE algorithm are used to extract features and visualize data. Extracted 2-dimensional features are clustered by Mini-batch K-Means algorithm.

Application of the model on step-type faults of TE process leads to meaningful result. Data of Fault-1,2,4,6,7 are distributed to clusters which only consist of data of their own types respectively, therefore the identification rate of Fault-1,2,4,6,7 can reach 100%.

But the model cannot distinguish data of Fault-3,5 and Normal state currently. In the near future, research will be focused on unsupervised states identification of these states by improving the feature extraction network and cluster algorithm.

References

Cho, Ji-hoon, Jong-min Lee, Sang Wook, Dongkwon Lee, and In-beum Lee. 2005. "Fault Identification for Process Monitoring Using Kernel Principal Component Analysis." *Chemical Engineering Science* 60:279–88.

Downs, J. J. and E. F. Vogel. 1993. "A Plant-Wide Industrial Problem Process." *Computers & Chemical Engineering* 17(3):245–55.

Fahad, Adil et al. 2014. "A Survey of Clustering Algorithms for Big Data : Taxonomy and Empirical Analysis." *IEEE* 2(3).

Hinton, Geoffrey E. 2006. "Reducing the Dimensionality of Data with Neural Networks." *Science* 313(5786):504–7.

Hinton, Geoffrey E. and Sam T. Roweis. 2002. "Stochastic Neighbor Embedding." *Advances in Neural Information Processing Systems* 833–40.

Hinton, Geoffrey and Geoffrey Hinton. 2010. "A Practical Guide to Training Restricted Boltzmann Machines." Retrieved (http://www.cs.toronto.edu/~hinton/absps/guideTR.pdf).

Van Der Maaten, L. J. P. and G. E. Hinton. 2008. "Visualizing High-Dimensional Data Using T-Sne." *Journal of Machine Learning Research* 9:2579–2605. Retrieved (https://lvdmaaten.github.io/publications/papers/JMLR_2008.pdf%0Ahttp://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=7911431479148734548related:VOiAgwMNy20J).

Rousseeuw, Peter J. 1987. "Silhouettes : A Graphical Aid to the Interpretation and Validation of Cluster Analysis." *Journal of Computational and Applied Mathematics* 20:53–65.

Venkatasubramanian, Venkat, Raghunathan Rengaswamy, and Surya N. Ka. 2003. "A Review of Process Fault Detection and Diagnosis Part III : Process History Based Methods." *Computers & Chemical Engineering* 27:327–46.

Zhang, Zhanpeng and Jinsong Zhao. 2017. "A Deep Belief Network Based Fault Diagnosis Model for Complex Chemical Processes." *Computers and Chemical Engineering*. Retrieved (http://dx.doi.org/10.1016/j.compchemeng.2017.02.041).

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50369-4

Deep Learning Based Soft Sensor and Its Application on a Pyrolysis Reactor for Compositions Predictions of Gas Phase Components

Wenbo Zhu[a], Yan Ma[a], Yizhong Zhou[b], Michael Benton[a], Jose Romagnoli[a*]

[a]*Department of Chemical Engineering, Louisiana State University, 3307 Patrick F. Taylor Hall, Baton Rouge 70803, USA*
[b]*Shanghai SupeZET Engineering Technology Corp,.Ltd, 268 Linxin Rd NO.3, Shanghai, 200335, China*
jose@lsu.edu

Abstract

In this work, we proposed a data-driven soft sensor based on deep learning techniques, namely the convolutional neural network (CNN). In the proposed soft sensor, instead of only building time-independent correlations among the key variable with other measurements, the moving window method is utilized to describe the most recent process dynamics, where the time-dependent correlation can be located. Beyond on that, a signal recovery scheme is developed to improve the model robustness when confronting common sensor faults. The proposed soft sensoring technique was tested on the composition data of gas-phase components from an ethylene pyrolysis reactor. The model was also verified through the manually introduced sensor faults.

Keywords: Soft sensor; Deep learning; Data-Driven; Big data

1. Introduction

In recent decades, increased accessibility of process measurement has boosted the development of data-driven approaches for process monitoring. The quality and quantity of the measurements are of the paramount importance for chemical processes. While in practice, some key variables can be difficult to measure in real-time due to physical or technical limitations. Therefore, soft sensoring techniques haven been developed for the supplement of real measurements. Generally speaking, soft sensor techniques are predictive models that can find correlations among key measurements with other variables. Two types of models that are commonly used as soft sensors are model-based models and data-driven models. Model-based approaches are designed based on physical features of the dynamical systems, while in data-driven approaches, features are obtained from process data solely through machine-learning. Although model-based approach contains physical meanings of the system, it is generally labour-intensive and difficult as it requires experienced researchers to develop the mathematical encoding, while it also requires massive experimental validations. On the contrary, machine learning algorithms are able to extract optimal features directly from raw data. Through such machine-learned features, feature selection algorithms are trained in order to select the most meaningful subset of the features to represent the system. The combination of feature learning and feature selection is the core of machine learning algorithms to learn the system through data.

Recent development in deep learning techniques (LeCun, 2015) has huge potential as a feature learner. Deep learning methods stack multiple layers of nonlinear estimators (neurons) to represent features in different levels from shallow to deep. Such architecture is able to represent complicated functions and correlations, and it has already been used to process images and natural languages as a state-of-the-art approach. Though deep learning methods have achieved huge success, training the deep learner is still challenging. In a typical deep network, there are always thousands of parameters that need tuning, which requires huge amount of training data. Moreover, activation functions e.g. sigmoid function that are used to enhance the nonlinearity expression in each layer can cause gradient vanishing or exploding problems, leading the gradient-based back propagation unable to tune the earlier layers in the network. The problem can then be solved by the development of the restricted Boltzmann machine (RBM) (Hinton G. E., 2006, 2010) and its application of greedy layer-wise training. Shang et al. (2014) proposed a deep learning based soft sensor based on RBM. They found that the deep soft sensor outperforms over many conventional methods with shallow architectures. Yan et al. (2017) applied stacked denoising autoencoders (SDAEs) to train the network, which estimated the oxygen content in flue gasses. Another approach using hierarchical extreme learning machine was proposed by Yao (2017).

However, finding time-correlation information remains a challenge in all above approaches. Hence, in our work, we proposed a soft sensor that is able to extract time-correlation information of the process as well as correlations among other measurements, so the dynamics of the system can be well established. Convolutional neural network (CNN) is used to obtain the process information from a moving window at each time point. Not only CNN has shown improved performance over traditional recurrent neural network (RNN) (Zhang, 2015), but it also reduces the training time over RNN (Baidu-Research, 2017). Besides, we also developed a signal recovery scheme in order to confront short-time sensor faults that happen in real processes.

2. Background

2.1. Rectified linear unit (Relu)

Relu is a bilinear activation function used in neural network, which is defined as:

$$f(x) = \max(0, x) \tag{1}$$

Where x is the input. Due to the simple nonlinear expression, the gradient computation is much more efficient than many other activation functions such as sigmoid. Besides, relu can effectively reduce the gradient vanishing problem since the derivative is independent of input domain. As of 2017, relu is widely used in deep learning for multiple types of tasks.

2.2. Convolutional neural network (CNN)

CNN is a deep neural network originally designed for image analysis. Recently, it was discovered that the CNN also has an excellent capacity in sequent data analysis such as natural language processing (Zhang, 2015). CNN always contains two basic operations, namely convolution and pooling. The convolution operation using multiple filters is able to extract features (feature map) from the data set, through which their corresponding spatial information can be preserved. The pooling operation, also called subsampling, is used to reduce the dimensionality of feature maps from the convolution

operation. Max pooling and average pooling are the most common pooling operations used in the CNN. Due to the complicity of CNN, relu is the common choice for the activation function to transfer gradient in training by backpropagation.

3. Methods

3.1. CNN soft sensor

Based on the recent success of the CNN in sequence data analysis (Zhang, 2015), we proposed a soft sensor using the similar idea. In the proposed soft sensor, convolutional filters with multiple region sizes are used to extract features at various granularities from moving windows. Hence, influence from different past time periods can be counted into future steps. Based on the extracted features, target measurement at the next time point can be estimated. The scheme of the proposed soft sensor is illustrated in Figure 1.

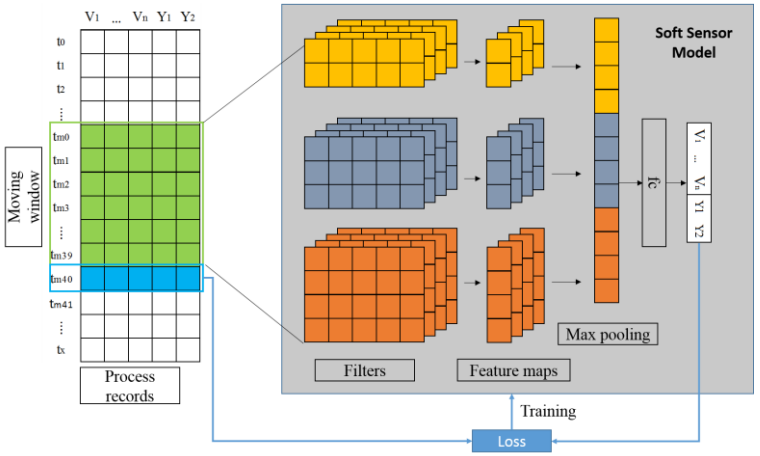


Figure 1. Illustration of overall architecture of the proposed soft sensor. *V* denotes the related variables to key variables *Y*. Their historical records are extracted by moving windows that are the model input. Feature maps are generated by convolutional 1D filters with multiple length. For the feature maps, max pooling and layer flatten operation (the fc layer) are used before final prediction.

3.2. Loss function

Since datasets from industrial processes are born with noise from measurements, conventional loss functions such as mean squared error (MSE) are not ideal, because such loss functions would just duplicate the noise into the model. Inspired by Salas's work (2017), we utilized the following loss function in order to reduce measurement noises for each variable.

$$\sum_{i=1}^{n}(y_i - \overline{y}_i)R^{-1}(y_i - \overline{y}_i)^T \tag{2}$$

where y_i is the predicted value, $\overline{y}_i$ is the target value, and R is the covariance matrix for measurements error which is estimated as $diag(\sigma_1, \sigma_2, \cdots, \sigma_i)$. It should be noted that the standard deviation used here should be taken from steady state, so the measurement errors can be properly estimated (Chen, 1997).

3.3. Implementation

Sensor faults are common in process operation. In practice, soft sensors still take data from real process measurements, and hence, faulty data can still downgrade the soft sensor performance. To improve the robustness of the soft sensor, we integrated the fault detection step with the soft sensor. In our work, we utilize a simple fault detection method that detect faults based on the mean and standard deviation values of the moving window. Besides, any proper fault detection methods can be replaced including any knowledge based methods to multivariable statistic methods (PCA T^2). By such method, the model can be robust for the sensor fault for a short time period.

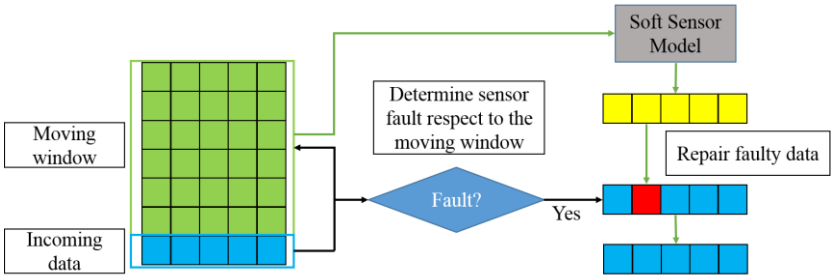


Figure 2. The scheme of integration of fault detection method with the soft sensor. The incoming data is tested with the current widow. Any faulty variables are replaced with the value generated from the sensor model.

4. Case Study

In this work, the proposed method is tested on an industrial pyrolysis reactor which is used to crack naphtha to ethylene. In the process, the naphtha is first mixed with steam before cracking in the furnace. In practice, due to the physical limitation, compositions of the gas-phase components in the outlet flow of reactor are unable to be measured in real time. The measurement takes about 16 minutes for each sample from the stream. Hence, the proposed soft sensor is applied in this process and the performance is tested on the historical data of the process.

Total 21 variables are selected, including 5 key variables (volume percentage measurements of methane, ethylene, ethane, propylene and propane) that require real-time prediction, and 16 related variables (e.g. feedstock compositions, feed flow rate, pressure, and temperature in different coils) that could provide extra process information to the model. Sensor faults are manually created into the dataset to verify the effectiveness of the proposed method.

5. Result and Discussion

5.1. Test result

The proposed model was trained with 9k data from an operation cycle whose recording interval is 1 minute. The model is validated by 1k data from the same operation cycle, and 500 data from another operation cycle was used for testing. Besides, sensor faults were introduced in the test data to verify the robustness of the proposed method. The overall test result is summarized in Figure 3, where the result was also compared with the partial least squares (PLS) model that did not account time-correlation information.

The result demonstrates the capability of the proposed soft sensor for key variables' prediction, and even for the drifting condition (Fig 3e), the proposed model can still trace the process trend. Additionally, when confronting sensor faults manually introduced into the dataset from time 190 to 210, the proposed signal recovery scheme proves its effectiveness. The comparison with the PLS model also indicates the superiority of the proposed model that outperforms the conventional linear and time-independent methods.

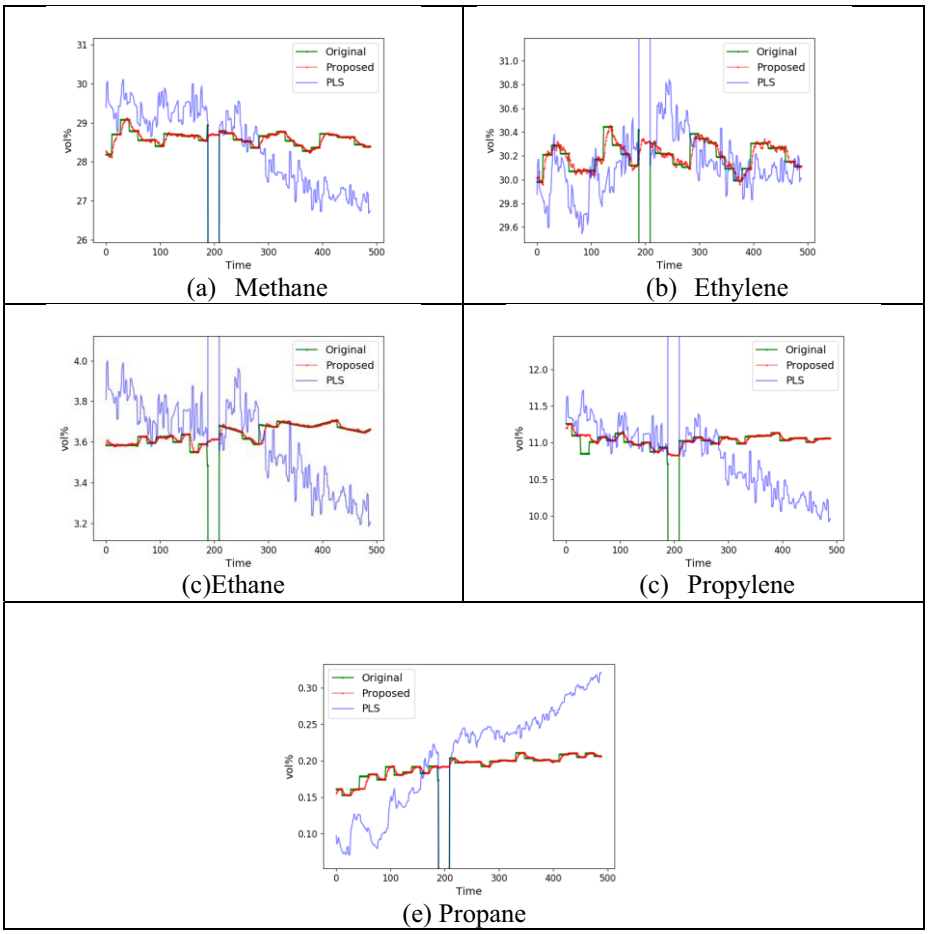


Figure 3. Composition prediction of gas-phase components under the condition of sensor faults that are introduced from time 190 to 210. The moving window size for the model is 10, and 60 filters for two filter region size of 5 and 10 are used in the convolutional layer.

5.2. Parameter selection

In the proposed soft sensor, the window size and filter configurations are important parameters that affect the ultimate performance. Hence, their effects are analysed (see Table 1). From the experimental results, a relative small window size is recommended. A large window length can lead the model less sensitive to the process changes and drifting, and hence an obvious delay can be observed, while a small length makes the model oversensitive to any fluctuations in the data including potential noises. Regarding the filter size selection, large filter size that covers the whole moving window seems

very necessary, since in all parallel experiments for different window lengths, such filter size setting can achieve better performance than others. Besides, multiple filters that can provide multi-granularity features can improve the model performance.

Table 1. Effect of window sizes and filter combinations to the model.

Window size	5		10				40			60		
Filter combination	2, 3	2, 5	3, 5	5, 10	3, 5, 10	10	20, 30	30, 40	20, 30, 40	20, 30	20, 30, 40	40, 50, 60
Loss	0.008	0.01	0.01	**0.004**	0.006	0.014	0.04	0.04	0.03	0.08	0.09	0.07

6. Conclusions

In this work, a soft sensor was proposed using the convolutional neural network, which predicts the measurements at next time step by extracting time-dependent correlations from a moving window. The proposed method was verified on the dataset from an industrial pyrolysis reactor. The model shows a remarkable performance on predicting the composition values of five gas-phase components. Additionally, through the experiment on manually introduced sensor faults, the proposed method proves its robustness to short-time sensor faults.

References

Baidu-Research. (2017, June). *DeepBench.* Retrieved from Github: https://github.com/baidu-research/DeepBench

Chen, J. B. (1997). Robust estimation of measurement error variance/covariance from process sampling data. *Computers & chemical engineering*, 593-600.

Hinton, G. (2010). A practical guide to training restricted Boltzmann machines. *Momentum*, 926.

Hinton, G. E. (2006). Reducing the dimensionality of data with neural networks. *science*, 504-507.

LeCun, Y. Y. (2015). Deep Learning. *Nature*, 436-444.

Salas, S. D. (2017). Online DEKF for State Estimation in Semi-Batch Free-Radical Polymerization Reactors. *Computer Aided Chemical Engineering*, 1465-1470.

Shang, C. Y. (2014). Data-driven soft sensor development based on deep learning technique. *Journal of Process Control*, 223-233.

Yan, W. T. (2017). A data-driven soft sensor modeling method based on deep learning and its application. *IEEE Transactions on Industrial Electronics*, 4237-4245.

Yao, L. &. (2017). Deep Learning of Semi-supervised Process Data with Hierarchical Extreme Learning Machine and Soft Sensor Application. *IEEE Transactions on Industrial Electronics*.

Zhang, Y. &. (2015). A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. *arXiv*.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50370-0

Estimating the Content of MAPD Contaminants in a Trickle Bed Reactor through a Virtual Analyser

Vicente Braga Barbosa[a,b*], Ana Rosa Massa[a], Karla P. S. Oliveira-Esquerre[a,b], Adelmo M. Aguiar Filho[b], Adonias M. S. Ferreira[a]

[a]*Industrial Engineering Program, Polytechnique Institute, Universidade Federal da Bahia, R. Prof. Aristides Novis, 2 , Federação, CEP 40210-630 Salvador, Brazil*
[b]*Chemical Engineering Department, Polytechnique Institute, R. Prof. Aristides Novis, 2, Federação, CEP 40210-630, Salvador, Brazil*
vicentebragab@gmail.com

Abstract

Historical data from thirteen monitored and controlled process variables (PVs) is used to build a Partial Least Square (PLS) model to estimate the content of methylacetylene and propadiene (MAPD) on the product of a propene production plant in Brazil. A technique called Variable Importance in Projection (VIP) is used to select the important PVs and the performances of the PLS-VIP models for two trickle bed reactors are compared with the performance of the PLS models for the same beds. An F test did not provide evidence for significant difference between the PLS and the PLS-VIP models in neither bed. The coefficient of determination of both PLS-VIP models were above 0.80 (0.81 and 0.87), an acceptable value given the industrial nature of the data. The developed models show are adequate for implementation as empirical models for virtual analysers that can estimate the content of MAPD contaminants in case on-line analysers fail to provide such.

Keywords: PLS, VIP, Virtual Analyser, MAPD content

1. Introduction

The petrochemical industry is constantly innovating its methods and searching for processes that are safer and more efficient. Strong competition among different companies calls for technology that can guarantee both the product specification and a solid yield on the invested capital. Such technological devices include on-line analysers such as gas chromatographers, which grant a faster answer on the composition of products when compared with laboratory analysis. However, the cost of such instrumentation is still high (thousands of dollars) and the constant monitoring of the material flow requires frequent skilled maintenance (Musaev, 2003). During those, the loss of vital information could lead to a halt in production, unless another device allows for such information to be carefully estimated. As such, virtual analysers claim a strategic position in modern industry due to its capacity to replace essential equipment, such as on-line analysers, for mathematical models when the former is unavailable.
On-line analysers are commonly used in the petrochemical industry. During the production of propene in a usual petrochemical plant, the removal of contaminants through catalytic hydrogenation employs gas chromatographers to evaluate catalytic conversion, selectivity and the content of contaminants, in order to guarantee product specification and process efficiency (Cohn, 2006). Other devices, such as temperature

probes and flow meters, may gather an immense amount of data about the process over the course of a campaign, which usually lasts between four to six months.
Such data can be used to predict the content of methylacetylene and propadiene (MAPD) contaminants in case the on-line analysers fail to provide information about the former, preventing a halt in production and a subsequent loss in capital. The prediction can be done through Partial Least Square (PLS) regression, which has been widely used in the monitoring of industrial processes, since it usually provides empirical models with reasonable prediction capacity (Indahl, 2014). However, as reported by Galindo-Prieto et al. (2015), highly computerized analytical instrumentation can produce datasets with a large amount of predicting variables. Reducing the dimensionality of the model can lead to easier interpretations and more robust predictions. One method to achieve such reduction is the Variable Importance in Projection (VIP), commonly used for selecting the important predicting variables in PLS models. (Liu and Chan, 2014).
Hence, the aim of this paper is to develop a virtual analyser to estimate the methylacetylene and propadiene (MAPD) content in the outlet of a trickle bed reactor on a propene production plant in Brazil. Such analyzer will use a PLS regression model in which the predicting variables are selected from thirteen process variables (PVs), currently used for monitoring the production process, through the Variable Importance in Projection (VIP) technique.

2. Catalytic Hydrogenation System

The hydrogenation reaction happens in a heterogeneous bed reactor, either bed A or bed B on Figure 01, which alternate as the primary reactor. If bed A is doing the conversion, bed B is bypassed. As the catalytic campaign of bed A ends, bed B starts its catalytic campaign and bed A is bypassed. As bed B ends its campaign, bed A starts a new one and the cycle repeats. The reactor feed is high on MAPD contaminants (up to 7% mol), which is out of market specification. The contaminants need to be converted to propene through the hydrogenation of MAPD. Concentration in the outlet of the bed reactor needs to reach a level lower than 0.35% mol before it goes into the next reactor (bed C), where the levels of MAPD drop to parts per million in mol. Levels of MAPD on the product are monitored by on-line gas chromatographers.

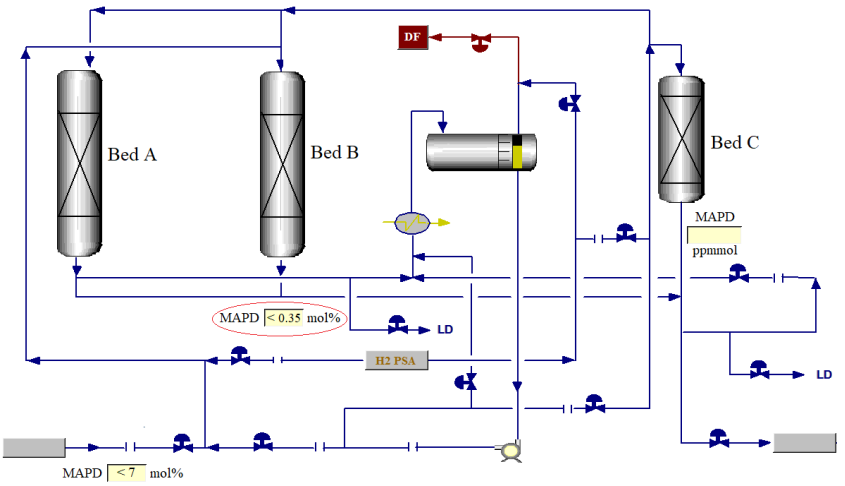


Figure 1. MAPD converter

Table 1. Process Variables

Process Variable	Symbol	Description
Fresh Feed Flow (C)	V1	Fresh propene flow, in ton/h
Recycle Flow (C)	V2	Recycle flow, from primary reactor, in ton/h
Fed H_2 flow (C)	V3	H2 fed into primary reactor, in ton/h
Bed Pressure (M)	V4	Primary reactor pressure, in bar
Combined Feed Temp. (M)	V5	Temperature of the combination of V1 (colder) and V2 (hotter), in ^oC
Bed Temp. (6 points) (M)	V6 to V11	Temperature of primary reactor across 6 different points, in ^oC
Outflow Temp. (M)	V12	Temperature of reactor outlet, in ^oC
MAPD in feed (M)	V13	MAPD content in feed flow, in % mol

The virtual analyser that is the focus of this paper will be monitoring the levels of MAPD currently monitored by the gas chromatographer circled on Figure 1. The thirteen monitored (M) and controlled (C) process variables of the trickle bed reactors are described in Table 1.

3. PLS-VIP Regression Model

The available dataset for modelling is divided into training (80%) and testing (20%) sets using the KenStone algorithm (Kennard and Stone, 1969). This algorithm starts selecting the testing set by choosing the two observations that have the largest geometric distance from one another. Among the remaining ones, it selects the observation with largest geometric distance from the closest observation already selected. This process is repeated until the testing set reaches 20% of the observations in the original dataset.

As reported by Indahl (2014), PLS regression has been widely used for empirical modelling where predicting variables are highly correlated. The predicting variables are organized in a data matrix X and the response variables in a data matrix Y. Both datasets can be decomposed respectively as $X = TP^T + E$ and $Y = TQ^T + F$ where T represents the score vectors of original dataset according to the first k terms of the Latent Variables (LVs), P and Q are the loading vectors of the data matrices X and Y, and E and F are residuals. (Liu and Chan, 2014). As such, a dataset $\hat{Y}$ can be predicted from a set of new observations X through Equation (1, where B is the pseudo-inverse of T.

$$\hat{Y} = \mathrm{X}P^T B Q^T \tag{1}$$

VIP scores provide the influence of each predicting variable j on the data matrix X_{IJ} on the PLS model of a given response vector $\hat{Y}$according to Equations (2, (3) and (4.

$$VIP_j = \sqrt{d \sum_{k=1}^{h} v_k \left(w_{kj}\right)^2 / \sum_{k=1}^{h} v_k} \tag{2}$$

$$\text{where } v_k = c_k^2 t_k' t_k \text{ and } c_k = \frac{t_k' y(k)}{t_k' t_k} \tag{3) and (4}$$

Where d is the number of variables and h the number of latent variables on the PLS model. The proportion of the fraction of the explained variance of X_{IJ}, for each variable j over all latent variables is expressed by v_k, which is weighted by the term w_{kj} that

represents the covariance between X and Y. The term c_k is calculated for each column of the PLS scores matrix T and for the predicted response Y.
As suggested by Akarachantachote et al. (2014), a threshold value for VIP of 1 is adopted as the criterion to define an important variable. After processes variables were selected, the same training and testing subsets used to develop the PLS model are used to develop a PLS-VIP model, where only the PVs selected by VIP are used.

4. Model Development

Data from one catalytic campaign of each bed was used to generate a PLS model for each bed after outlier removal using Principal Component Analysis (PCA). The VIP scores of each process variable from Table 1 is shown on Figure 2, where the threshold value of one is highlighted. For bed A, the selected PVs were the recycle flow (V2), the hydrogen fed flow (V3), bed pressure (V4), three temperature points across the bed (V8 to V11) and MAPD content in feed (V13). For bed B, the selected variables were recycle flow (V2), bed pressure (V4) and five temperatures points across the bed (V7 to V11). Although the fed hydrogen flow (V3) did not reach the threshold value of one for bed B, its importance to the conversion process along with the proximity of its VIP score (0.985) to the threshold (one) prevented its exclusion from the PLS-VIP model for bed B. The difference between the VIP scores for the same PVs on beds A and B suggests that, although there must be some similarities of the overall process behavior on both beds, a single empirical model describing both is not adequate.
Figure 3 displays the predicted MAPD content against the observed MAPD content (measured by the gas chromatographer) for bed A for the PLS and PLS-VIP models. Both show an adequate distribution of points around the support line, suggesting the models perform well when predicting MAPD content for bed A. The points are more dispersed around the support line for the PLS-VIP model and there is a drop of R^2 (from 0.84 to 0.81) and an increase in the Root Mean Square Error of Prediction, RMSEP, (from 1125 to 1207 ppm), when compared with the PLS model, as seen on Table 2. However, an F test did not show any significant difference between the RMSEP of the models, with an F_{cal} of 1.07. At $\alpha = 5\%$, the critical F for the dataset is 1.30. There is no significance difference between the biases of the models neither, and their low absolute values (23 and 28 ppm) suggest the lack of systematic errors.
A similar behavior is seen on Figure 4, where the predicted MAPD content against the observed MAPD content for bed B is shown. The points are still well distributed along the support line and the performance of the PLS and PLS-VIP models are even more suitable than on bed A, with an R^2 of 0.92 and 0.87, respectively.

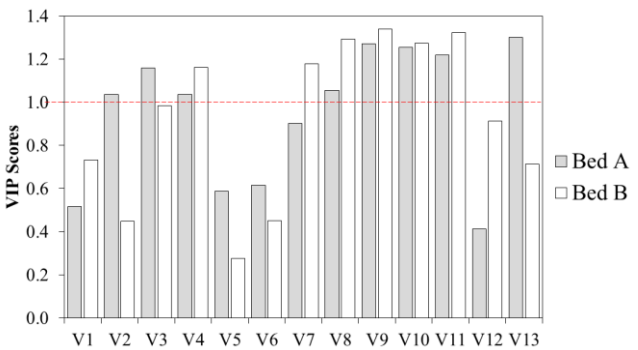


Figure 2. VIP Scores for beds A and B

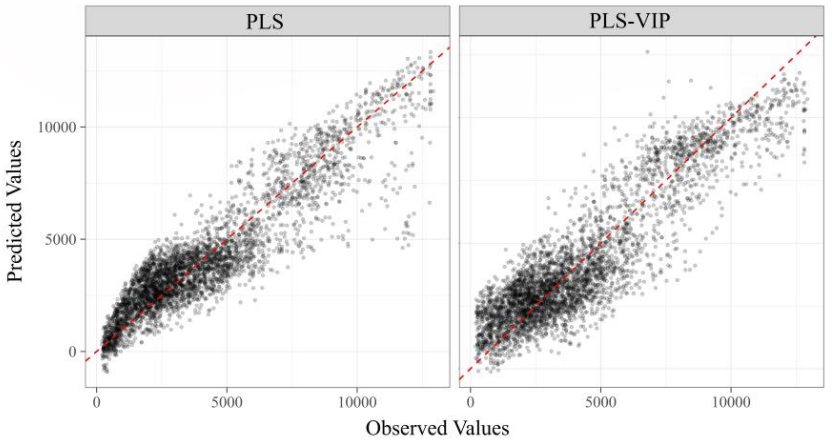


Figure 3. Predicted vs. Observed MAPD content - bed A

There is also an increase in the dispersion of points around the support line from the PLS to PLS-VIP model for bed B. There is a more noticeable increase in RMSEP (from 878 to 1131 ppm) and BIAS (from 1.38 to 32 ppm). Yet, an F test with $\alpha = 5\%$ did not support a significant difference of the models, with the F_{calc} at 1.28 and $F_{critical}$ at 1.30.
Table 2 compares the performances of the models before and after variable selection by VIP for both beds. All models have adequate performances and there is no evidence supporting significant difference between the PLS and the PLS-VIP models for neither bed. A reduction in the number of process variables required to estimate the MAPD content is important not only because it lowers the costs associated with the model, but also because it makes the model more robust. In case there is a failure of the on-line analyser, the virtual analyser is only able to guarantee the product is within specification if data for the input variables is available on real time. A model that requires less input implies that a less number of devices are required to keep the prediction running. As a result, in a case where both the gas chromatographer and the fresh feed flow meter fail, the PLS-VIP model would be able to predict the MAPD content whereas the PLS model would not. Hence, there is a much lower chance of there being a halt in production due to the lack of information on the concentration of MAPD on the product.

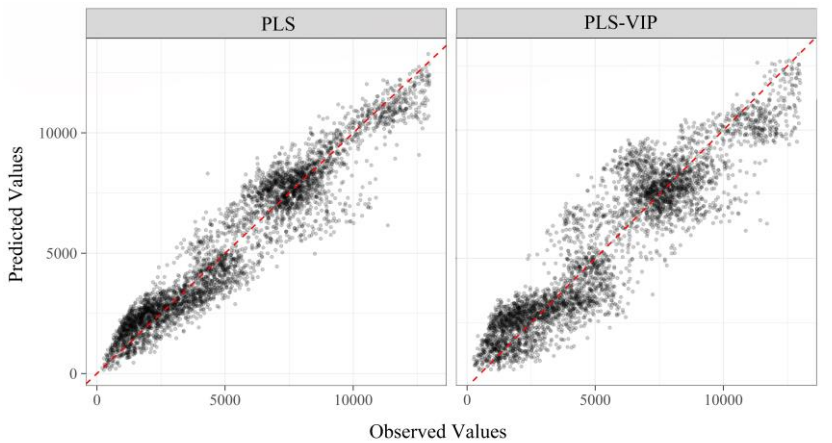


Figure 4. Predicted vs. Observed MAPD content - bed B

Table 2. Figures of Merit for PLS before and after VIP

	Before Variable Selection				After Variable Selection			
Bed	PVs	R^2	RMSEP / ppm	BIAS / ppm	PVs	R^2	RMSEP / ppm	BIAS / ppm
A	13	0.84	1125	23	9	0.81	1207	28
B	13	0.92	878	1.8	8	0.87	1131	32

5. Conclusions

The developed PLS-VIP models on this paper show that it is possible to achieve a reliable virtual analyser that predicts the MAPD content in the outflow of a trickle-bed reactor based on collected data from available process variables. Comparison between PLS and PLS-VIP models indicate that it is possible to reduce the number of required process variables as input on the model while still keeping suitable prediction capacity. The implementation of the established models does not require effort as there is already a control system in place, onto which the model could be adapted. Hence, the virtual analyser could be easily used in case the current on-line analyser fail to provide information on MAPD content of the product, which could potentially halt production of propene and cause an enormous loss in invested capital. Moreover, the methodology applied to arrive at the mathematical model for this virtual analyser can be easily replicated in other important reactors of the same plant, potentially making different processes more efficient and robust.

References

A. Akarachantachote, N.; Chadcham, S.; Saithanu, K. "Cutoff Threshold Of Variable Importance In Projection For Variable Selection" *International Journal of Pure and Apllied Mathematics*, vol 94, no. 3, 2014, pp. 307-322.

B. Cohn, P.F. Analisadores Industriais. 1st ed. Rio De Janeiro: Interciência, 2006, pp. 251-58. ISBN: 857193147X

C. Galindo-Prieto, B.; Eriksson, L.; Trygg, J. “Variable Influence on Projection (VIP) for OPLS Models and Its Applicability in Multivariate Time Series Analysis.” *Chemometrics and Intelligent Laboratory Systems*, vol. 146, 2015, pp. 297–304., doi:10.1016/j.chemolab.2015.05.001.

D. Indahl, U. G. "The Geometry of PLS1 Explained Properly: 10 Key Notes on Mathematical Properties of and Some Alternative Algorithmic Approaches to PLS1 Modelling." *Journal of Chemometrics*, vol. 28, no.3, 2014, pp. 168-80.

E. Kennard, R. W.; Stone, L. A. “Computer Aided Design of Experiments.” *Technometrics*, vol. 11, no. 1, 1969, pp. 137–148., doi:10.1080/00401706.1969.10490666.

F. Liu, J.; Chen, D. "Developing a Soft Sensor for an Air Separation Process Based on Variable Selection in Dynamic Partial Least Squares." *Computer Aided Chemical Engineering*, 24th European Symposium on Computer Aided Process Engineering 33, 2014, pp. 685-90.

G. Musaev, A. A. “Virtual Analyzers: The Concept of Design and Use in the Problems of Continuous Process Control” *Avtomatiz. Promyshl.*, no. 8, 2003, pp. 28-33.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50371-2

Root cause analysis of estimation error of a soft-sensor

Narutomo Ishikawa[a], Sanghong Kim[a*], Manabu Kano[b], Shinji Hasebe[a]

[a]*Department of Chemical Enginerring, Kyoto University, Nishikyo-ku Kyoto 615-8510, Japan*
[b]*Department of Systems Science, Graduate School of Informatics, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan*
kim@cheme.kyoto-u.ac.jp

Abstract

In many industrial processes, soft-sensors have been used for estimating a product quality or other key variables. Many kinds of modeling methods have been researched to deal with the problems such as small number of samples, the nonlinearity, and the changes in process characteristics. Because no modeling methods is perfect to all industrial processes, it is difficult to select the best modeling method and to appropriately tune parameters at the early stage of soft-sensor design. Thus, the estimation error of soft-sensors often becomes unacceptable. In such a case, another modeling method should be used, however, there is no systematic way to classify the root cause of bad estimation performance and select the alternative method. In this research, we focus on main two root causes, the shortage of samples and the process nonlinearity, and propose a method for evaluating the effect of each root cause. The proposed method provides important information to improve the estimation performance of soft-sensors. The usefulness of the proposed method is demonstrated through case studies of twelve actual process data.

Keywords: soft-sensor, root cause analysis, small number of samples, process nonlinearity

1. Introduction

In many industrial processes, on-line measurements of the quality variables are not always available because of the unacceptable cost of the measurement equipment, and the measurement delay. To solve this problem, statistical soft-sensors, which estimate difficult-to-measure variables from easy-to-measure ones, have been widely used. Various modeling methods such as partial least squares (PLS), support vector regression (SVR), and artificial neural network (ANN) have been developed to improve the estimation accuracy of soft-sensors. When the estimation accuracy of such modeling methods is poor, however, its reason is difficult to clarify. Generally, building a high performance soft-sensor is very laborious, because the input variables and the samples for model construction have to be carefully selected and parameters have to be tuned appropriately (Ogawa et al., 2008). From the practical viewpoint, time-saving for building a good soft-sensor is very important, and it is useful to clarify the root causes of the estimation error. In this research, we propose a method for evaluating effect of the number of model construction samples and the strength of the process nonlinearity on the estimation performance. In the proposed method, the estimation errors obtained by different number of model construction samples are used to evaluate the effect of number

of samples on the estimation performance, using PLS and locally weighted PLS (LW-PLS) (Kim et al., 2011) as modeling methods. Additionally, the strength of process nonlinearity estimated by these methods is used to judge whether the introduction of a nonlinear modeling method is appropriate for improving the estimation error or not. The usefulness of the proposed method is demonstrated through applying the proposed method to twelve process data.

2. Analysis methods for root causes of soft-sensors' estimation error

In industrial processes, main causes associated with soft-sensors' estimation error are as follows:

- Insufficient number of samples for model construction
- Strong process nonlinearity
- Inappropriate selection of the input variables
- Inappropriate selection of the samples for model construction
- Changes in process characteristics

In this research, we focus on "number of samples for model construction" and "process nonlinearity".

2.1. Analysis method for the effect of the number of samples and process nonlinearity on the estimation error

In the proposed method, the estimation error is calculated by using different number of samples and two modeling methods to evaluate the effect of the number of samples and the process nonlinearity on the estimation error. PLS and LW-PLS are used as the modeling methods. In LW-PLS, the localization parameter φ is used to adjust the weight of each sample. When φ is larger, smaller samples are prioritized, and when φ is zero the LW-PLS becomes the same as the conventional PLS (Kim et al., 2011). The proposed method consists of the following steps:

1. Set the localization parameter φ of LW-PLS to 0.
2. Set the ratio of samples used for model construction p to 0.4.
3. Set the iteration number r to 1.
4. Divide the samples of a process are randomly at a ratio of 7:3, and set 100p % of the former data as model construction samples and the latter data as model verification samples.
5. Construct an LW-PLS model by using the model construction data. The number of latent variables is selected by leave-one-out cross validation.
6. Apply the LW-PLS model is to the model verification samples, and calculate the root mean square error of prediction RMSEP:

$$\mathrm{RMSEP}_{p,r}^{\varphi} = \sqrt{\frac{1}{N_{\mathrm{v}}}\sum_{n=1}^{N_{\mathrm{v}}}\left(\hat{y}_{\mathrm{v},n}^{\varphi} - y_{\mathrm{v},n}\right)^2} \tag{1}$$

   where, N_{v} is the number of model verification samples, $y_{\mathrm{v},n}$ is the output of the n-th model verification sample, and $\hat{y}_{\mathrm{v},n}^{\varphi}$ is the estimated value of $y_{\mathrm{v},n}$ by LW-PLS model.
7. Update r to $r + 1$. If r is smaller than or equal to the maximum iteration number R, return to step 4. In this research, R is set to 50.
8. Calculate the average value of RMSEP MRMSEP, by

$$\mathrm{MRMSEP}_{p}^{\varphi} = \frac{1}{R}\sum_{r=1}^{R}\mathrm{RMSEP}_{p,r}^{\varphi} \tag{2}$$

9. Change the ratio of samples used for model construction p, and return to step 3 until all candidates of p are selected. In this study, the candidates of p are 0.4, 0.6, 0.8, and 1.
10. Change the localization parameter φ and return to step 3 until all candidates of φ are selected. In this study, the candidates of φ are 0, 0.5, 1, 1.5, 2, 2.5, and 3.

2.2. The evaluation of the effect of the number of samples

Using obtained $\mathrm{MRMSEP}_p^{\varphi}$, it is possible to evaluate whether the given number of samples is sufficient or not. The value of p corresponds to the number of samples used for model construction. Thus, it is expected that $\mathrm{MRMSEP}_p^{\varphi}$ with the same φ decreases with the increase of p. From the gradient of the $\mathrm{MRMSEP}_p^{\varphi}$, it is possible to estimate the number of the additional samples which are necessary to make the estimation performance acceptable.

2.3. The evaluation for the effect of process nonlinearity

Generally, when the process is close to linear, a smaller value of φ shows a better performance, and the nonlinearity of the process becomes stronger a larger value of φ shows a better performance. When φ is extremely large, a small number of samples are used to construct the model. Thus, usually $\mathrm{MRMSEP}_p^{\varphi}$ with the same p becomes convex with φ. By using this feature of the localization parameter φ, the degree of process nonlinearity is evaluated. The degree of process nonlinearity can be evaluated from the gradient of the $\mathrm{MRMSEP}_p^{\varphi}$. For example, if the $\mathrm{MRMSEP}_p^{\varphi}$ decreases with φ at the wide range of φ, the nonlinearity of the process should be strong. For such a process, we can expect that the performance of a soft-sensor will improve if we use a nonlinear model. On the other hand, if the gradient of $\mathrm{MRMSEP}_p^{\varphi}$ is almost zero or positive, it is expected that the nonlinearity of the process should be weak. In such a case, it can be stated that application of various nonlinear models will not greatly contribute to the improvement of accuracy of soft-sensor.

3. Application to actual process data

In this section, we describe the results of applying the cause analysis methods shown in the previous section to actual process data.

3.1. Overview of actual p*rocess data*

In this research, 12 actual process data from as chemical, pharmaceutical, and food processes are used. Table 1 outlines the data. The detailed information of datasets of processes 1 – 9 is confidential since the datasets are from private companies. Those of processes 10 – 12 are opened to public and located at the following URL: http://www.eigenvector.com/data/.

Table 1: Contents of data in each process

No.	Number of the samples	Number of the inputs	Estimation target
1	284	114	API content in a tablet
2	113	91	The amount of MgSt after cleaning
3	93	377	The amount of Ibuprofen after cleaning
4	776	40	The residual amount of NaOH after cleaning acid gas
5	634	9	Aroma concentration in the product of distillation column 1
6	606	70	95% boiling point of the product of distillation column 1
7	122	16	Impurity (chloride) amount in the the product of distillation column 2
8	405	23	Impurity (water) amount in the the product of distillation column 2
9	157	18	Impurity (water) amount in the the product of distillation column 3
10	460	650	A quality attribute of a tables
11	395	401	90% boiling point of diesel fuel
12	80	700	A quality attribute of a corn

Results and discussion

The results of applying the analysis method described in section 2 to each process data shown in Table 1 are illustrated in Figures 1 and 2.

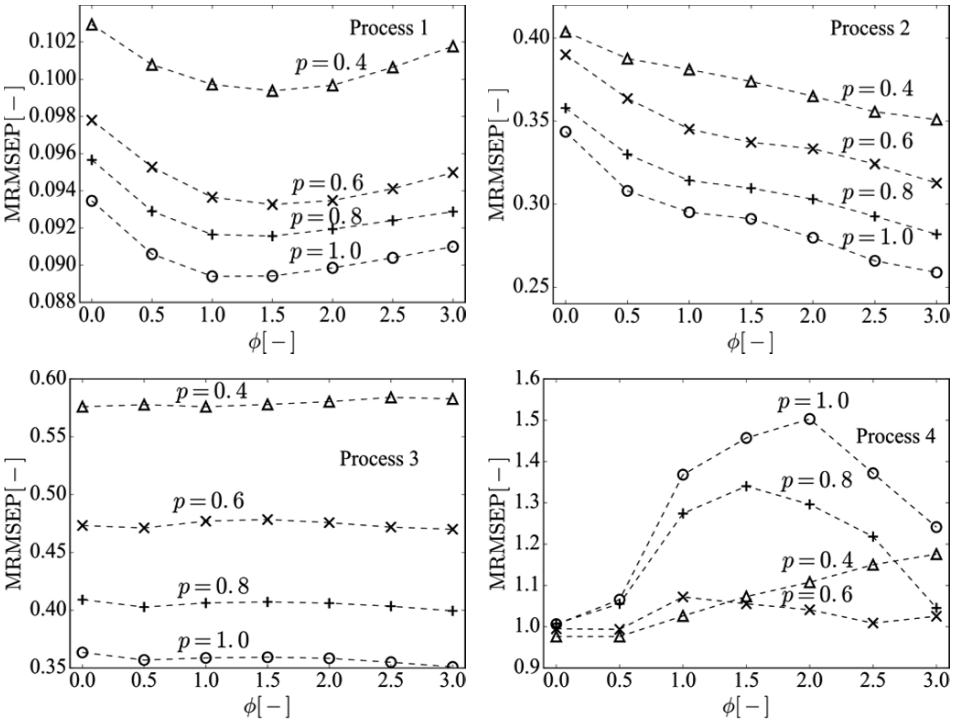


Figure 1: Estimation error in each process (Process 1-4)

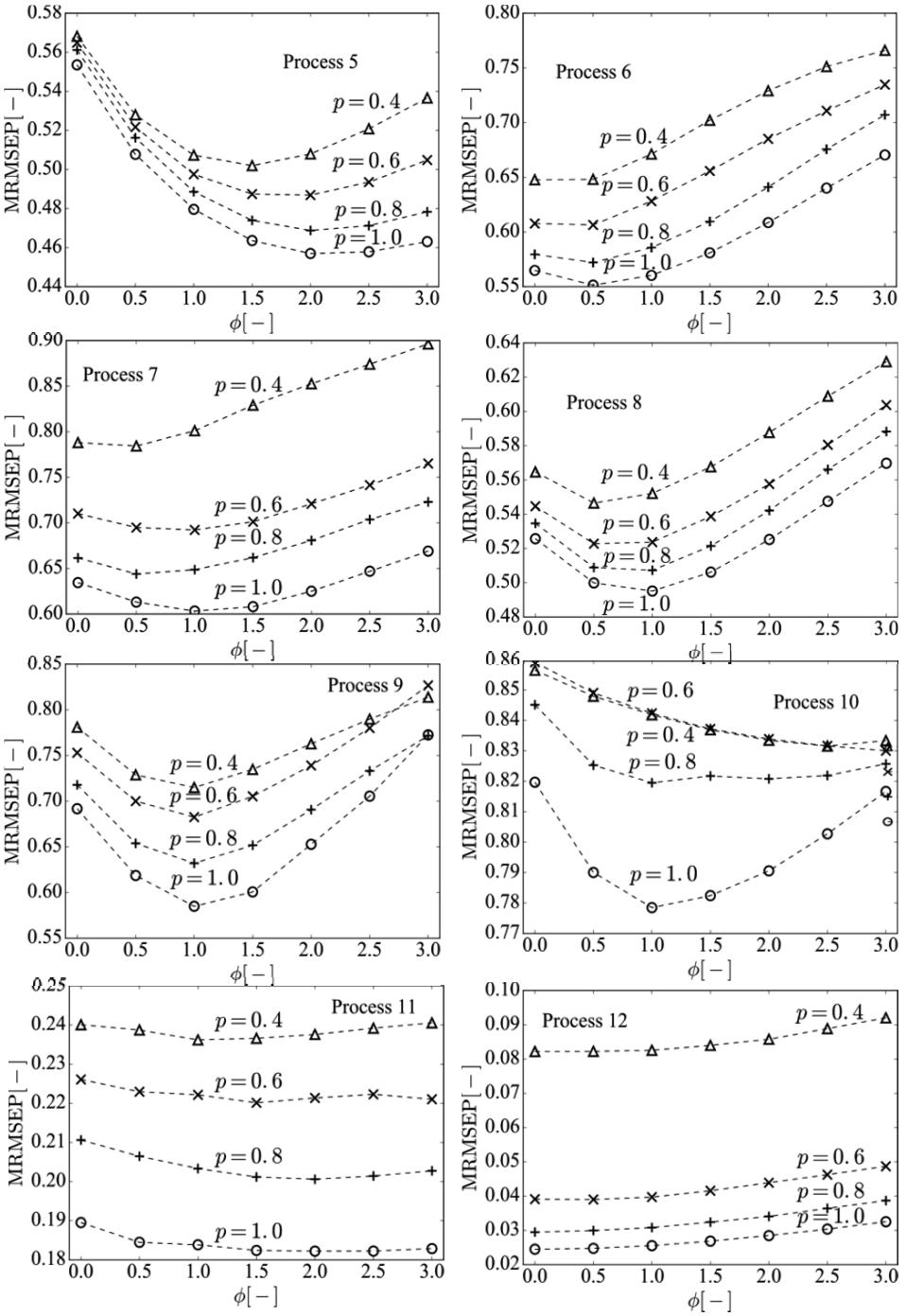


Figure 2: Estimation error in each process (Process 5-12)

3.1.1. Evaluation result of the effect of the number of model construction samples

In Figures 1 and 2, the estimation error decreases with the number of samples for model construction except for process 4. From this result, the estimation error when the number of model construction samples is increased can be estimated. For example, in process 1, it is expected that MRMSEP of linear PLS (φ = 0) will be 0.091 when the number of samples is increased by 20% (p = 1.2). By using this information, the number of samples required to reach an allowable estimation error is estimated, and engineers can make decisions as to whether or not to actually increase the number of samples. Moreover, since the same tendency is shown not only in PLS but also in LW-PLS, it can be expected that this method can be useful regardless of the modeling method.

3.1.2. Evaluation result of the effect of degree of p*rocess nonlinearity*

In process 2, the gradient of $\mathrm{MRMSEP}_p^{\varphi}$ with the same p has the tendency of decrease in wide range. In processes 1, 5, 8, 9, and 10, the $\mathrm{MRMSEP}_p^{\varphi}$ clearly has convexity with φ. In other words, the $\mathrm{MRMSEP}_p^{\varphi}$ decreases when φ is small, while the $\mathrm{MRMSEP}_p^{\varphi}$ increases when φ is large. This is because a small number of samples are used to construct the model. For these processes, we can judge that nonlinear modeling methods are effective to improve estimation accuracy because these processes can be regarded as highly nonlinear ones. On the other hand, in processes 3, 4, 6, 7, 11, and 12, the gradient of $\mathrm{MRMSEP}_p^{\varphi}$ is almost zero or positive. In such a case, it can be said that applying various nonlinear models will not greatly contribute to the improvement of the accuracy of soft-sensors since the nonlinearity of the process is expected to be weak.

4. Conclusions

In this research, we proposed a method for evaluating the root causes of the estimation error of the soft-sensor. The effectiveness of the proposed methods about insufficient number of samples for model construction and the degree of process nonlinearity was demonstrated through applying it to 12 process data. By using the proposed method, it becomes easier to decide whether to increase the number of samples for construction or whether to apply another nonlinear model as a model construction method. It is expected that the proposed method can contribute to efficient soft-sensor design at various industrial processes.

References

M. Ogawa, and M. Kano, 2008, “Practice and challenges in chemical process control applications in Japan.” *IFAC Proceedings Volumes*, 41, 2, 10608-10613.

S. Kim, et al., 2011, "Estimation of active pharmaceutical ingredients content using locally weighted partial least squares and statistical wavelength selection.", *International journal of* p*harmaceutics*, 421, 2, 269-274.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50372-4

Data mining-based algorithm for pre-processing biopharmaceutical manufacturing records

Gioele Casola[a*], Christian Siegmund[b], Markus Mattern[b], Hirokazu Sugiyama[a]

[a]*Department of Chemical System Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, 113-8656 Tokyo, Japan*
[b]*Sterile Manufacturing, F. Hoffmann–La Roche Ltd, Wurmisweg, 4303 Kaiseraugst, Switzerland*
casola@pse.t.u-tokyo.ac.jp

Abstract

In this work, a data mining-based algorithm is presented for the pre-processing—e.g., noise removal, batch isolation—of continuously measured historical records of biopharmaceutical manufacturing. The algorithm applies approximate string match and decision tree classifiers to remove noise from commercial production data automatically. Single batches are isolated using k-means clustering, after which algebraic semantic is used to characterize whether the data points within a batch describe the normal process execution or failures. The algorithm was applied to a dataset containing two months of manufacturing data in a cleaning and sterilization process. The performance of the algorithm was evaluated, resulting in a yield of 95 %, a mean time deviation of 1.3±4.6 %, and a rate of misclassification of 1.5 %, which showed a high performance of the algorithm. This study supports the introduction of data-driven automation approaches as well as smart manufacturing in pharmaceutical manufacturing.

Keywords: GMP, Natural language recognition, Supervised machine learning, Unsupervised machine learning.

1. Introduction

Pharmaceutical companies are facing an increasing competition and cost pressure on the market. In fact, the cost and time necessary to discover and develop new drugs are increasing with time (DiMasi et al., 2016), and for this reason, “big” companies are asked to contrast the increasing offer of generics—i.e., biosimilars and small molecule generics—in the market.

The optimization of processes and operations is of primary importance to ensure competitiveness while maintaining the high-quality standards and the compliance to good manufacturing practice (GMP). Process system engineering (PSE) approaches such as computational process design and retrofitting have been applied to achieve optimal design through Quality by Design (QbD) in the synthesis of pharmaceutical processes. In a review work, Rantanen and Khinast (2015) emphasized the importance of QbD-based thinking in the industry, and showed various technics and methods that are essential for the process modeling and design in the manufacturing science. While first-principle approaches tackle process design, synthesis, and optimization, data science approaches, such as data mining, can be used in the improvement of operations (Köksal et al., 2011).

In the same study, Köksal et al. (2011), showed that the quality of mining activities strongly depended on the quality of the data; data quality, in this study, is defined as integrity and noisiness—i.e., how much noise is comprised of the data. Before applying any intelligent algorithm, the data usually requires a preliminary "cleaning"; hence, in this study, the translation of raw data—e.g., timestamp, ID, and text—into systematized and noise-free numerical data is reviewed. Recently, in a proof-of-concept study, Meneghetti et al. (2016) presented a data mining-based algorithm for recognizing batches and process phases using data historians in drug product manufacturing of pharmaceuticals. However, as it was also pointed out in the study, challenges, such as the presence of noise in the data, could encumber the reliable translation of information into real knowledge, which is often the case in the pharmaceutical industry. Kemppainen (2017) discusses the challenges and opportunities in transforming the pharmaceutical manufacturing industry by increasing digitalization and introducing smart manufacturing concepts.

In this work, we present a data mining-based algorithm for pre-processing the continuously measured manufacturing records of biopharmaceutical batch processes in a robust manner. The algorithm combines natural language recognition and supervised machine learning approaches to classify and remove noise from commercial production data efficiently. Unsupervised machine learning as well as algebraic semantic are then applied to isolate single batches and characterize uncertainty, such as process failures. The application of the algorithm delivers data that can be utilized for the improvement of the operations—e.g., process runtime reduction—through the current industrial practices without manual interactions with the original records. The process that is the object of the study, where the records are to be pre-processed, is the so-called cleaning-in-place and sterilizing-in-place (CIP/SIP) process. Such a process consists of a sequence of independent tasks and is one of the change-over operations that contribute to producing high-quality drugs the most.

2. Pre-processing algorithm

The algorithm consists of six steps: import raw data, sequence raw data, train noise classifier, filter noise, cluster data by batch, and characterize batch. The algorithm shown in Figure 1 is used to identify noise and isolate single batches in a continuously recorded dataset while considering for process failures.

2.1. Import raw data (step 1)

The analysis timeframe—i.e., from initial to final time—and the resulting number of data points N_0 $[-]$ (with n as the counter) are selected to retrieve the correspondent data from the manufacturing database $\boldsymbol{M}$. The step imports all the time stamps, here defined as data tuples consisting of the time of execution t_n $[h]$, the task identity document (ID) id_n $[-]$, and a binary variable called process-noise (PN) class. The PN class classifies the timestamp as "process"—i.e., $PN = 0$—or "noise"—i.e., $PN = 1$—and is used in the steps 4 and 5. The resulting dataset is hereafter referred to as the raw dataset $\boldsymbol{D_1}$.

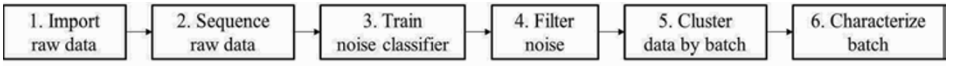


Figure 1 Sequence of steps necessary to pre-process manufacturing records.

2.2. Sequence raw data (step 2)

The sequencing step consists of four main sub-steps: (2.1) find extremity task sequence (ETS), (2.2) calculate primer size, (2.3) calculate heterogeneity coefficient H_n.

2.2.1. Find extremity task sequence (ETS) (step 2.1)

In this step, $\boldsymbol{D_1}$ is screened to locate the position of the ETSs ($p^{\mathrm{ETS}}[-]$). Start-ETSs (sETS) and end-ETSs (eETS) are the sequences of s-number of tasks located at the start and end of the process recipe. The positions p^{ETS} is found when the Wagner-Fischer distance $D_{n,n+s}^{\mathrm{ETS}}\ [-]$ (Ukkonen, 1985) between the id_n of an ETS and that of a same-size sequence of tasks belonging to $\boldsymbol{D_1}$ and starting from p_n is equal to zero. The step is concluded by gathering the execution times of the task relative to p^{EST} of each ETS in the set T^{ETS}.

2.2.2. Calculate primer size s_n^p (step 2.2)

The primer size $s_n^{\mathrm{p}}[-]$ is calculated dynamically for each point of $\boldsymbol{D_1}$ depending on three conditions, here shown by Eq. (1):

$$\begin{cases} s_n^{\mathrm{p}} = 1 & if\, d_n^{\mathrm{ETS}} \le s \\ s_n^{\mathrm{p}} = |d_n^{\mathrm{ETS}}| + 1 & if\, \mathrm{d}_n^{\mathrm{ETS}} > s\ s.t.\ p_n > p^{\mathrm{ETS}} \\ s_n^{\mathrm{p}} = \left(1 - e^{\frac{-d_n^{\mathrm{ETS}}}{h}}\right) s^{\mathrm{p,max}} & else \end{cases} \tag{1}$$

The parameters $d_n^{\mathrm{ETS}}\ [-]$, $s^{\mathrm{p,max}}\ [-]$, and $h[-]$ are distance between p_n and the nearest p^{ETS}, the maximal primer size, and a size-change velocity factor, respectively.

2.2.3. Calculate heterogeneity coefficient H_n (step 2.3)

First, the primer of size s_n^{p} at position p_n is defined as the fragment of the process recipe from k to $k + s_n^{\mathrm{p}} - 1$, where $k \in [1, K]$ is the counter of the tasks and K is total number of tasks in the recipe. Second, the minimum WF distance $D_n^{\mathrm{WF}}[-]$ between the primer and the fragment of $\boldsymbol{D_1}$ is calculated following Eq. (2).

$$D_n^{\mathrm{WF}} = \min_k D_{n,k}^{\mathrm{WF}}\left(s_n^p\right) \tag{2}$$

The distance $D_{n,k}^{\mathrm{WF}}[-]$ is the k-depedent distance, and is a function of s_n^{p}. Third, the heterogeneity coefficient $H_n\ [-]$ is calculated. H_n is defined in Eq. (3) and quantifies the degree of difference between the primer and the fragment of $\boldsymbol{D_1}$.

$$H_n = \frac{D_n^{\mathrm{WF}}}{s_n^{\mathrm{p}}} \tag{3}$$

2.3. Train noise classifier (step 3)

Supervised machine learning is used to classify noise; here, noise is defined as the recorded tasks that are not executed within a commercial batch—i.e., a task that does not follow the process recipe. Because of the industry—i.e., the industry is the final user of the algorithm—decision tree classifier was chosen for its simplicity and flexibility. The classifier was trained using the parameters H_n and s_n^{p}, and PN as predictors and response, respectively; the class PN of the data points contained in the training and the validation datasets is evaluated manually.

2.4. Filter noise (step 4)

In this step, the data points classified as noise—i.e., $PN = 1$—are removed from the dataset $\boldsymbol{D_1}$, resulting in a smaller and noise-free dataset, hereafter named as $\boldsymbol{D_2}$.

2.5. Cluster data by batch (step 5)

The data from $\boldsymbol{D_2}$ is clustered; the k-means algorithm is applied to isolate the clusters, data in which are then re-ordered to enforce a clear split between batches. The algorithm iteratively identifies the number of clusters, namely the number of batches, by evaluating how many batches are comprised in each cluster, and whether there are empty clusters. The clustering dimensions are defined as the process coordinate π_n $[-]$ and t_n; the parameter π_n locates the task id_n inside the process recipe with respect to the end of the process, where $\pi_n = 0$ and $\pi_n = 1$ stand for the beginning and the end of the process, respectively. The re-ordering algorithm, which moves data points within the clusters, uses the execution time t_n of the sETS stored in T^{ETS} as criteria; in practice, it enforces each cluster to start with an sETS. After determining the clusters, the time extrema of each cluster, which determine the beginning and the end of each batch, are used to summon all the data points contained in each batch from $\boldsymbol{D_1}$.

2.6. Characterize batch (step 6)

The data within one batch is analyzed to identify the presence of failures. Because GMP dictates the actions to be taken in case of failure, the failure-related data points are isolated by means of algebraic semantics. As an outcome, failure-related data points are labeled as *failure*, whereas the others are labeled as *normal*.

3. Implementation of the algorithm on commercial records

The algorithm was applied to a raw dataset containing two months of commercial records of a CIP/SIP process in a drug product manufacturing plant of F. Hoffmann-La Roche Ltd., Switzerland.

3.1. Preliminary studies

The design of the algorithm was analyzed by performing a multi-objective evaluation. As a preliminary study, an analysis was performed on the influence of the parameters s and $s^{\mathrm{p,max}}$ on the total execution time and prediction F-score of the algorithm. Another study was performed regarding the stability over time of the classifier. From Figure 2 (left), which shows a cumulative cost function—i.e., normalized weighted sum of the execution time and F-score— as function of s and $s^{\mathrm{p,max}}$, the first study could identify $s = 4$ and $s^{\mathrm{p,max}} = 20$ as the best constellation of the parameters. From Figure 2 (right), which represents the stability over time, the algorithm could deliver stable results for 1,500 h, after which the variance of the F-score started the increase. Because of the increasing instability of the classifier, the model was re-trained every 1,500 h.

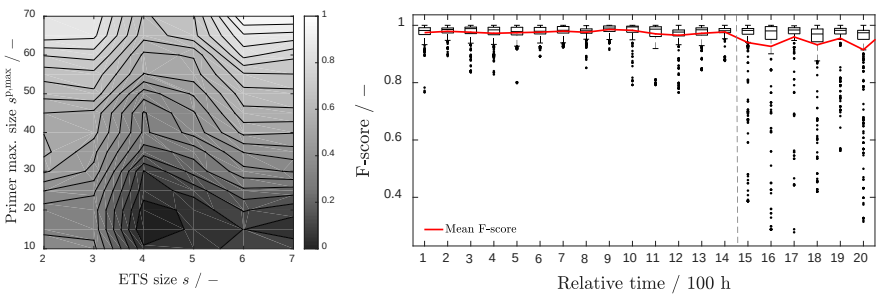


Figure 2 Results of the sensitivity analysis (left) and stability over time analysis (right).

3.2. Implementation

First, in step 1, manufacturing records of the CIP/SIP process were exported from the database, the dataset of which represented 1,451 h of the production data, and was defined as the raw dataset $\boldsymbol{D_1}$. The raw dataset consisted of the time series of $N_0 =$ 8,762 triplets of the variables t_n, id_n and PN.

Second, in step 2, the raw dataset was sequenced. After finding the sETS and eETS, 43 and 35 counts, respectively, the size of the primer s_n^{p} (see Eq. 1) and the heterogeneity coefficients H_n (see Eq. 3) were calculated. Sequencing, in a practical sense, was a transformation of dimension, namely, a translation from natural language in the id_n to a dimension describing the difference between the recorded data points from the process recipe.

Third, the decision tree classifier (max. number of split = 10, cost matrix $\begin{bmatrix} 0 & 1 \\ 10 & 0 \end{bmatrix}$) was trained and cross-validated (10-fold), where the parameters s_n^{p} and H_n were used as the predictors and the class PN as response. The classifier showed high prediction quality, yielding an F-score of 99.52 %; the result of the training, namely, the decision boundary, is highlighted in Figure 3 (left). In Figure 3, the decision boundary divides the combination of predictors in the two classes: noise (crosses) and process (circles).

Fourth, the classifier was used to evaluate the PN classes of the dataset $\boldsymbol{D_1}$; the data points that were classified as noise were eliminated. The filtering of the noise resulted in the noise-free dataset $\boldsymbol{D_2}$, containing 6,596 data points; hence, the 39% of the recorded points did not contain essential information, namely, was potentially hampering the coming steps of the algorithm.

Fifth, the two-dimensional k-means algorithm was used to identify, inside the dataset $\boldsymbol{D_2}$, the number and the domains of the clusters, which corresponded to a CIP/SIP batch. The iterative procedure, which was initiated by setting the initial guess to 43—i.e., the number of sETS—identified 40 different clusters. The result of the clustering can be seen in Figure 3 (right), where each grey zone represents a cluster, namely one batch. The points located within each cluster belong to the same batch.

Sixth, failures within each cluster were identified by isolating the failure specific task and alarms using semantic. The result of this step, namely the points categorized as *failure*, is highlighted in red (Figure 3, right); these data points can be used to evaluate the relative impact of uncertainty on the total time invested in the CIP/SIP. Because CIP/SIP is a support process—i.e., process which does not produce products—, the shorter the runtime is, the higher the production capacity of the manufacturing plant becomes, resulting in potentially higher revenues.

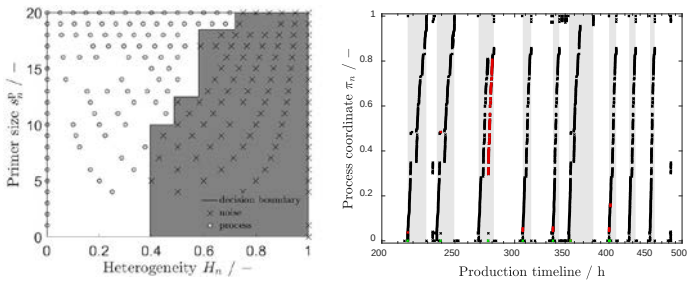


Figure 3 Results of the noise-classification (left) and batch-clustering (right) operations

The performance of the algorithm was evaluated by comparing the outcome of the algorithm with the manual pre-processing using three indicators: yield, mean time deviation and misclassification rate. The evaluation resulted in a yield of 95 %, a mean time deviation of 1.3±4.6 %, and a rate of misclassification of 1.5 %, which showed the high performance of the algorithm.

4. Conclusion and outlook

In this study, we presented an algorithmic approach for the pre-processing of biopharmaceutical manufacturing records considering the presence of uncertainty. The algorithm is used to automate the isolation of batches within a continuously measured historical record. DNA-like sequencing technics, an example of which is the approximate string matching algorithm, were used to transform natural language data into quantitative data. Such data was subsequently used for the classification of noise by decision tree models. After clustering and filtering the dataset into single batches, the impact of uncertainty was characterized. The outcome of characterization was the isolation of the data points describing operations that are not expected in the normal operational framework, and that can be attributed to the presence of failures in the process. A limitation is the relatively high algorithm execution time, which was attributed to the sequencing step dealing with text-recognition algorithms. Another limitation is the fact that the heterogeneity coefficients can only be calculated in the presence of a significant amount of the data, rendering online applications difficult. In future works, the algorithm and the resulting data will be applied in a framework for implementing predictive maintenance in biopharmaceutical drug product manufacturing. The current and the future study support the introduction of data-driven automation approaches as well as industry 4.0 and smart manufacturing in pharmaceutical manufacturing.

Acknowledgement

The authors acknowledge the financial support from a Grant-in-Aid for Young Scientists (A) No. 17H04964 from the Japan Society for the Promotion of Science.

References

DiMasi, J.A., Grabowski, H.G., Hansen, R.W., 2016. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J. Health Econ.* 47, 20–33.

Kemppainen, P., 2017. Pharma Digitalisation: Challenges and opportunities in transforming the pharma industry. *Eur. Pharm. Rev.* https://www.europeanpharmaceuticalreview.com/news/51733/pharma-digitalisation-challenges/ (accessed 8.3.17).

Köksal, G., Batmaz, I., Testik, M.C., 2011. A review of data mining applications for quality improvement in manufacturing industry. *Expert Syst. Appl.*

Meneghetti, N., Facco, P., Bezzo, F., Himawan, C., Zomer, S., Barolo, M., 2016. Knowledge management in secondary pharmaceutical manufacturing by mining of data historians"A proof-of-concept study. *Int. J. Pharm.* 505, 394–408.

Rantanen, J., Khinast, J., 2015. The Future of Pharmaceutical Manufacturing Sciences. *J. Pharm. Sci.* 104, 3612–3638.

Ukkonen, E., 1985. Algorithms for approximate string matching. *Inf. Control* 64, 100–118.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50373-6

Soft-Sensor Design for a Crude Distillation Unit Using Statistical Learning Methods

Aysun Urhan[a], Nevin Gerek Ince[b], Ronald Bondy[b], Burak Alakent[a*]

[a]*Department of Chemical Engineering, Bogazici University, Istanbul, 34342, Turkey.*

[b]*SimSci by Schneider Electric, Houston, TX, USA.*

[*]*burak.alakent@boun.edu.tr*

Abstract

Data-driven soft-sensors are statistical models constructed from historical data, and expected to perform well both in normal and novel process conditions. Numerous adaptive mechanisms for soft-sensors have been developed, but more work is required to develop appropriate statistical modeling tools for chemical processes, which yield highly collinear measurements in changing operating conditions. In the current study, numerous predictor subset selection and statistical learning tools are employed on 12 different data sets from a crude distillation unit simulated in SimSci PROII™, and the generalization capacities of the models are evaluated using the test points inside and outside the operating regions of the training sets. Relevance vector machine is found to be promising for adaptive modeling, with accuracies close to those of neural network models.

Keywords: adaptive modeling, cross-validation, predictive modeling, predictor selection.

1. Introduction

Data driven soft-sensors for chemical processes are usually constructed using global modeling methods, such as partial least squares (PLS), artificial neural networks (NNs), support vector machines, or adaptive local models (Yan et al., 2004, Kadlec et al., 2011). Proper tuning of parameters in these statistical models is essential for achieving a high generalization capacity, for which cross-validation (CV) is commonly used. The current CV practice involves partitioning the training data into k (usually taken to be between 5-20) folds, each of which is predicted using models constructed from the concatenation of the remaining folds (Kohavi, 1995). Online automatic sampling of processes, producing tens to a few hundred collinear process variables, creates a large number of redundant predictors, inflating the variance of the quality variable predictions. Predictor subset selection algorithms may be used to improve the accuracy and reduce the computational complexity of the statistical models (Hastie, et al., 2011), and to determine a convenient dimension of the neighborhood space, i.e. the subspace in which nearest neighbor search is performed, in order to combat "curse of dimensionality". It should, however, be noted that predictor selection, a form of parameter tuning itself, must be performed in an "outer" loop in CV to prevent biased prediction error (PE) estimates (Reunanen, 2003).

In the chemical process literature, various tools, such as mutual information or NNs, are used to reduce the dimension of predictors (Sun et al., 2016, Jin et al., 2014), yet

experience-based subset selection is still not uncommon, bringing doubt on the generalizability of the suggested modeling techniques. Additionally, abrupt changes in operational conditions necessitates availability of modeling methods, which, trained on small samples, can make accurate predictions outside the operating region; hence, statistical learning algorithms convenient for soft sensors in chemical/biochemical industries are to be developed. The current study aims to (i) construct "automated" unbiased statistical learning methods using various model building tools (Section 2.1), (ii) implement predictor selection methods (Section 2.2), and (iii) assess their performances in predicting T95 (distillation temperature of 95% vol. of oil) measurements from a crude distillation unit (CDU), simulated by SimSci ProII^{TM} (Section 2.3), under small/large sizes of training data and interpolation/extrapolation test data. The promising learning methods for adaptive soft-sensing are discussed Section 3.

2. Statistical Learning Methods

2.1. Linear and Nonlinear Modeling Approaches

A linear statistical model can be represented by, $y_i = \mathbf{x_i}\boldsymbol{\beta} + \epsilon_i$; response and predictor variables are denoted with the mean centered $y_i \in \mathbb{R}$, and $\mathbf{x_i} \in \mathbb{R}^p$ for $i = 1, \ldots N$ samples, respectively, and ϵ_i is the uncorrelated random error. Here, $\boldsymbol{\beta} \in \mathbb{R}^p$ is estimated using training data $\mathbf{X} = [\mathbf{x_1} \ \ldots \ \mathbf{x_N}]^T$. Since least squares (LS) estimates have high variance in the presence of highly redundant predictors, PLS extracts orthogonal components, which yield the maximum covariance between the response and the predictor variables, and linear inner relations are built between each score ($\mathbf{t_j}$) and unexplained (residual) response ($\mathbf{y_j}$) vector in a stepwise manner (Wold et al., 2001). Ridge regression (RR) and Lasso minimize the sum of squared residuals subject to a constraint on the norm of $\boldsymbol{\beta}$, equivalently stated as $\min_{\boldsymbol{\beta}} \left\{ \frac{1}{2} \sum_{i=1}^{N} (y_i - \mathbf{x_i}\boldsymbol{\beta})^2 + \lambda \sum_{j=1}^{p} \|\beta_j\|^q \right\}$; the amount of shrinkage (λ) of the LS parameter estimates, is equal to 1 for Lasso, and 2 for RR (Hastie, et al., 2011). In Bayesian linear regression (BLR), assigning a zero mean Gaussian prior on the weight parameter yields a posterior distribution with mean values equivalent to the ridge estimates, while defining hyperparameters for each weight in relevance vector machine (RVM) yields sparsity due to narrow zero-mean posterior distribution of most of the weights (Tipping, 2001).

A nonlinear statistical model can be represented by $y_i = f(\mathbf{x_i}) + \epsilon_i$, in which $f(x)$ can be approximated with either linear functions, as discussed above, or nonlinear modeling methods, such as nonparametric models, and NN. Kernel PLS may be used to construct a non-linear PLS model, in which a nonlinear mapping is performed from the original data space to the feature space using a kernel function, while the inner relation remains linear (Rosipal and Trejo, 2001). Nonlinear PLS modeling (PLS_{QI}) may also be performed via replacing the linear inner relation between $\mathbf{t_j}$ and $\mathbf{y_j}$ with a quadratic function (Wold, et al., 1989). NN is a collection of interconnected artificial neurons, which pass the weighted input signals through threshold transfer functions, and a NN model with at least one hidden layer may be used to approximate any continuous function (Cybenko, 1989).

2.2. Predictor Selection

The simplest alternative to exhaustive search is sequentially adding or removing predictors, based on correlation coefficient ($\hat{\rho}_i$) and mutual information (I_i) between the

i^{th} predictor and the response variable. At each step of forward selection (FS), predictor improving the fit the most significantly is added to the model (Hastie, et al., 2011). In jack-knife estimation scheme on PLS, PLS(JK), backward selection is employed in eliminating insignificant predictors (Martens and Martens, 2000). In minimum redundancy-maximum relevance (MRMR) method, predictor yielding the maximum ratio of relevance to redundancy is selected in stepwise fashion (Ding and Peng, 2005). In the current study, relevance and redundancy are measured via $\hat{\rho}$, as in $RC_i = \widehat{\rho_i}/VIF_i^{0.75}$, or $\hat{I}$, as in $RM_i = I_i/VIF_i^{0.75}$, and variance inflation factor (VIF), respectively.

3. Model Construction and Assessment using Synthetic CDU Data

SimSci PROII™ is used to model a CDU, to which crude oil (API gravity 29.5) at 390°C and 2 atm enters at a flowrate of 13.5 m^3/min (Figure 1A). Approximate fractional factorial design and perturbations on eigenvectors around the nominal conditions were employed on 16 inputs, such as crude flowrate (F_c) and reflux ratio (R), to represent two different operational modes, from which various datasets were produced at steady state. Columns of the predictor matrix (**X**) consist of the measured values of 70 process variables, such as inputs, column and side-stripper tray temperatures (T_i), condenser duty (C_D) (dashed lines in Figure 1), and T95 values of kerosene (y_1), diesel (y_2) and atmospheric gas oil (y_3) outlet streams are the response variables, to which uncorrelated measurement noise ~$\mathcal{N}$(0°C, 2.5^2°C^2) was added. Test Sets were formed in two different simulation settings: interpolation data (D_{int}) was collected from the same operating region sampled during the training set, while perturbations in input variables covered a wider range and different subspaces for producing extrapolation test data (D_{ext}). σ_1^2/σ_2^2 and σ_1^2/σ_{10}^2 (σ_i is the i^{th} singular value of **X** matrix) lie between 4-8 and 25-400, respectively, showing the high redundancy of the predictors.

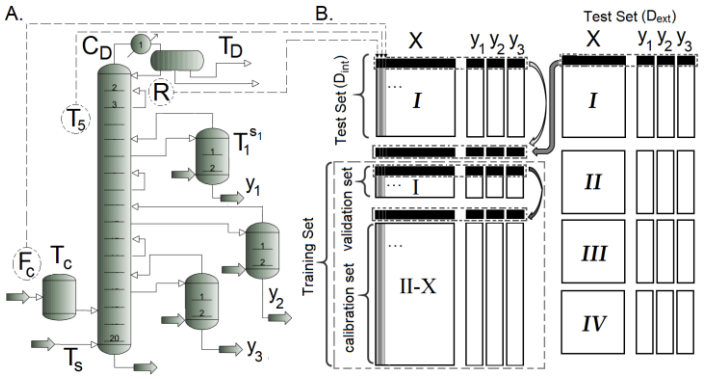


Figure 1. Schematic representation of (A) the CDU; (B) model training/testing algorithm. Construction of the predictor matrix is demonstrated using representative process variables (dashed arrows). Solid arrows (between the blocks of data) represent the inclusion of validation/test data to calibration/training data in stepwise manner for the adaptive local RVM method.

For each response variable, 37 predictive models (P) (Table 1) were trained on D_{int} for N=225 and 50 (Training Set in dashed lines in Figure 1B), and tested using repeated double CV (Krstajic, et al., 2014). Linear models were trained also on data including

squares of predictors (SP). The inner loop, consisting of 10 partitions of validation-calibration sets (normal Roman numbers in Figure 1B) and 20 random repetitions, was used for tuning model parameters and determining number of predictors in 1 to 3-D grid searches. The tuned parameters were then used in fitting the suggested model structure to the whole Training Set and predicting the Test Set. Additionally, models obtained for each calibration set was used to form an ensemble of predictive models (crogging) for PLS and NN models (Barrow, 2013). The outer loop was partitioned into four folds, each consisting of a Test Set of 75 observations (italic Roman numbers in Figure 1B), to determine unbiased PE estimates ($\widehat{\text{PE}}$), computed by root mean square error: $\text{RMSE}_k = \left(\frac{1}{75}\sum_{i=1}^{75}\left(y_{k,i} - \hat{y}_{k,i}\right)^2\right)^{0.5}$, k = 1, 2, 3. In both inner and outer loops, calibration sets and training sets were autoscaled (except for PLS/Sc), and validation and test sets were scaled accordingly. Repeating the outer loop 10-15 times yields 50 $\widehat{\text{PE}}$ values, which were averaged ($\text{avg}\widehat{\text{PE}}$). D_{ext} was not used during parameter tuning and training; a similar four-fold outer loop architecture was also used on D_{ext} (the second column in Figure 1B) to decrease sampling variability in comparing the prediction performances on D_{int} and D_{ext}. While batch training was used for the first 36 methods, an adaptive learning framework was adopted in local RVM (RVM_{L}); optimum dimension of neighborhood space and number of nearest neighbors, measured by Euclidean norm, were determined by CV on calibration sets, with stepwise inclusion of each validation sample to calibration set and each test sample to Training Set (shown with solid arrows between blocks in Figure 1B).

Learning and predictor selection methods are labeled with the name of the former method followed by the latter in parentheses, while crogging and predictor scaling based on PLS **X** residuals are denoted by /Cr and /Sc, respectively, e.g. $\text{PLS}_{\text{SP}}(Red_{cor})$ uses Red_{cor} for predictor selection on **X** including the quadratic terms and PLS for parameter estimation. Predictive accuracies of models are assessed using the following efficiency metric on P:

$$\text{Eff}_i(k,l) = \frac{\min_{j\in P}\left(\text{avg}\widehat{\text{PE}}_j(k,\{N=225,\text{Dint}\})\right)}{\text{avg}\widehat{\text{PE}}_j(k,l)},\ k \in K \text{ and } l \in L \quad (1)$$

Here, K represents three T95 measurements and two modes of data production, and L consists of D_{int} and D_{ext} for N=50, 225. $\text{Eff}_i(k,l)$ taking a value close to unity shows that accuracy of the i^{th} method is close to that of the best predictor on D_{int} for N=225. $\text{Eff}_i(k,l)$ values are averaged and minimized over K to obtain $\text{avgEff}_i(l)$ and $\text{minEff}_i(l)$, respectively, which are indicators of the relative generalization capacities of the models.

Table 1. Predictor selection and statistical learning methods used in the current study.

LS, RR, RR_{SP}, BLR, BLR_{SP}, PLS, PLS_{SP}, PLS/Sc, PLS_{SP}/Sc, PLS/Cr, PLS_{SP}/Cr, PLS (FS), PLS_{SP}(FS), PLS(FS_{val}), PLS(Red_{cor}), $\text{PLS}_{\text{SP}}(\text{Red}_{\text{cor}})$, PLS(JK), PLS_{SP}(JK) PLS(JK)/Cr, PLS_{SP}(JK)/Cr, RVM, RVM_{SP}, Lasso, Lasso_{SP}, PLS_{QI}, PLS_{QI}(FS), $\text{PLS}_{\text{QI}}(\text{Red}_{\text{MI}})$, Kernel PLS, NN, NN/Cr, NN(FS), NN(FS)/Cr, $\text{NN}(\text{Red}_{\text{cor}})$, $\text{NN}(\text{Red}_{\text{cor}})$/Cr, $\text{NN}(\text{Red}_{\text{MI}})$, $\text{NN}(\text{Red}_{\text{MI}})$/Cr, RVM_{L}(FS)

4. Results & Discussion

Representative examples of the prediction methods are placed in Figure 2A with respect to avgEff and minEff values. For NN(FS-MRMR)/Crogging and RVM_{L}(FS) models,

minEff on D_{int} is > ~95%, followed by NN(FS), $PLS_{QI}(Red_{MI})$, PLS(FS), PLS_{SP}(FS), RVM, PLS(Red_{cor}) methods, while minEff drops down to < ~90% for Lasso, PLS, PLS_{SP} and NN. Extrapolation efficiencies (Figure 2B) reveal a clearer picture of the relative performances: the lowest prediction accuracy of local RVM_L(FS) method on extrapolation set can reach 70% of the highest prediction accuracy by a global method on interpolation test set.

Extrapolation efficiency of the global RVM exceeds 50%, followed by NN(FS/MRMR)/Crogging methods. PLS(Red_{cor}) has the highest efficiency among the remaining predictors, while PLS(FS) and PLS_{SP}(FS), contrary to their small PEs for D_{int}, show low extrapolation efficiencies (too small to display in Figure 2). Average efficiency of Kernel PLS is unacceptably low (Table 2), and occasional high PEs prevent Lasso from being a reliable estimator. Efficiencies of PLS, RR, BLR and NN lag significantly behind those of the estimators discussed above, and including quadratic term does not generally improve efficiency. PLS(FS_{val}), in which predictor selection is guided by the inner CV loop, yields worse efficiency compared to that of PLS(FS).

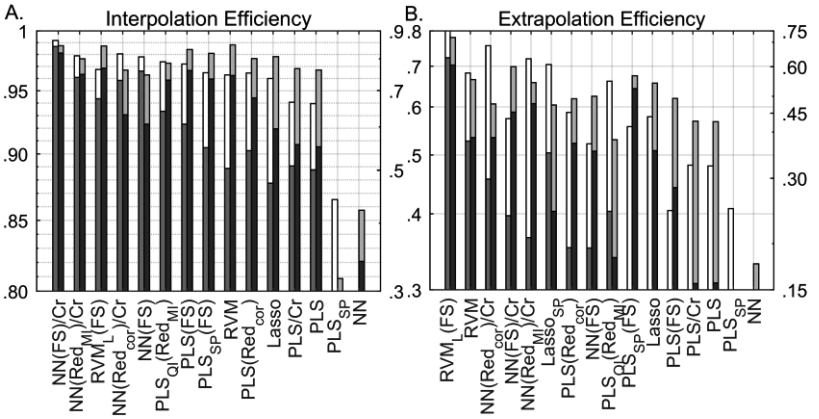


Figure 2. Prediction efficiencies of the best and benchmark methods for (A) D_{int} (B) D_{ext} data: Bars on the left and right side of each tick corresponds to N=225 (left axis) and 50 (right axis), respectively; the light and dark colored bars correspond to avgEff and minEff values, respectively.

FS-based methods yield marginally better predictions for N=225 on both D_{int} and D_{ext}, while predictions using MRMR are highly superior on D_{ext} with N=50. Number of predictors selected by MRMR is found to exceed that by FS, particularly for N=50, hence resulting models are likely to possess a higher amount of information on the predictor-response relation, possibly improving extrapolation predictions.

Table 2. RMSE values of selected prediction methods, averaged over datasets.

Name	RMSE (°C)	Name	RMSE (°C)
RVM_L(FS)	2.66, 3.06, 3.19, 3.60*	Lasso	2.68, 3.24, 8.14, 5.10
NN(Red_{MI})/Cr	2.62, 3.22, 3.91, 4.80	PLS	2.74, 3.44, 6.61, 7.66
RVM	2.67, 3.04, 3.90, 4.80	Kernel PLS	3.13, 4.34, 7.33, 8.70
PLS(Red_{cor})	2.67, 3.23, 4.98, 5.43	NN	3.41, 6.17, 47.0, 17.0

*: Numbers separated by comma are PEs for N=225 and 50 on D_{int} and N=225 and 50 on D_{ext}, respectively.

5. Conclusions

Raw industrial data usually consists of large but redundant set of predictors, so an automated predictor selection and model construction method is required for unbiased predictions. In the current study, a large number of predictor selection and modeling methods, mostly tuned by repeated k-fold CV, are tested on various synthetic data, sampled both from the interior and exterior regions of operation history. The modeling methods are coupled with statistical tests when necessary to prevent inflation of variance in an automated modeling setting. Relatively low prediction efficiencies are encountered for LS, PLS, RR; BLR, Kernel PLS, NN and Lasso. Predictor selection has significant benefit in linear/nonlinear PLS and NN models, particularly with crogging. PLS(Red_{cor}) carries potential for adaptive modeling, and the only non-linear PLS model with moderate extrapolation efficiency is constructed via Red_{MI}. RVM, being >10 times faster than NN, has a high generalization capacity, and using RVM in an adaptive local frame improves the prediction accuracy significantly. Future study aims to implement the best methods in the current study to chemical/biochemical processes in dynamic adaptive frameworks.

References

Barrow, D.K., 2013. Crogging (cross-validation aggregation) for forecasting—A novel algorithm of neural network ensembles on time series subsamples, IJCNN, 1-8.

Cybenko, G., 1989, Approximation by superpositions of a sigmoidal function, MCSS, 2, 303-314.

Ding, C.H.Q. and Peng, H., 2005, Minimum Redundancy Feature Selection from Microarray Gene Expression Data. J Bioinform. Comput. Biol., 3, 185-205.

Hastie, T., Tibshirani, R. and Friedman, J. H., 2011. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York, USA.

Kadlec, P., Grbic, R., Gabyrs, B., 2011, Review of adaptation mechanism for data-driven soft sensors, Comput. Chem. Eng., 35, 1-24.

Krstajic, D., Buturovic, L.J., Leahy, D.E. and Thomas, S., 2014. Cross-validation pitfalls when selecting and assessing regression and classification models. J. Cheminform., 6, 1-15.

Jin, H., Chen, X., Yang, J. and Wu, L., 2014. Adaptive soft sensor modeling framework based on just-in-time learning and kernel partial least squares regression for nonlinear multiphase batch processes, Comput. Chem. Eng., 71, 77-93.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. IJCAI, 14, 1137-1145.

Martens, H. and Martens, M., 2000. Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). Food Qual. Pref., 11, 5-16.

Reunanen, J., 2003. Overfitting in making comparisons between predictor selection methods. JMLR, 3, 1371-1382.

Rosipal, R. and Trejo, L. J., 2001. Kernel Partial Least Squares Regression in Reproducing Kernel Hilbert Space. J. Mach. Learn. Res., 2, 97-123.

Sun, K., Jang, S., Wong, D.S., 2016, Development of soft sensor with neural network and nonlinear variable selection for crude distillation unit process. Comput. Aided Chem. Eng. 38, 337-342.

Tipping, M. E., 2001. Sparse Bayesian Learning and the Relevance Vector Machine. J. Mach. Learn. Res., 1, 211-244.

Wold, S., Wold , N. and Skagerberg, B., 1989. Nonlinear PLS Modeling. Chemom. Intell. Lab. Syst., 7, 53-65.

Wold, S., Sjöström, M. and Eriksson, a. L., 2001. PLS-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst, 58(2), pp. 109-130.

Yan, W., Shao, H. and Wang, X., 2004. Soft sensing modeling based on support vector machine and Bayesian model selection. Comput. Chem. Eng., 28, 1489-1498.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50374-8

A Bi-objective Optimization Approach to Reducing Uncertainty in Pipeline Erosion Predictions

Wei Dai[a], Selen Cremaschi[a*], Hariprasad J. Subramani[b], Haijing Gao[b]

[a]*Department of Chemical Engineering, Auburn University, Auburn, AL 36849, USA*
[b]*Chevron Energy Technology Company, Houston, TX 77002, USA*
selen-cremaschi@auburn.edu

Abstract

Uncertainty presented in pipeline erosion predictions may limit the field application of these models. Especially when the to-be-tested regions highly extrapolate the prediction model, large prediction uncertainties can be witnessed. We previously developed a Gaussian process (GP) model based framework to estimate the erosion prediction uncertainty. There are two major goals in the quantification of erosion prediction uncertainty. The first is to minimize the prediction discrepancy of the model uncertainty. The second is to enhance the reliability of the predicted model uncertainty. GP modeling, as a kernel-based approach, relies on the proper selection of hyperparameters. The hyperparameters are generally optimized by maximum marginal likelihood (MLE) using conjugated gradient approach. However, using MLE as the objective function may not satisfy both goals. Furthermore, for non-convex marginal likelihood functions, optimization approaches like conjugated gradient is sensitive to the selection of initial values, and may lead to local optima. Here, we present a bi-objective optimization approach for training the GP model of erosion prediction uncertainty. GP models trained using bi-objective optimization outperform the GP model trained using the MLE based approach in terms precision and reliability for the prediction of erosion model uncertainty.

Keywords: ε-constrained approach; Bayesian optimization; Gaussian process modeling

1. Introduction

The unplanned extraction of solids in oil and gas production leads to erosion in the pipeline, especially under gas-dominated conditions. The formulation of erosion is a complex process and depends on many factors. Several empirical or semi-empirical models have been developed to predict erosion rate as a guideline for equipment design, repair, and replacement. Due to the complexity of the erosion process and the extreme cost in collecting field data, the discrepancies between erosion models' predictions and actual erosion rates can be very significant, e.g., several orders of magnitude for some operation conditions (Dai and Cremaschi, 2015). We hypothesize that most of the discrepancy can be explained by the uncertainty of the model predictions. We used Gaussian process method (Rasmussen and Williams, 2006) to quantify the model uncertainty and the prediction variance of erosion-rate models (Dai and Cremaschi, 2015). Gaussian process method, as a machine learning approach, is able to provide the underlying distribution of the prediction uncertainty. The Gaussian process (GP) model estimations of erosion-rate prediction uncertainty can be evaluated using two criteria: (1) the mean prediction of GP model, which quantifies the prediction accuracy and is generally defined by the squared residual between the mean prediction of GP model and

the target; and (2) the confidence on the GP model predictions, which is indicated by the covariance function and can be quantified using the overall prediction variance. An accurate GP model should yield the minimum mean squared error (MSE) and the minimum prediction variance (VAR). In our previous work (Dai and Cremaschi, 2015), the hyperparameters of GP models were obtained by maximizing the marginal likelihood function using the Maximum Likelihood Estimation approach (MLE, Giunta et al., 2006). Although MLE is considered the gold standard for statistical parameter estimation, it cannot guarantee that the two aforementioned objectives are satisfied at the same time. Here, we train GP models using a bi-objective optimization approach that minimizes MSE and VAR, and apply ε-constrained approach for generating the Pareto optimal set. Bayesian optimization algorithm is used to solve the resulting optimization problems. Spectral mixture kernel with a special initiation scheme is found to be the best kernel for the GP model to estimate model uncertainty of erosion-rate predictions. The results reveal that the GP model trained using MLE yields a solution that is not Pareto dominant.

2. Methodology: quantification of erosion-rate uncertainty

For uncertainty quantification of erosion rate, the experimental response, $\mathrm{y^e}$, can be expressed as the summation of erosion model's prediction of erosion rate, $\mathrm{y^m}$, erosion model bias, α, and the experimental error, ε (Kennedy and O'Hagan, 2001). The model bias includes the uncertainties associated with estimated model parameters in erosion model, the numerical errors and the model form discrepancies, and can be expressed as a Gaussian random process (Jiang et al., 2013). The experimental variability is shown to follow a normal distribution (Dai et al., 2016). Therefore, the model bias and the experimental variability are combined into one term that defines the erosion model uncertainty, and is referred to as model discrepancy (δ) for the remainder of the paper.

2.1. Gaussian Process Modeling (GPM)

In Gaussian process modeling, the distribution of a function is determined by a Gaussian process, which is composed of a set of random variables, where any finite number of random variables from the set follows a joint probability distribution (Rasmussen and Williams, 2006). Therefore, we can express the function using a multivariate Gaussian distribution with its mean and covariance functions (Eq. (1)).

$$f(x) \sim GP(m(x), K(x, x')) \tag{1}$$

where $x, x' \in R^d$, $m(x) = \mathrm{E}[f(x)]$, $K(x, x') = \mathrm{E}\left[\left(f(x) - m(x)\right)\left(f(x') - m(x')\right)^T\right]$ for measurements taken at $D = \{(x_i, f_i), i = 1, \ldots N\}$, and $f_i = f(x_i)$. The function values f_* at N_* untested locations (x^*) are predicted using $f_* \sim N\left(m_*, K(x_*, x_*)\right)$. The joint distribution of the observed f and the to-be-predicted f_* is

$$\begin{bmatrix} f \\ f_* \end{bmatrix} \sim GP\left(\begin{pmatrix} m(x) \\ m(x_*) \end{pmatrix}, \begin{bmatrix} K & K_* \\ K_*^T & K_{**} \end{bmatrix}\right) \tag{2}$$

where $K = \{k_{ij}\}$ is an $N \times N$ symmetric positive definite covariance matrix, K_* is an $N \times 1$ covariance matrix, and K_{**} is an $N_* \times N_*$ matrix. The elements $k_{ij} = k(x_i, x_j)$ measure the correlation between variables x_i and x_j. The posterior distribution of f_* is:

$$P(f_*|x_*, x, f) = N(\mu_*, \Sigma_*) \tag{3}$$

where $\mu_* = m(x_*) + K_*^T K^{-1}\big(f - m(x)\big), \Sigma_* = K_{**} - K_*^T K^{-1} K_*$.

Since the covariance function is a positive definite matrix satisfying the Mercer Theorem, it is also a kernel function. The performance of GP prediction relies on the selection of the kernel function (Wilson and Adams, 2013) (Eq. (3)). Squared exponential, Matern and rational quadratic kernels are among the commonly used kernel functions (Rasmussen and Williams, 2006). The kernel function should be selected based on knowledge of the underlying function and characteristic of the data. Once it is set, its hyperparameters are optimized using maximum marginal likelihood. However, optimization approaches like conjugated gradient rely on choice of initial values and are prone to be trapped in local minima. In this paper, to obtain the optimum set of hyperparameters, a global optimization approach, Bayesian optimization, is used.

2.2. Bayesian optimization

Bayesian optimization, a method used with black-box models with moderate dimensions, is well-suited for the optimization of hyperparameters in machine learning approaches (Shahriari et al., 2016), and provides the global solution without multiple random initializations (Bull, 2011). Bayesian optimization approach sequentially evaluates the objective function as a black-box deterministic function. The approach uses acquisition function (α_n) to search for new data points to evaluate the objective function. Acquisition function provides a balanced sampling between exploration and exploitation. In this paper, the objective function is approximated using a Gaussian process model. In Figure 1, the GP model used as the objective function in the Bayesian optimization is referred as GP2 to differentiate it from the GP model used for estimating erosion model discrepancy. The posterior distribution of GP2 reveals areas not modeled well (with large uncertainty) and the objective function values at sampled points reveal the region where the minimum value lies. The algorithm starts by training GP2 given initial n sample points. A new sample point with the maximum α_k is selected. The new sample and its corresponding objective function value are then used to update the distribution of GP2. The acquisition function based on $n+1$ data points is calculated and utilized to determine the next sampling point. The whole process is repeated until the difference of the objective function values at the latest two samples is smaller than a threshold ϵ (Fig. 1).

2.3. Kernel function used in the GPM

Spectral Mixture (SM) kernel (Eq.(4)) is used for the erosion model discrepancy GP model. It contains Q scaled mixture of Gaussian distributions. It can cover a wide range of spectral density and can approximate any stationary kernels with high accuracy.

$$k_{SM}(x, x') = \sum_{q=1}^{Q} w_q cos\,(2\pi(x - x')\mu_q) \prod_{d=1}^{D} exp\big(-2\pi^2(x - x')^2 v_q^{(d)}\big) \tag{4}$$

where w_q's are weights indicating the relative importance of mixture components, inverse means, $1/\mu_q$, are the periods of components, and inverse standard deviation, $1/\sqrt{v_q}$, is the length scale that controls the influence of each input dimension. The hyperparameters (w_q, $1/\mu_q$, $1/\sqrt{v_q}$) are initialized via data-dominated priors (Wilson

and Adams, 2013). The hyperparameters in the kernel function for the covariance and the constant value used in the mean function are the set of hyperparameters (θ) that are optimized based on the data for training the GP model.

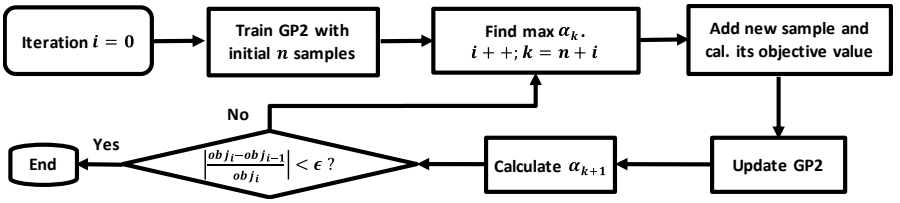


Figure 1. Bayesian optimization algorithm used to determine the hyperparameters

3. Optimization objectives for determining GPM hyperparameters

Here, the commonly used objective formulation to optimize the GPM hyperparameters is introduced in addition to two other formulations, which represent the two major goals in the quantification of erosion prediction uncertainty.

3.1. The Maximum Likelihood Estimation (MLE)

From Section 2.1, the measurements can be expressed using Eq. (5), which is also called the marginal likelihood function. The log marginal likelihood is given in Eq. (6).

$$p(f|x,\theta) = N(m(x), K) \tag{5}$$

$$\log\big(p(f|x,\theta)\big) = -\frac{1}{2}f^T K^{-1} f - \frac{1}{2}log\det|K| - \frac{n}{2}log2\pi \tag{6}$$

Optimal set of hyperparameters are obtained when the log marginal likelihood function is maximized. The conjugated gradient approach is commonly used to solve the partial derivatives of the log marginal likelihood with respect to hyperparameters (Rasmussen and Williams, 2006). This is the traditional approach for constructing GPMs.

3.2. Minimization of mean squared error (MSE)

The GP model hyperparameters can be estimated by minimizing mean squared error (MSE) using Eq.(7), where δ_i and $\hat{\delta}_i$ are the calculated and estimated model discrepancies for the i^{th} data point (in unit of mils/lb), and N is number of data points.

$$\min_{\theta} \frac{\sum_{i=1}^{N}\left(\delta_i - \hat{\delta}_i\right)^2}{N} \tag{7}$$

3.3. Minimization of prediction variance (VAR)

Performance of GPM is also assessed using prediction reliability. A smaller prediction variance represents a more reliable estimation. Equation (8) can be used to estimate the GPM hyperparameters by minimizing the total prediction variance (VAR), where $\hat{\sigma}_i^2$ stands for the estimated variance of the model discrepancy for the i^{th} data point.

$$\min_{\theta} \frac{\sum_{i=1}^{N}\hat{\sigma}_i^2}{N} \tag{8}$$

3.4. Bi-Objective Optimization for GPM Training: ε-constrained approach

In this paper, MSE and VAR are used as two objectives for estimating GPM hyperparameters. The Pareto optimal set for the bi-objective optimization problem is obtained using the ε-constrained approach (Miettinen, 1999).

4. Results

The data used for training and testing the GPM contains 585 samples (erosion rate measurements) with 10 dimensions with a large input space. A MATLAB® based toolbox, Gaussian Process for Machine Learning – GPML (Rasmussen and Nickisch, 2010), is used to train the GP models and make predictions based on the GP models. The function `bayesopt` in MATLAB is used to develop the Bayesian optimization approach. Cross validation is used to evaluate the generality of the developed GPMs. The optimization problem is solved using training data set constructed by a 4-fold cross validation. The hyperparameter values obtained with the highest marginal likelihood from the Spectral mixture kernel initialization (based on data-dominated priors) are used as the initial values for the θ to be optimized by Bayesian optimization approach. To evaluate the objective function efficiently in Bayesian optimization process, expected improvement with time-weighting is used as the acquisition function (Snoek, Larochelle, and Adams, 2012). The Bayesian optimization process is continued until the difference of the objective function values at the latest two evaluations is smaller than 1×10^{-3}. The minimum MSE (Eq. (7)) and the corresponding VAR are 6.03×10^{-7} $(\text{mils/lb})^2$ and 1.41×10^{-1} $(\text{mils/lb})^2$. The minimum VAR (Eq. (8)) and the corresponding MSE are 1.74×10^{-8} $(\text{mils/lb})^2$ and 3.48×10^{-6} $(\text{mils/lb})^2$. The minimization of MSE with VAR as the ε-constraint yields the Pareto front shown in Figure 2.

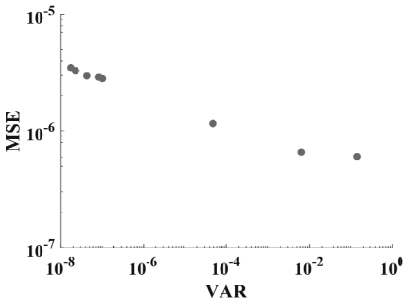


Figure 2. Pareto front of the bi-objective problem

The corresponding MSE and VAR values obtained when hyperparameters are estimated using MLE are also calculated for comparison, and they are 8.66×10^{-4} $(\text{mils/lb})^2$ and 3.40×10^{-3} $(\text{mils/lb})^2$. These values sit far above the Pareto front, revealing the advantage of constructing the problem as a bi-objective one and a potential disadvantage of using MLE to achieve both precision and reliability for prediction of erosion model discrepancy. The root mean square error (RMSE) of erosion model discrepancy for the database is 3.32 mils/lb, indicating for per pound of sand particles transported in the pipeline, there are 3.32 mils (thousandth of an inch) length loss of pipe inner wall incorrectly predicted by the erosion prediction model. Considering the case with natural gas transported at a rate of 3.0×10^{8} standard cubic feet per day (a very typical production rate in oil and gas transportation), the yearly thickness loss calculated is less

than 20 mils. According to ANSI/ASME Standard B31.3 (applies to major facilities onshore and offshore worldwide), the minimum design wall thickness is 0.11 inch (110 mils) (ANSI/ASME 2002). Therefore, this miscalculation is significant, and it will increase for predicting thickness loss for longer periods. Thus, we utilized the bi-objective framework to develop GP models for estimating erosion model discrepancy. The RMSEs of the erosion model discrepancy using the optimum GP models range from 7.76×10^{-4} to 1.86×10^{-3} mils/lb. Therefore, for unit sand particles transported, the uncertainty of erosion rate is at most off by 1.86×10^{-3} mils on average. It can be concluded the estimations obtained from the GP trained using the bi-objective optimization approach have small prediction biases.

5. Conclusions

In this paper, Gaussian process modeling is adopted to estimate the uncertainty of erosion-rate model predictions, which is defined by model discrepancy. To reduce the variance of model discrepancy estimates while maintaining precision, a bi-objective optimization problem is formulated using mean squared error and total variance. Bayesian optimization, as a global optimization algorithm, is implemented to solve the bi-objective problem using ε-constraint approach. The results reveal that the Pareto front solutions dominate the traditional approach, Maximum Likelihood Estimation.

Acknowledgements

Authors acknowledge financial support from Chevron Energy Technology Company.

References

A.D. Bull, 2011, Convergence Rates of Efficient Global Optimization Algorithms. Journal of Machine Learning Research 12: 2879–2904.

ANSI/ASME Standard B31.3, Standard for Chemical Plant and Petroleum Refinery Piping. 2002.

W. Dai, and S. Cremaschi, 2015, Quantifying Model Uncertainty in Scarce Data Regions - A Case Study of Particle Erosion in Pipelines. Comp. Aided Chem. Eng. Vol. 37.

W. Dai, S. Cremaschi, M.A. Islam, R.T. Nukala, H.J. Subramani, G.E. Kouba, and H. Gao. 2016, Uncertainty Analysis of Multiphase Flow - Case Studies from Erosion, Sand Transport, Liquid Entrainment Models. 10th North American Conference on Multiphase Technology.

A.A. Giunta, J. M. McFarland, L. P. Swiler, and M. S. Eldred, 2006, The Promise and Peril of Uncertainty Quantification Using Response Surface Approximations. Structure and Infrastructure Engineering 2 (3–4): 175–89. doi:10.1080/15732470600590507.

Z. Jiang, W. Chen, Y. Fu, and R.J. Yang, 2013, Reliability-Based Design Optimization with Model Bias and Data Uncertainty. SAE Int. J. of Materials and Manufacturing 6 (3): 502–16.

M.C. Kennedy, and A. O'Hagan, 2001, Bayesian Calibration of Computer Models. Journal of the Royal Statistical Society. Series B (Statistical Methodology) 63 (3): 425–64.

K. Miettinen, Kaisa, 1999. Concepts, Nonlinear MultiObjective Optimization.

C.E. Rasmussen, and H. Nickisch, 2010, Gaussian Processes for Machine Learning (GPML) Toolbox. Journal of Machine Learning Research 11: 3011–15.

C.E. Rasmussen, and C. K Williams. 2006. Gaussian Processes in Machine Learning. MIT Press.

B. Shahriari, K. Swersky, Z.Y. Wang, R.P. Adams, and N.D. Freitas, 2016, Taking the Human out of the Loop: A Review of Bayesian Optimization. Proceedings of the IEEE.

J. Snoek, H. Larochelle, and R.P. Adams, 2012, Practical Bayesian Optimization of Machine Learning Algorithms. Advances in Neural Information Processing Systems, 1–9.

A.G. Wilson, and R.P. Adams, 2013, Gaussian Process Covariance Kernels for Pattern Discovery and Extrapolation. In Proceedings of the 30th International Conference on Machine Learning, 28:15.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50375-X

Information Entropy based Indices for Variable Selection Performance Assessment

Q. Peter He*

Department of Chemical Engineering, Auburn University, Auburn AL 36849, USA
qhe@auburn.edu

Abstract

Variable selection plays an important role in data-driven modeling and other applications, and becomes increasingly important as big data become ubiquitous. In the last a few years, many different variable selection methods have been reported. However, how to directly evaluate the performance of variable selection methods has received limited attention. The common criteria used to assess variable selection performance either indirectly measure the effects of variable selection, such as through prediction performance of a model, or require ground truth of variable relevancy, which is usually unavailable, incomplete, or unverified in industrial applications. To address this limitation, two information entropy based consistency indices are proposed to directly evaluate the performance of variable selection methods: one does not require ground truth of variable relevancy, the other can make use of such information if available. A simulated case study (with ground truth) and an industrial case study (without ground truth) are provide to compare the proposed indices with the existing methods.

Keywords: variable selection, soft sensor, consistency index, information entropy.

1. Introduction

With ever-accelerating advancement of information, communication, sensing and characterization technologies, tremendous amount of data are generated and stored every day. Those so called “Big Data” are often extremely high-dimensional, contaminated by noise, and interspersed with a large number of irrelevant or redundant features, making it a challenging task to retrieve useful information from the data (Comminges and Dalalyan, 2011; Ting et al., 2010). Variable selection has been one of the practical approaches to reducing data dimensionality prior to data interpretation or modeling. Even for projection-based dimension reduction methods such as principal component analysis (PCA) and partial least squares (PLS), variable selection is often applied as a pre-processing step to further improve the modeling performance. Among various applications where variable selection is applied, soft sensor development is arguably the one that saw the largest impact from variable selection – it has been shown that the performance of soft sensor can be tremendously improved if only vital variables are included in model development (Andersen and Bro, 2010; Reunanen, 2003).

In the past few decades, many different variable selection approaches have been reported for various applications. However, how to directly evaluate their performance has received limited attention. Currently, variable selection performance is evaluated through prediction performance with indirect metrics such as the average mean absolute percentage error ($MAPE$), coefficient of determination (R^2). It can also be evaluated

directly with metrics such as the geometric mean of selection sensitivity and specificity (G). Their definitions are given below:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i-\hat{y}_i}{y_i}\right| \times 100\% \quad (1)$$

where y_i is the measurement and $\hat{y}_i$ is the prediction.

$$R^2 = 1 - \frac{SSR}{SST} \quad (2)$$

where $SSR = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2$ is the sum of squared residual, and $SSR = \sum_{i=1}^{N}(y_i - \bar{y})^2$ is the total sum of squares, with $\bar{y}$ being the average of y. R^2 measures how well the data fits the model.

$$G = (Sensitivity \times Specificity)^{1/2} \quad (3)$$

where sensitivity and specificity are the true positive (*i.e.*, relevant variables selected) and true negative (*i.e.*, irrelevant variables not selected) rates, respectively. The value of G ranges from 0 to 1. $G = 1$ indicates that all the predictors are classified correctly.

Among these three commonly used criteria, $MAPE$ and R^2 can only indirectly measure the effects of variable selection through the prediction performance of a model. Although G directly measures the accuracy of variable selection, it is only applicable when the ground truth of the variable relevancy is known, such as in the simulated case studies. For most industrial applications, the information on the true relevant variables could be unavailable, incomplete, or unverified. In these cases, G cannot be computed.

In order to address these limitations, in this work we propose two information entropy based consistency indices to directly evaluate the performance of different variable selection methods. Based on the assumption that the underlying process model of a given data set stays the same, it is expected that the variables being selected using different training data set (such as through performing Monte Carlo (MC) runs) should be consistent if truly relevant variables are being selected every time. On the other hand, if a variable selection method is not capable of identifying truly relevant variables, different training data sets (containing different disturbances and noises) will likely result in different variables being selected. Therefore, the consistency indices proposed in this work evaluate the performance of a variable selection method by quantifying the consistency of the variable selection results among different MC runs.

2. Information Entropy and Variable Selection Consistency Indices

2.1. Shannon's information entropy

In the physical sciences, or thermodynamics and statistical mechanics, the entropy is a measure of the degree of randomness – the higher entropy means the choices are made more randomly, or less consistently. According to Shannon, information is a measure of one's freedom of choice. The amount of information is defined to be measured by the logarithm of the number of available choices:

$$H_s = -\sum_{i=1}^{n} p_i log_b(p_i) \quad (4)$$

where p_i is the probability of occurrence of event i and the base b can be 2, e, or 10. H_s plays a central role in information theory as measures of information, choice and uncertainty (Shannon and Weaver, 2015).

2.2. The proposed information entropy based consistency indices

For a random variable in the case of two possibilities with probability p and $q = 1 - p$, the Shannon's information entropy can be defined as

$$H_s = -(plog_2 p + qlog_2 q) \tag{5}$$

H_s is plotted in Figure 1 (a) as a function of p, which shows that $H_s = 0$ if and only if $p = 0$ or $p = 1$. In other words, H_s vanishes if we are certain of the outcome. H_s equals the maximum of 1 if $p = 0.5$, which is the most uncertain situation.

To propose a variable selection consistency index, namely I_c, it is desired that I_c is at the maximum if the outcome is certain, or the selection is the most consistent, *i.e.*, either $p = 0$ or $p = 1$. On the other hand, it is desired that I_c is at the minimum if the outcome is the least certain, or the selection is the least consistent (*i.e.*, $p = 0.5$, 50% chance being selected, 50% chance being not selected).

Based on the above considerations, we propose a variable selection consistency index I_c based on Shannon's information entropy (H_s) of two possibilities (which is true for variable selection – either selected or not selected). Specifically, I_c is defined as

$$I_c = 1 - H_s, \text{ or } I_c = 1 + plog_2 p + qlog_2 q \tag{6}$$

I_c is plotted in Figure 1 (b) as a function of p, and it can be seen that I_c has the desired properties discussed above. For the general case of multiple variables, I_c is the average consistency indices of all variables:

$$I_c = \frac{\sum_{i=1}^{n}(1+p_i log_2 p_i + q_i log_2 q_i)}{n} \tag{7}$$

where p_i is the probability of variable i being selected, $q_i = 1 - p_i$, the probability of variable i not being selected, n the total number of variables. It can be seen that $I_c \in [0,1]$. Eqs. (6) and (7) are defined based on the assumption that the true relevance of the variable(s) is unknown. Therefore the consistency index is solely based on the certainty or predictability of a variable being selected in the future events.

When the true relevance of a variable is available, this knowledge should be taken into account. In order to do so, let us consider a truly relevant variable. If the variable is not selected at all, *i.e.*, $p = 0$, it is the most consistent case so $I_c = 1$ based on Eq.(6). However, it is consistently wrong as it is not selected, so its corresponding I_c should be penalized. Similarly for all cases of $p < 0.5$, I_c should be smaller than that for cases of $p > 0.5$. Because of the symmetry of Eqn. (10) at $p = 0.5$, $I_c(p) = I_c(1 - p)$.

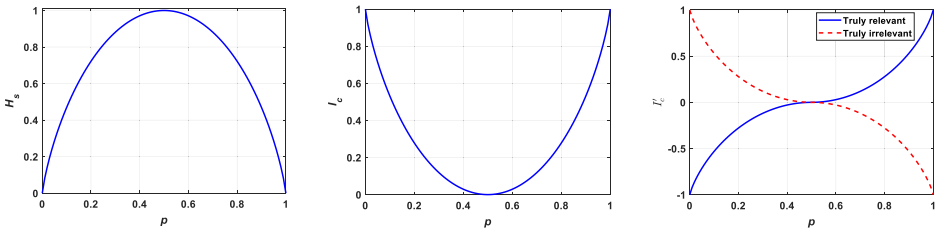


Figure 1. (a) H_s as a function of p; (b) I_c as a function of p; (c) $I_c^{'}$ as a function of p.

Therefore, it makes sense to flip the left half of the curve (*i.e.*, $p < 0.5$) in Figure 1 (b) down around the x-axis, which will result in the modified consistency index (I_c') with penalty, *i.e.* $I_c' = -1$ for $p = 0$. Similarly, for a truly irrelevant variable, the right half of the curve (*i.e.*, $p > 0.5$) in Figure 1 (b) should be flipped down around the x-axis.

$$I_c' = r \cdot \rho \cdot I_c, \text{ or } I_c' = r \cdot \rho \cdot (1 + plog_2 p + qlog_2 q) \tag{8}$$

where $r = 1$ if the variable is truly relevant, otherwise $r = -1$; $\rho = 1$ if $p \leq 0.5$, otherwise $\rho = -1$. It can be seen that I_c' ranges from -1 to 1.

Figure 1 (c) plots I_c' for a truly relevant variable (solid line) and for a truly irrelevant variable (dashed line) as a function of p. As can be seen from Figure 1 (c), if a truly relevant variable mistakenly never gets selected (*i.e.*, $p = 0$), $I_c' = -1$; if it correctly always gets selected (*i.e.*, $p = 1$), $I_c' = 1$. On the other hand, if a truly irrelevant variable correctly never gets selected (*i.e.*, $p = 0$), $I_c' = 1$; if it mistakenly always gets selected (*i.e.*, $p = 1$), $I_c' = -1$. It is worth noting that I_c' applies equal penalties on wrongly omitted truly relevant variable and wrongly selected irrelevant variable. However, the definitions of I_c' can be adapted to provide asymmetric penalties.

For the general multivariate case, I_c' is defined as the average of the consistency indices of all variables, which will still guarantee that $I_c' \in [-1,1]$:

$$I_c' = \frac{\sum_{i=1}^{n}\{r \cdot \rho \cdot (1 + p_i log_2 p_i + q_i log_2 q_i)\}}{n} \tag{9}$$

It is worth noting that we previously proposed a consistency index based on Shannon's information entropy (Wang et al., 2015). Due to the limited space, it is not discussed in this work. However, there is a serious flaw in that the index is not symmetric at $p = 0.5$. The $\mathrm{I_c}$ proposed in this work is completely different from it.

3. A Simulated Case Study

In this section, we will use a simulated case where the ground truth of variable relevancy is known to compare the variable selection consistency indices proposed in this work (*i.e.*, $\mathrm{I_c}$ and I_c') with the traditional index G. The simulation case study was introduced in (Chong and Jun, 2005) and was used in our previous work (Wang et al., 2015). There are totally 40 variables but only 10 variables are true relevant variables. Details can be found in (Wang et al., 2015). In one case study, the variable selection results of two methods, PLS with variable importance in projection (PLS-VIP) (Wold et al., 2001) and competitive adaptive reweighted sampling with PLS (CARS-PLS) (Li et al., 2009), are shown in Figure 2 as the frequencies of all variables being selected. The frequencies are based on 100 Monte Carlo (MC) runs of each method where the training data were randomly selected for each MC run. In Figure 2 the frequencies of the true relevant variables (*i.e.*, variables 1-5 and 36-40) are shown as the black bars, while those of the irrelevant variables (*i.e.*, variables 6-35) are shown as cyan (or gray in monochrome). Various indices discussed previously are calculated and listed in Table 1.

As can be seen from Table 1, all indices show that PLS-VIP is better than CARS-PLS in this case, which agrees with our intuition. For CARS-PLS, I_c' is significantly lower than $\mathrm{I_c}$ because $\mathrm{I_c}$ does not make use of the ground truth of variable relevancy while I_c' penalizes the wrongly selected irrelevant variables.

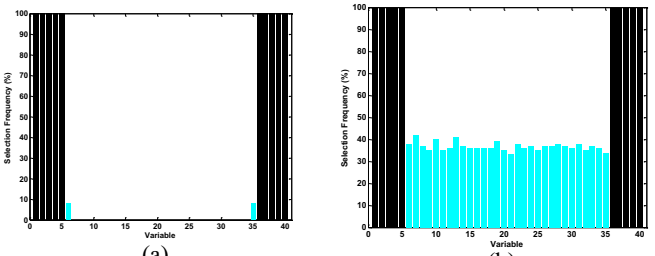


Figure 2. Frequencies of variables being selected by two methods: (a) PLS-VIP; (b) CARS-PLS

Table 1. Comparison of variable selection indices

Method	G	I_c	I_c'
PLS-VIP	.997	.990	.980
CARS-PLS	.795	.645	.289

The evaluation from the $\boldsymbol{G}$ value is the highest, which seems overly optimistic: a closer look at the differences between Figure 2 (a) and (b) suggests that (b) performs much worse than (a), as all irrelevant variables were selected around 40% of the time.

4. An Industrial Case Study

The industrial case study is the production of polyester resin using in the manufacturing of coatings via batch poly-condensation between a diol and a long-chain dicarboxylic acid. Thirty-four variables are routinely measured and recorded. There is no information on the true relevance of the variables to the final product quality, which is assessed by acidity number and viscosity (Facco et al., 2009). Due to limited space, only PLS-VIP and genetic algorithm with PLS (GA-PLS) are discussed in this work. Variables are selected based on their relevance to acidity number, as shown in Figure 3 (a) and (b), and to viscosity, Figure 3 (c) and (d). The two methods are compared in Table 2. Neither G nor I_c' is applicable in this case because the ground truth of variable relevancy is unknown. It can been seen from Figure 3 and Table 2 that the two variable selection methods perform drastically differently in terms of variable selection consistency. PLS-VIP has significantly more consistent variable section for both acidity number and viscosity than GA-PLS.

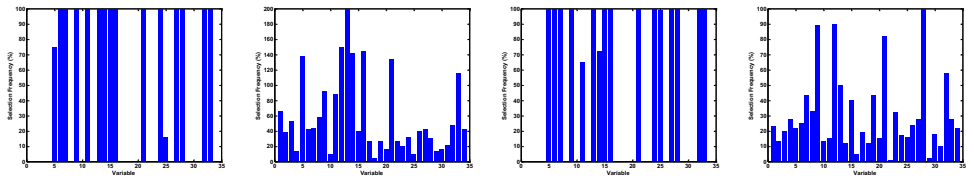


Figure 3. Industrial case study: frequencies of variables being selected by PLS-VIP and GA-PLS for acidity number – (a) and (b); and for viscosity – (c) and (d).

Table 2 also shows that I_c is in good agreement with other performance indices, including percentage improvement in $MAPE$ compared to full model without variable selection (%Imp), R^2, and prediction bias. This case study shows that variable selection consistency is highly correlated with the performance of the resulted soft sensor, validating the need and benefits of directly assessing variable selection consistency.

Table 2.Variable selection and other performance indices: industrial case study

Property	Method	I_c	%Imp	R^2	Bias
Acidity number	PLS-VIP	.958	28.4	.95	-.03
	GA-PLS	.354	21.9	.45	.07
Viscosity	PLS-VIP	.945	30.5	.92	-.01
	GA-PLS	.334	19.9	.87	-.13

5. Conclusions

In this work two variable selection consistency indices, I_c and I_c', are proposed to directly assess the performance of different variable selection methods. I_c' is used when the ground truth of variable relevancy is known, such as in the simulated cases. I_c is used when the knowledge of variable relevancy is unavailable, such as in most industrial cases. I_c and I_c' overcome some of the limitations of existing indices such as G. The simulated case study in this work shows that I_c and I_c' give more objective assessments than G. The industrial case study shows that I_c is highly correlated with the performance of the resulted soft sensors, validating the need and benefits of directly assessing variable selection consistency. More cases are needed to validate the consistency between I_c (I_c') and resulted soft sensor prediction performance and whether it is application dependent.

References

Andersen CM, Bro R. 2010. Variable selection in regression—a tutorial. *J. Chemom.* **24**:728–737.

Chong I-G, Jun C-H. 2005. Performance of some variable selection methods when multicollinearity is present. *Chemom. Intell. Lab. Syst.* **78**:103–112.

Comminges L, Dalalyan AS. 2011. Tight conditions for consistent variable selection in high dimensional nonparametric regression. In: . *COLT*, pp. 187–206.

Facco P, Doplicher F, Bezzo F, Barolo M. 2009. Moving average PLS soft sensor for online product quality estimation in an industrial batch polymerization process. *J. Process. Control.* **19**:520–529.

Li H, Liang Y, Xu Q, Cao D. 2009. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Anal. Chim. Acta* **648**:77–84.

Reunanen J. 2003. Overfitting in making comparisons between variable selection methods. *J. Mach. Learn. Res.* **3**:1371–1382.

Shannon CE, Weaver W. 2015. The mathematical theory of communication. University of Illinois press.

Ting J-A, D'Souza A, Vijayakumar S, Schaal S. 2010. Efficient learning and feature selection in high-dimensional regression. *Neural Comput.* **22**:831–886.

Wang ZX, He QP, Wang J. 2015. Comparison of variable selection methods for PLS-based soft sensor modeling. *J. Process. Control.* **26**:56–72.

Wold S, Sjöström M, Eriksson L. 2001. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **58**:109-130.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018,San Diego,California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50376-1

Implementation Costs and Output Reliability on Virtual On-line Analysers

Vicente Braga Barbosa[a,b*], Adelmo M. Aguiar Filho[b], Ana Rosa Massa[a], Karla Oliveira Esquerre[a,b]

[a]*Industrial Engineering Program, Polytechnique Institute, Universidade Federal da Bahia, R. Prof. Aristídes Novis,2, Federação, CEP 40210-630, Salvador, Brazil*
[b]*Chemical Engineering Department, Polytechnique Institute,Universidade Federal da Bahia, R. Prof. Aristídes Novis, 2, Federação,CEP 40210-630,Salvador, Brazil*
vicentebragab@gmail.com

Abstract

Virtual Analysers are used in industrial processes due to its capacity to replace essential equipment by mathematical models. In order to guarantee the precision of the model output, one has to assure the model input will be available and reliable, implying greater efforts to sustain a high maintainability of the measuring devices used to generate model input. Thus, there is a significant trade-off between model costs and output reliability. This work introduces a new index, called *Ana-Braga-Barbosa-Adelmo* (ABBA), which embodies such economic trait when selecting an empirical model for virtual analysers. Associated costs of different Partial-Least-Square (PLS) models used to estimate the content of contaminants in a trickle bed reactor are simulated through Monte Carlo assuming a Poisson distribution for instrument failure. The performance of different PLS models are analysed in light of the new proposed index, which indicates the most adequate one. Such is different from the PLS model developed based on selection of variables through Variable Importance in Projection (VIP) as the latter relies on a greater number of input variables for similar performance.

Keywords: PLS, Virtual Analysers, VIP, Model Costs

1. Introduction

The petrochemical industry is constantly innovating its methods and searching for processes that are safer and more efficient. Strong competition among different companies calls for technology that can guarantee both the product specification and a solid yield on the invested capital. Such technological devices include on-line analysers such as gas chromatographers, which grant a faster answer on the composition of products when compared with laboratory analysis. However, the cost of such instrumentation is still high (thousands of dollars) and the constant monitoring of the material flow requires frequent skilled maintenance (Musaev, 2003). During those, the loss of vital information could lead to a halt in production, unless another device allows for such information to be carefully estimated. As such, virtual analysers claim a strategic position in modern industry due to its capacity to replace essential equipment, such as on-line analysers, for mathematical models when the former is unavailable.
As described by Bakhtadze (2003), virtual analysers are also used as the algorithmic basis of on-line control systems. Such could be done in one of two ways: either the system only monitors process variables (PVs) and supports the decision making by the operator to guarantee product specification, or the system does not only monitor the

process variables but also offers correction actions. Recent papers by Nogueira *et al.* (2017) and Yadykin (2016) highlight such functionalities. The former monitors the quality of a polymerization process to support the decision making on the operational level in a process plant. The latter evaluates power system stability and suggests the threat of a cascading failure.

Although developing a process representative model for a virtual analyser is not trivial, once it is established, its implementation and integration do not demand much energy. Yet, in order to guarantee the precision of the model output, it is necessary to ensure that data for the model input will be available and reliable. Such imply greater efforts in sustaining a high maintainability of the measuring devices which are used to generate the model input data. As such, there is a significant trade-off between the costs associated with the virtual analyser model and the precision of its output, an economic feature that is often disregarded in a usual purely statistical model development. Hence, the aim of this paper is to introduce a new index, called *Ana-Braga-Barbosa-Adelmo* (ABBA), which attaches an economic trait when deciding the mathematical model used on the development of virtual analysers. In a case study, such index will support the decision of the most adequate Partial-Least-Square (PLS) model of a virtual analyser which estimates the content of methylacetylene and propadiene (MAPD) contaminants in the outlet stream of a trickle bed reactor on a propene plant located in Brazil.

2. The proposed index

This section describes the parameters used to define the new proposed index. Subsections 2.1 to 2.3 describe such parameters and subsection 2.4 defines the index.

2.1. The coefficient of determination

A figure of merit used for characterizing the output reliability of empirical models is the coefficient of determination (R^2), which represents how much of the variability of the output variable is predicted by the input variables based on sample values. It can be defined by Equation (1).

$$R^2 = 1 - \frac{\sum_i^n (y_i - \hat{y}_i)^2}{\sum_i^n (y_i - \acute{y})^2} \tag{1}$$

Where n is the number of observations on the dataset, y_i is the measured output variable at time$i$, $\hat{y}_i$ is the model estimated values of the output variable and $\acute{y}_i$ is the mean of the observed values of the output variable.

2.2. Implementation Costs

The total implementation costs associated with an empirical model over the course of t years is defined in this paper by Equation (2).

$$C_t = \sum_{j=1}^{j} \left(C_{sj} \acute{k}_j \right) \tag{2}$$

Where j is the number of input variables, C_{sj} is the cost of replacing the device responsible for providing the numerical value of input j and $\acute{k}_j$ is the number of failures of such device over the course of t years.

The number of failures of each device is assumed to follow a Poisson distribution, which is often used to describe counting processes associated with failure in industries (Ross, 2014). Equation (3) describes the probability mass function of such distribution.

$$P(k) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{3}$$

Where k is the number of events in a given time interval, λ is the average number of events per interval (also known as failure rate) and e is Euler's number.

2.3. The definition of model complexity

For data collected from an m number of process variables, a PLS model can be developed with j of those variables, where$j \leq m$. Among the available techniques to select the process variables that will be part a model, Chongc and Jun (2005) suggest VIP, which provides a score related to the importance of each input variable on a PLS model built with all m variables. The technique develops a new PLS model with all input variables that have a VIP score greater than one. Besides that, the VIP technique allows the input variables to be ranked from least important to most important.
The model complexity (h) is defined in this paper as the difference between the number of input variables used on the development of any PLS model (j) and the number of input variables selected by the VIP technique (j_{VIP}), as shown by Equation (4).

$$h = j - j_{VIP} \tag{4}$$

2.4. The definition of ABBA

The *ABBA* index for a model of complexity h is defined by Equation (5)

$$ABBA_h = \frac{{R^2}_h - {R^2}_{h-1}}{{c_T}_h - {c_T}_{h-1}} \tag{5}$$

Where ${c_T}_h$ is the standardized cost of a model with complexity h. The standard used is the model with maximum cost, where $h = m$.

2.5. Employing ABBA

ABBA will be employed on the selection of the most adequate mathematical model through the computation of the VIP scores for each PV used as a predicting variable. For an m number of available PVs, there are m potential PLS models. The first PLS model is developed with v latent variables, where all PVs are used. Their respective VIP scores are determined and the variables are organized by increasing order of VIP. The lowest score variable is removed and a new PLS model (with complexity $h = m - 1$) is developed with the same number of latent variables as before. This process continues until a PLS model with only one PV ($h = j_{VIP} - m$) is finally developed. For every developed PLS model there is a respective value of model complexity (h), coefficient of determination R^2 and standardized cost c_T. As both R^2 and c_T tend to increase with h, the model with the best trade-off between implementation costs and output reliability will be represented by a peak in an *ABBA* vs. h graph.

3. Case study

The introduced ABBA index will be applied to support the decision of the most adequate PLS model used on a virtual analyser that estimates the concentration of MAPD contaminants in a trickle-bed reactor at a propene production plant in Brazil. Such concentration is currently estimated by a gas chromatographer (on-line analyser) that requires frequent skilled maintenance. During those, there is no real-time monitoring of the MAPD content on the product, since laboratory analysis does not

provide an answer at the same rate of production. This could potentially halt the output of propene, as product specification is not guaranteed, and cause an enormous loss in invested capital.

Table 1 - Devices used for monitoring PVs - Costs and Failure Rate

Model ID	PV	Monitoring Device	Cost / U$	λ
V1	Fresh feed flow	Orifice plate	300	0.20
V2	Recycle flow	Orifice plate	300	0.20
V3	Fed Hydrogen flow	Orifice plate	300	0.20
V4	Combined feed temperature	Thermocouple	2500	0.20
V5	Bed Pressure	Manometer	100	0.50
V6-V11	Bed Temperature Across 6 points	Thermocouple	2500	0.20
V12	Outflow temperature	Thermocouple	2500	0.20
V13	MAPD content in feed	Chromatographic column	4000	0.30
		Thermal Detector	6000	0.25

Historical data from thirteen monitored and controlled process variables (PVs) is available as input for modelling. The measuring devices of the PVs are described together with their respective implementation cost and failure rate (λ), for a time spam of 5 years, on Table 1. Implementation costs include both the cost of new equipment, since it is assumed that when a device fails it is not repaired but replaced, and costs associated with the failing of transmitters, since the measured values of the process variables need to reach the control room in order to be monitored.

4. Results and Discussion

The VIP scores of the PLS model with all thirteen process variables ($j = 13$) are presented on Figure 1. There are six PVs with greater scores than the threshold of one ($j_{VIP} = 6$). Hence, the maximum model complexity is $h = 7$ and the minimum is $h = -4$. A purely statistical approach would develop a PLS model with the six selected variables. However, the way the model performance (measured by R^2) is changing as a function of the number of selected PVs (measured by c_T) would be completely disregarded. The coefficient of determination is plotted against the standardized cost (where the standard is the median of the full model) for all developed PLS models (from $h = -4$ to $h = 7$). The model complexity is displayed over the data points on Figure 2.

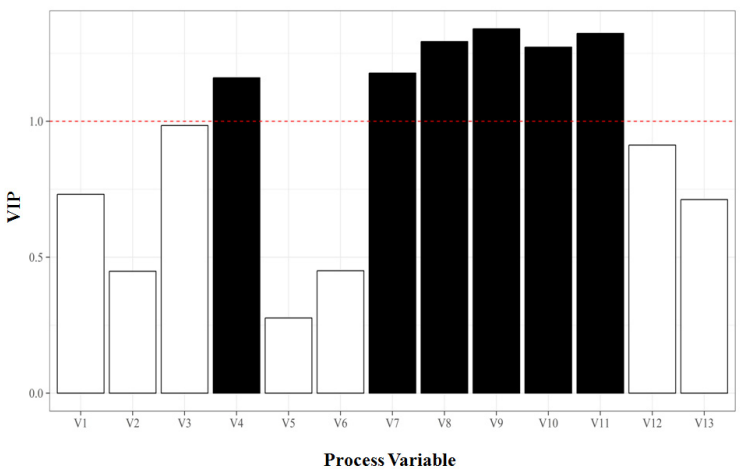


Figure 1 – VIP scores for Process Variables.

It is important to notice that there could be a model where no PV is used as input and the model output is estimated as the historical average MAPD content. Such model performs extremely poorly since the process is dynamic, and it is not a suitable candidate for the virtual analyser. Thus, it has been removed from this analysis.

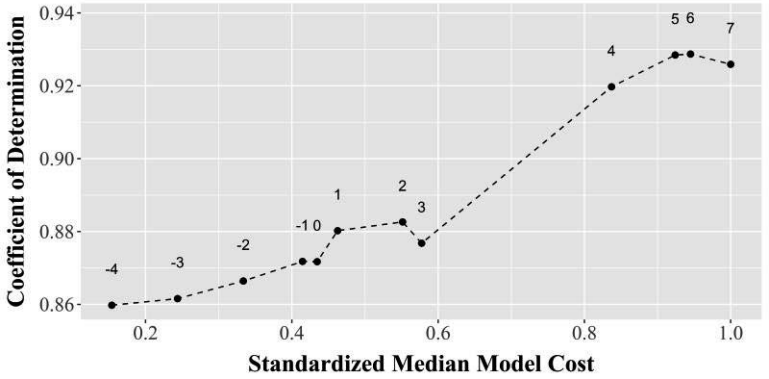


Figure 2 – Evaluation of model R^2 versus standardized model cost.

On a simple multiple linear regression, an increase in the coefficient of determination is expected when a predicting variable is added to the mathematical model. In case the addition of such variable causes a drop in performance, its coefficient would be zero and R^2 would, at the least, remain the same. However, such relationship is not always seen on PLS regression models. As see on Figure 2, an increase in model complexity does not always cause an increase in R^2. Going from $h = 2$ to $h = 3$, there is actually a decrease in R^2. This can be explained by the fact that PLS models maximize the covariance between the latent variables and the response variable. As such, the addition of a predicting variable could cause a lower model prediction capacity while covariance is being maximized. Such has been previously reported by Brereton (2009). Besides that, Figure 2 also indicates that the suggested threshold of one for VIP scores is not valid for all datasets. In fact, the drop in performance observed with PLS models of lower complexity ($h < 0$) does not seem to be significant and, although no statistical test was performed, such has been reported before by Akarachantachote *et al*. (2014). The trade-off between implantation costs and output estimate is evaluated through ABBA scores on Figure 3.

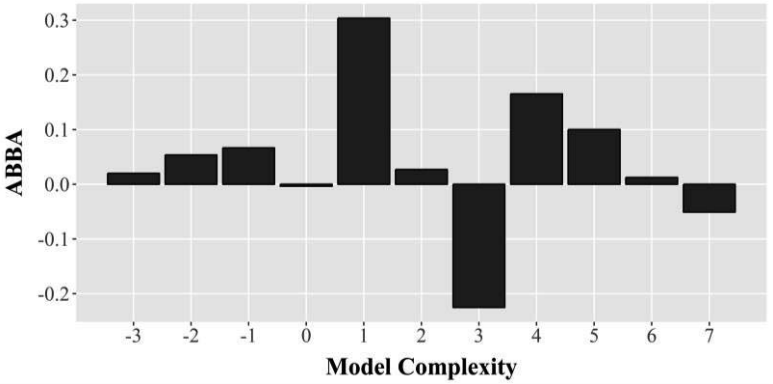


Figure 3 – ABBA index versus model complexity.

The trade-off between the associated cost and model fitting presented by the ABBA index (Figure 2) indicates the model with complexity $h = 1$ as the most adequate one for implementation on a virtual analyzer. Its peak in 0.3 can be translated as the maximum gain in output reliability due to an increase in the associated model cost. It is important to note that models where ABBA is negative must be avoided, since an investment in said models mean a reduction in output reliability.
In some cases, there may be a need to define a minimum value for the model reliability (R^2) or a maximum value for its cost. As the ABBA index provides a relative measure of the impact on output reliability due to an increase in model complexity, regions of acceptance might need to be defined if restrictions are in place. As an example, for an investment restriction of up to 0.6 the maximum cost, the model with complexity $h = 1$ should be implemented. If the restriction is on a R^2 higher than 0.9, Figure 2 indicates a region of acceptance which includes models with complexity higher than $h = 5$. The ABBA index on Figure 3 suggests that the model with $h = 4$ is the most adequate, as investment on more complex model do not generate any higher gains.

5. Conclusions

This paper shows that the analysis of the trade-off between implementation costs and model output reliability when deciding the most adequate model for a virtual analyser is an asset for the success of its implementation. Models of higher complexity require higher maintainability of the measuring devices that provide numerical values for input variables, meaning higher associated costs. Purely statistical techniques, such as VIP, do not consider the economics involved on the decision of the most adequate model and might lead to virtual analysers where the associated costs are not translated into model reliability. The use of ABBA in the case studied provided evidence to justify the implementation of a model that, albeit its higher complexity, possesses economic foundations for its selection, indicating the high potential of the new proposed index.

References

A. Akarachantachote, N.; Chadcham, S.; Saithanu, K. "Cutoff Threshold Of Variable Importance In Projection For Variable Selection" *International Journal of Pure and Apllied Mathematics*, vol 94, no. 3, 2014, pp. 307-322.

B. Bakhtadze, N. N. “Virtual Analyzers: Identification Approach.” *Automation and Remote Control*, vol. 65, no. 11, 2004, pp. 1691–1709., doi:10.1023/b:aurc.0000047885.52816.c7.

C. Brereton, R. G.; Ch4: Preprocessing, in “Chemometrics for Pattern Recognition”. John Wiley & Sons, Ltd, Chichester, UK. doi: 10.1002/9780470746462.

D. Chong, I.; Jun, C.; “Performance of Some Variable Selection Methods when Multicollinearity Is Present.” *Chemometrics and Intelligent Laboratory Systems*, vol. 78, no. 1-2, 2005, pp. 103–112., doi:10.1016/j.chemolab.2004.12.011.

E. Musaev, A. A. “Virtual Analyzers: The Concept of Design and Use in the Problems of Continuous Process Control” *Avtomatiz. Promyshl.*, no. 8, 2003, pp. 28-33.

F. Ross, S. M. “Simulation Analysis of System Life When Component Lives Are Determined by a Marked Point Process.” *Journal of Applied Probability*, vol. 51, no. 2, 2014, pp. 377–386., doi:10.1239/jap/1402578631.

G. Yadykin, I. B.; Grobovoy, A. A.; Iskakov, A. B.; Kataev, D. E.; Khmelik, M. S. “Stability Analysis of Electric Power Systems Using Finite Gramians” *IFAC-PapersOnLine*, vol. 48, no. 30, 2015, pp. 548–553., doi:10.1016/j.ifacol.2015.12.437.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50377-3

Using Big Data in Industrial Milk Powder Process Systems

Irina Boiarkina[ab], Nick Depree[b], Arrian Prince-Pike[b], Wei Yu[ab], David I. Wilson[bc], Brent R. Young[ab*]

[a]*Chemical and Materials Engineering, University of Auckland, 2-6 Park Avenue, Grafton, Auckland 1023, New Zealand*
[b]*Industrial Information and Control Center, University of Auckland, 2-6 Park Avenue, Grafton, Auckland 1023, New Zealand*
[c]*Electrical and Eectronic Engineering Department, Auckland University of Technology, 34 St Paul Street, Auckland Central 1010, New Zealand*
b.young@auckland.ac.nz

Abstract

This work looks at the application of big data in the milk powder processing industry, where the focus is on improving product quality and preventing off-specification product. This results in an increased focus on the quality attributes, which may be infrequently measured, and have low repeatability, thus resulting in challenges to the veracity of the data. Combined with the low frequency of failures, the data set is also 'unbalanced', making it difficult to analyse using normal black-box big data techniques. However, these can be addressed using tailored algorithms for elucidating the effects of specific parts of the process on the quality attributes of interest, and by using techniques such as bootstrapping with up- and down-sampling of data to address the issue of an unbalanced dataset.

Keywords: big data, dairy, processing, modelling

1. Introduction

The term "Big Data", characterised by volume, velocity, variety and veracity (White, 2016), is rapidly gaining attention, both by vendors providing novel IT solutions, and manufacturers wondering how best to take advantage of these new offerings. While the application of big data is not unusual in the dairy industry, to date it has been primarily restricted to farm management (Wolfert 2017), the management of the dairy herds using genetics to increase milk production (Cole, 2012), and machine learning for establishing culling rules for herd management (McQueen, 1995). The application of big data mining techniques and analytics to dairy processing at the factory is still relatively new.

Milk powder is a complex powdered product with a wide range of chemical, microbiological, physical and functional properties set down by the customer (Sharma, 2012). Functional attributes, such as different attributes of rehydration and solubility performance, are important quality parameters for end use. However, they are not well understood as they are defined by individually tailored testing procedures, with often complex links to the physical and chemical properties of the powder. This means that the effect of processing conditions on these properties is not well understood. At first glance this is precisely the type of applications that big data analysis is suited for.

Big data approaches are many and varied, including principal component analysis, partial least squares regression, neural networks, random forests, and machine learning (Chiang, 2017). These can be linear or non-linear, and are all black box in nature, and therefore ignore any pre-existing knowledge of the process. However, a crucial distinction between herd and farm management and processing is that personnel in the latter domain wish to know "why" a particular phenomenon occurs. A 'pure' black-box model approach is unlikely to articulate this, and is one reason why traditionally engineers have been uneasy to embrace such approaches. In our case, it is important to know why the milk powder is occasionally off-specification for particular attributes, and therefore how to subsequently correct this.

This paper explores the specific challenges for big data analytics in milk powder processing and presents two case studies for overcoming some of the challenges.

2. Challenges in the Milk Powder Industry

Of the four characteristic "V"s mentioned above, we argue that veracity and velocity are the key concerns for large scale powder production processes. For example in our application what we really want to know is how the powder performs in terms of rehydration, such as how it disperses, wets and produces any sediment. As mentioned earlier, these attributes are measured using individually tailored tests, such as the ISO/TS 17758:2014 standard (IDF 87:2014) for the determination of the dispersibility and wettability of instant dried milk, and are often laborious. Therefore, unlike standard process measurements such as pressure, flow or temperature, this quality data is infrequent, somewhat subjective, and often measured with a significant time lag and is difficult to trace through the process.

This means that although the velocity of the process data is high, the quality of the data is low. In carrying out the work presented in Section 3, this resulted in the need for laborious data reconciliation, to be able to apply `big data analytics' in this area. This is not unusual in the process industry – for example Cauvin et al. (2008) created a unified database for information management for a petrochemical site as the data storage within the institute was found to be highly heterogeneous and unsuitable for data mining.

This is especially important for short, infrequent processing events, such as line swaps during continuous operation, which rely on accurate traceability, something not handled well by the standard statistical algorithms. In Section 3.1 we describe the effect of powder re-blend on product quality, an infrequent operation when powder manufactured during start-up is held back, and then subsequently blended into powder made during stable production periods. This required a purpose-built, bespoke, algorithm in order to isolate these periods. This bespoke algorithm approach was also successfully used for analysing ingredient addition variation and off-spec results by Boiarkina (2017).

This introduces the second challenge in the dairy industry: veracity. Powders are notoriously difficult to sample reliably and many of the functional tests have poor repeatability. The quality tests are designed with grading in mind, not for diagnosing the effect of process parameters. For example, the functional properties of dispersibility and slowly dissolving particles are quantised into single digit numbers of discrete levels, seven and five levels respectively. This means the quality test results are not suitable for standard modern analytical algorithms to tease out any useful underlying correlations. Coupled with the fact that specification failures are rare means that even the large data set is information poor, and therefore makes it difficult to generate accurate models for

these rare, but important events. This means that techniques, such as up-sampling of off-spec data (Section 3.2), during model building become important in overcoming these limitations, as standard application of techniques, such as PLS, did not work.

3. Case Studies

3.1. Bespoke Algorithms for Process Data Analysis

The quality of milk powder produced cannot be guaranteed during start up, shut down and transitory periods (such as equipment line swaps). During this time, the milk powder is diverted into specific "start-stop" hoppers from which the powder can be re-blended at a low rate into the powder to be packed out or packed out as a lower grade powder. However, the effect of re-blend on the powder quality produced at an industrial plant was unknown.

Although a dataset suitable for multivariate analysis had been created previously, this dataset could not be used for this analysis as reblend events were not demarcated. Furthermore, due to mixing in the powder storage hoppers, two start-stop hoppers and different rates of re-blend being used, there was a fuzzy boundary around whether a quality sample was definitively associated with a specific re-blend event. For example, two bins could be used for packing out powder, but only one of them used for re-blending into, and information around which bin the quality sample came from was not available. This was a big data veracity challenge.

A bespoke algorithm was created to scan through a season worth of powder hopper weight data and identify periods of 're-blending' and 'packing out' taking place. This was based on the rate of change in the bin level, with a cut-off for what is considered a re-blend event versus packing out being set visually, and by discussions with plant operators. It was verified manually. The gradient detection algorithm had to be able to handle noisy data, so using straight differencing methods alone for the change in bin level was ineffective, and erroneously picked up a significant number of non-events. An example of the re-blend data is shown in Figure 1.

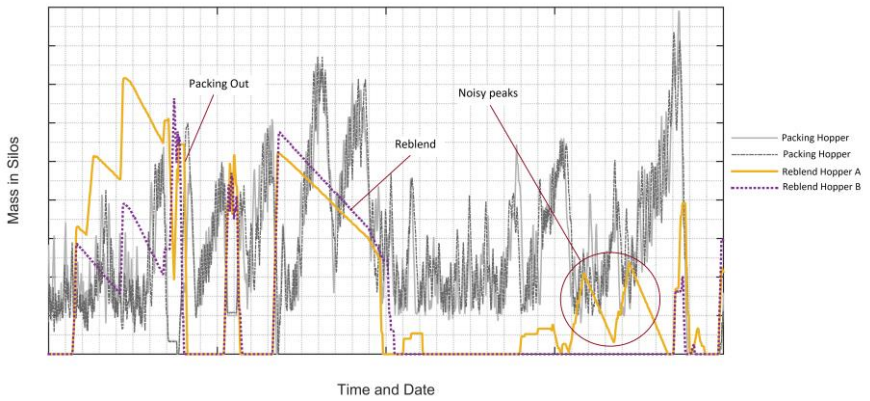


Figure 1 - An example of hopper weight fluctuations during a production run showing re-blend, packing out and noise in the data that makes it difficult to extract events.

In this case, the interest was to determine if the quality of the powder produced during re-blend was lower than the quality of the powder produced without re-blend. An analysis of variance (ANOVA) was used to compare the means of the two groups (with and

without re-blend), as well as to evaluate if there was a difference in the quality of the powder produced with a low and high rates of re-blend. A statistical difference was only found for two quality parameters, the sifter bulk density and a wettability test (shown in Figure 2). However, the difference was very small, and it was confirmed that the plant could continue to re-blend powder safely at a low rate without impacting the quality parameters studied.

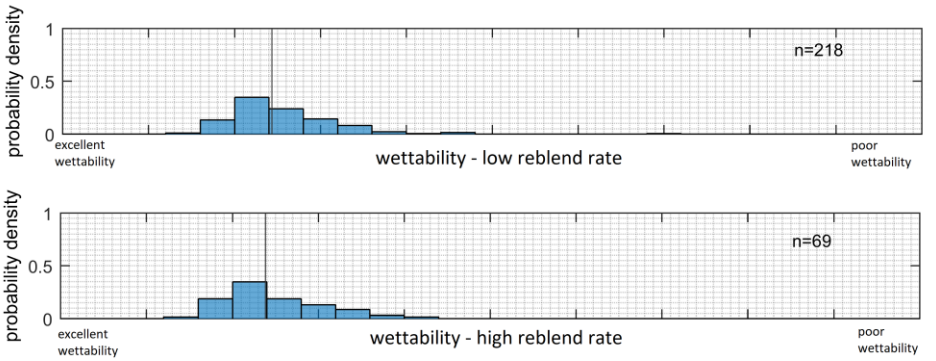


Figure 2 - Effect of re-blend rate on the wettability of instant whole milk powder, with the number of samples analysed shown in the top right hand corner.

3.2. Down-sampling and Up-sampling for Information Poor Data

In many cases in industrial data, no matter how big the volume, there may be very little information in the key areas of interest, such as times when product quality is poor. In the case of a well-controlled process, there may be so few occurrences of product failure that it is difficult or impossible to find predictive relationships by statistical regression, in order to avoid operational regions causing product failure in future. This is particularly important in the case of milk powder, where the production rate is very high, and quality tests are delayed, such that even the rare occurrences result in a high cost of quality failure.

A key functional property studied involved a proprietary test, where rehydrated powders were visually manually graded on a scale where a rating of 5 or higher constitutes a quality failure. The scale was such that some powders received a 3 rating, almost all powders received a 4 rating, and very few (off spec powders) were rated 5. Furthermore, this is an integer scale, where fractional values cannot exist, and hence categorical modelling was used, rather than linear regression. The training data included no values of 1 or 2, or greater than 5, and hence they can never be predicted from this data, but obviously it is ideal if the plant can avoid process operations likely leading to a score of 5.

Random forest models (James, 2013) were used to predict functional properties, from a big data set covering a full season of production, spanning approximately 30 plant process variables including temperatures, pressures and differential pressures, and flowrates around key unit operations, utilising expert domain knowledge to create this subset from the hundreds of measurements available. This process data is available at 30-second intervals from the plant historian, but it must be aligned against physical property measurements of fat, protein, moisture, and bulk density which are sampled hourly, and then against the desired functional tests which occur 3-5 times per day. This results in significant reduction of the original dataset, where several hundred thousand rows of process data result in only 392 samples, of the correct product type.

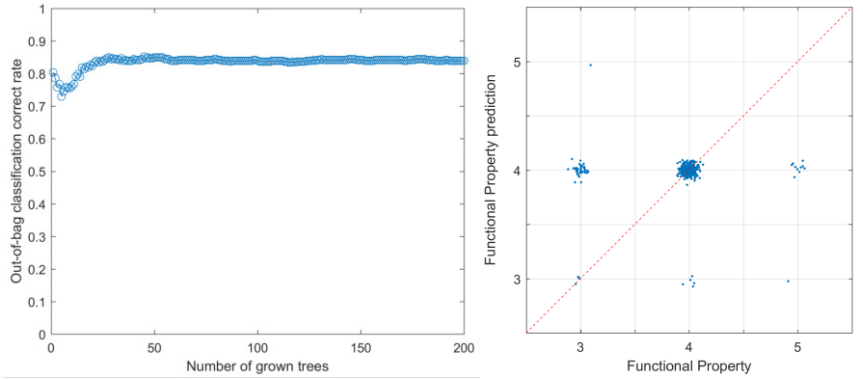


Figure 3 - Random forest prediction without resampling (points jittered for visibility)

Bootstrapping is used in the Random Forest modelling to improve the model fit (R^2), where the dataset is repeatedly randomly sampled, and a new tree grown each time, to see how the model fit improves as more trees are used. The initial model results shown in Figure 3 appear very promising, with R^2 well over 0.8, with only a few trees needed. On closer inspection however, it was clear that the model need only ever predict a (constant) value of 4, and it will be correct over 80% of the time due to the distribution of input data. The number of predictions of any other value were few, with only 3 out of 50 values of 3 predicted correctly, and 0 out of 11 values of 5. This clearly illustrates the potential downfall of modelling where the good/bad data distribution is unbalanced.

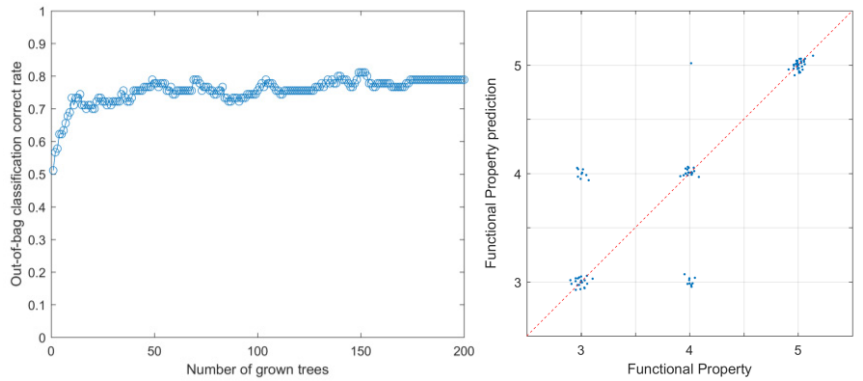


Figure 4 - Random forest prediction with resampling (points jittered for visibility)

Resampling was a key approach to improve the model in this case, where both down-sampling and up-sampling were used, to reduce the class imbalances (Kuhn, 2013). The 50 existing samples of 3, and 331 samples of 4 were randomly selected to form a sample size of 30 each, and the 11 existing values of 5 were randomly up-sampled with replacement, until it also had 30 occurrences. The model results in Figure 4 show a lower R^2 of around 0.75 after a similar number of trees, however the prediction is significantly improved in practice. In this case, around 70% of each of the 3 & 4 values were correctly predicted, with some 3's predicted as 4's and 4's predicted as 3's, but more importantly for the application of preventing off-specification powder, all 30 values of 5 were

predicted correctly, and only 1 of all 90 points was incorrectly predicted as a 5 i.e. there were no missed-alarms, and only one false-alarm.

4. Conclusions

The utilisation of big data in the milk powder processing industry, or indeed any processing industry, has a very different purpose compared to many other big data applications. The focus in processing is on the 'why' something is off-specification. This results in a larger focus on the quality attributes, rather than the process data. Given that quality attributes are collected at a lower rate than process data, and the tests themselves are often complex, with poor repeatability and coarse quantized results the veracity and velocity of this data becomes paramount. Furthermore, the data is often information poor, with few results of interest, partly of course because industries primarily produce on-spec product. This means that bespoke algorithms and techniques such as bootstrapping and up- and down-sampling during modelling become critical to extract useful information.

5. Acknowledgements

The authors would like to acknowledge the Primary Growth Partnership program from the New Zealand Ministry of Primary Industries for funding the project and would also like to thank Fonterra staff, specifically Richard Croy, James Winchester, Nigel Russell and Steve Holroyd for providing resources and support throughout the project.

References

I. Boiarkina, N. Depree, W. Yu, D. I. Wilson, B. R. Young, 2017, Fault diagnosis of an industrial plant using a Monte Carlo analysis coupled with systematic troubleshooting, Food Control*, 78*, 247-255

S. Cauvin, M. Barbieux, L. Carrié, B. Celse, 2008, A Generic Scientific Information Management System for Process Engineering, Computer Aided Chemical Engineering, 25, 931-936

L. Chiang, B. Lu, I. Castillo, 2017, Big Data Analytics in Chemical Engineering, Annual Review of Chemical and Biomolecular Engineering, 8, 63-85

J. B. Cole, S. Newman, F. Foertter, I. Aguilar, M. Coffey, M., 2012. Breeding and Genetics Symposium: Really Big Data: Processing and Analysis of Very Large Data Sets, Journal of Animal Sicence, 90, 723–733

G.James, D.Witten, T.Hastie, R.Tibshirani, 2013, An Introduction to Statistical Learning. Springer. ISBN 978-1-4614-7138-7

M. Kuhn, K.Johnson, 2013. Applied Predictive Modelling. Springer. ISBN 978-1-4614-6848-6

R. McQueen, S. Garner, C. G. Nevill-Manning, I. H. Witten, 1995. Applying Machine Learning to Agricultural Data, Computers and Electronics in Agriculture, 12, 1

A. Sharma, A. H. Jana, R. S. Chavan, 2012, Functionality of Milk Powders and Milk Based Powders for End Use Applications – A Review, Comprehensive Reviews in Food Science and Food Safety, 11, 518-528

D. White, 2016, Big Data What is it?, CEP Magazine, March 2016, 32-35

S. Wolfert, L. Ge, C. Verdouw, M.-J. Bogaardt, 2017, Big Data in Smart Farming, Agricultural Systems, 153, 69-80

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50378-5

Automated visual helmet identification based on deep convolutional neural networks

Hao Wu[a], Jinsong Zhao[a,b,*]
[a] *Department of Chemical Engineering, Tsinghua University, Beijing, China, 100084*
[b] *Beijing Key Laboratory of Industrial Big Data System and Application, Tsinghua University, Beijing, China, 100084*
jinsongzhao@tsinghua.edu.cn

Abstract

Helmets are an important piece of personal protective equipment on many worksites. In spite that wearing a safety helmet is mandatory for each worker, some do not always obey the regulation. In order to ensure the safety of workers and promote the automation of safety management, in this paper, an automated visual helmet identification method based on deep convolutional neural networks (CNN) is proposed. First, motion detection is applied to capture the segments of moving objects from videos. Then pedestrian detection based on deep CNN is utilized to detect whether there are humans or not in the image segments. Next the head regions are captured from human images. Finally, with red, yellow, blue, white and non-helmet image samples, helmet identification based on deep CNN is used to detect whether humans are wearing helmets or not and to identify the colors of helmets if helmets exist.

Keywords: Accident Prevention; helmet identification; work safety; deep convolutional neural networks.

1. Introduction

Caused by slips, falls, collisions and struck by falling objects, many high-risk industries are suffering from high injury and fatality rate. In order to protect workers from the danger, safety helmet is an important piece of personal protective equipment on many worksites, such as construction, chemical engineering, manufacturing, mining industries. In spite that wearing a safety helmet is mandatory for each worker, some do not always obey the regulation. Presently, most worksites already deployed video surveillance systems to supervise behaviors of workers for work safety. However, considering a large number of surveillance screens for safety helmet supervision, it is a tough task for humans to detect violators during long periods of time. It was proved that as the duration of monitoring of videos increases, the errors made by humans also increase (Hu et al., 2004).

With the advent of the era of smart manufacturing, the automation of safety management has equal importance with manufacturing process. To make factories more intelligent and automated, it is necessary to replace humans to supervise the behaviors of workers through machines like video surveillance systems. Based on computer vision, automated visual approaches develop rapidly in the manufacturing industry. Over the recent years, some approaches have been proposed for motorcycle helmet identification. Chiverton et al. (2012) used Histogram of Oriented Gradients (HOG) to extract features of motorcyclists and their heads and Support Vector Machines (SVM) for feature classification, which achieved 85 % accuracy. Waranusast et al. (2013)

studied color based features for the helmet identification task. However, the aforementioned approaches only used one descriptor for feature extraction. Silva et al. (2013) developed a hybrid descriptor, combined with Circle Hough Transform (CHT), HOG and Local Binary Patterns (LBP). By using Random Forest as the classifier, the accuracy for helmet identification reached 94.23 %. This result showed that the combination of different types of descriptors could achieve better performance than a single one. Recently, Dahiya et al. (2016) proposed a helmet identification system for bike-riders. It compared the performances of three widely used feature descriptors, HOG, Scale-Invariant Feature Transform (SIFT) and LBP using SVM classifier.

The above helmet identification methods all used hand engineered features, which suffer from challenges such as occlusion of objects and illumination effects, which cause the difficulty to accurately classify workers between wearing a helmet and not wearing a helmet. To achieve higher accuracy of motorcycle helmet identification, Vishnu et al. (2017) used deep convolutional neural network (CNN) to extract discriminative features of helmets. This study showed that using features extracted by deep CNN is better than using hand engineered features. CNN was first proposed in the late 1980s (LeCun et al., 1989), and brought about a revolution in computer vision and pattern recognition with the emergence of deep CNN in 2012 (Krizhevsky et al., 2012). Over the past few years, deep CNN has become the dominant algorithm in the computer vision and pattern recognition tasks. Vishnu et al. (2017) used the AlexNet architecture (Krizhevsky et al., 2012) for two classification tasks, one is for classification between motorcyclists and non-motorcyclists and the other is for classification between helmets and heads.

In spite that many motorcycle helmet identification methods with the satisfactory performance were developed, the helmet identification task for workers have been paid less attention. Different from motorcycle helmets covering the entire head, safety helmets for workers only cover the upper part of the head. Besides, helmet colors have different meanings in different industries. For example, in chemical engineering, workers wearing red safety helmets are operators while supervisors wear yellow safety helmets. Therefore, it is also useful to identify the helmet colors of workers. Park et al. (2012) developed an approach based on color histograms for construction workers. A recent work was conducted by Rubaiyat et al. (2017). They combined color-based and CHT for automatic helmet detection in the construction industry. However, they both used hand engineered features for feature extraction.

In this paper, we present an automated visual method based on deep CNN to improve the accuracy of helmet identification for workers in industries. First, motion detection is applied to capture the segments of moving objects from videos. Then, we collect human images as positive dataset and non-human images as negative dataset. Here pedestrian detection based on deep CNN is utilized to detect whether there are humans or not in the image segments. Next the head regions are captured from human images. Finally, with red, yellow, blue, white and non-helmet image samples, helmet identification based on deep CNN is used to detect whether humans are wearing helmets or not and to identify the colors of helmets if helmets exist.

The rest of this paper is organized as follows: Section 2 describes the proposed automated visual helmet identification method. Section 3 presents the experimental evaluation and finally section 4 summaries this paper.

2. The proposed method

Figure 1 shows the flowsheet of the proposed method. This method contains four steps: motion detection, pedestrian detection, head location and helmet identification.

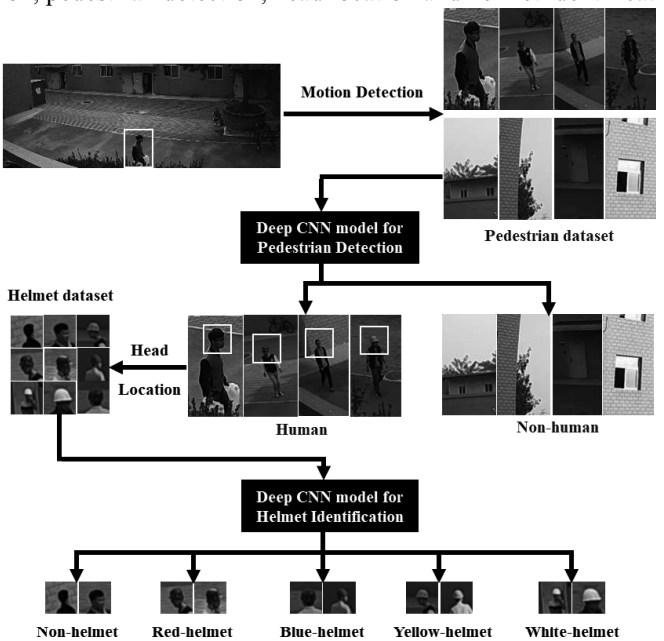


Figure 1-Flowsheet of the automated visual helmet identification method.

2.1. Motion detection

The dataset is collected from cameras and the frame size is 1920 × 1080 pixels. First, we apply K-Nearest Neighbors (KNN) background subtraction method (Zivkovic and Van Der Heijden, 2006) to capture the regions of moving objects from the static background. The moving objects include humans and non-human objects including shaking trees, moving clouds, opening or closing doors. If the area of the moving object is over 1000 pixels, it will be regarded as human-like objects and resized to 320 × 160 pixels.

2.2. Pedestrian detection

To distinguish human images from non-human images, the human-like image segments are classified by a deep CNN model. CNN is a stack of several convolutional layers and pooling layers, which is aimed to extract features from images. Following the part of feature extraction, fully connected layers are used to classify the human-like image segments into human class and non-human class. In this way, human image segments are detected from videos.

2.3. Head location

Before the helmet identification, the head position should be located in the corresponding human image segment. To solve this problem, we need to use the results of the background subtraction. Figure 2(a) is the human image segment and Figure 2(b) is the corresponding image after the process of the background subtraction. The point

(x, y) is the top pixel of the human and the point (x, y+5) is set as the center pixel of the head region (see Figure 2(c)). The head region is then located with a fixed size, 50× 50 pixels (see Figure 2(d)).

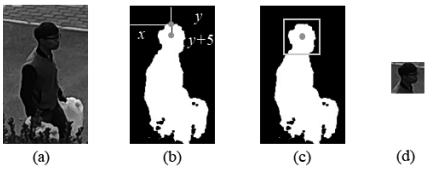


Figure 2-Process of the head location.

2.4. Helmet identification

To detect whether humans are wearing safety helmets or not and to identify the colors of helmets if helmets exist, the head image segments are classified by a deep CNN model. Different from the pedestrian detection, the last fully connected layer of the model for helmet identification is aimed to classify the head image segments into five classes (red, yellow, blue, white and non-helmet).

3. Experimental evaluation

3.1. Dataset

Pedestrian dataset : Using the motion detection to handle the videos, we collect 20504 human-like image segments (320 × 160 pixels) including 4646 human samples and 15858 non-human samples.

Helmet dataset : Using the head location to handle the human image segments, we collect 4648 head image segments (50 × 50 pixels) including 720 red-helmet, 706 yellow-helmet, 700 blue-helmet, 451 white-helmet and 2071 non-helmet samples.

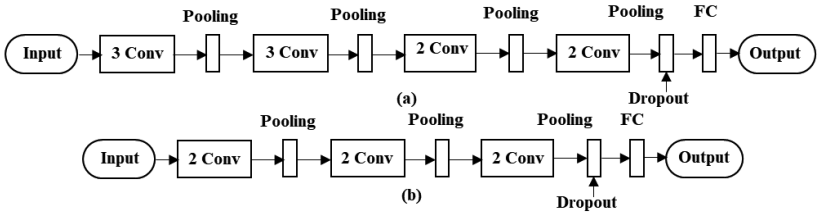


Figure 3-Deep CNN models: (a) For pedestrian detection; (b) For helmet identification.

3.2. Pedestrian detection

A deep CNN model is built for the pedestrian detection, which consists of 10 convolutional layers, 4 max pooling layers, and 1 fully connected layer for classification (see Figure 3(a)). The dropout is utilized for the last max pooling layer, which is aimed to avoid "overfitting". Based on the pedestrian dataset, 5-fold cross validation is implemented to train and test the model. The accuracy (ACC) results of the classification task is listed in Table 1. Figure 4 illustrates the 2D visualization of sample distributions of the input and output using t-SNE. Here "0" represents non-human images and "1" represents human images. The output t-SNE map shows that the deep CNN model has great performance for distinguishing the human and non-human image samples.

Table 1-Results of pedestrian detection.

	1	2	3	4	5	Average
Train ACC (%)	99.8	99.8	99.8	99.8	99.8	99.8
Test ACC (%)	99.6	99.6	99.5	99.4	99.7	99.6

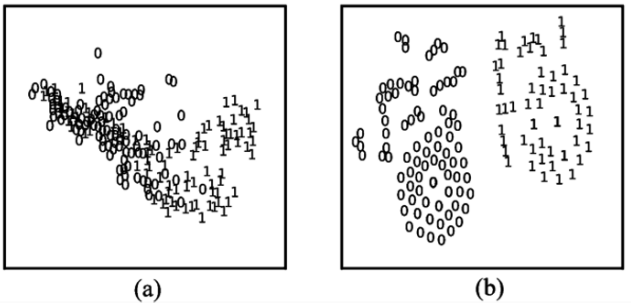


Figure 4-2D visualization of sample distributions for pedestrian detection using t-SNE.

3.3. Helmet identification

A deep CNN model is built for the helmet identification, which consists of 6 convolutional layers, 3 max pooling layers, and 1 fully connected layer for classification (see Figure 3(b)). The dropout is utilized for the last max pooling layer. Based on the helmet dataset, 5-fold cross validation is implemented to train and test the model. The accuracy results of the classification task is listed in Table 2. Figure 5 illustrates the 2D visualization of sample distributions of the input and the output using t-distributed stochastic neighbor embedding (t-SNE) (Maaten and Hinton, 2008). Here "1" represents "red", "2" represents "yellow", "3" represents "blue", "4" represents "white" and "0" represents "non" helmet image samples. In the input t-SNE map, the samples of 5 classes are mixed up. After the calculation of the deep CNN model, the output t-SNE map shows clear clustering of samples, which demonstrates the validation of our proposed method.

Table 2-Results of pedestrian detection.

	1	2	3	4	5	Average
Train ACC (%)	99.2	99.1	99.1	99.2	99.0	99.1
Test ACC (%)	98.9	99.0	99.3	98.9	98.1	98.8

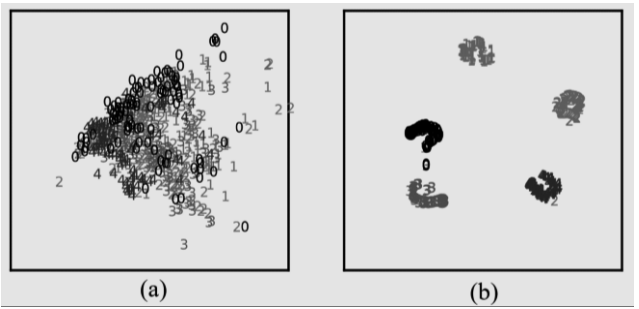


Figure 5-2D visualization of sample distributions for helmet identification using t-SNE.

4. Conclusions

In this paper, we present an automated visual helmet identification method based on deep convolutional neural networks. The deep CNN models show outstanding performances on the pedestrian detection and helmet identification tasks. The average accuracy for pedestrian detection reaches 99.6 % and for helmet identification reaches 98.8 %. This method can be applied on many high-risk work sites, which will help reduce the risk of casualties and promote the automation of safety management for smart factories.

References

Chiverton, J., 2012. Helmet presence classification with motorcycle detection and tracking. IET Intell. Transp. Syst. 6, 259. doi:10.1049/iet-its.2011.0138

Dahiya, K., Singh, D., Mohan, C.K., 2016. Automatic detection of bike-riders without helmet using surveillance videos in real-time, in: Proceedings of the International Joint Conference on Neural Networks. pp. 3046–3051. doi:10.1109/IJCNN.2016.7727586

Hu, W., Tan, T., Wang, L., Maybank, S., 2004. A survey on visual surveillance of object motion and behaviors. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 34, 334–352. doi:10.1109/TSMCC.2004.829274

Krizhevsky, A., Sutskever, I., Geoffrey E., H., 2012. ImageNet Classification with Deep Convolutional Neural Networks. Adv. Neural Inf. Process. Syst. 25 1–9. doi:10.1109/5.726791

LeCun, Y., Jackel, L.D., Boser, B., Denker, J.S., Graf, H.P., Guyon, I., Henderson, D., Howard, R.E., Hubbard, W., 1989. Handwritten digit recognition: applications of neural network chips and automatic learning. IEEE Commun. Mag. 27, 41–46. doi:10.1109/35.41400

Maaten, L. Van Der, Hinton, G., 2008. Visualizing Data using t-SNE. J. Mach. Learn. Res. 1 620, 267–84. doi:10.1007/s10479-011-0841-3

Park, M., Palinginis, E., Brilakis, I., 2012. Detection of Construction Workers in Video Frames for Automatic Initialization of Vision Trackers. Constr. Res. Congr. … 940–949. doi:10.1061/9780784412329.095

Rubaiyat, A.H.M., Toma, T.T., Kalantari-Khandani, M., Rahman, S.A., Chen, L., Ye, Y., Pan, C.S., 2017. Automatic detection of helmet uses for construction safety, in: Proceedings - 2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops, WIW 2016. pp. 135–142. doi:10.1109/WIW.2016.10

Silva, R., Aires, K., Santos, T., Abdala, K., Veras, R., Soares, A., 2013. Automatic detection of motorcyclists without helmet, in: Proceedings of the 2013 39th Latin American Computing Conference, CLEI 2013. doi:10.1109/CLEI.2013.6670613

Vishnu, C., Singh, D., Mohan, C. K., & Babu, S., 2017. Detection of motorcyclists without helmet in videos using convolutional neural network. IEEE 2017 International Joint Conference on Neural Networks (IJCNN). pp. 3036-3041.

Waranusast, R., Bundon, N., Timtong, V., Tangnoi, C., Pattanathaburt, P., 2013. Machine vision techniques for motorcycle safety helmet detection, in: International Conference Image and Vision Computing New Zealand. pp. 35–40. doi:10.1109/IVCNZ.2013.6726989

Zivkovic, Z., Van Der Heijden, F., 2006. Efficient adaptive density estimation per image pixel for the task of background subtraction. Pattern Recognit. Lett. 27, 773–780. doi:10.1016/j.patrec.2005.11.005

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50379-7

Towards the economic optimisation of European supply chains for CO_2 capture, transport and sequestration, including societal risk analysis

Federico d'Amore[a], Paolo Mocellin[b], Chiara Vianello[b], Giuseppe Maschio[b], Fabrizio Bezzo[a*]

[a]*CAPE-Lab – Computer-Aided Process Engineering Laboratory, Department of Industrial Engineering, University of Padova, via Marzolo 9, 35131, Padova (PD), Italy*
[b]*Department of Industrial Engineering, University of Padova, via Marzolo 9, 35131, Padova (PD), Italy*
fabrizio.bezzo@unipd.it

Abstract

European large stationary sources are currently emitting more than 1.4 Gtons of CO_2 every year. A significant decrease in greenhouse gases generation cannot be achieved without relying on carbon capture and sequestration technologies. However, although being practiced for over 30 years, CO_2 transportation is intrinsically characterised by the risk of leakage. To assess and tackle this issue, this study proposes a spatially mixed-integer linear programming approach for the economic optimisation of a European supply chain for carbon capture, transport and geological storage, in which a risk analysis is carried out and incorporated within the modelling framework. Post-combustion, oxy-fuel combustion and pre-combustion are considered as technological options for CO_2 capture, whereas both pipelines (inshore and offshore) and ships are taken into account as transport means. Societal risk assessment and risk mitigation measures are included in the design of the transport network. The overall supply chain is economically optimised, in such a way as to guarantee that the local level of risk complies with a pre-defined threshold.

Keywords: supply chain optimisation, carbon capture and storage, societal risk analysis, risk mitigation measures, hazardous CO_2 transport.

1. Introduction

Although being practiced for over 30 years, transport of CO_2 is nowadays still rising public concern, especially when the possibility of leakage is taken into account in the vicinity of densely populated areas. In fact, pipelines corrosion and external sources may affect the transport safety and, when transported for long distances through densely populated regions, CO_2 exposes to risk a number of people that may be significantly greater than the number of those exposed to risk from either capture or storage sites (IPCC, 2005). Although being a non-toxic and non-flammable gas, CO_2 exhibit a greater density than air, and could therefore locally accumulate to highly dangerous concentrations in certain areas. To assess and tackle this issue, the implementation of a carbon capture and sequestration (CCS) network cannot be disjointed from a proper optimisation-driven analysis that aims at risk minimisation, even at early supply chain (SC) design stage (Onyebuchi et al., 2017). Similarly to what is commonly proposed for gas pipelines, some papers deal with risk-constrained CO_2 SC optimisation, but just a few

were focused on hazards related to pipeline transport systems and on the quantification and prevention of their potential loss. Han and Lee (2013) proposed a fuzzy optimisation for the simultaneous techno-economic, environmental and risk assessment of a CCS SC in South Korea, where risk related to transport modes was treated in terms of cost and technical loss. Khan and Tee (2016) described a general mathematical framework for risk-cost optimisation of underground pipelines networks, where both pipeline corrosion and other failure modes were taken into account, so as to define the life-cycle costs and propose a maintenance strategy. Knoope et al. (2014) addressed risk for the appropriate design and routing of a CO_2 pipeline, including geographic features and considering different measures of risk mitigation. In particular, as regard methods for risk reduction, several contributions were specifically focused either on the optimal positioning of safety valves, or on the selection of suitable transportation paths (e.g., Vianello et al., 2016).
As a result, a lack of knowledge can be identified in the field of large-scale CCS SC optimisation, i.e. the case in which risk and safety measures are simultaneously included in the modelling framework along with the description of those economic aspects related to the installation and operation of such complex networks. In particular, given the high-scale of the investigated problem, societal risk appears to be the best option to estimate the measure of incident size, since it represents the risk related to a population located in the effect region of one or more incidents (AIChE, 2000). Therefore, this study proposes a mixed-integer linear programming (MILP) model for the economic optimisation of a European SC for carbon capture, transport and geological storage, in which a risk analysis is carried out and incorporated within the modelling framework.

2. Material and methods

This paper proposes a single objective, multi-echelon, time-static, spatially-explicit MILP model for the strategic design of a risk-constrained CCS SC within the European geographical context. The objective is the total cost minimisation, such that all feasibility and logical constraints for the optimal SC design are satisfied as in the model reported in d'Amore and Bezzo (2017), while simultaneously including societal risk analysis on CO_2 transport modes and implementing the possibility of installing specific risk mitigation measures on the resulting logistic network, according to the locally accepted levels of societal risk. All the SC stages (i.e., capture, transport and sequestration) are simultaneously optimised within the mathematical framework. In order to decrease the computational burden, here the SC is not described as a multi-period time evolving infrastructure, but only the final configuration (defined in terms of capability for CO_2 sequestration) is taken into account. The spatial resolution is achieved as in d'Amore and Bezzo (2017) by means of a grid $g = \{1, 2, 3, \ldots, 122, 123, 124\}$ of European squared cells. Capture technologies are then introduced through a set of options k that includes: (*i*) post-combustion absorption from the flue gasses of coal-fired power plants; (*ii*) post-combustion absorption from the flue gasses of gas-fired power plants; (*iii*) oxy-fuel technology applied to coal-fired power plants, in which a pre-combustion air separation is obtained for a nearly pure oxygen combustion; and (*iv*) pre-combustion, which entails the production of a hydrogen-rich syngas to fuel gas-fired facilities. Transport is described through a set of modes $l = \{$*inshore pipeline*, *offshore pipeline*, *ship*$\}$; flowrates are discretised according to a capacity set p to speed up the problem computing. The risk analysis is characterised by the implementation of different classes of scenarios $h = \{i, ii, iii, iv\}$ according to the magnitude of the hazard, whereas risk mitigation measures, whose selection may affect both societal risk and total cost of specific transportation paths, are included within a set of possible options m.

Societal risk is defined as the risk of health consequences related to the population that may be affected by hazardous incidents *h* in a certain region *g*. Therefore, this work assesses the risk analysis for inshore pipelines only, being the offshore and the ship options clearly practiced far from densely populated areas. Furthermore, risk related to docking operations was not considered in this model. Three options were chosen to describe the damage surface: (*i*) a strong impact (LC50), in which results a toxic dose of 100000ppm for 15 min; (*ii*) an intermediate situation with 60000ppm for 10 min; and (*iii*) an area of irreversible damage (IDLH) generated by a dose of 40000ppm for 30 min (Vianello et al., 2016).

Within the framework of pipeline transport systems, a risk mitigation measure can be defined as either a physical or a planning safety option that, if implemented, produces (*i*) a decrease in the failure frequency or in the quantity of CO_2 released, i.e. a beneficial influence on societal risk, but (*ii*) also an increase in the costs of the transport system, i.e. a disadvantageous effect on economics. Following the indications by Knoope et al. (2014), the following risk mitigation measures were included in this study within the set of options *m*: (*i*) the installation of marker tape above the pipeline; (*ii*) the installation of protective concrete slabs that may prevent an excavator driver from punching a hole in the pipeline; (*iii*) the deep burying of pipelines (at 2.0 m), which may reduce the possibility to hit the pipeline; and (*iv*) the establishment of a weekly surveillance interval to detect illicit excavations and therefore prevent damages. In addition, the "none" mitigation option was included in set *m* as well, in order to consider the "no-action" choice. On the other hand, other mitigation options that are related to the specific flowrate characteristics (e.g., the optimisation of block valves distance, or the choice of the design factor for calculating walls thickness), were not considered in this study, given the large scale of a European SC optimisation and the spatial resolution that is here considered according to the grid *g*.

3. Mathematical formulation

The objective is the minimisation of total costs (*TC* [€]) that occur to install and operate the CCS SC, in such a way as to guarantee that the local level of risk is lower than a pre-set threshold. *TC* is given by the contributions of the cost for capture (*TCC* [€]), the cost for transport (*TTC* [€]), and that for sequestration (*TSC* [€]):

$$TC = TCC + TTC + TSC \tag{1}$$

The capture and the sequestration problems are described as in d'Amore and Bezzo (2017). On the other hand, the transport problem is here modified to introduce the risk analysis on inshore pipeline transport for both inter- and intra-connection system. As regards inter-connection between region *g* and *g'*, the regional societal risk SR_g^{inter} [events] in origin region *g* is given by the sum of local societal risk $SR_{p,g,l,g'}^{inter}$ [people·events]) produced by each flowrate *p* that is transported from region *g* through mean *l* to region *g'* (oppositely, $SR_{g'}^{inter}$ [events] refers to destination region *g'*):

$$SR_g^{inter} = \sum_{p,l,g'} SR_{p,g,l,g'}^{inter} / P_g \qquad SR_{g'}^{inter} = \sum_{p,l,g} SR_{p,g,l,g'}^{inter} / P_{g'} \qquad \forall g, g' \tag{2}$$

where P_g [people] and $P_{g'}$ [people] are the population in region g and g', respectively. $SR^{inter}_{p,g,l,g'}$ can be evaluated according to the optional implementation of a mitigation option:

$$SR^{inter}_{p,g,l,g'} = \sum_m MF_m \cdot \delta^{inter}_{m,p,g,l,g'} \cdot \sum_h \overline{Pd}_{g,g'} \cdot Sh^{inter}_{h,g,g'} \cdot Pf_h \cdot LD_{g,g'} \qquad \forall p,g,l,g' \qquad (3)$$

$\overline{Pd}_{g,g'}$ [people/km^2] represents the average value of population density between region g and g', while $Sh^{inter}_{h,g,g'}$ [km^2] is the surface affected by hazardous case h between region g and g'. Each incident typology h is then weighted according to the probability Pf_h [events/km] that incident h through mode l generates an event (Table 1).

Then, $LD_{g,g'}$ [km] is the distance between region g and g' (i.e., the transportation length). The binary variable $\delta^{inter}_{m,p,g,l,g'}$ of Eq.(3) is a decision variable representing whether a mitigation option m is taken into account on a specific inter-connection path, or not, and it is then coupled with mitigation factors MF_m [%] (Table 2) so as to take into account the applicable measures for risk mitigation. $Sh^{inter}_{h,g,g'}$ is calculated by multiplying the safety distance from pipeline L_h [km] (Table 1) by the overall transport length $LD_{g,g'}$. Eq.(3) therefore defines local societal risk as a result of population, transport modes and incident typology as well as considering the possible implementation of a mitigation option. Conversely, as regards intra-connection within cell g, the regional societal risk SR^{intra}_g [events] in region g is given by the contribution of local societal risk ($SR^{intra}_{g,l}$ [people·events]) generated by mode l, that is then weighted for the local population P_g:

$$SR^{intra}_g = \sum_l SR^{intra}_{g,l} / P_g \qquad \forall g \qquad (4)$$

where $SR^{intra}_{g,l}$ can be estimated according to the possible implementation of mitigation measures (i.e., through a binary variable $\delta^{intra}_{m,g,l}$, representing whether a mitigation option m is implemented in region g for intra-connection mode l, or not), and on the basis of the potential impact surface ($Sh^{intra}_{h,g}$ [km^2]) that may be affected:

$$SR^{intra}_{g,l} = \sum_m MF_m \cdot \delta^{intra}_{m,g,l} \cdot \sum_h Pd_g \cdot Sh^{intra}_{h,g} \cdot Pf_h \cdot LD_g \qquad \forall g,l \qquad (5)$$

Table 1. Probability (Pf_h [events/km]) and liquid release distance (L_h [km]) of hazard h.

		Pf_h			L_h		
h	hazard	LC50	60000ppm	IDLH	LC50	60000ppm	IDLH
i	Rupture	0.002600	0.000780	0.000026	1.800	2.300	2.700
ii	large puncture	0.002600	0.000780	0.000026	0.450	0.587	0.725
iii	med. Puncture	0.007800	0.002340	0.000078	0.108	0.139	0.171
iv	small puncture	0.013000	0.003900	0.000130	0.027	0.035	0.044

Table 2. Mitigation factor (MF_m [%]) and unitary cost (UMC_m [€/km]) of mitigation m.

m	none	marker tape	concrete slab	deep burying	surveillance
MF_m	100.0	59.9	20.0	29.0	70.0
UMC_m	0	220	110000	51500	1337

Pd_g [people/km^2] is the population density in region g, LD_g [km] is the maximum size of cell g, $Sh^{intra}_{h,g}$ is calculated by multiplying the safety distance L_h by LD_g. Finally, the total

regional societal risk SR_g [events] is constrained to be lower than the maximum allowable regional societal risk (SR_g^{max} [events]), that is imposed by national standards:

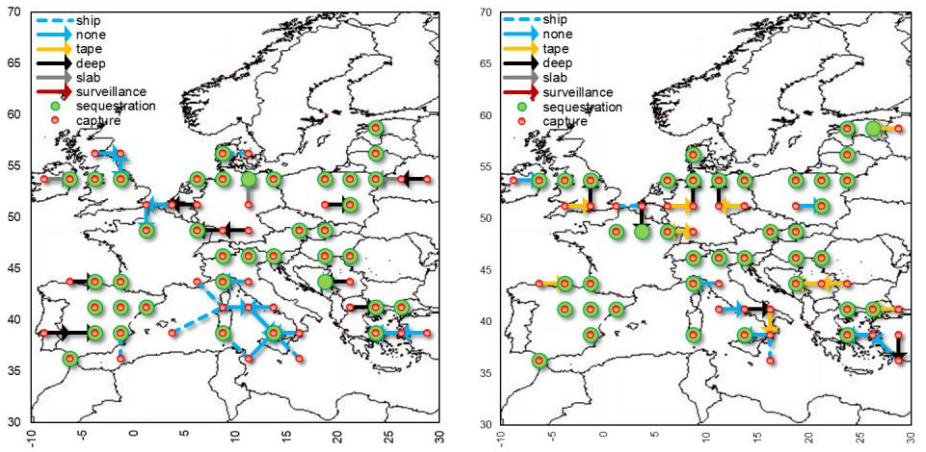
$$SR_g = SR_g^{inter} + SR_{g'}^{inter} + SR_g^{intra} \leq SR_g^{max} \qquad \forall g \tag{6}$$

and the total cost for mitigation TTC^m [€] is evaluated for the selected and installed transport network, according to the unitary cost of mitigation (UMC_m [€/km]) (Table 2):

$$TTC^m = \sum_{m,p,g,l,g'} \delta_{m,p,g,l,g'}^{inter} \cdot UMC_m \cdot LD_{g,g'} + \sum_{m,g,l} \delta_{m,g,l}^{intra} \cdot UMC_m \cdot LD_g \tag{7}$$

4. Results and discussion

The modelling framework was optimised in less then 1 hour by aim of the software GAMS® through the CPLEX solver on a 32 GB RAM laptop. The regional upper bound for societal risk SR_g^{max} calculation was set equal to 0.01 events (i.e., following the indications provided by United Kingdom regulations, AIChE, 2000), whereas the lower bound for minimum European carbon reduction target was imposed equal to 50 % of overall yearly CO_2 emission from large stationary sources (d'Amore and Bezzo, 2017). Furthermore, three scenarios will be here investigated according to the employed methodology for failure evaluation: (*i*) Scenario A - LC50; (*ii*) Scenario B - 60000ppm; and (*iii*) Scenario C - IDLH criteria (Table 3).



(a) (b)

Figure 1. SC configuration for Scenario A (a) and Scenario B (b).

Scenario A exhibits a total cost *TC* of 20.9 G€ (the SC is sequestering more than 600 Mtons of CO_2), of which *TTC* is 1.8 G€ (i.e., the 8.7 % of *TC*). In particular, the SC entails the installation of strong mitigation measures (deep burying and concrete slabs) on the inshore pipeline system (Figure 1a), in order to keep SR_g always lower than SR_g^{max}. This produces a total cost for mitigation TTC^m of 187 M€ (10.3 % of *TTC*). Furthermore, a noteworthy part of the transport infrastructure is developed within offshore areas, in order to exploit the beneficial effect of low population density on societal risk. As regards Scenario B (Figure 1b), the implementation of the 60000 ppm methodology generates a lower mitigation cost (TTC^m is here 89 M€, since mostly marker tape is now installed as mitigation option), which consequently produces a lower value of *TTC* (1.5 G€) and also

a decrease in *TC* (20.5 G€), despite an identical amount of sequestered CO_2 with respect to Scenario A. Within Scenario B, the majority of the transport infrastructure is developed in the mainland, with some deep burying operated in densely populated areas. Conversely, the IDLH approach (Scenario C) is even less cautionary then other scenarios, since it suggests the installation of no mitigation measures, thus producing the best economic performance (*TC* is 20.3 G€), and lower cost for transport (*TTC* is 1.3 G€). Therefore, according to the IDLH methodology for safety distance calculation, it is possible to design an economically optimal European transport network, which already satisfies the regional risk constraints without implementing any mitigation measure on the inshore system.

Table 3. Economic results and computational information for all scenarios.

scenario	criteria	*TTC* [M€]	TTC^m [M€]	sol. time [s]	opt. gap [%]
A	LC50	1814.8	186.6	792	1.9
B	60000ppm	1531.2	89.6	569	1.3
C	IDLH	1292.1	0.0	627	0.4

5. Conclusions

A wide-scale European risk-constrained economic optimisation has been here presented, in which the minimisation of total cost for capturing, transporting and sequestering CO_2 from large stationary sources has been coupled with a societal risk analysis and with the possible implementation of risk mitigation measures on the transport infrastructure. In particular, different system configurations emerged according to the chosen methodology for failure evaluation. In the most conservative scenario (i.e., when the LC50 criteria is employed), mitigation actions represented 10 % of total cost for installing and operating the transport network. For the two additional scenarios being considered, mitigation costs never represented more than 6 % of transport costs.

References

AIChE – Center for Chemical Process Safety, 2000. Chemical Process Quantitative Risk Analysis. Wiley-Interscience, New York.

d'Amore, F., Bezzo, F., 2017. Economic optimisation of European supply chains for CO_2 capture, transport and sequestration. Int. J. Greenh. Gas Con. 65, 99-116.

Han, J.H., Lee, I.B., 2013. A comprehensive infrastructure assessment model for carbon capture and storage responding to climate change under uncertainty. Ind. Eng. Chem. Res. 52, 3805-3815.

IPCC, 2005. IPCC Special Report on Carbon Dioxide Capture and Storage. Prepared by Working Group III of the Intergovernmental Panel on Climate Change, Cambridge University Press.

Khan, L.R., Tee, K.F., 2016. Risk-cost optimization of buried pipelines using subset simulation. J. Infrastruct. Syst. 22, 1–9.

Knoope, M.M.J., Raben, I.M.E., Ramìrez, A., Spruijt, M.P.N., Faaij, A.P.C., 2014. The influence of risk mitigation measures on the risks, costs and routing of CO_2 pipelines. Int. J. Greenh. Gas Con. 29, 104-124.

Onyebuchi, V.E., Kolios, A., Hanak, D.P., Biliyok, C., Manovic, V., 2017. A systematic review of key challenges of CO_2 transport via pipelines. Renew. Sust. Energ. Rev., 81, 2563-2583.

Vianello, C., Mocellin, P., Macchietto, S., Maschio, G., 2016. Risk assessment in a hypothetical network pipeline in UK transporting carbon dioxide. J. Loss Prevent. Proc. 44, 515-527.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50380-3

Alarm Data Analysis for Safe Plant Operations: Case Study of Ethylene Plant

Zhexing Wang and Masaru Noda*

Department of Chemical Engineering, Fukuoka University, Fukuoka 814-0180, Japan
**mnoda@fukuoka-u.ac.jp*

Abstract

Sequential alarms are sets of alarms occurring in succession within a short period of time after triggering an initial warning alarm about an abnormality. These types of alarms reduce the ability of plant operators to cope with operation abnormalities because critical alarms are often lost in numerous other correlated ones. Previously, we proposed an identification method for sequential alarms buried in noisy plant operation data using dot matrix. The dot matrix method is a sequence alignment method for identifying similar regions in DNA or RNA, which may be a consequence of functional, structural, or evolutionary relationships between sequences. In this case study, we use this dot matrix method with operation data from an industrial ethylene plant. The results revealed that sequential alarms in a high volume of operation data from the industrial ethylene plant could effectively be identified using the dot matrix method.

Keywords: Plant Alarm System; Sequential Alarm; Dot Matrix Method; Plant Operation Data, Ethylene Plant

1. Introduction

Advances in distributed control systems (DCS) in the chemical industry have made it possible to inexpensively and easily install numerous alarms in these DCS. While most alarms help operators detect an abnormality and identify its cause, some do not. A poor alarm system might cause sequential alarms, which are a series of alarms that generally occur quickly in specifically-timed succession. These sequential alarms reduce the operators' ability to cope with plant abnormalities because the critical alarms get buried under many unnecessary ones. Therefore, it is very important to identify sequential alarms in plant operation data to ensure safe operations.

Event correlation analysis (Nishiguchi *et al.*, 2010) was first proposed to identify sequential alarms in noisy plant operation data. This method used plant operation data and a cross correlation function to quantify the degree of similarity on the basis of the time lag between two alarms. Sequential alarms were identified by grouping correlated alarms and operations in accordance with their degree of similarity. This event correlation analysis was applied to the operation data of an industrial ethylene plant, previously, and correctly identified similarities between correlated sequential alarms (Higuchi *et al.*, 2010, Takai *et al.*, 2012). However, the event correlation analysis occasionally failed to detect similarities between two physically related sequential alarms when deletions, substitutions, and/or transpositions occurred in the alarm sequence.

A method for evaluating similarities between sequential alarms by using the normalized Levenshtein distance metric was proposed by Akatsuka *et al.* in 2013. The Levenshtein distance (Levenshtein 1966, Yujian and Bo, 2007) is a string metric for measuring the

difference between two sequences defined as the minimum number of edit operations, such as insertion, deletion, and substitution of a single character, needed to transform one string into another.

In our previous study (Wang *et al.*, 2017), we proposed an identification method of sequential alarms by applying a dot matrix method to plant operation data. The Levenshtein distance was applied to the simulation data of an azeotropic distillation column, and the results revealed that the method was able to correctly identify similarities between correlated sequential alarms even when the event correlation analysis failed due to deletions, substitutions, and/or transpositions in the alarm sequence. However, the method was not able to identify sequential alarms hidden in plant operation data. To address this, we applied the dot matrix method to the operation data of an industrial ethylene plant in this study.

2. Dot Matrix Method

A dot matrix method (Mount, 2004) is a sequence alignment method for identifying similar regions in DNA or RNA. Figure 1 shows an example of two DNA sequences. Similar regions in DNA or RNA may be a consequence of functional, structural, or evolutionary relationships between the sequences.

In the dot matrix method, one sequence (S_1) is listed across the bottom of the graph, and the other sequence (S_2) is listed down the left side, as illustrated in Fig. 2. Starting with the first character in S_2, the comparison moves across the graph in the first row and places a dot in any column where the character in S_1 is the same. The second character in S_2 is then compared to the entire S_1 sequence, and a dot is placed in row 2 wherever a match occurs. This process is continued until the graph is filled with dots representing all the matches of S_2 characters with S_1 characters. A diagonal row of dots reveals the similarity between these two sequences. Dots not on a diagonal row represent random matches that are probably not related to any significant alignment.

S_1: A G C T A G G A

S_2: G A C T A G G C

Fig.1 Dot matrix comparison of two DNA sequences

The major advantage of the dot matrix analysis for finding sequence alignments is that all possible matches between two sequences are found, leaving the engineers the choice of identifying the most significant matches through an examination of the dot matrix for long runs of matches, which appear as diagonals.

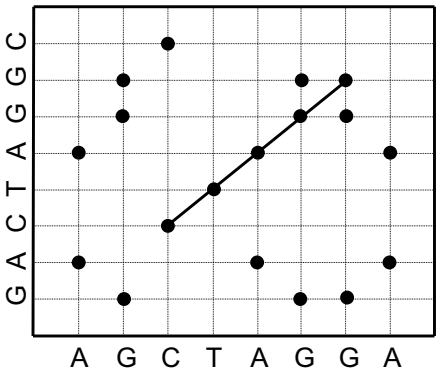


Fig.2 Dot matrix comparison of two DNA sequences

3. Identification Method for Sequential Alarms

The plant operation data recorded in DCS generally consisted of the alarm timings and their tag names, as listed in Table 1. Plant operation data was first converted into a single alarm sequence by putting them in order of the time the alarm occurred. Then, similar regions in the alarm sequences were identified by comparing the alarm alignment with the converted sequence. Figure 3 shows an example of the plant operation data dot matrix analysis. Finally, the identified similar regions, which were assumed to be sequential alarms, were classified into smaller sets of similar sequential alarms in accordance with the similarities between them.

Table 1 Example of plant operation data

Date/Time	Alarm tag name
2013/01/01 00:08:53	A_1
2013/01/01 00:09:36	A_4
2013/01/01 00:11:42	A_2
:	:

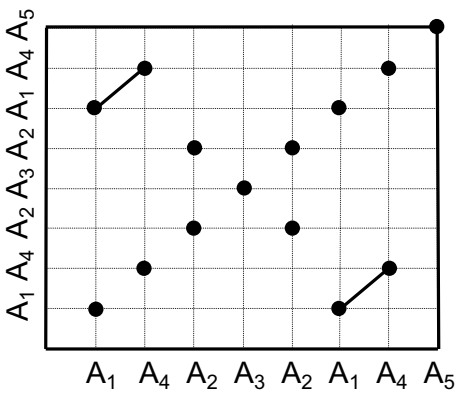


Fig. 3 Dot matrix analysis example from plant operation data

4. Case Study of Ethylene Plant

We focused our study on an Idemitsu Kosan Co. Ltd. ethylene plant in Chiba, which started operations in 1985. Figure 4 is a process flow diagram for the ethylene plant, which has two board operators using DCS. The plant IDs in Fig. 4 indicate the unit identification number, which are summarized in Table 2.

The total number of alarms in the DCS was 3236 for process control and monitoring. When an alarm occurred, the alarm name and time were recorded in the operation log data.

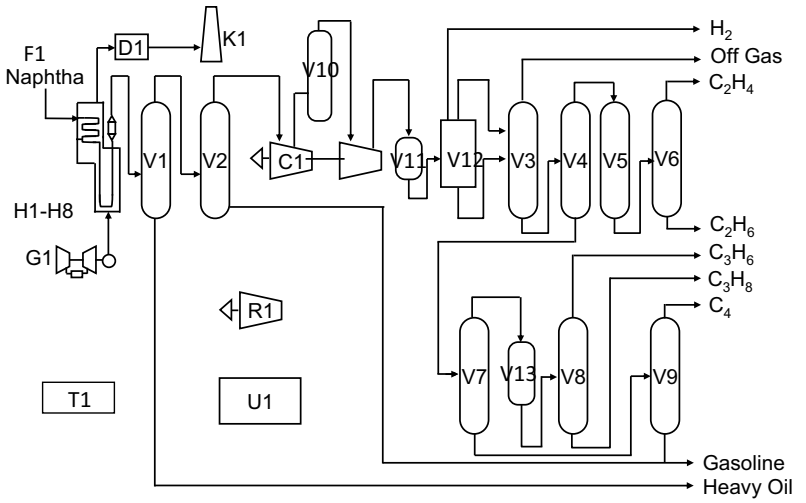


Fig. 4 Process flow diagram for ethylene plant (Higuchi *et al.*, 2010)

Table 2 Units in ethylene plant

No.	Unit name	No.	Unit name
C1	Cracked gas compressor	V2	Quench water tower
D1	DeNOx section	V3	Demethanizer
F1	Feed	V4	Deethanizer
G1	Gas turbine	V5	Acetylene absorber
H1–H8	Cracking furnaces 1–8	V6	Ethylene fractionator
K1	Exhaust gas stack	V7	Depropanizer
P1	Product processing unit	V8	Propylene fractionator
R1	Refrigeration compressor	V9	Debutanizer
T1	Tank	V11	Dryer
U1	Utility section	V12	Chill train
V1	Primary fractionator	V13	Hydrogenation Reactor

The plant log data gathered in one month included 914 different types of alarms. A total of 16803 alarms were generated. Figure 5 shows the times at which the 914 types of

alarms and operation events occurred. It was difficult to identify sequential alarms from just the data shown in Fig. 5.

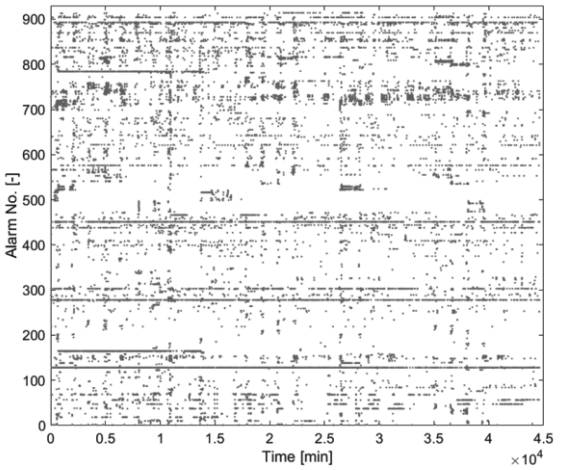


Fig. 5 Plant operation data of ethylene plant (Higuchi *et al.*, 2010)

Figure 6 shows the result of a dot matrix analysis of the plant operation data. A large number of sequential alarms in the plant operation data can be inferred from the multiple diagonal lines represented in black. Table 3 lists the top 10 longest sequential alarms identified by our method. The results revealed that our method is able to correctly identify similar sequential alarms in plant operation data.

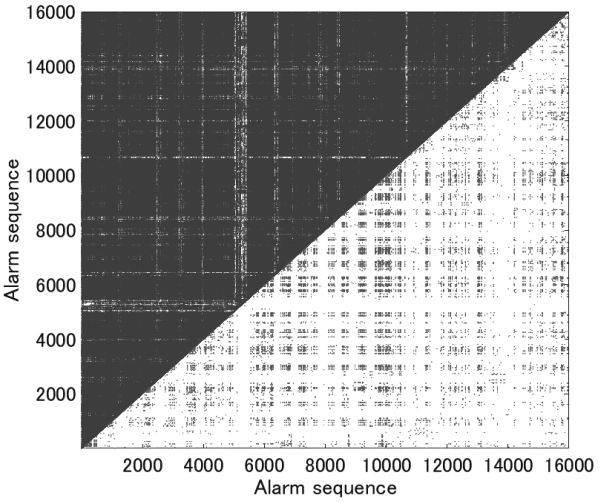


Fig. 6 Results of dot matrix analysis

Table 3 Top 10 longest sequential alarms

Rank	Fre.	Len.	Alarm sequence	Related units
1	2	8	$A_{451}\rightarrow A_{128}\rightarrow A_{451}\rightarrow A_{278}\rightarrow A_{451}\rightarrow A_{128}\rightarrow A_{451}\rightarrow A_{128}$	V2, U1
2	2	7	$A_{450}\rightarrow A_{451}\rightarrow A_{128}\rightarrow A_{516}\rightarrow A_{450}\rightarrow A_{451}\rightarrow A_{128}$	H2, U1
3	2	7	$A_{128}\rightarrow A_{451}\rightarrow A_{128}\rightarrow A_{451}\rightarrow A_{128}\rightarrow A_{165}\rightarrow A_{784}$	H2, U1
4	2	6	$A_{893}\rightarrow A_{128}\rightarrow A_{812}\rightarrow A_{813}\rightarrow A_{814}\rightarrow A_{815}$	H6
5	2	6	$A_{826}\rightarrow A_{164}\rightarrow A_{165}\rightarrow A_{784}\rightarrow A_{451}\rightarrow A_{128}$	H2, U1
6	2	6	$A_{516}\rightarrow A_{451}\rightarrow A_{128}\rightarrow A_{451}\rightarrow A_{516}\rightarrow A_{128}$	H2, U1
7	2	6	$A_{451}\rightarrow A_{816}\rightarrow A_{451}\rightarrow A_{816}\rightarrow A_{128}\rightarrow A_{816}$	H6, U1
8	2	6	$A_{451}\rightarrow A_{812}\rightarrow A_{813}\rightarrow A_{814}\rightarrow A_{815}\rightarrow A_{128}$	H6, U1
9	2	6	$A_{165}\rightarrow A_{784}\rightarrow A_{278}\rightarrow A_{451}\rightarrow A_{128}\rightarrow A_{893}$	U1, U2
10	2	6	$A_{128}\rightarrow A_{451}\rightarrow A_{128}\rightarrow A_{278}\rightarrow A_{451}\rightarrow A_{128}$	U1

5. Conclusions

A dot matrix method using the Levenshtein distance metric was applied to an ethylene plant operation data. The results revealed that the method was able to correctly identify similar sequential alarms in the data. By classifying sequential alarms into smaller groups, a dot matrix method using the Levenshtein distance metric effectively identified similar types of alarms. These results revealed that this method can correctly identify similarities, despite changes to sequences, in plant operation data to ensure safer operations at an industrial chemical plant.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 17K06909.

References

Akatsuka, S. and Noda, M.: Similarity Analysis of Sequential Alarms in Plant Operation Data by Using Levenshtein Distance, *Proc. of PSE Asia 2013*, Kuala Lumpur (2013)
Higuchi, F., Noda, M., and Nishitani, H.: Alarm Reduction of Ethylene Plant using Event Correlation Analysis (in Japanese), *Kagaku Kogaku Ronbunshu*, 36(6), 576-581 (2010)
Levenshtein, V.I.: Binary Codes Capable of Correcting Deletions, Insertions, and Rever-sals, *Cybernetics and Control Theory*, 10(8), 707-710 (1966)
Mount, D.W.: Bioinformatics Sequence and Genome Analysis Second Edition, Cold Spring Harbor Laboratory Press, New York (2004)
Nishiguchi, J. and Takai, T.: IPL2&3 Performance Improvement Method for Process Safety Using the Event Correlation Analysis, *Computers & Chemical Engineering*, 34, 2007-2013 (2010).
Takai, T., Noda, M., and Higuchi, F.: Identification of Nuisance Alarms in Operation Log Data of Ethylene Plant by Event Correlation Analysis (in Japanese), *Kagaku Kogaku Ronbunshu*, 38(2), 110-116 (2012)
Wang, Z. and Noda, M.: Identification of Repeated Sequential Alarms in Noisy Plant Operation Data Using Dot Matrix Method with Sliding Window, *Journal of Chemical Engineering of Japan*, 50(6), 445-449 (2017)
Yujian, L. and Bo, L.: A Normalized Levenshtein Distance Metric, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6), 1091-1095 (2007)

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50381-5

Comparison of Fire Extinguishing Approaches in Engine Control Rooms

Julian Enrique Mendez Alvarez*, Felipe Muñoz Giraldo

Product and Process Design Group Chemical Engineering Department, Universidad de Los Andes, Bogotá 111711, Colombia
je.mendez188@uniandes.edu.co

Abstract

Electric cabinets are a major contributor to electrical fire hazards, with an occurrence of more than 10% of all fire events. This project addresses the comparison of various fire suppression approaches in engine control rooms, using Fire Dynamics Simulator as support tool. The project studies three different scenarios, one of them without a protection system, a second one with a clean agent as extinguishing tool and a final one, which implements an inherently safer design. For the first case, the gas temperature in the room reached a limit of 620 °C. This condition may cause structural damages to the room, as well as injuries to the personnel in charge of fire extinction. For the second scenario, it was determined that the extinguishing agent may decrease the negative effects of the fire, although it may not protect other important structures. Finally, the inherently safer design limits the negative effects to the cabinet where the fire occurs, protecting the structure of the remaining cubicles, as well as the personnel.

Keywords: Compartment fire, Inherently Safer Design, Computational Fluid Dynamics, Process Safety, Fire Dynamics Simulator

1. Introduction

The study of electrical fires is a major concern for fire safety in the industry, particularly for oil industry. In petrochemical facilities, the fire hazard may cause severe consequences on safety functions inside the fire compartment and may lead to the loss of one or several compartments (room for electrical control equipment, battery room, etc.), as just as structural damage on the facilities. From the OECD Fire Database the majority of the fire ignition is due to electrical sources (close to 50%) (W. Werner et al, 2011; W. Werner et al. 2009).

Electric fires mean economic losses for companies because they cause partial or full stops of the processes, economic sanctions and several injuries or fatalities (mainly related with severe burns and exposure to toxic combustion products) (Department for Communities and Local Government, 2013; Society of Fire Protection Engineers, 2002). Considering the possible consequences, it is necessary to implement systems or strategies that reduce the likelihood or mitigate the above-mentioned consequences. These strategies are usually implemented following either a prescriptive approach or a performance-based approach (International Maritime Organization, 2009). Prescriptive methods include the compliance of international preventive standards which may mitigate the consequences, yet they use correction factors that can overestimate the system specifications (creating extra costs) or underestimate those requirements (worsening the consequences). On the other hand, performance-based approaches rely on the simulation of the combustion

dynamics in order to estimate the effects of the fire over the structure. The combustion dynamics can be simulated using different computational fluid dynamics tools, considering its great accuracy. One of the most used computational tools is Fire Dynamics Simulator (FDS, developed by the National Institute of Standards and Technology). The simulation of fire protection techniques allows to compare several performance-based techniques such as the active technologies (use of a clean extinguishing agent), as well as a passive technique (inherently safer design).

2. Methodology

Three different scenarios were defined for the comparison of different fire extinguishing approaches. The three scenarios share the cubicles distribution, based on NFPA 70 (National Fire Protection Association, 2014). However, the compartments have frontal and on top ventilation for the first and second scenarios, while the inherently safer system suppresses this element because of its incompatibility with the standards used in the design of this kind of applications (Striebel & John GmbH & Co. KG, 2007). Moreover, for the unprotected scenario (based on an existing industrial facility), an equivalent ventilation area of 15% was defined for the room. These vents were suppressed for the protected scenarios in order to avoid the loss of extinguishing material and a flashback scenario in the inherently safer scheme.

A heat release ratio of 1600 kW/m^2 was imposed, considering typical fires in industrial facilities. The combustion material was assumed as a mix of typical materials present in cubicles, such as cellulose, polyester, polyurethane and PVC. This fictional fuel was defined with the lumped species model proposed by FDS, using the properties reported in Table 1 .

Table 1. Fuel properties

Fuel properties	
Heat of Combustion [J/kg]	29,850,000
Molecular Formula [-]	$C_1 H_{1.6} O_{0.285} N_{0.03}$
CO yield [mol/mol]	0.505
Soot yield [mol/mol]	0.159

Table 2. Fire suppression technologies

Fuel properties		
Material	Clean Agent	Fireproof coating
Molecular Formula [-]	$C_6 F_{12} O$	N/A
Specific Heat [J/kg.K]	987	1,200
Conductivity [W/kg.K]	0.00035	0.15
Emissivity [W/m^2K^4]	N/A	0.7

A clean extinguishing agent was selected for the second scenario. Several sprinkler systems were installed in the room, in order to inject the extinguisher. A volumetric concentration of 5% was considered for a 10 *s* injection time (considering the technology requirement). On the other hand, a fireproof coating was selected for the inherently safer design case. The material was selected according to the DIN 4102-1 code (Striebel &

John GmbH & Co. KG, 2007). Both clean agent and coating properties are reported in Table 2.

A span of one minute was imposed for the simulations, considering the time that takes to the different technologies to mitigate the fire. Furthermore, the grid size was set at 0.20 m for every direction, considering the characteristic diameter criteria, proposed by FDS, depicted in (1). This approach considers the heat release ratio of the fire and the density, specific heat and temperature of the environment.

$$D^* = \left(\frac{HRR}{\rho_\infty Cp T_\infty \sqrt{g}}\right)^{\frac{2}{3}} \tag{1}$$

Additionally, in order to accurately predict the flame extinction, the adiabatic flame temperature of a lower flammability limit mixture was considered. The software algorithm determined if enough heat was transferred to every cell in order to generate a combustion reaction. This temperature is determined as it follows.

$$T_{CFT} = T_0 + \left(\frac{LFL}{100}\right)\frac{\Delta H_c}{nCp} \tag{2}$$

In order to determine the walls' temperature, as well as the cubicle temperature, some temperature recording devices were installed. However, the tool estimates the adiabatic wall temperature that must be corrected in order to obtain temperatures in the same magnitude order of the experiments (Ulf Wickström, Dat Duthinh, & McGrattan, 2009). Hence, the following models were implemented in the post processing stage for the unprotected scenario and the clean agent scenario.

$$\frac{dT}{dt} = \frac{1}{\rho C_p l}\left(h(T_A - T) + \sigma\epsilon(T_A^4 - T^4)\right) \tag{3}$$

As for the surfaces receiving a heat flux as a product of conduction, convection and radiation mechanisms, the proposed model in (3) turns into the following set of equations.

$$\frac{dT_1}{dt} = \frac{h\left(T_{A_1} - T_1\right) + \sigma\epsilon\left(T_{A_1}^4 - T_1^4\right)}{\rho C_p l} + \frac{\alpha}{2l^2}(T_2 - T_1) \tag{4}$$

$$\frac{dT_2}{dt} = \frac{h\left(T_{A_2} - T_2\right) + \sigma\epsilon\left(T_{A_2}^4 - T_2^4\right)}{\rho C_p l} + \frac{\alpha}{2l^2}(T_1 - T_2) \tag{5}$$

where the sub index "1" stands for the cubicle where the fire occurs and the sub index "2" corresponds to the adjacent cubicle. Finally, for the third case, the fireproof coating must be considered. Thus, the model is modified as it follows.

$$\frac{dT_1}{dt} = \frac{1}{\rho C_p l}\left(h\left(T_{A_1} - T_1\right) + \sigma\epsilon\left(T_{A_1}^4 - T_1^4\right) + \frac{kk_r}{2(k_r l + kl_r)}(T_2 - T_1)\right) \tag{6}$$

$$\frac{dT_2}{dt} = \frac{1}{\rho C_p l}\left(h\left(T_{A_2} - T_2\right) + \sigma\epsilon\left(T_{A_2}^4 - T_1^4\right) + \frac{kk_r}{2(k_r l + kl_r)}(T_1 - T_2)\right) \tag{7}$$

3. Results

After implementing the post-processing algorithm, cubicle wall temperatures were obtained. The temperature profiles are depicted in Figure 1.

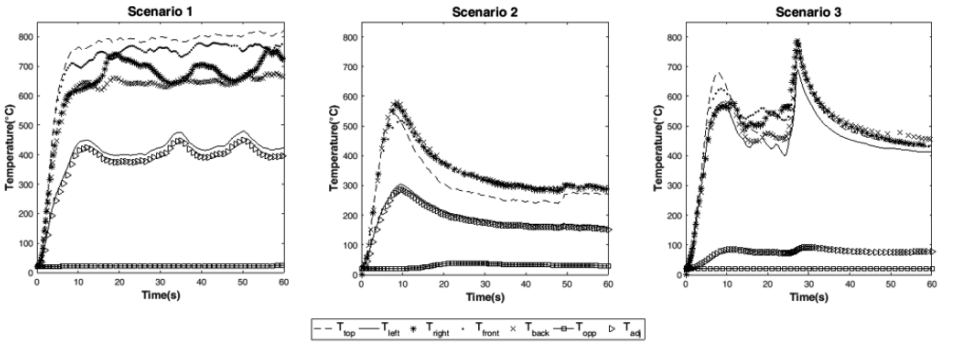


Figure 1. Cubicle Temperature Profile

For the scenario with no fire protection technique, the compartment reached a peak of approximately 820°C at the top of the cubicle, and temperatures close to 800°C for the left, right, frontal and rear walls. This temperature magnitude represents potential risks for the wiring and mechanical integrity of the cubicle where the fire is present and the adjacent one. However, it does not affect the cubicle located in front, given that its temperature does not rise above 50°, which indicates that the prescriptive approach (NFPA70) distribution is effective. In the second scenario, the extinguishing agent lowers the temperature inside the burning cabinet. However, the agent injection spreads the heat of the fire to the adjacent cubicle and may cause potential affectations. The inherently safer design scenario confines the fire to the original cubicle so it prevents potential risk for other cabinets in the room.

In addition, the evolution of the temperature profiles are correlated to the oxygen concentration inside the cubicle. In the first case, there is a constant oxygen feed because of the top and frontal ventilation, that causes the re ignition of the fire and all the peaks depicted for Scenario 1. In the second case, the extinguishing agent displaces the available oxygen in the compartment which causes the combustion process to stop and contributes to a lower temperature. The third case presents a particular behavior; two different temperature peaks can be recognized around 8 *s* and 30 *s*. The second peak corresponds to a re-ignition of the flame due to the entrance of extra oxygen via the frontal ventilation. This shows that a frontal ventilation might not be suitable for an inherently safer design application.

Moreover, net incident radiation was analyzed in order to perform a consequence calculation. As displayed in Figure 2, the first scenario reaches a value of 3.7 kW/m^2 for the surroundings of the cubicle. This incident heat flux could cause first degree burns in 70% of the exposed personnel for an exposure period of 60 *s* and even, second degree burns (in 27% of the exposed personnel). For the protected scenarios, the incident heat flux does not reach a value that may constitute a risk for human health. For a 60 *s* period of exposure, an incident heat flux of 1.5 kW/m^2 or 1 kW/m^2 does not cause any burn to the attending personnel.

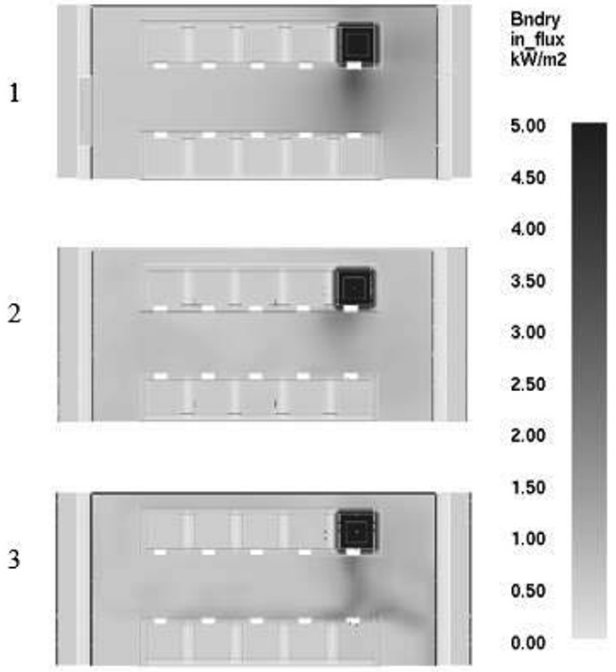


Figure 2 Incident Heat flux

4. Conclusions and future work

The comparison of the scenarios and techniques using FDS was successful. It was determined that even though the use of a clean agent extinguishes the fire, it may propagate the consequences of the conflagration to other compartments. On the other hand, an inherently safer design approach gives up the mechanical and electrical integrity of the affected cubicle but saves the mechanical integrity of the remaining compartments. An inherently safer design implementation is recommended when the heat release ratio is so high that an active system (such as a sprinkler system) cannot guarantee the mechanical integrity of a facility. However, if the forecasted heat release ratio is low, an extinguisher injection system should be implemented, only if it is linked to a smoke detection system. This allows an early detection of the incident and an effective mitigation of the consequences.

Although valuable conclusions were drawn from this comparison, different factors should be considered in order to perform deeper and more valuable analysis. For instance, the cabinets' distribution should be considered in a sensitivity analysis, as well as the ventilation of the room and the compartments. Likewise, the studied technologies should be varied in order to have a wider panorama (different wetting agent, different coating material with varying thickness). Furthermore, economic and technical assessment parameters could be included (cubicle replacement cost, capital cost of the alternatives, economic losses).

Additionally, it is suggested to implement a performance-based analysis for the inherently safer design case, removing the frontal vent in order to determine its suitability for this

kind of approach. Finally, it is recommended to perform an "after simulation" analysis of the grid, in order to determine a balance between computational resources demand and results accuracy.

References

P. Aanderson, P. Van Hees, 2005, Performance of Cables Subjected to Elevated Temperatures, Fire Technol., pp. 1121–1132.

M. Coutin, W. Plumecocq, P. Zavaleta, and L. Audouin, 2015, Characterisation of open-door electrical cabinet fires in compartments, Nucl. Eng. Des., vol. 286, pp. 104–115.

Department for Communities and Local Government., 2013, Fire statistics: Great Britain.

International Maritime Organization., 2009, Guidelines for Measures to Prevent Fires in Engine-Rooms and Cargo Pump-Rooms.

National Fire Protection Association., 2014, NFPA 70.

W. Plumecocq, M. Coutin, S. Melis, and L. Rigollet, 2011, Characterization of closed-doors electrical cabinet fires in compartments, Fire Saf. J., vol. 46, no. 5, pp. 243–253.

Society of Fire Protection Engineers., 2002, SFPE Handbook of Fire Protection Engineering.

Striebel & John GmbH & Co. KG., 2007, Fire Protection Cabinets for Fire Prevention.

W. Werner, J.S. Hyslop, N. Melly, R. Bertrand, M. Röwekamp & A. Huerta., 2011, Enhancements in the OECD fire database – fire frequencies and severity ofevents.

W. Werner, M.Röwekamp, J. Gauvain & A. Angner, A. , 2009. «The OECD fire database– conclusions from phase 2 and outlook. Helsinki (Finland).

U. Wickström, D. Duthinh, & K. McGrattan., 2009, Adiabatic Surface Temperature For Calculating Heat Transfer to Fire Exposed Structures.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50382-7

A novel method for real-time alarm root cause analysis

Feifan Cheng, Jinsong Zhao*

Department of Chemical Engineering, Tsinghua University, Beijing 100084, China
chengff15@mails.tsinghua.edu.cn

Abstract

With complexities in the control systems, it is natural that the chemical plants may become vulnerable to faults in practice. Alarm systems were designed to raise an alarm when a fault is detected. But modern alarm systems often produce large amounts of false or nuisance alarms which leads to alarm floods. Operators receive far more alarms than they can handle. To reduce the alarm floods, we developed a novel method which combined the Time-delayed Convergent Cross Mapping (TCCM) and Bayesian Network (BN). In this paper, the BN is constructed with application of TCCM to process data and expert knowledge. The alarm data is used to estimate the conditional probability among the nodes of the BN. The BN is used to find the alarms propagation paths and root cause at real-time. Finally, the method is applied to the Tennessee Eastman model to illustrate the applicability.

Keywords: Root Cause; Alarms; Bayesian network;

1. Introduction

Process control has made a great contribution to the production efficiency, product quality, process safety and environmental protection in the modern chemical process industry (Shu and Zhao 2015). With complexities in the control systems, it is natural that the chemical plants may become vulnerable to faults in practice. Alarm systems were designed to raise an alarm when a fault is detected. As a process consists of a large amount of different process devices that provide an alert when the process shows an abnormal operating situation, most of the alarms are not meaningful and can be ignored by the operators (Rodrigo, Chioua et al. 2016). The operators are distracted and might ignore critical alarms. So it is necessary to develop a methodology to select the most meaningful alarms and bring the operator's attention to the root cause of the alarms.

A lot of research are focused on the alarm management. A large number of unimportant alarms fall within three categories: standing alarms, repetitive alarms, and consequence alarms. For standing and repetitive alarms, such widely implemented techniques as delay-timer, deadband and shelving are suggested. These methods were successfully implemented on the chemical process. Compared with standing and repetitive alarms, consequence alarms are more complicated and more difficult to handle. To handle the consequence alarms, it is important to obtain the correlations of different alarms. For example, a modified Smith–Waterman algorithm was used for alarm flood pattern matching (Cheng, Izadi et al. 2013). But these methods have the limitation that only alarm data is used for analysis. The BN can be used for root cause analysis of alarms. However, the BN built by expert experience is time-consuming (Wang, Khan et al. 2015). And the root cause detected by the BN may be not the true root cause. Because the BN detects the root cause by the probability of variables at abnormal state. In

present work, these drawbacks are overcome. The BN is constructed with application of TCCM to process data and expert knowledge. The alarm data is used to estimate the conditional probability among the nodes of the BN. After an important variable generates an alarm, the BN is used to find all the possible alarm propagation paths to the important variable. A root cause variable can be found at each path. TCCM is applied to find the causality between these root cause variables. Finally, a true root cause variable is found.

Our research aims to develop a comprehensive method that can precisely find the root cause of alarms. In the following section, we first introduce the basics of TCCM and BN. In Section 3, the proposed method is presented. In section 4, the proposed method is applied to the Tennessee Eastman model to illustrate the applicability. Finally, a conclusion is made.

2. TCCM and BN

2.1. Time-delayed Convergent Cross mapping

At first, the mathematical mechanism of Convergent Cross Mapping (CCM) is described. CCM is a powerful new methodological approach that can help distinguish causality from spurious correlation in time series from dynamical systems (Sugihara, May et al. 2012). The technique is based on the idea that causation can be established if states of the causal variable can be recovered from time series of the affected variable. CCM uses Takens' idea to detect if two variables belong to the same dynamical system.

CCM appears to be limited by the fact that it may not be able to distinguish between bidirectional causality and strong unidirectional causality that leads to synchrony. TCCM makes an extension to CCM that can resolve this problem: by explicitly considering different lags for cross mapping, it is possible to determine whether a driving variable acts with some time delay on a response variable (Ye, Deyle et al. 2015). The main idea is showed as Figure 1. The model is (1). The parameter embedding dimension E and time lag τ are important for TCCM. The value E is estimated by applying simplex projection to the individual time series and choosing the optimal E. Time lag τ is estimated by average mutual information.

$$\begin{cases} x_t = 0.1 * x_{t-1} + rnorm(0,1) \\ y_t = 2 * \mathrm{x}_{t-lag}{}^2 \end{cases} \tag{1}$$

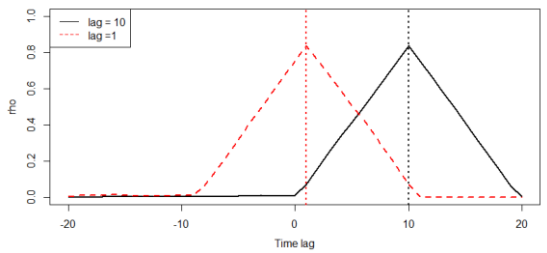


Figure 1: Calculate the correlation coefficient rho by shifting the lag between x and y. To the simple system, the time delay lag is the time lag at which the rho is maximum.

2.2. Bayesian network

A BN graphically shows the relationship among variables of a probability distribution. The BN contains two parts, a graphical structure that defines the qualitative representation and the conditional probabilities that define the quantitative representation. The BN is a directed acyclic graph (DAG). But there may be some loops at chemical plants. Cyclic networks are dealt with by transformation of cyclic networks to acyclic networks using dummy nodes.

3. Method

The overall picture of the methodology is in Figure 2. TCCM is applied for detecting causal relationship between variables and for construction of network. Knowledge of process is used to verify the constructed network. If the network contains a loop, a dummy node is created for one of the variables involved in the loop. Conditional probabilities among different nodes are the quantitative part of the network which is calculated using maximum likelihood method. More details on each part is presented in the application.

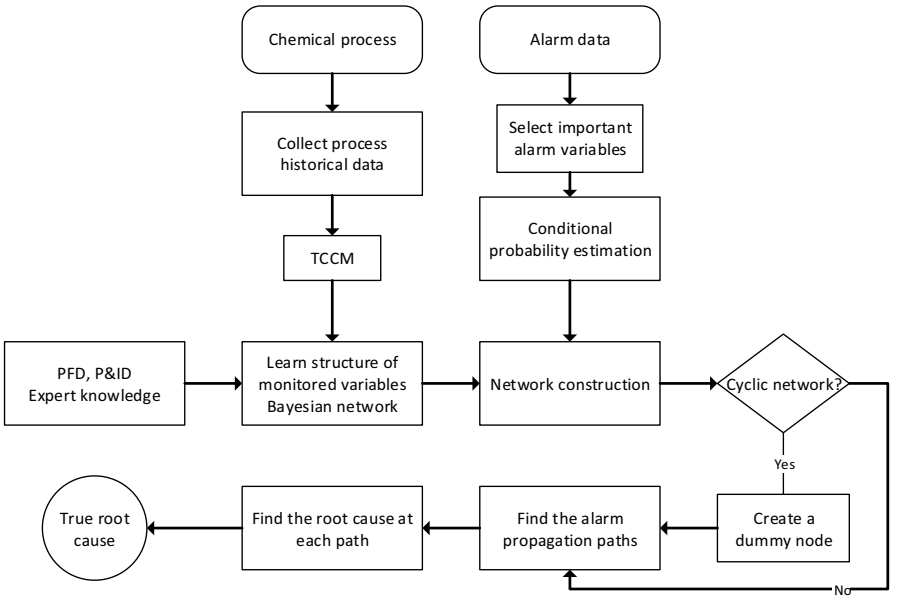


Figure 2: Proposed method for the root cause analysis of alarms

4. Application

To illustrate the applicability of the proposed method, the method is applied to the Tennessee Eastman model. The model consists of five major units: a reactor, condenser, compressor, separator, and stripper; and, it contains eight streams: A, B, C, D, E, F, G, and H. The flow diagram of this process is shown in Figure 3. It consists 41 measured variables and 12 manipulated variables. Among measured variables, 22 variables are continuous process variables and 19 variables are related to composition measurements. The 22 continuous process variables are shown in Table 1 that are the main focus of the

research (Ricker 1996). There are 20 potential faults in this model, a fault scenarios is studied at this paper.

The faulty scenario is a loss in feed A at 20 h. The sample interval is 0.01 h. This variation may change the process parameters in downstream units. The simulation stopped at 27.19 h. The data set consists of 2000 samples from normal data and 719 samples from the abnormal data.

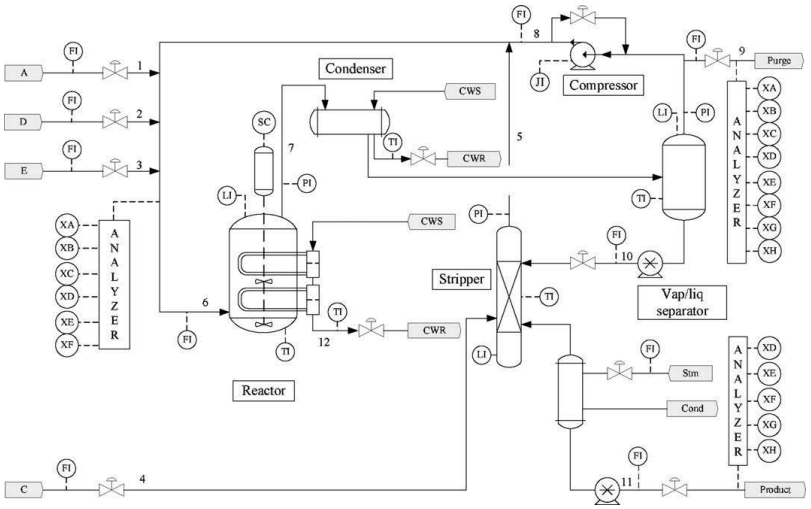


Figure 3: Schematic diagram of Tennessee Eastman model

Table1: Measured Variables in Tennessee Eastman model

Variable	Description	Variable	Description
XMEAS1	A Feed (stream 1)	**XMEAS12**	Separator level
XMEAS2	D Feed (Stream2)	**XMEAS13**	Separator pressure
XMEAS3	E Feed (Stream 3)	**XMEAS14**	Separator Underflow (Stream 10)
XMEAS4	Total Feed (Stream 4)	**XMEAS15**	Stripper Level
XMEAS5	Recycle Flow (Stream 8)	**XMEAS16**	Stripper Pressure
XMEAS6	Reactor Feed Rate (Stream6)	**XMEAS17**	Stripper Underflow (Stream 11)
XMEAS7	Reactor Pressure	**XMEAS18**	Stripper Temperature
XMEAS8	Reactor Level	**XMEAS19**	Stripper Steam Flow
XMEAS9	Reactor Temperature	**XMEAS20**	Compressor Work
XMEAS10	Purge Rate (Stream 9)	**XMEAS21**	Reactor Cooling Water Outlet Temperature
XMEAS11	Separator temperature	**XMEAS22**	Separator Cooling Water Outlet Temperature

The history process data is used to calculate the structure of BN by TCCM. The high alarm limit of each variable is the mean plus three times standard deviations of normal data. The low alarm limit of each variable is the mean minus three times standard deviations of normal data. The process variable can be transferred into three levels: low, normal and high. Reactor Temperature (XMEAS9) is selected as the important alarm variable. The samples when XMEAS9 is at abnormal state are used to estimate the conditional probability of BN. Tennessee Eastman model contains a loop. To deal with this issue, a dummy node is designed for the Recycle Flow (XMEAS5) involved in the

loop. GeNIe software is used to display and update the BN. The constructed network is showed in Figure 4. In this fault scenario, the XMEAS9 is at low level. So the probability of XMEA9 at low level is set 100%. Two alarm propagation paths found at BN is showed at Figure 5. A feed (XMEAS1) and Compress Work (XMEAS20) are the root cause at each path. To find the true root cause of alarms, TCCM is used to detect the causality between XMEAS1 and XMEAS20. The result is showed in Figure 6. From the Figure 6, XMEAS1 is the driving variable, and XMEAS20 is the response variable. XMEAS1 at low level is the true root cause of the alarms.

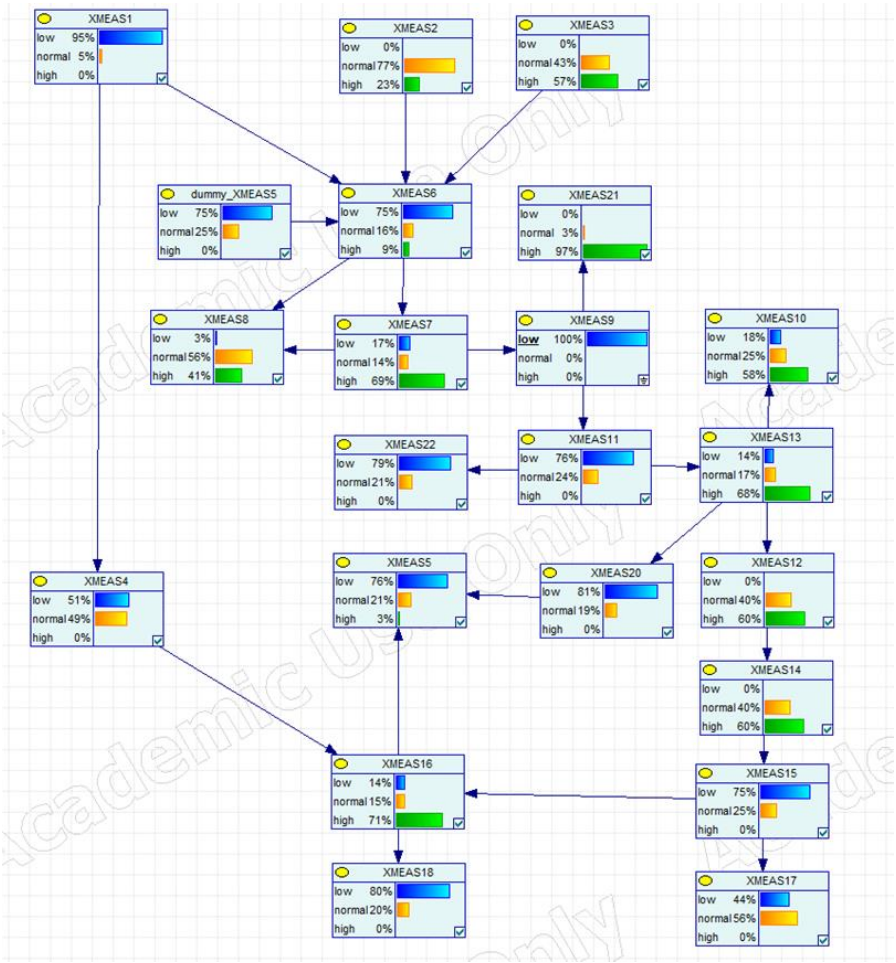


Figure 4: BN for Tennessee Eastman model

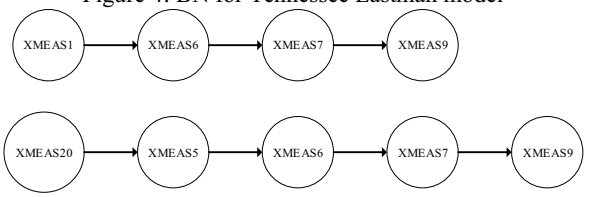


Figure 5: Two alarm propagation paths

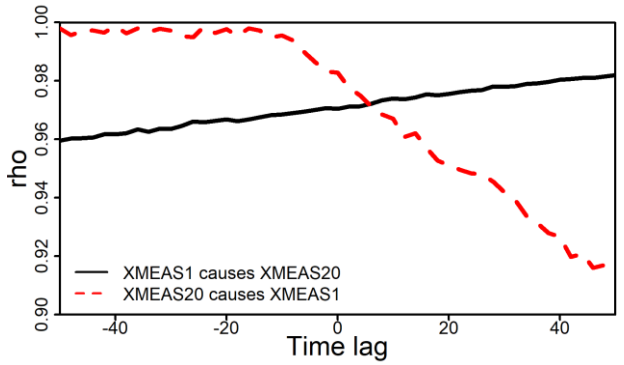


Figure 6: Calculate the correlation coefficient rho by shifting the time lag. In the full line, the time lag maximizing the rho is positive, which means XMEAS1 is the driving variable. In the dotted line, the time lag maximizing the rho is negative, which means XMEAS 20 is the response variable.

5. Conclusion

This paper integrates TCCM and BN to generate a comprehensive method for root cause analysis of alarms. The structure of BN was learned from the process data by TCCM. The proposed method was applied to the Tennessee Eastman model, and it showed a powerful diagnostic capability. The strength of proposed method is that it diagnoses root cause of alarm as well as shows the propagation pathway of the fault. This information will help operators to take correct action and recover the process quickly. However, most of chemical processes are working in dynamic condition. In such cases conditional probability of the BN will shift with time. So the dynamic BN should be implemented to handle this problem. This method needs more test with industrial data sets.

References

Cheng, Y., I. Izadi and T. Chen (2013). "Pattern matching of alarm flood sequences by a modified Smith–Waterman algorithm." Chemical Engineering Research and Design **91**(6): 1085-1094.

Ricker, N. L. (1996). "Decentralized control of the Tennessee Eastman challenge process." Journal of Process Control **6**(4): 205-221.

Rodrigo, V., M. Chioua, T. Hagglund and M. Hollender (2016). "Causal analysis for alarm flood reduction." IFAC-PapersOnLine **49**(7): 723-728.

Shu, Y. and J. Zhao (2015). "Fault diagnosis of chemical processes using artificial immune system with vaccine transplant." Industrial & Engineering Chemistry Research **55**(12): 3360-3371.

Sugihara, G., R. May, H. Ye, C.-h. Hsieh, E. Deyle, M. Fogarty and S. Munch (2012). "Detecting Causality in Complex Ecosystems." Science **338**(6106): 496-500.

Wang, H., F. Khan and S. Ahmed (2015). "Design of scenario-based early warning system for process operations." Industrial & Engineering Chemistry Research **54**(33): 8255-8265.

Ye, H., E. R. Deyle, L. J. Gilarranz and G. Sugihara (2015). "Distinguishing time-delayed causal interactions using convergent cross mapping." Scientific Reports **5**: 14750.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50383-9

Convolution Neural Network based Chemical Leakage Identification

Lening Li[a], Xuqing Jia[a], Wende Tian[a*], Suli Sun[b], and Wu Cao[a]

[a]*College of Chemical Engineering, Qingdao University of Science & Technology, Qingdao 266042, China*
[b]*College of Marine Science and Bioengineering, Qingdao University of Science & Technology, Qingdao 266042, China*
tianwd@qust.edu.cn

Abstract

As leakage seriously threaten the safety of chemical processes, the technology of image recognition based on Convolutional Neural Network (CNN) is used in this paper to probe the chemical leakage. First, the three-dimensional scenes of chemical process are built under normal working conditions and leakage conditions. Then, a CNN with 15 layers is established, including 3 convolutional layers, 4 activation layers, 3 pooling layers, 2 local response normalization layers, 2 fully connected layers, 1 dropout layer, respectively. After the network performance is optimized by adjusting the parameters and functions in each layer, the experimental results with accuracy of 95.82 % are obtained.

Keywords: convolution neural network; leakage; image recognition.

1. Introduction

During the process of chemical production, the corrosion widely exists in many materials, which will lead to not only the material leakage of production, storage, and transportation equipment in long-period operation, but also environmental pollution and material property losses. In some serious circumstances, it will cause fire, explosion and poisoning, endangering the safety of personnel. In practice, the monitoring equipment of chemical plant is widely used, which provides favorable terms for the identification of leakage condition by using image recognition technology.

Convolutional neural network (CNN) is one of the most effective methods for image recognition. CNN originates from the proposal of BP algorithm. In 1989, LeCun used it in multilayer neural network (LeCun et al., 1989). In 1998, LeCun proposed LeNet-5 model (LeCun et al., 1998) to complete the prototype of neural network. In the following ten years, the related research of CNN tends to stagnate. There are two reasons: the first one is that researchers realized that calculation of multilayer neural network in training BP is too large to complete by the limited computing ability of hardware; the second one is that the shallow machine learning algorithms like support vector machine (SVM) emerge gradually. From 2006, the CNN achieves rapid development again (Hinton and Salakhutdinov, 2006). In 2012, CNN won the ImageNet Championship, so the superior performance of ReLU activation functions in the network attracts people's attention (Krizhevsky and Sutskever, 2012). In 2014, Google developed a 20-layer VGGNet model. In the same year, the DeepFace and DeepID models were proposed, and the correct rate of face recognition and face authentication

on the LFW database was enhanced to 99.75 % (Sun et al., 2014; Sun et al., 2015). In 2015, LeCun, Bengio, and Hinton, who are famous in the field of deep learning, published a review of Deep Learning on "Nature" (LeCun et al., 2015). In 2016 AlphaGo defeated Li Shishi, and in 2017 AlphaGo defeated Ke Jie (Silver et al., 2016).

Due to the outstanding performance of CNN in image feature extraction, this paper applies CNN to identify chemical process leak. We used it to identify leakage in the chemical processes and obtained superior classification accuracy by adjusting its parameters.

2. The architecture of the CNN

CNN is a kind of multilayer neural network specially designed for processing two-dimensional data. CNN is considered to be the first truly successful deep learning method based on a multi-tiered hierarchical network. The close connection and spatial information between layers in CNN make it especially suitable for image processing and understanding. So it can automatically extract rich correlation characteristics from images.

CNN consists of convolutional layer, active layer, pooling layer, fully connected layer, and input-output layer. It owns structural characteristics like local connection, weight sharing, temporal or spatial sub-sampling, which make it superior in image recognition.

In this work, a CNN with 15 layers (except the input layer) is established, including 3 convolution layers, 4 activation layers, 3 pooling layers, 2 local response normalization layers, 2 fully connected layers, 1 dropout layer, and 1 output layer. The architecture of the CNN is shown in Fig. 1.

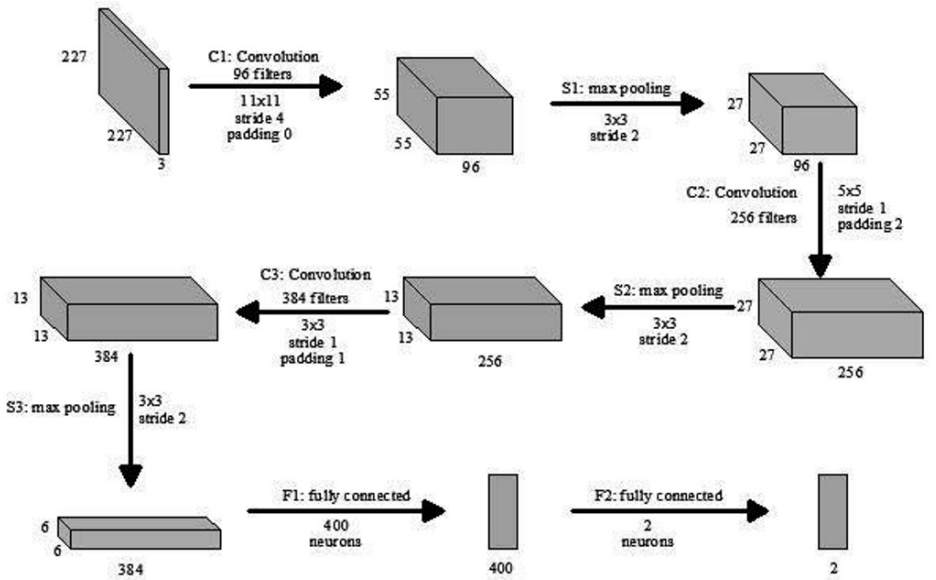


Figure 1. The architecture of the CNN

2.1. The input, output layer and fully connected layer

All the images are normalized into 227×227 pixels. Because the images are RGB images (3 channels), there are 227×227×3 inputs.

In Fig. 1, F1 layer is a fully connected one. It has 400 neurons, which are fully connected with S3 layer.

F2 layer is the output layer, also a fully connected layer. Because there are 2 kinds of images in total, the number of the output neurons amounts to 2. These 2 neurons are used to connect the 400 neurons in F1. We can then get 2 float typed prediction probability through Gauss filter.

2.2. Activation layer

There are four activation layers after C1, C2, C3 and F1, respectively. The activation layer injects nonlinear factors into the neural network to make the data distinguishable under nonlinear conditions (Nair and Hintion, 2010). In order to optimize the training result, we compared the effect of different activation functions on the classification accuracy.

2.3. Pooling layer

There are three pooling layers after the first three activation layers, as shown in Fig. 1. The first one is S1 layer. Its kernel size is 3×3， stride is 2. We get 96 feature maps with a size of 27×27. The second one is S2 layer, which kernel size is 3×3 and stride is 2. We can then get 256 feature maps with a size of 13×13. The third one is S3 layer, which kernel size is 3×3 and stride is 2. We can get 384 feature maps with a size of 6×6. In order to optimize the training result, the effect of different pooling methods on the classification accuracy should be compared.

2.4. Local response normalization layer and dropout layer

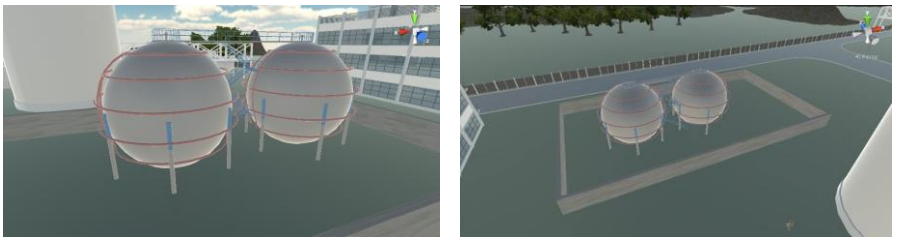
There are two local response normalization layers after S1 and S2, respectively. The role of the local response normalization layer lies in improving the recognition rate and increasing the generalization ability through smooth processing. The dropout layer is added before F2 to prevent the occurrence of over fitting.

3. Experimental results

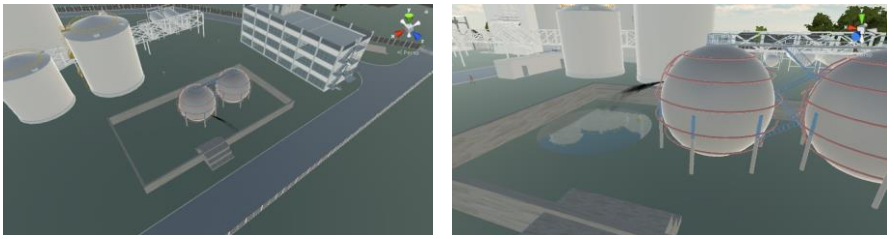
This paper mainly simulates the occurrence of a leakage accident caused by a crack in a liquefied hydrocarbon spherical tank. In practice, the microcrack source in the spherical tank gradually develops into macroscopic crack or penetrating crack, resulting in low-stress brittle fracture of the spherical tank. The leaked liquefied hydrocarbon absorbs the heat around it and vaporizes rapidly. It is easy to be ignited and leads to overpressure explosions at last.

3.1. Dataset

The scene picture of material leakage in chemical process is not easy to obtain, so this paper draws the scene picture in the chemical production process by using a 3D drawing software. We take the image interception from different angles, as shown in Fig. 2. A total of 1,477 pictures were intercepted with 1,190 (595 for normal working conditions and 595 for leakage conditions) as a training set and 287 (141 for normal working conditions and 146 for leakage conditions) as a test set. Then we normalize all the images to 227×227 pixels to enter the input layer.



(a) Normal working conditions



(b) Leakage conditions

Fig. 2 Schematic diagram of normal and leakage conditions

Table 1. The accuracy of the maximum pool method and the mean pool method

The number of iterations	MAX	AVE	The number of iterations	MAX	AVE
100	0.7561	0.7073	1100	0.9268	0.7456
200	0.7666	0.7282	1200	0.9373	0.777
300	0.8014	0.7456	1300	0.9477	0.7561
400	0.8014	0.7317	1400	0.9338	0.7666
500	0.8118	0.7596	1500	0.9477	0.7491
600	0.8676	0.7143	1600	0.9547	0.77
700	0.8293	0.7352	1700	0.9512	0.7735
800	0.885	0.7491	1800	0.9512++	0.7526
900	0.9303	0.7666	1900	0.9547	0.777
1000	0.9443	0.7526	2000	0.9582	0.7909

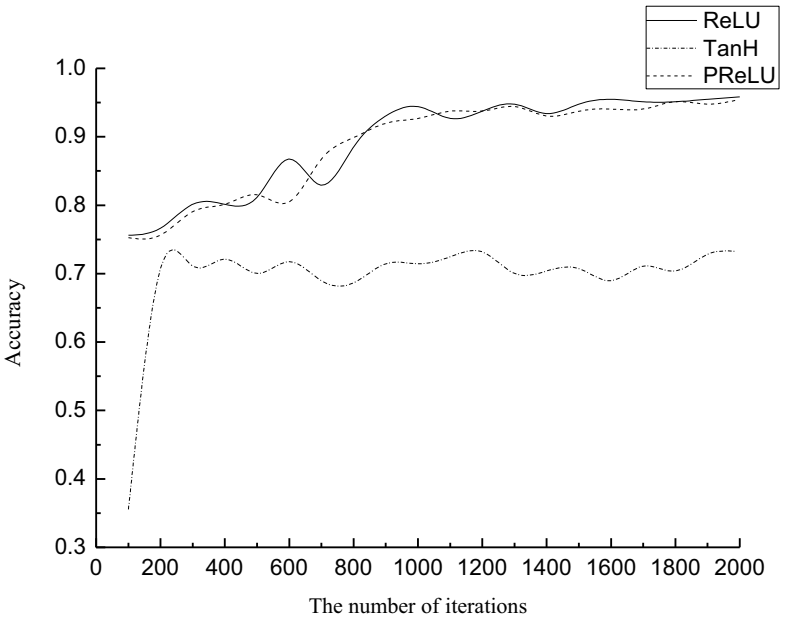


Fig. 3 Comparison between ReLU, PReLU and TanH

3.2. Network optimization

In order to optimize the training result, the effect of different pooling methods and activation functions on the classification accuracy are compared. As shown in Table 1, when the iteration number goes to 2,000 times, the accuracy of the maximum pooling method has reached 95.82 %, and the mean pooling method is only 79.09%. As can be seen from Fig. 3, the final accuracy of ReLU reached 95.82 %, PReLU reached 95.48 %, while TanH is only 73.17 %. With the increase of iteration times, the convergence rate of PReLU and ReLU is almost the same. Consequently, the maximum pooling method and ReLU activation function could make the classification accuracy higher.

4. Conclusions

This paper applies convolution neural network to image recognition of chemical process leakage. A 15 level convolution neural network is constructed. The influence of different pooling methods and activation functions on classification results is studied. The maximum pooling method and ReLU activation function was identified to make the classification accuracy higher. However, the image in this paper obtained by the 3D drawing software, is some different from the locale scene and needs to be further improved.

Acknowledgement

Financial support for carrying out this work was provided by the National Natural Science Foundation of China (Grant No. 21576143).

References

Y LeCun, B Boser, J S Denker, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural computation, 1989, 1(4): 541-551.

Y LeCun, L Bottou, Y Bengio, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

Y Sun, X Wang, X Tang. Deep Learning Face Representation by Joint Identification-Verification[J]. Advances in Neural Information Processing Systems, 2014, 27:1988-1996.

Y Sun, X Wang, X Tang. Deeply learned face representations are sparse, selective, and robust[C]// Computer Vision and Pattern Recognition. IEEE, 2015:2892-2900.

Y Lecun, Y Bengio, G Hinton. Deep learning[J]. Nature, 2015, 521(7553):436-444.

V Nair, G E Hinton. Rectified linear units improve restricted boltzmann machines[C]// International Conference on International Conference on Machine Learning. Omnipress, 2010:807-814.

G E Hinton, R R Salakhutdinov. Reducing the dimensionality of data with neural networks[J]. science, 2006, 313(5786): 504-507.

D Silver, A Huang, C J Maddison, et al. Mastering the game of Go with deep neural networks and tree search[J]. Nature, 2016, 529(7587):484.

A Krizhevsky, I Sutskever, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the Acm, 2012, 60(2):2012.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50384-0

An application of STAMP to safety and cyber security for ICS

Shun Kondo [a*], Hiroto Sakashita [a], Souta Sato [a], Takashi Hamaguchi [a], Yoshihiro Hashimoto [a]

[a]*Graduate School of Engineering, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya-shi, 466-8555, Japan*
s.kondo910@nitech.jp

Abstract

Cyber security is currently a serious problem for not only office automation (OA) systems but also industrial control systems (ICSs). If ICSs are cyber-attacked, serious accidents such as explosions and leakage of harmful substances may occur. Therefore, cyber-security is a very important factor in discussing safety, and should not be considered separately. Risk analyses have been conducted in plant design and operation stage so as to avoid the risk to the operation of ICSs. However, since conventional analysis methods focus on the "equipment" and "products" of a plant, it is difficult to consider the upper layer system. In this paper, we suggest a risk assessment for safety that it is possible to link the physical processes and ICS. We propose a risk analysis method using the Systems Theoretic Accident Model and Processes (STAMP).

Keywords: Risk Analysis, Safety, Cyber Security, STAMP

1. Introduction

Industrial control systems (ICSs) are used for various kinds of social infrastructure, and they play important roles in realizing their control functions and ensuring their safety. ICSs currently use open architectures and they are often connected to external systems such as office systems. Most communication protocols for ICSs are designed without consideration of cyber security. By the appearance of Stuxnet in 2010, physical attacks due to cyber-attacks have become a reality.

Cyber-attacks against ICSs can directly affect physical objects, unlike attacks directed solely against information systems. Therefore, they have a high risk of causing accidents that are directly connected to disasters and that affect human lives. The accidents may occur not only in processes controlled by ICSs but may stop the services provided by them and also supply products that do not meet product requirements. For example, cyber-attacks against chemical plants can cause serious accidents such as explosions. Furthermore, accidents such as a loss of opportunity due to the interruption of the production and shipment of products containing harmful substances may happen. Therefore, to prevent such accidents, strengthening security against cyber-attacks that affect safety is an important issue. Therefore, safety and cyber-security must not be considered separately. Also, cyber-security must be considered based on countermeasure to prevent hazardous situations. To deal with such problem, risk analysis for safety must be expanded for enterprise systems.

2. Risk Analysis for ICSs

Problems that may occur from cyber-attacks against ICSs require consideration and countermeasures; these problems are "Inappropriate settings are given because of the rewriting of target values in automatic control systems," "Information related to automatic control and monitoring is tampered with and concealed by cyber-attackers," "Console operators and field operators cannot take appropriate actions because of falsification and concealment," and "Actions taken to remedy problems inadvertently cause more accidents."

Such problems may be caused by cyber-attacks not only at the site level where operators and field workers are but also at the upper layers of ICSs such as enterprise resource planning (ERP) and manufacturing execution systems (MESs). These problems may also affect other companies and regions through products and services.

In order not to cause the previously mentioned problems, the controlled process must also be protected from cyber-attacks. It is important to improve security by introducing security tools as countermeasures against cyber-attacks. However, we must consider the existence of numerous cyber-attack methods and zero-day vulnerabilities. Even if security tools are introduced, it is impossible to deny that the controlled processes are affected by expected attacks and that the countermeasures for security don't work well.

Therefore, on the premise that ICSs are vulnerable to cyber-attacks, maintaining system functions and business continuity for the whole structure of ICSs including their operation and management (which we call "whole ICS" in this paper) is necessary. It is also important to construct a whole ICS that can stop safely even when production is stopped for any reason. In order to construct such a structure, risk assessment for safety and security of the whole ICS is necessary.

3. Risk Analysis using STAMP

From the design stage to the operation stage in the plant life cycle, risk analysis methods such as hazard and operability (HAZOP) study and fault tree analysis (FTA) are used to analyze the latent risks in the plant. However, these risk analysis methods focus on "products" and "equipment" in the plants, it is difficult to expand the analysis targets to human systems and organizations with these methods. For example, in HAZOP study, due to characteristic that many experts analyze over a long period of time using P&ID, the discussion of countermeasures often ends with "Sound an alarm when detecting an abnormality." However, such countermeasures are scarce in terms of business continuity such as what is an alarm indicating what to do and who to respond. Therefore, it is insufficient to discuss business continuity when ICS is cyber-attacked.

To solve the problems described in Section 2, we consider the Systems Theoretic Accident Model and Processes (STAMP), which is the accident model proposed by Nancy G. Leveson et al., is useful for the purpose [1]. In STAMP, a system has components called a controller and its target process (controlled process). An instruction (control action) is sent from the controller to the controlled process, and reports (feedback data) from the controlled process to the controller are transmitted; thereby, a control loop is formed by the relationship. The control structure of a contemporary system whose control loop has accumulated many layers is described in the diagram. Furthermore, it is said that human systems and organizations can also be analyzed by dividing them into controllers and controlled processes.

When expressing whole ICS using STAMP, factories and plants are located at the lowest level of the control structure diagram. The control system and the production planning system are located at the higher level; furthermore, the management layer and regulations for the company are located at the higher level. With such a description, it is possible to hierarchically represent the mechanism of the entire system. In the conventional risk analysis methods, accidents are assumed to occur because of simple component problems such as a component failure or a single malfunction. On the other hand, STAMP extends the conventional analysis methods in which accidents can occur even if the behavior of components and mutual interference between components violate the safety constraints of the system.

However, detailed description of the component is not set in the original STAMP. Therefore, it is difficult to express clearly the concrete control structure of ICSs and relationship between physical processes, such as process flow. Also, it is difficult to make concrete discussions when considering the scenario leading up to the accident and measures to prevent it. We propose to express control structure by assigning "Possible Actions," "Accessible Information", "Prepared Information" as a role to each component. Possible Actions means a control actions that the component can do. Possible Actions is based on Accessible Information requested from other components and Prepared Information keeping itself. This proposal makes it possible to track interactions between components.

In this paper, we apply STAMP in order to improve the business continuity ability of the whole system including many hierarchies from the control site to management. We propose the following procedure for risk analysis using the extended STAMP which is previously mentioned.

Step 1: Identify risks that can occur in the control system to be analyzed

Step 2: Express the target control system using STAMP

Step 3: Assign "Possible Actions," "Accessible Information," "Prepared Information" as role to each component

Step 4: Derivation of scenarios leading up to accident and planning countermeasures

In Step 1, the analysts identify the risk of causing undesired events in the ICS to be analyzed. The result of the conventional risk analyses may be available. In Step 2, the structure of the entire system including management and operation of the control system to be analyzed is represented by the STAMP. Subsequently, the role is described in the diagram and the control structure is expressed in more detail. In Step 4, discuss scenarios leading to accidents from detailed diagrams and consider countermeasures. After that, the analyst adds countermeasures to the diagram.

We set the analysis target as a batch plant on the premise of flexible operation and organize the analysis procedure in the next section.

4. Case Study with Batch Plant

Even if the batch plant to be analyzed is connected by piping, it is possible to manufacture multiple products at the same time by using different tanks and transfer paths. However, by performing such flexible operation, a risk of contamination due to the mixing of different raw materials for manufacturing into tanks and transfer paths that are under different manufacturing process may happen. Also, contamination occurs even if the

material passes through tanks and transfer paths that are insufficiently cleaned. These contaminations not only cause problems for the quality of manufactured products, but also generate hazardous substances such as toxic gases because of mixing. In addition, the multipurpose batch plant is also required to realize new production by mastering general purpose equipment. On the basis of the above preconditions, we proceed with the procedure proposed in Section 3.

Step 1: Identify risks that can occur in the control system to be analyzed

In this step, analysts identify potential risks. For example, the risk of contamination can be considered as follows.

A) Contamination due to insufficient cleaning inside of tanks and piping

B) Contamination due to incomplete blocking between processes

Step 2: Express the target control system using STAMP

In Step 2, analysts represent the batch plant and the control structure of the whole system that uses it in the STAMP. In this paper, with reference to the S95 standard [2] [3], we created a diagram of the whole system of a batch plant including management (as shown in Figure 1).

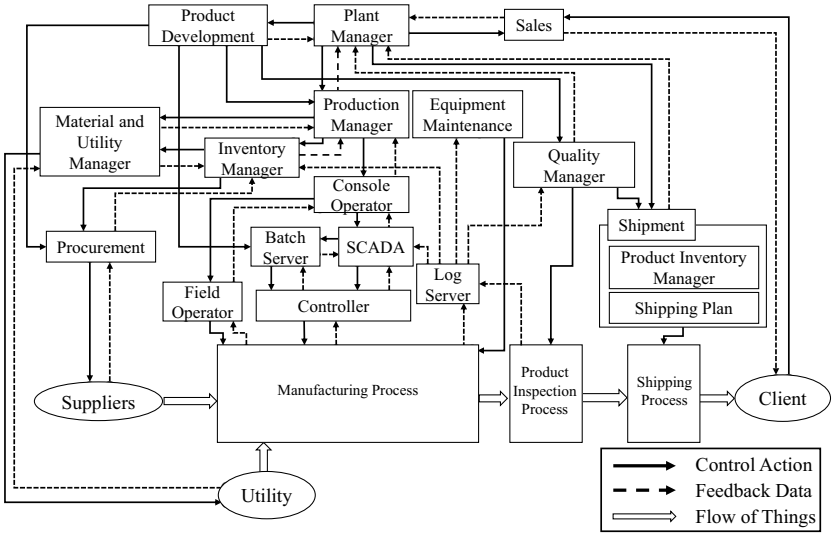


Figure 1: Control Structure Diagram of Batch Plant

Step 3: Assign "Possible Actions," "Accessible Information," "Prepared Information" as role to each component

In this paper, we focus on the range shown in Figure 3 among the control structure and discuss it. However, we thought that a more detailed diagram is necessary for analyzing a batch plant controlled based on recipe, and described the components of the Controller and the Batch Server in more detail (as shown in Figure 2).

Step 4: Derivation of scenarios leading up to accident and planning countermeasures

With reference to Figure 2, we explain the derivation of the scenarios leading up to the

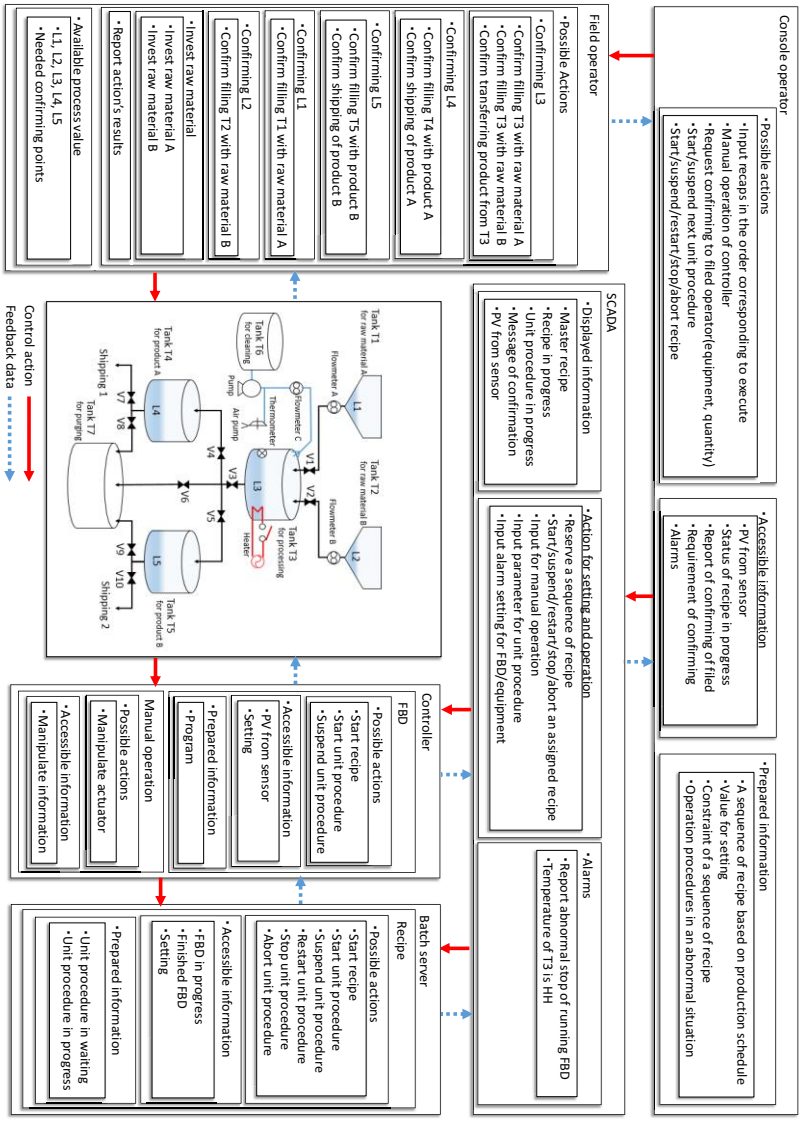


Figure 2: Control Structure Diagram Assigning Roles to Components

accident and countermeasures using “A) Contamination due to insufficient cleaning inside of tanks and piping” as an example. If there is an inadequacy in the execution of the recipe to be cleaned, the Controller recognizes it and the report it to the Batch Server. The Batch Server compares the report with “Unit procedure in progress” as Prepared Information, and reports it to SCADA. SCADA, which is the monitoring screen, alarms the Console Operator for inaccuracies. The Console Operator can execute “Manual Operation of controller” as Possible Action by comparing “alarm” and “recipe execution state” as Accessible Information, and furthermore, “emergency response operation procedure” which is Prepared Information. Therefore, from Figure 2 and the scenario, it can be interpreted that an accident will occur if any one of these interactions fails. In this way, by implementing the suggested expression of roles, it is possible to discuss the

scenarios that the operation for safety is performed and the ones leading to accidents. After that, countermeasures are planned from the scenario. In this case, it is effective to monitor the behavior by comparing the execution status of the cleaning recipe with the sensor's value, in consideration of the situation where the cleaning incomplete alarm does not function properly. This countermeasure builds a system that console operators can notice abnormality and respond.

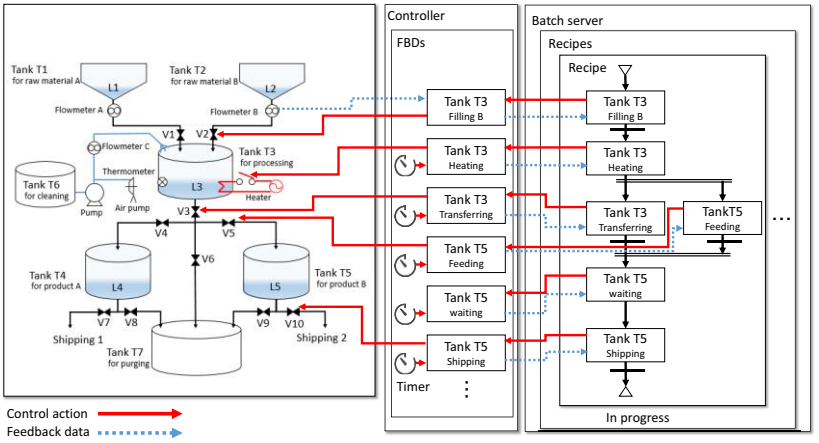


Figure 3: Control Structure Diagram in More Detail

5. Conclusion

In this paper, we suggest that risk assessment for safety and business continuity is important on the premise that ICS systems are threatened by and vulnerable to cyber-attacks. It is necessary to extend the conventional risk analysis methods so that upper layers of management and operation can be analysed. We proposed a procedure for creating a diagram and discussing it by using the STAMP. Also, we are expecting that it is possible to discuss cyber-security by extending this method. If we use the diagram where physical process and ICS are linked like this, we should be able to discuss appropriate network division against cyber-attacks [4].

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 25282101.

References

[1] Leveson, N. Engineering a Safer World: Systems thinking applied to safety. Cambridge, MA, MIT Press, 2012.

[2] IEC62264-1(2003) Enterprise-control system integration – Part1 : Models and technolog

[3] Kluge, A. The acquisition of knowledge and skills for taskwork and teamwork to control complex technical systems: A cognitive and macroergonomics perspective. Dordrecht, Springer, 2014.

[4] Y. Hashimoto, T. Toyoshima, S. Yogo, M. Koike, T. Hamaguchi, S. Jing, I. Koshijima, Safety Securing Approach against Cyber-Attacks for Process Control System, Computers and Chemical Engineering, 57, pp.181-186, (2013)

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50385-2

Parameters determination of Convergent Cross Mapping for causal analysis of chemical variables

Si Meng Li, Tong Qiu*

Department of Chemical Engineering, Tsinghua University, 100084 Beijing, China
** qiutong@tsinghua.edu.cn*

Abstract

Causal analysis of chemical variables is important for safe and efficient operation of chemical processes. Convergent cross mapping (CCM) proposed recently is suitable for nonlinear systems and can calculate the time delay and causal relationship between variables accurately. However, CCM has not been directly applied for causal analysis of chemical variables, because it does not have an algorithm for calculating phase space reconstruction parameters. In this paper, the optimal parameter algorithms are selected to improve CCM, greatly expanding CCM for analyzing causal relationship of chemical variables. Finally, using Tennessee Eastman model, it's proved that CCM with optimized parameters can basically get the causal relationship network more efficient than the original CCM and transfer entropy method.

Keywords: Convergent Cross Mapping; Phase space reconstruction; Takens theory; Embedded dimension; Time delay interval

1. Introduction

Causality inference and root cause analysis are important for fault diagnosis in the chemical industry. Many methods have been discussed to resolve causal analysis between variables, such as Directed Transfer Function (DTF), Bayesian Nets (BN), Transfer Entropy, Granger Test and etc.-DTF proposed in 1991 only applies to linear systems. BN has high stability requirements for the data at the same time, limiting the scope of application in engineering. Granger Test requires a linear relationship between variables. Transfer Entropy relative to these algorithms can be used for nonlinear system data analysis, but it is computationally expensive, inefficient and highly dependent on the probability distribution density of data (Luo L et al., 2017). Chemical processes are non-linear, multivariate processes, so above methods are not suitable enough.

Convergent cross mapping (CCM) (G. Sugihara et al., 2012) introduced by Sugihara et al. is a technique for computing specific kinds of correlations between sets of time series and is reported to be "a necessary condition for causation" capable of distinguishing causality from standard correlation, intended to be useful in situations where Granger causality is known to be invalid (Luo L et al., 2017). However, it cannot be directly applied to chemical processes. Causal relationship between chemical variables is mostly non-linear because chemical process system is a multi-dimensional and nonlinear causal system. And data sampling intervals of different chemical plants are different, resulting in the diversity of chemical data. These two characteristics limit CCM applied to the chemical industry variables.

The main principle of CCM algorithm is to reconstruct the phase space of the time series according to Takens reconstruction manifold affine theorem, and then make a predictive

analysis of the causal relationship between the time series. In this paper, it'll be shown that the selection of phase space reconstruction parameters which are embedding dimension-E and time delay interval-τ has a great influence on the prediction effect of CCM, and it's important to specify the method for determining these parameters.

2. The reconstruction parameters of CCM

Takens reconstruction theorem is a delay reconstruction (Takens F, 1981). The principle of time delay reconstruction is to project each column $X(t), Y(t)$ in the original manifold $M = [X(t), Y(t)]$ to a one-dimensional space, and then take out the individual element at τ intervals and form the E-dimensional matrix to reconstruct the shadow manifold. E and τ mentioned here are two important parameters. The expression for reconstructing the vector is as follows:

$$M_x = [X(t), X(t+\tau), X(t+2\tau), \cdots, X(t+(E-1)\tau)] \tag{1}$$

$$M_y = [Y(t), Y(t+\tau), Y(t+2\tau), \cdots, Y(t+(E-1)\tau)] \tag{2}$$

Taking the Lorenz equation as an example, take the parameters $b = \frac{8}{3}, r = 28, \sigma = 10$,

$$\frac{dx}{dt} = -\sigma(x-y) \quad \frac{dy}{dt} = rx - y - xz \quad \frac{dz}{dt} = xy - bz \tag{3}$$

The following figure is the original manifold and shadow manifold. It's shown that the original manifold and shadow manifolds are diffeomorphisms. There is a continuous injective relationship between the shadow manifolds M_x and M_y. If all $x_{t,i}(i = 1 \cdots k)$ converge on a particular $x_{t,0}$, then $y_{t,i}(i = 1 \cdots k)$ also converge on the corresponding $y_{t,o}$.

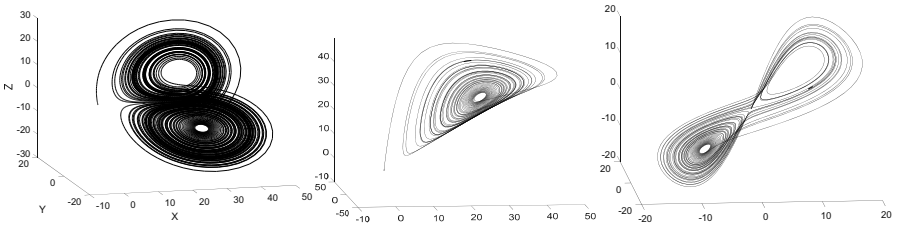


Figure 1. The original manifold and shadow manifold

CCM predicts causality by measuring points on the shadow manifold (G. Sugihara et al., 2012). So the effect of reconstruction phase space directly affects the accuracy and computational efficiency of CCM calculation results. And the effect of reconstructing phase space is related to the parameters E and τ.

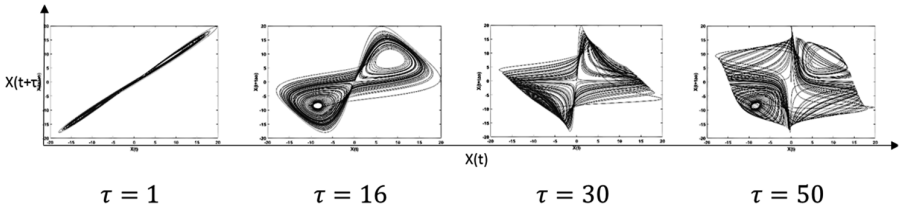


Figure 2. Reconstruction manifold changes with τ

Fig. 2 shows the influence of time delay interval τ on the manifold. As the value of τ increases, the change of the attractor of the reconstructed manifold is as shown. Whether the value of τ is too large or too small, it will make the reconstruction of manifold can't reflect the mutual information of original system well.

As for embedded dimension E, it will increase the amount of computation if the value of E is too much. And it will reduce calculation accuracy if the value of E is too small. As mentioned earlier, the chemical data is multi-dimensional, the data interval varies with the sampling interval. But in the original CCM algorithm, the values of the parameters are fixed and do not change with the input data, which are not always the best for each pair of data. The work of this paper is to propose a method to calculate the optimal parameter E and τ.

3. Algorithm to determine the parameters

In this paper, algorithm to determine the parameters of CCM is improved to make them more suitable for calculating chemical variables.

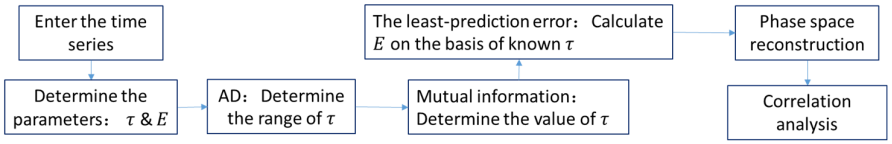


Figure 3. Flow chart of algorithm to determine the parameters

3.1. Average Distance: determine the range of τ

In our approach, Average Distance (AD) is mainly used as the initial treatment of time series (Harman E A et al., 1989). The basic principle of choosing τ is that the original sample sequence $X(t)$ and its delay sequence $X(t+\tau)$ are independent but not irrelevant completely, so that $X(t)$ and $X(t+\tau)$ can be treated as independent coordinates in phase space reconstruction. By measuring the degree to which the trajectory of the reconstruction state space opens from the main diagonal of the state space, quantitative reduction of the redundancy error as τ increases is quantitatively described, ie, the change of the average displacement

$$S(E,\tau) = \frac{1}{N}\sum_{E,n=1}^{N_E}\sqrt{\sum_{l=1}^{E-1}\left(x(n+l\tau)-x(n)\right)^2} \tag{4}$$

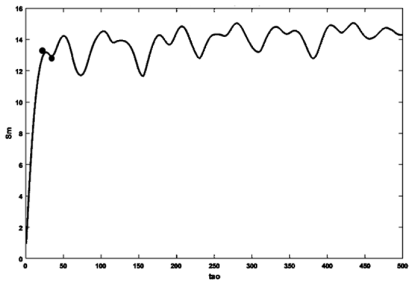


Figure 4. The average displacement curve.

Take the first extremal interval where the slope of $S(E,\tau)$ changing with τ as the feasible interval of delay time, that is, the τ value is in this interval. Fig. 4 is the average

displacement curve made using Lorenz system data, which is the feasible interval between two circles.

3.2. Mutual Information: Determine the value of τ

The related entropy between two random variables is called mutual information (Fraser A M, 1986), which is a measure of the randomness of Y when X is known, or a measure of the randomness of X when Y is known. Mutual information function:

$$I(X,Y) = H(X) + H(Y) - H(X,Y) \tag{5}$$

After reconstructing the phase space, the mutual information curve with the change of τ between the adjacent vectors is made. When the curve is reduced to the minimum for the first time, τ is the best result.

We only need to calculate the mutual information corresponding to the value of τ in the extremal interval determined by AD method, so as to improve the calculation computational efficiency and accuracy.

3.3. Least-prediction Error : Calculate E on the basis of known τ

The algorithm that is chosen to determine E is the Least-prediction Error method (Cao L, 1997), which is selected by comparing the results of various algorithms in our study (Simeng Li, 2017). The principle of the algorithm is to avoid the intersection of the orbit of the phase space and itself when reconstructing. The algorithm is more computationally efficient than the other methods [6]. It can be used to calculated E after knowing time delay interval τ, the specific calculation process reference literature (Cao L, 1997).

The parameters of the reconstructed phase space have all been determined, and phase space reconstruction and correlation analysis can be performed. In this paper, the CCM using the parameters calculated by above algorithm to reconstruct the phase space is called the improved CCM.

4. Example systems

This paper uses Tennessee Eastman (TE) model to verify the improved CCM. TE simulation process characterizes the condition by 41 measured variables, detailed description of the process and model establishment can be found in Tian et al (1993).

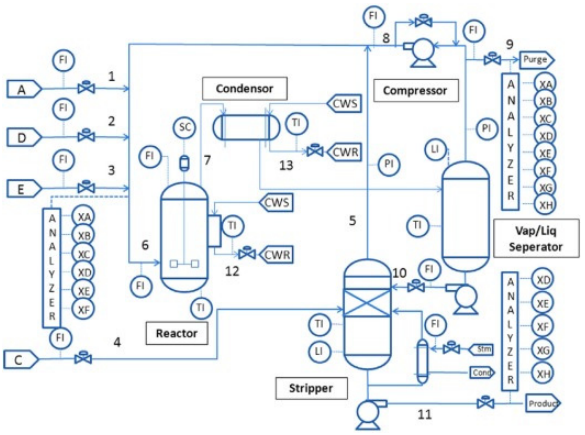


Figure 5. Flowchart of the Tennessee Eastman process

Measurements 1 to 22, which are process variables such as flowrates, the temperature, the pressure, and the liquid level, are considered in this case. In the steady state, a normal-distribution random number is added to the steady-state data to simulate the noise in the actual working condition. The figure below is the outputs of the data in steady state, each variable output by the TE model has its own weak fluctuations because of the added noise. The existence of these fluctuations makes the causal relationship between each other can be detected by the improved CCM algorithm. The algorithm can obtain the causality between variables in TE model according to the steady-state data with noise, which shows the sensitivity of the improved CCM algorithm.

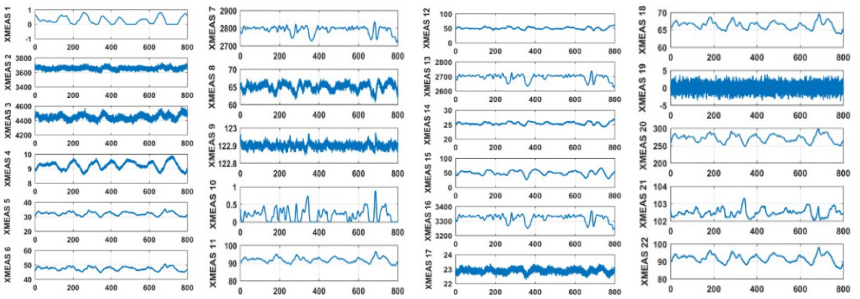


Figure 6. Outputs of the Tennessee Eastman process

We performed a correlation test on steady-state data and compared the results with those of the original CCM program. The data named causality in the table show the existence and strength of causality between variable A and variable B. The causality shows the correlation calculation results, the larger the value of causality, the stronger the correlation. And the causality is over 0.5 indicating that there is a correlation between this two variables (G. Sugihara et al, 2012). As for results of the improved CCM, there are 231 results of the calculation, of which 189 are over 0.5, indicating there are causal relations in 189 pairs of data.

Table 1. Causal relations calculated in TE process

Variable $\boldsymbol{A}$	Variable $\boldsymbol{B}$	$\boldsymbol{E}$	$\boldsymbol{\tau}$	Causality $\boldsymbol{A \rightarrow B}$	Causality $\boldsymbol{B \rightarrow A}$	Causality (the original CCM)
A feed (stream 1)	Reactor feed rate (steam 6)	6	13	0.875	0.349	0.021
D feed (stream 2)	Reactor feed rate (steam 6)	6	13	0.819	0.425	0.037
E feed (stream 3)	Reactor feed rate (steam 6)	6	21	0.833	0.695	0.101
Reactor feed rate (steam 6)	Reactor feed rate (steam 6)	6	25	0.546	0.255	0.002
Reactor feed rate (steam 6)	Reactor level	5	13	0.734	0.352	0.009
Reactor temperature	Reactor cooling water outlet temperature	5	20	0.632	0.655	0.013
Product separator under flow (stream 10)	Stripper level	6	19	0.726	0.407	0.243
Stripper underflow (stream 11)	Stripper level	5	23	0.652	0.294	0.044
Stripper steam flow	Stripper temperature	6	30	0.622	0.388	0.000

Table 1 are part of results calculated by the improved CCM and the original CCM. The original CCM takes $E = 3, \tau = 1$. Shu (2013) points out that the causal relations list in

table 1 should exist according to inference but have not been captured by Transfer Entropy. The method proposed in this paper successfully solves this problem. All of these relationships can be calculated by the improved CCM algorithm from Table 1.

And compared with the calculation results of the original program, it can be seen that the causality calculated by the original CCM program is nearly 0, which is that the correlation is not detected at all. It indicates the improvement of the original CCM in this paper is successful and meaningful.

The improved CCM algorithm in this paper is more suitable for the chemical system than both the original CCM program and the transfer entropy algorithm. What's more, it is worth mentioning that the computational efficiency of the original CCM program is roughly 6 times that of the improved program.

5. Conclusions

As a correlation detection algorithm, CCM has been widely used in many fields. Chemical process is a time-lag and non-linear system and can't directly use the CCM program. The innovation of this paper lies in the combination of AD algorithm, mutual information and the least-prediction error to calculate the parameters of reconstructed phase space. The CCM algorithm is optimized to improve the accuracy and efficiency of the calculation, so that the correlation between chemical variables can be measured. Finally, an example of TE process verifies the effectiveness of this method, which shows that this method can calculate the correlations of steady-state data which can't calculate by the original CCM and Transfer entropy.

References

Luo L, Cheng F, Qiu T, et al. Refined convergent cross-mapping for disturbance propagation analysis of chemical processes[J]. Computers & Chemical Engineering, 2017, 106:1-16.

G. Sugihara et al., "Detecting Causality in Complex Ecosystems," Science (80-.)., vol. 338, no. 6106, pp. 496–500, 2012.\

Takens F.Detecting strange attractor in turbulence.in Dynamical systems and turbulence, Warwich, 1980, Lecture Notes in Mathematics [C] , Vol 898, 1981, Rand and Young eds:366-381

Harman E A, Rosenstein R M, Frykman P N, et al. Effects of a belt on intra-abdominal pressure during weight lifting.[J]. Medicine & Science in Sports & Exercise, 1989, 21(2):186-190.

Fraser A M, Swinney H L. Independent coordinates for strange attractors from mutual information[J]. Physical Review A General Physics, 1986, 33(2):1134.

Simeng Li, Qiu T. Application and Parameter Determination of Convergent Cross Mapping Method.[D].Tsinghua University, 2017:1-58

Cao L. Practical method for determining the minimum embedding dimension of a scalar time series[J]. Physica D-nonlinear Phenomena, 1997, 110(1-2):43-50.

Downs J J, Vogel E F. A plant-wide industrial process control problem[J]. Computers & Chemical Engineering, 1993, 17(3):245-255.

Shu Y, Zhao J. Data-driven causal inference based on a modified transfer entropy[J]. Computers & Chemical Engineering, 2013, 57(5):173-180.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50386-4

Eye Tracking as a Tool to Enhance Operator Learning in Safety Critical Domains

Punitkumar Bhavsar[a], Babji Srinivasan[a,*], Rajagopalan Srinivasan[b*]

[a]*Indian institute of Technology Gandhiniagar, Gujarat, 382355, India*
[b]*Indian institute of Technology Madras, Tamil Nadu, 600036, India*
raj@iitm.ac.in

Abstract

Human operator performance is a key factor for proper operation of safety critical industries like process control, nuclear power plant and aviation. Several incident reports identify human error as one of the causes responsible for accidents. One way to reduce human error is to impart proper operator training which requires understanding how operators learn during training. Traditionally, learning is evaluated on the basis of subjective assessments and outcomes of the task execution. It obscures the cognitive aspect of the learning process like understanding how operator gives attention to various information sources and develops the process of decision making during the training program. In this study, we use pupil size measurement from eye tracking system to study and understand learning as participants undergo repetitive trials of simulated flight operations tasks. A derived measure based on time-frequency analysis of pupil size variation is observed to be sensitive to the process of learning during task repetition. We observe improvement in the task performance accompanied by the decrease in the proposed measure.

Keywords: Eye tracking, training, learning, pupillometry

1. Introduction

Safety has been one of the prime areas of interest for industries like process control, nuclear power plant and aviation as they are aggravated persistently by large-scale accidents despite many interventions. For instance, on a statistical basis, of the 20 accidents that have led to the largest property damage losses in the hydrocarbon industry in the 40 year period from 1974 - 2013, 25% have occurred in the last 5 years from 2009 (Marsh, 2014). Flight businesses relating to the UK fixed-wing general avionics reported 1007 incidents in the years from 2005 to 2011. The analysis of which revealed 49.7% resulted from the loss of control (LOC), 69.2% of these occurring during the landing phase (Taylor et al., 2014). A detailed analysis of various incidents reveals that human error is one of the principal causes for failure to control the situation during abnormal condition.

Human operator's ability to successfully handle an abnormal scenario plays a significant role in deciding the magnitude of such industrial accidents. Hence, the decision making and the cognitive information processing skills of an operator needs to be advanced to reduce or eliminate effects of such incidents. Experts in safety critical domains suggest the need for appropriate training for operating personnel to manage the abnormal situation (Bullemer and Nimmo, 1994). It is therefore important to develop training programs with an emphasis on the cognitive abilities of the operator, and tailor

training tasks for the operators so as to impart an adequate mental model of the situation that can help in their decision-making process. Traditionally, operator training simulators (OTS) provide the trainee operator to understand and be acquainted with the simulated environment that mimics the "look and feel" of the situation. Research attention has been focused on developing comprehensive training simulators that can mimic plant behavior in different operating conditions and abnormalities (Patle et al., 2014). However, there are hardly few approaches that focus on understanding their learning process, identified as crucial to enhancing operators' skills and abilities (Bullemer and Nimmo, 1994). The conventional approach to evaluate learning is through means of subjective assessment of the imparted knowledge about the task and scenario. The process lacks understanding of how operator gives attention to various information sources and develops the process of decision making during the training program. Recent studies emphasize that use of virtual reality for training operators has been claimed to improve operator's cognitive readiness (Manca et al, 2013).

Much like control room operators in process plants, pilots in aviation industry are required to continuously monitor the state of the system, analyze the information and take necessary control actions. In both the settings, the operator/pilot processes information acquired from the HMI, assesses the situation and decides on the appropriate future actions. Similar issues of improper cognitive processing and resulting human error are also found in pilots (Miller and Parasuraman, 2007). The pilot error (also known as cockpit error), usually defined as the error caused due decision, action or inaction by the pilot is determined to be the major cause for loss of control during flight operation (Belcastro and Fostar, 2010). Sarter et al. (2007) have used eye tracking to understand shortcomings in pilot's automation monitoring strategies and corresponding performance. They confirm that monitoring failures are one major contributor to breakdowns in pilot-automation interaction. Di Nocera et al. (2000) used the distribution of fixation on HMI to evaluate pilot's mental workload during the different phase of flight. The study reveals landing phase of the flight operation requires high mental workload demand which may prone to error during these phases. Kilingaru et al. (2013) analyzed the visual scanning behavior using gaze attentional strategies for the expert and novice pilot. The study indicates the existence of the difference in visual attention behavior for novice and expert pilots. The experts (highly trained) tend to focus on the more relevant instrument with frequent and small dwell compared to the novice (less trained) failing to give proper attention to the relevant instrument. Many other studies have utilized eye tracking to understand the cognitive behavior of the pilots. However, less effort has invested to understand the learning process as the novice pilot repeatedly undergo the flight operation. In this study, we used the derived measure from time-frequency analysis of pupil size variations to understand the pilot's learning behavior with the repetition of the experimental trials. The magnitude of power in the low-frequency spectrum corresponds to workload and various affective states like stress, anxiety, fatigue which potentially affects the human operator's performance. (Yamanaka and Kawakami, 2009). The high power level in the lower spectrum (0-0.7 Hz) corresponds to higher workload and presence of affect states (Lüdtke et al.,1998). The repetition of trials reduces the workload as the participants learn and adapt to the situation present in the scenario (Foroughy C., 2017). We analyze the power level in the low frequency spectrum of pupil size variation

as the participant performs the repeated trials of landing procedures with varying level of difficulties.

2. Experimental methodology and data analysis

We conducted an experiment with two participants who voluntarily participated in the study and played the role of a pilot for the flight simulator. Participants are the graduate students who have no prior experience of the flight simulator and its operations. The simulator consists of one training task followed by actual experimental tasks. Participants can repeat the training tasks until they become ready to perform the actual experimental tasks. The experimental tasks involve take-off phase, cruise phase and landing phase with varying level of difficulties. The difficult tasks involve landing in the hilly region (task 2 and task 3) or urban region with a sheer position of runways (task 4) whereas easies task contains the landing in the coastal area with relatively less requirement of proper altitude control (task 1). Each participant has performed all the four tasks five times with eye tracking. Standard nine-point calibration is presented before the start of the experimental tasks in order to ensure the proper quality of gaze and pupil data collection. The eye gaze data and pupil data are collected at 120 Hz with the use of Tobii TX300 eye tracking system.

The pupil diameter collected as a time series is first processed to remove the blink artifacts. The consecutive invalid entries of the pupil size for more than 100ms are treated as the blinks and removed for further analysis. This is followed by the calculation of spectrogram of the valid entries of the pupil data using Matlab. Moving window of 720 samples with an overlap of 700 samples is selected to calculate the spectrogram. The values of the power levels at different frequency components in each time window are considered to calculate the area under the power spectral density. The power levels in the frequency bands between 0-0.7 Hz are added to manipulate the value of the measure for each window. The measure is then normalized with respect to its maximum and minimum values across all the tasks performed by a particular participant. The measure of normalized area under the power spectral density (APSDn| f=0-0.7 Hz) is considered to evaluate the learning in this experiment.

3. Illustrative Example

Figure 1(a) shows the spectrogram for a typical task containing take-off phase, cruise phase and the landing phase. It contains the majority of the power level within 0-0.7 Hz for all the windows. The normalized area under the power spectral density ($APSDn_{| f=0\text{-}0.7\ Hz}$) of the first participant while carrying out the first trial of task 2 is shown in figure 1(b). The maximum value of the measure (1.00) is observed to occur during the landing phase between t=100 s and t =110 s. The participant would have faced high workload demand since they were not aware of the situation present in task 2 during the first trial which has resulted in the failure. Figure 2 shows the values of the measure with time for the same participant performing the same task during fifth trial. The participant succeeds in the fifth trial of task 2 and the corresponding maximum value of the measure reduces to 0.78 between t=80 s and t=90 s. It is expected that the participant would have learned how to handle the scenario of task 2 with repetitive trials culminating in successful completion in the fifth trial accompanied by the reduction in the maximum by 22%. However, single task of success and failure are not sufficient to indicate the process of learning hence all the task performed by both the participants are analyzed in detail in the next section.

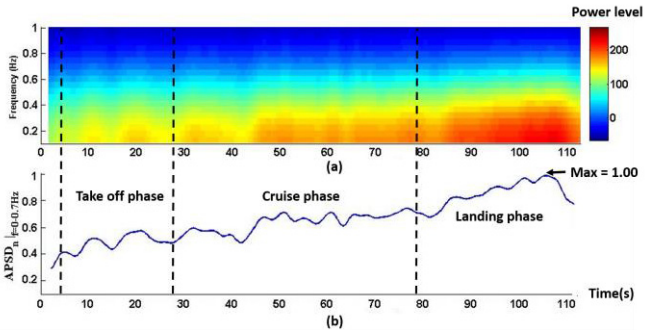


Figure 1:(a) A typical spectrogram of the pupil size variations during task execution (b) Normalized area under the power spectral density during trial 1 of task 2 for participant 1

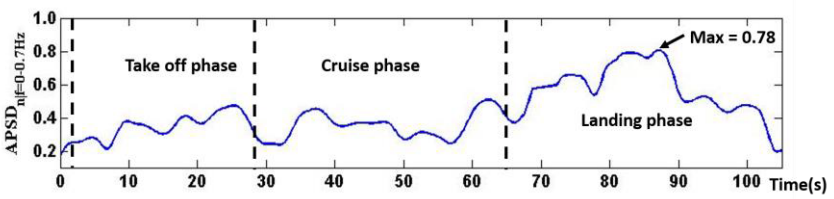


Figure 2: Normalized area under power spectral density during trial 5 of task 2 for participant 1

4. Analysis of all tasks

In this study, the landing phase is considered and the maximum value of the measure during this phase is analyzed to observe the effect of learning through task repetition. The maximum value of the measure represents the peak demand exhibited by the participant during the course of task execution. The peak demand decreases as individual learns through repetition of the same task. It leads to decrease in the maximum value of the measure during the task execution. Foroughy C. et al (2017) have observed the decrease in the maximum value of the pupil size to analyze the learning through repetition in case of target identification task with the unmanned aerial vehicle.

Figure 3(a) and figure 3(b) show the bar graph representing the maximum of the measure in each of the trials for participant 1 and participant 2 respectively. The character "S" and "F" on the top of each bar represent the success and failure of the task respectively. The improved success rate in consecutive trials of a particular task has been considered as an indication of learning. The success rate is compared with the maximum of the measure to verify the sensitivity of the measure in the evaluation of learning process.

Analysis for participant 1:

Participant 1 has not succeeded in any of the tasks during the first trial and correspondingly the maximum value of the measure is also observed to be the highest during these trials. However, task 3 does not exhibit this pattern. This participant has succeeded in the subsequent trials of task 1 and task 4 and this is reflected as a reduction in the maximum value of the measure during these tasks. The trend of the proposed measure, therefore, provides an indication of the participant's learning to handle the scenario for these two tasks. The maximum value of the measure is observed to fall from 0.85 in the first trial to 0.58 in the fifth trial of task 1. Likewise, the maximum value of the measure has reduced from 1.00 in the first trial to 0.81 in the last trial of task 4. The maximum is observed to increase for the third trial of task 2 which can be attributed to

the difficulty faced by the participant in learning and handling the scenario. The participant has succeeded in the subsequent trials which are accompanied by the fall in the maximum values. For task 3, the participant has failed in the first three trials and the highest has occurred at trial 3. Though the participant has succeeded in handling the task 3 in fourth and fifth trials, the maximum of the measure does not observe to reduce which may indicate high workload demand for him to handle the scenario of task 3. This indicates the sensitivity of the measure for the participant's ability to adapt to the scenario which may not be revealed directly through the final outcome of success or failure.

(a) (b)

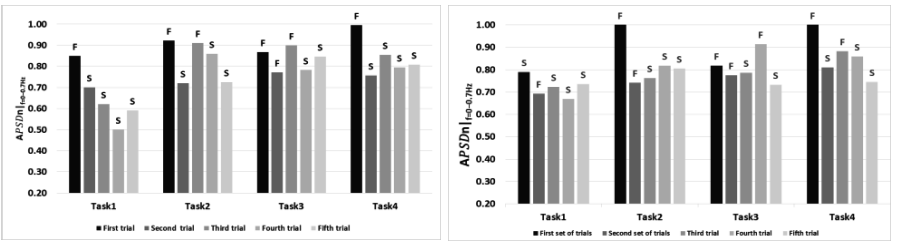


Figure 3: Maximum of the measure during each task for participant 2

Analysis for participant 2:

The maximum of the measure exhibits similar behavior for participant 2 for the first trial tasks. For task 1, the participant has succeeded in 4 out of 5 trials and the corresponding maximum value has also reduced from 0.79 in trial 1 to 0.74 in trial 5. The maximum of the measure in task 2 has also observed to be reduced from 1.00 during trial 1 to 0.80 in trial 5. Similarly, for task 4, the participant has succeeded in last two trials and the corresponding maximum value has reduced to 0.74 during trial 5 compared to 1.00 during trial 1. Like participant 1, task 3 has also more failures (3 out of 5 tasks) and correspondingly the maximums of the measure do not contain the consistency with the performance in successive trails.

Due to the small sample size, a paired t-test is performed to compare the mean difference between the first and the last trial of each task ($t=4.79$, $DOF = 7$, $p<0.01$). The decrease in the mean value (Mean $APSD_n$ during the first trials = 0.90, Mean $APSD_n$ during the last trials = 0.74) of $APSD_n$ during the last trial indicative of the effectiveness of the technique to evaluate learning.

5. Conclusion

The effectiveness of operator training is a crucial aspect of safer operation in safety critical domains. Traditional techniques to evaluate learning process involve subjective measures based on the knowledge and ability to perform the certain task successfully which obscures the cognitive aspect of the task performance. Use of eye tracking measures provide the almost real-time window of operator cognition and hence reveal the dynamics of the learning process. In this study, we analyzed the pilot's ability to learn with repetitive trial and observed the corresponding change in the derived measure of pupillometry. The maximum value of the measure is observed to decrease with repetition accompanied by the increase in the success rate. However, further investigation with relatively large population is required in order to claim the statistical significance on sensitivity and validity of the measure.

The present study can be expanded in several directions. The proposed measure can be combined with gaze position to get more insight into the process of learning. An increased value of the measure while visual attention on a specific information source

during repeated trials may indicate difficulty in comprehending the information from a particular source. The similar analysis in other safety critical fields like process control and nuclear power plant may provide insight for developing the more effective training scenarios or learning strategies for the operators.

References

Belcastro, C.M.; Foster, J.V. (2010), August. Aircraft loss-of-control accident analysis. In Proceedings of AIAA Guidance, Navigation and Control Conference, Toronto, Canada, Paper No. AIAA-2010-8004.

Bullemer, P. T.; Nimmo, I. (1994). Understanding and supporting abnormal situation management in industrial process control environments: a new approach to training. *In Systems, Man, and Cybernetics. IEEE International Conference on Humans, Information and Technology*.1, 391-396.

Di Nocera, F.; Camilli, M.; Terenzi, M. (2007). A random glance at the flight deck: Pilots' scanning strategies and the real-time assessment of mental workload. *Journal of Cognitive Engineering and Decision Making*, 1(3), 271-285.

Foroughi C.; Sibley C.; Coyne J. (2017). Pupil size as a measure of within-task learning. *Psychophysiology 54(10),1436-1443*

Kilingaru, K.; Tweedale, J.W.; Thatcher, S.; Jain, L.C. (2013). Monitoring pilot "situation awareness". *Journal of Intelligent & Fuzzy Systems*, 24(3), pp.457-466.

Lüdtke, H.; Wilhelm, B.; Adler, M.; Schaeffel, F.; Wilhelm, H. (1998). Mathematical procedures in data recording and processing of pupillary fatigue waves. *Vision research*, 38(19), 2889-2896.

Marsh (2014). The 100 largest losses 1974-2013. Marsh and McLennan Companies.

Manca, D.; Brambilla, S.; Colombo, S. (2013). Bridging between virtual reality and accident simulation for training of process-industry operators. *Advances in Engineering Software*, 55, 1-9.

Miller, C.A.; Parasuraman, R. (2007). Designing for flexible interaction between humans and automation: Delegation interfaces for supervisory control. Human factors, 49(1), 57-75.

Research attention has been focused on developing comprehensive training simulators that can mimic plant behavior in different operating conditions and abnormalities (Patle et al., 2014).

Sarter, N.B.; Mumaw, R.J.; Wickens, C.D. (2007). Pilots' monitoring strategies and performance on automated flight decks: An empirical study combining behavioral and eye-tracking data. *Human factors*, 49(3), 347-357.

Sibley, C.; Coyne, J.; Baldwin, C. (2011). Pupil dilation as an index of learning. *In Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 55 (1), 237-241. Sage CA: Los Angeles, CA: SAGE Publications.

Taylor, A.; Dixon-Hardy, D.W.; Wright, S.J. (2014). Simulation training in UK general aviation: An undervalued aid to reducing loss of control accidents. *The International Journal of Aviation Psychology*, 24(2), 141-152.

Yamanaka, K.; Kawakami, M. (2009). Convenient evaluation of mental stress with pupil diameter. *International journal of occupational safety and ergonomics*, 15(4), 447-450.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50387-6

Safety, sustainability and economic assessment in conceptual design stages for chemical processes

Karen de Jesús Guillén-Cuevas[a], Ecem Ozinan[b], Andrea P. Ortiz-Espinoza[a], Nikolaos K. Kazantzis[c], Mahmoud M. El-Halwagi[b], Arturo Jiménez-Gutiérrez[a]*

[a]*Departamento de Ingeniería Química, Instituto Tecnológico de Celaya, Celaya, Gto., 38010, México*
[b]*Artie McFerrin Chemical Engineering Department, Texas A&M University, College Station, TX, 77843, USA*
[c]*Department of Chemical Engineering, Worcester Polytechnic Institute, Worcester, MA, 01609-2280, USA*
arturo@iqcelaya.itc.mx

Abstract

Safety and sustainability are commonly considered once the conceptual design of a process has been carried out. As an effect, modifications to the process to address safety and sustainability issues are more expensive and difficult to make. Several efforts to include safety and sustainability criteria during the process conceptual design have been made. The main challenge is to find a systematic way to reconcile these criteria with the economic objectives to provide an outcome easy to interpret by decision makers. The objective of this paper is to introduce a new safety and sustainability weighted return on investment metric, a tool that is intended to achieve such reconciliation. A case study dealing with the production of methanol from shale gas is presented. The results show how a design with suitable operating conditions that reconcile economic, environmental and safety criteria can be identified with the use of the new metric.

Keywords: safety, sustainability, process design.

1. Introduction

Typically, the decision making process in the chemical industry relies on technical and economic analysis. Safety and sustainability are usually considered after the conceptual design has been carried out. The effect is that changes in the design due to such factors become more difficult to make at that point in time.

Recently, some approaches have been proposed to include safety in addition to economic considerations during process design, such as the one by Martinez-Gomez et al. (2017) in which the optimization of the production of syngas using economic and safety considerations was developed. El-Halwagi et al. (2013) introduced an approach to incorporate safety objectives in the selection, location and sizing of a biorefinery.

The environmental impact is another criterion to be considered during the design stage of the process. Gong et al. (2014) introduced an optimization model for the sustainable design and synthesis of an algae production network with economic and environmental criteria. Julián-Durán et al. (2014) analyzed the economic and environmental impact of different reforming technologies for methanol production.

Recently, El-Halwagi (2017) presented a metric that includes sustainability criteria in the conventional return on investment (ROI) analysis, called the sustainable weighted return on investment metric (SWROIM). The SWROIM includes sustainability indicators that are compared with target values using a weighted factor, determined by the decision maker. The targets represent the best performance that a process can have for a given indicator; such targets can be determined for instance from the effect of mass and energy integration.

The SWROIM serves as the basis for a new metric presented in this work, called safety and sustainability weighted return on investment metric (SASWROIM). When safety is to be included in the metric, it is not possible to use the target comparison because there is not such a thing as the safest process, a process can only be made safer. Then, the objective of this new metric is to include safety and sustainability criteria in the conventional ROI concept and to apply the metric to a case study to illustrate its potential use as part of the conceptual process design.

2. Description of the metric

The SASWROIM is based on the conventional ROI,

$$ROI = \frac{AEP}{TCI} \tag{1}$$

where AEP is the annual after-tax economic profit and TCI is the total capital investment. A set of design alternatives is to be considered, and the index P is used to indicate each project. The projects are assessed using safety and sustainability indicators given by the index $\{i|i=1,2,\ldots,N_{Indicators}\}$. To perform the comparison, target and base-case values for the indicators should be determined. The target is the best possible value of an indicator in the range studied; the base-case is the actual state of the process that must be included in the range. The importance of each indicator is considered through a weighted factor w_i, which represents the relative significance of a given safety or sustainability indicator compared with the economic profit; the values for the factors are established based on company objectives, consequence analysis or government restrictions.

The SASWROIM is calculated using Eq.(2) and Eq.(3) where ASSP is the annual safety and sustainability profit.

$$SASWROIM = \frac{ASSP}{TCI} \tag{2}$$

$$ASSP_P = AEP_P \left[1 + \sum_{i=1}^{N} w_i \left(\frac{indicator_{Base,i} - indicator_{P,i}}{indicator_{Base,i} - indicator_i^{Target}} \right) \right] \tag{3}$$

In the ratio $\frac{indicator_{Base,i} - indicator_{P,i}}{indicator_{Base,i} - indicator_i^{Target}}$ the denominator indicates the maximum possible improvement and the numerator indicates the improvement or deterioration of project P for the indicator i.

The SASWROIM is especially useful when the value of the ROI is close to the threshold value established by the company. In such cases, the new metric can promote projects that are, apparently, not sufficiently profitable, but that are safer or have a positive impact on the environment. Conversely, projects that are just a little over the threshold value, but that are more pollutant or hazardous than the base-case, may show a level below the threshold value based on the new metric. Therefore, this metric provides an easy and systematic approach to reconcile safety and sustainability criteria with the profitability objectives of the company.

3. Case study

The discovery of shale gas reserves has promoted the development of gas-to-liquid technologies; one of the most attractive applications for the chemical industry is the production of methanol. The production of methanol from shale gas has been analyzed by some authors (Julián-Durán et al., 2014, Ortiz-Espinoza et al., 2017a, Ortiz-Espinoza et al., 2017b), who have found conflicts when evaluating the environmental and safety aspects of the process. In this work, we take a process reported in Ortiz-Espinoza et al. (2017a) to expand the analysis and solve the conflicts among economic, environmental and safety criteria with the use of the SASWROIM.

The production process for methanol involves three main stages, reforming, methanol synthesis and purification. There are different reforming technologies, such as partial oxidation, steam methane reforming, auto thermal reforming, and combined reforming (Julián-Durán et al., 2014). For this case study, only the partial oxidation (POX) reforming method is considered. Partial oxidation is used to produce syngas from hydrocarbons and oxygen. The advantage of this technology is the low amount of carbon dioxide, which has to be removed before the methanol synthesis step. (Julián-Durán et al., 2014). Figure 1 shows the flowsheet of the process.

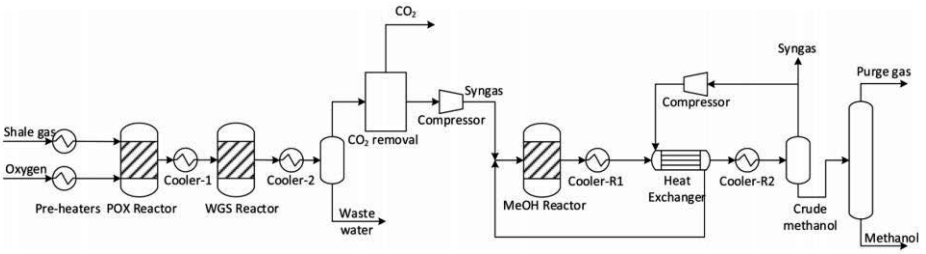


Figure 1. Methanol production flowsheet using POX (Julián-Durán et al., 2014)

3.1. Safety indicator

The process route index (PRI) is used to evaluate safety; this index quantifies the inherent safety level for each process route according to its level of explosiveness. One of the most important criteria for the quantification of the level of explosiveness is the combustibility (ΔFL_{mix}), which is calculated from the difference between the lower (LFL) and upper flammability limits (UFL),

$$LFL_T = LFL_{25}\left[1 - \frac{0.75(T-25)}{\Delta H_C}\right] \tag{4}$$

$$UFL_T = UFL_{25}\left[1 - \frac{0.75(T-25)}{\Delta H_C}\right] \quad (5)$$

The objective of the PRI is to represent the overall process inherent safety level by accounting for the average value of stream parameters such as density, pressure, energy and combustibility; the process with the lowest PRI value in the range is considered inherently safer (Leong and Shariff, 2009). Eq.(6) shows how to calculate the PRI index.

$$PRI = \frac{[(average\ mass\ heating\ value) \times (average\ fluid\ density) \times (average\ pressure) \times (average\ \Delta FL_{mix})]}{10^8} \quad (6)$$

3.2. Sustainability indicator

The environmental impact of the process is assessed through GHG emissions. The estimation of emissions was made in a life cycle basis, considering the acquisition and processing of shale gas, the electricity consumption, the fuel for the heating utilities and emissions due to the outlet streams. A combination of emissions factors (Laurenzi and Jersey, 2013; EPA, 2015; EPA, 2008) and Aspen Plus simulations were used to determine the total emissions of the different process configurations. The results are reported as CO_2 equivalents (CO_2-e).

4. Results

The methanol process has two main drawbacks, namely its high CO_2 emissions and its rather adverse inherent safety levels. A way to decrease the emissions of CO_2 is to increase the recycling fraction of the unreacted syngas in the methanol synthesis loop. On the other hand, the main hazard in this process is the high pressure in the methanol synthesis reactor; hence, a way to reduce the hazard is to reduce the operating pressure of the reactor. For this case study the recycling fraction and the pressure of the reactor were varied from 0.5 to 0.9 and from 50 to 83 bar, respectively.

The first step is to calculate the conventional ROI and the values of the indicators. The results are shown in Table 1.

Table 1. ROI, safety and sustainability indicators for the different process configurations

R / P (bar)	0.5			0.6			0.7			0.8			0.9		
	ROI	CO_2-e	PRI	ROI	CO_2-e	PRI	ROI	CO_2-e	PRI	ROI	CO_2-e	PRI	ROI	CO_2-e	PRI
83	46.90	14.90	17.56	50.04	14.06	17.56	53.62	13.27	17.51	57.42	12.49	17.35	61.62	11.71	17.08
70	43.06	15.89	13.51	46.81	14.85	13.49	50.95	13.87	13.47	55.53	12.88	13.42	60.63	11.89	13.27
60	38.97	17.07	10.72	43.21	15.85	10.64	47.94	14.63	10.61	53.26	13.40	10.57	59.35	12.16	10.47
50	33.03	19.02	8.11	37.87	17.41	8.10	43.37	15.80	8.09	49.71	14.19	8.07	57.20	12.56	8.00

Before applying the SASWROIM, we need to determine the base-case and the target values. For this case, the targets for CO_2-e and PRI are 11.71 and 8.00 respectively, and the base-case is at 60 bar with a recycling fraction of 0.7. The values of the weight factors for this case study were taken as 0.1. The metric was then applied, and the

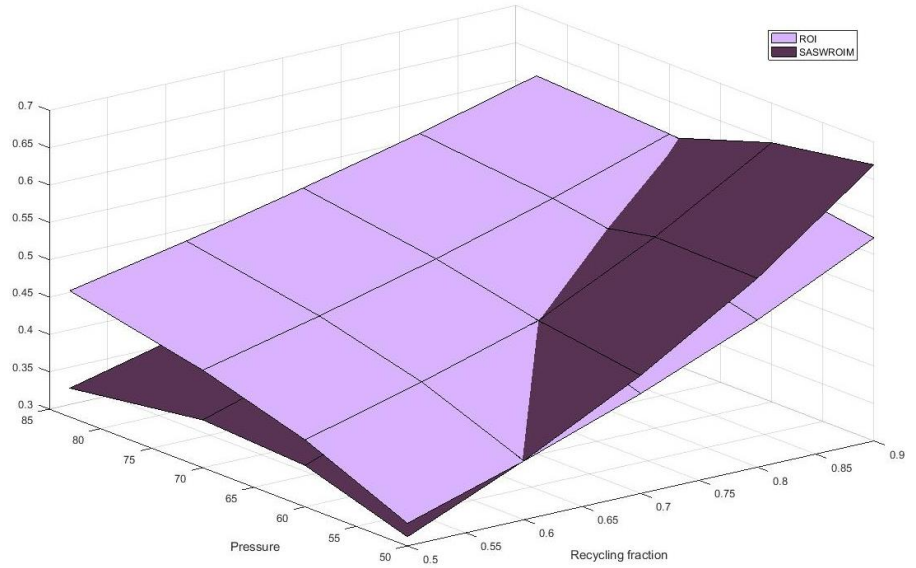


Figure 2. Comparison between SASWROIM and ROI

values of ROI and SASWROIM are plotted in Figure 2. It can be observed that, while the highest ROI was found at a pressure of 90 bar and a recycling fraction of 0.9, the highest SASWROIM resulted at 50 bar and a recycling fraction of 0.9.

The values of the metric can change depending on the weight factors for each indicator. Different factors can be considered to select the weight factors such as the company's core values, penalties due to regulations or consequence analysis. A sensitivity analysis of the impact of the weight factors in the metric can help to select the right values for the factors.

5. Conclusions

The SASWROIM metric provides a systematic method to reconcile safety and sustainability criteria with the economic objectives during the conceptual process design, and it gives an outcome easy to interpret by decision makers. This new metric has been applied to a case study dealing with the production of methanol from shale gas, using a partial oxidation reforming technique. The main hazard was identified as the high pressure of the methanol synthesis reactor. As for the environmental implications of the design, the synthesis gas recycling ratio was found to be the most critical part of the process. Different projects were considered and evaluated using the new metric, and a configuration that reconciles safety, sustainability and economic criteria was found with the reactor operating at 50 bar, and with the design based on a recycling ratio of 0.9.

References

A. El-Halwagi, C. Rosas, J. M. Ponce-Ortega, A. Jiménez-Gutiérrez, M. S. Mannan, M. M. El-Halwagi, 2013, Multi-Objective Optimization of Biorefineries with Economic and Safety Objectives, AIChE J., 59, 7, 2427–2434.

M. M. El-Halwagi, 2017, A Return on Investment Metric for Incorporating Sustainability in Process Integration and Improvement Projects, Clean Technologies and Environmental Policy, 19, 611-617.

EPA, 2008, EPA430-K-08-003 Direct Emissions from Stationary Combustion Sources, U.S. Washington, DC: Environmental Protection Agency.

EPA, 2015, Emission Factors for Greenhouse Gases Inventories, U.S. Washington, DC: Environmental Protection Agency, https://www.epa.gov/sites/production/files/2015-12/documents/emissionfactors_nov_2015.pdf [Accessed October 9, 2017].

L. M. Julián-Durán, A. P. Ortiz-Espinoza, M. M. El-Halwagi, A. Jiménez-Gutiérrez, 2014, Techno-economic assessment and environmental impact of shale gas alternatives to methanol, ACS Sustainable Chemistry and Engineering, 2, 10, 2338–2344.

I. J. Laurenzi, G. R. Jersey, 2013, Life cycle greenhouse gas emissions and freshwater consumption of marcellus shale gas, Environ. Sci. Technol., 47, 4896-4903.

C. T. Leong, A. M. Shariff, 2009, Process route index (PRI) to assess level of explosiveness for inherent safety quantification, Journal of Loss Prevention in the Process Industries, 22, 216-221.

J. Martínez -Gomez, F. Nápoles-Rivera, J. M. Ponce-Ortega, M. M. El-Halwagi, 2017, Optimization of the Production of Syngas from Shale Gas with Economic and Safety Considerations, Applied Thermal Engineering, 110, 678-685.

A. P. Ortiz-Espinoza, A. Jiménez-Gutiérrez, M. M. El-Halwagi, 2017a, Including Inherent Safety in the Design of Chemical Processes, Ind. Eng. Chem. Res., Just Accepted Manuscript, DOI: 10.1021/acs.iecr.7b02164

A. P. Ortiz-Espinoza, A. Jiménez-Gutiérrez, M. M. El-Halwagi, 2017b, Inherent Safety Evaluation for Process Flowsheets of Natura/Shale Gas Processes, Editor(s): A. Espuña, M. Graells, L. Puigjaner, In Computer Aided Chemical Engineering, Elsevier, 40, 1243-1248.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50388-8

Deep Neural Networks for Source Tracking of Chemical Leaks and Improved Chemical Process Safety

Hyunseung Kim[a], En Sup Yoon[b], Dongil Shin[a*]

[a]*Dept. of Chemical Eng., Myongji University, Yongin, Gyeonggido 17058, Korea*
[b]*Eng. Development Research Center, Seoul National University, Seoul 08826, Korea*
[*]*dongil@mju.ac.kr*

Abstract

Leak accidents of chemical plant belong to a major industrial accident that can cause huge damage to human and facilities. Accurate detection and prompt response to leaks, earlier in the event of accidents, is crucial to alleviate damage. This research presents an approach to predict the location of leak source in real-time, by using the measured data from a sensor network as an input data to a pre-trained deep neural network. Computational Fluid Dynamics (CFD) simulations for various leak scenarios are conducted for a real plant geometry, and the massive data derived from the simulations are used for training and verification of the neural networks. Proposed deep neural network models perform well as a solution to the problem of leak-source tracking, one of the challenging inverse problems in chemical process safety.

Keywords: chemical leak, source tracking, deep-learning, big data, process safety.

1. Introduction

Leak accidents of chemical plants are serious industrial accidents that cause enormous human and material damage. The extent of the damage depends on whether the initial response is appropriate or not. To reduce the scale of the damage, it is necessary to establish a system that can identify the leak source(s) precisely at the beginning of the accident. It makes it possible to respond promptly and systematically.

For this reason, several attempts have been made to develop a methodology to identify the source of chemical leak(s). In the robotics field, behavior-based algorithms are embedded in mobile robot(s), and the location of leak source is tracked by executing a pre-input unit action according to the situation that robots are to be encountered in the tracking process (Holland and Melhuish, 1996; Russell, 2004). These methods require expensive mobile sensors and a long time demand for tracking. On the other hand, the approach of predicting the leak point through the data analysis, based on the information collected from a sensor network using only the fixed sensors, can overcome the above-mentioned disadvantages; lower cost of the sensor and reduced tracking time.

In this study, we tried to develop a model that can predict leak location(s) using an artificial neural network when leak substances are detected in the sensors located in plant fence. CFD simulations were conducted on a real chemical plant geometry to generate data for training and verification of neural networks. The concentration data of simulation were extracted at the sensor installation points and the neural networks were trained with the data of concentration, wind speed, and wind direction.

2. Design of leak-source tracking system

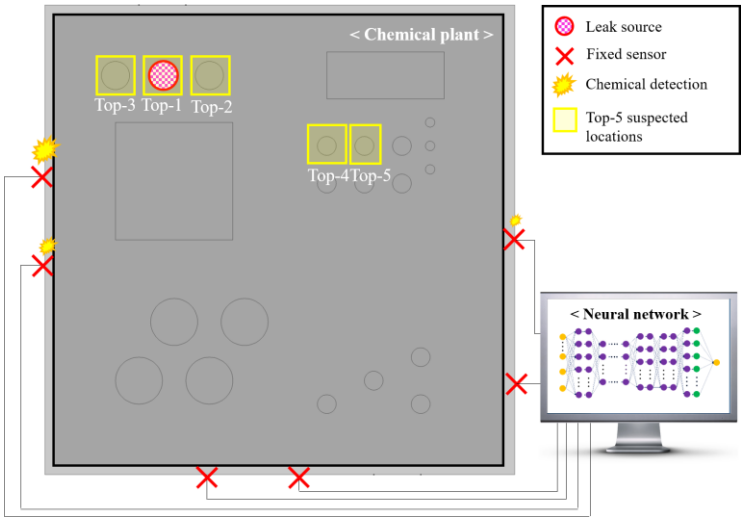


Fig. 1 Schematic diagram of the leak-source tracking system using fence-monitoring data.

The leak-source tracking model inputs the data to the neural network and propagates the data through the complex network. The computed values at the last hidden layer are transformed into the likelihood of each storage tank as potential leak points through softmax function. Fig. 1 shows the schematic concept of proposed model. The model outputs the top five points with high likelihood.

In order to train and validate neural networks, sensor detection database for various leak accident scenarios is needed. However, safety and cost problems are too huge to conduct experiments on various leak accident scenarios for real chemical plants. For this reason, CFD simulations for a real chemical plant geometry were conducted to generate data. However, the detection errors of the data measured by the sensor are not considered, and it is necessary to assume that the CFD and the actual phenomenon coincide very closely. The development of the predictive model was done as follows:

a. Select target: the target chemical plant should be selected first because the leak flow pattern varies greatly depending on the geometry.

b. Generate scenarios of chemical leak accident: the representative scenarios of potential leak accidents should be generated.

c. Conduct CFD simulation and extract data from the sensor installation points.

d. Preprocess and split the data: the extracted data must be preprocessed and split into training data and test data.

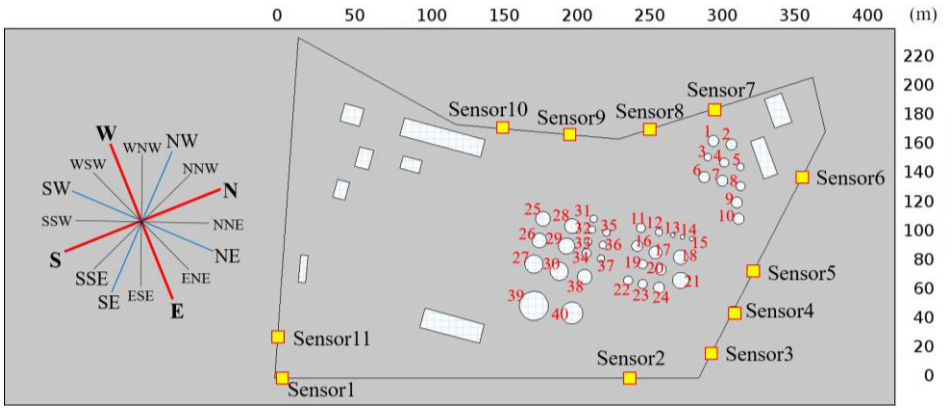
e. Train the networks and check the performance: the performance of the trained models should be verified using test data.

f. Derive the best model: it is needed to experiment with several structures and experimentally and empirically determine the optimal model because no clear methodology for network structure optimization has been presented yet.

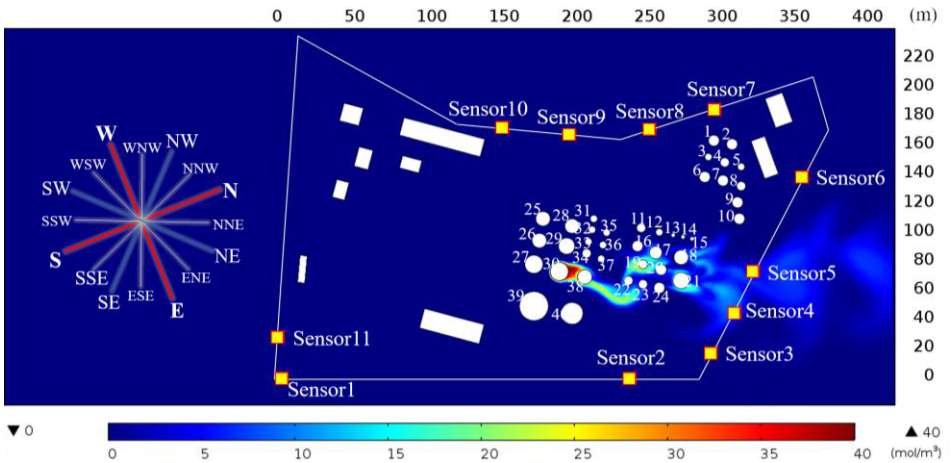
3. CFD simulations for database generation

Originally, CFD simulations were conducted to optimize the sensor-placement, using COMSOL Multiphysics 5.0 for "D" chemical plant in Yeosu, Korea by Cho (2017). Toluene was applied as a leak substance. Fig. 2a shows the information of target and considered wind fields. 40 storage tanks inside the chemical plant were set as potential leak candidates and 16 wind fields are applied to each tank; a total of 640 scenarios. The storage tanks are numbered from 1 to 40. For all accident scenarios, real-time simulations were performed from 0 to 750 seconds in the 2D-horizontal plane. Fig. 2b is an example of the simulation result.

After conducting CFD simulation, concentration data of leak material were extracted at the corresponding points in second. The collected concentration data were saved in the database including the wind speed and direction data used in each accident scenario. The wind direction was converted into the degrees rotated clockwise from the north and saved in the database.



(a) Schematic diagram of "D" chemical plant in Yeosu, Korea.



(b) An example of CFD simulation result: NNE wind, leakage of the 30-storage tank.

Fig. 2 Schematic diagram for the target and toluene dispersion simulations.

4. Modeling of neural network

4.1. Experiment for designing input layer structure

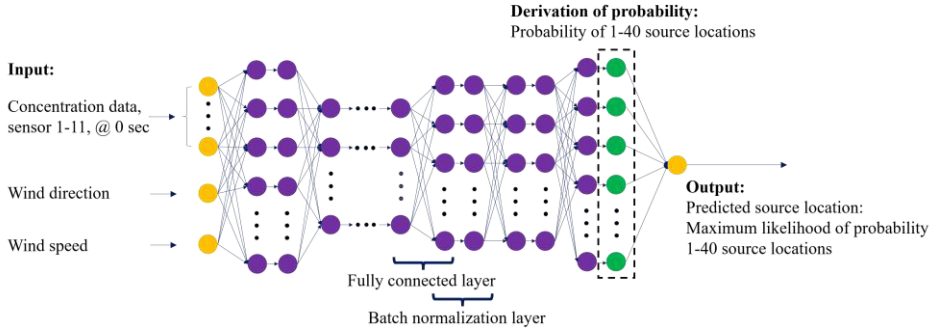


Fig. 3 Feed-forward neural network using the current time point data as inputs.

A simple approach when solving the leak-source tracking problem using the artificial neural network is to train the network using datasets that consist of the concentration, wind direction, and wind speed data collected in one-second. Fig. 3 shows a schematic diagram of the network structure.

Each data is labeled as the corresponding storage tank number if the maximum value of concentration measured by 11 sensors exceeds ERPG-2. Here, ERPG-2 refers to the maximum concentration in the air where almost everyone is exposed for an hour, causing irreparable injury, irreparable or without serious health effects.

A Feed-forward Neural Network (FNN) of 13-15*b*-20*b*-25*b*-20*b*-25*b*-30*b*-35*b*-40 was used to check the approximate performance, but there is no mathematical base on the hidden layer structure. Where *b* is the layer where batch normalization (Ioffe and Szegedy, 2015) is used. Softmax activation function was used for the last hidden layer and ReLU activation function (Nair and Hinton, 2010) was used for the other hidden layers. Adam optimizer (Kingma and Ba, 2014) was used for training.

As a result, low accuracy was obtained; the training accuracy and the test accuracy were 37.8% and 37.2%, respectively. This is because the amount of information was insufficient to predict the location.

In general, the leaking material forms a plume along the wind and is dispersed. The wind fields which move the plume are fluctuated by the complex geometry. Therefore, the concentration graphs measured on the fixed sensors also show the fluctuating patterns according to time. These phenomena have been found in most scenarios. Considering these facts, it was judged that it is better to create a dataset with continuous time series data with a certain sequence length, not only one-second length data. Fig. 4 shows the network structure.

The data were reshaped in a sequence length of 60 seconds. The length of 60 seconds is an amount that would allow a sufficient amount of data to be taken into account in the predictive model, even though it is a short time to predict the leak location with a human perception

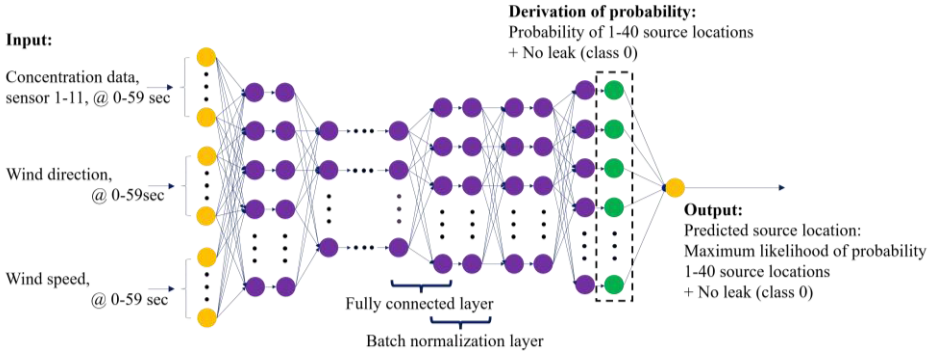


Fig. 4 Feed-forward neural network using 60-second time series data.

In this case, the concentration data, wind direction and wind speed measured from 11 sensors are composed of 60-sequence columns; a total of 780 variables constitute one dataset. It means that each dataset includes 660 variables of concentration. If 10 or more of them were above the ERPG-2 concentration, we labeled the datasets with the storage tank number of the corresponding scenario. For datasets that did not meet these criteria, they were labeled as class 0 (no leak).

4.2. Experiment for designing hidden layer structure

There are four models used in the experiment. As shown in Table 1, the number of parameters is increased by 2, 3, and 4 times based on FNN 1. In all four models, the number of neurons in the input layer, output layer, and the last hidden layer are the same. The batch size, the position that batch normalization applied, activation function, and optimizer are all the same. Through this, we tried to find the proper network size to solve the problem by observing the performance according to the size of the network rather than the structural difference.

As shown in Fig. 5, the greater the number of parameters, the wider the gap between the test loss and the training loss. It means that overfitting problem is getting worse because test loss does not decrease, even though training loss decreases continuously as the number of parameters increases. This means that it is no longer meaningful to increase the number of training epochs or network capacity.

Table 1. Summary of results obtained

	Network structure	Epoch	Number of parameters	Maximum accuracy training/test
FNN 1	780-25*b*-28*b*-28*b*-28*b*-28*b*-28*b*-41	5,000	25,020	71.2 % / 71.1 %
FNN 2	780-50*b*-48*b*-45*b*-42*b*-37*b*-28*b*-41	5,000	49,979	81.5 % / 78.7 %
FNN 3	780-75*b*-67*b*-57*b*-47*b*-38*b*-28*b*-41	5,000	74,998	87.0 % / 82.4 %
FNN 4	780-100*b*-82*b*-65*b*-50*b*-38*b*-28*b*-41	5,000	100,022	89.5 % / 84.1 %

where b: batch normalization is used

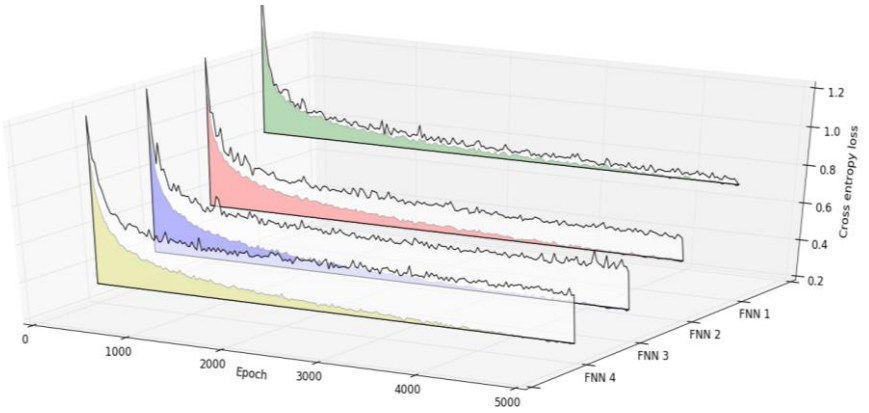


Fig. 5 Cross entropy loss of feed-forward neural networks for training data and test data.

As shown in Fig. 5, the greater the number of parameters, the wider the gap between the test loss and the training loss. It means that overfitting problem is getting worse because test loss does not decrease, even though training loss decreases continuously as the number of parameters increases. This means that it is no longer meaningful to increase the number of training epochs or network capacity.

Conclusions

In this study, for big-data based source tracking, CFD simulations were performed for leak accident scenarios in a real chemical plant, and leak-source tracking models were developed by training the leak dispersion data in FNNs. It was found that constructing input data to have a certain sequence length is suitable for the nature of leak-source tracking problem of chemical plants, a challenging safety inverse problem. For the testbed problem, the actual leak points were predicted with the test accuracy of 84.1% for 40 leak candidates. Considering the high-prediction performance of the proposed leak-source tracking method based on the deep neural net, it is expected to be deployed as a solution to the problems of identifying the leak source(s) out of chemical plants.

References

D. Kingma, J. Ba, 2014, Adam: A method for stochastic optimization, arXiv: 1412.6980

J. Cho, 2017, *Placement Optimization and Reliability Analysis of Stationary and Mobile Sensors for Chemical Plant Fence Monitoring*, M.S. Thesis, Myongji University

O. Holland, O. Melhuish, 1996, Some adaptive movements of animats with single symmetrical sensors, *From Animals to Animats*, **4**, 55-64

R. A. Russel, 2004, Robotic location of underground chemical sources, *Robotica*, **22**, 109-115

S. Ioffe, C. Szegedy, 2015, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, arXiv:1412.1441

V. Nair, G. E. Hinton, 2010, Rectified Linear Units Improve Restricted Boltzmann Machines, *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 807-814

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50389-X

Optimal design of a COEX process for spent nuclear fuel reprocessing using small channels

Davide Bascone[a*], Panagiota Angeli[a], Eric S. Fraga[a]

[a]*CPSE, University College London, London WC1E 6BT, UK*
davide.bascone.14@ucl.ac.uk

Abstract

The use of small channels for liquid-liquid extraction seems to be promising, since hydrodynamics are more easily controlled and mass transfer is enhanced. The use of small scale contactors may be particularly beneficial for spent nuclear fuel (SNF) reprocessing, due to the small size and consequent safety improvements. In this paper, a mathematical model for the intensified liquid-liquid extraction in small channels has been developed. A multi-component system is considered, including most of the components in the SNF. The resulting model is a large scale mixed integer nonlinear optimisation problem. A novel design of the COEX (co-extraction of uranium and plutonium) process, employing small channels, is presented. Results show that several improvements are obtained, compared to traditional technologies, in terms of safety and solvent degradation.

Keywords: Liquid-liquid extraction; small channels; spent nuclear fuel reprocessing; optimisation; process design.

1. Introduction

Current commercial plants for Spent Nuclear Fuel (SNF) reprocessing use the PUREX (Plutonium URanium EXtraction) process. The aim is to recover U and Pu, which can be reused as MOX (Mixed OXide) fuel in appropriate nuclear reactors. These components are approximately the 97% of the SNF, and can be recovered by Liquid-Liquid Extraction (LLE) by a selective organic solvent, the Tributyl Phosphate (TBP), after dissolution of SNF in nitric acid, according to the reactions:

$$UO_2^{+2} + 2NO_3^- + 2TBP \leftrightharpoons [UO_2(NO_3)_2] \cdot 2TBP \quad (1)$$

$$Pu^{4+} + 4NO_3^- + 2TBP \leftrightharpoons [Pu(NO_3)_4] \cdot 2TBP \quad (2)$$

This LLE is mostly carried out in pulsed columns and mixer settlers. However, both conventional process and technologies have several disadvantages. For instance, pulsed columns and mixer-settlers require long residence time of the aqueous-organic mixture, which inevitably leads to solvent degradation. A promising alternative to these traditional technologies may be the intensified LLE in small channels, exploiting the advantages of improved mass transfer in small scale contactors. The concept of decreasing plant size but not performance, which is in fact improved in small scale contactors, is known as process intensification. This technology may be particularly interesting for nuclear applications. Some benefits may be improvements in safety, due to the small volume of

hazardous materials involved, reduced solvent degradation due to short residence time, easy control of hydrodynamics to improve mass transfer and low pressure drop.

The PUREX process delivers a pure Pu stream, implying nuclear proliferation risk. Several alternatives to this process have been proposed, such as UREX, UREX+2, NPEX, COEX (Nash and Lumetta, 2011). The latter flowsheet may be the most promising, since it reduces extraction steps. It also improves proliferation resistance: a mixed U/Pu is achieved, providing a better homogeneity of U/Pu distribution within the MOX fuel, compared to the current approaches.

The COEX process has the same first section of PUREX process, i.e. the codecontamination section, where U and Pu are separated from most of the SNF. The product stream of the codecontamination section, which is an organic phase containing U, Pu, HNO_3 and very small amounts of Tc, Np, Zr and Ru undergoes a first stripping step. This step is called "co-strip" (step 5 in Figure 1), the goal is to strip (or back extract) simultaneously U and Pu from the organic stream. However, the Pu/U ratio in the organic stream outgoing the codecontamination section is too low to produce MOX fuel from this stream. Hence, the idea is to strip all Pu and only part of U to achieve an adequate Pu/U ratio for MOX fabrication, which is around 10%. To back extract Pu and only part of U, the oxidation states must be changed to exploit their different equilibrium behaviour. Pu is originally present in the mixture as Pu(IV), U as U(VI). A reducing agent is used to reduce Pu(IV) to Pu(III). Pu(III) is not very extractable, contrarily to Pu(IV) and U(VI), therefore is it simple to strip. Reductants used in nuclear industry for plutonium stripping are hydroxylamine nitrate (HAN) and U(IV). Also, a scavenger agent is used to reduce re-oxidation from Pu(III) to Pu(IV) and improve Pu stripping. Typically, hydrazine nitrate $N2H4{\cdot}HNO_3$ is used for this purpose.

In this paper, a mathematical model for intensified extractions in small channels, suitable for SNF reprocessing, has been developed. A differential equation has been used to describe the mass balance of each component. The model involves a 12-component LLE. Attempts to model LLE with "nonlinear equilibrium", i.e. when the distribution coefficient is estimated as function of the concentration of components, for a large number of components and using differential mass balances are limited in the literature. The most complex and detailed model for SNF reprocessing available is the model developed by Gonda and Matsuda (1986). They have modelled a 5-component LLE, with U(VI), U(IV), Pu(IV), Pu(III) and HNO_3, using a partial differential equation to describe the dynamic behaviour in pulsed columns, which also included six chemical reactions.

Uranium extraction in small channels for nuclear application has been investigated by Tsaoulidis and Angeli (2015). Empirical correlations for mass transfer and hydrodynamic parameters have been obtained from experimental data. They have also investigated mass transfer through CFD simulations, but no models for optimisation and design of

intensified LLE in small channels have been developed, and no multi-component LLE has been investigated.

2. Mathematical modelling

The segmented two-phase flow pattern obtained in small channels, i.e. dispersed phase plugs separated by continuous phase slugs, has been approximated in this work with a simple ideal plug flow model to describe mass balance of the *i* component in the *k* phase:

$$-v_k \frac{dC_{i,k}}{dL} - K_L a\left(C_{i,k} - C_{i,k}^{eq}\right) + \Sigma R_{i,k}\varphi = 0 \tag{3}$$

Ideal plug flow conditions are typically used in the literature to estimate mass transfer coefficients in small channels. Steady state conditions have been considered. The mass transfer coefficient has been calculated according to Kashid *et al*. (2010):

$$K_L a = 0.88 \frac{v_{mix}}{L} Ca^{-0.09} Re^{-0.09} \left(\frac{D}{L}\right)^{-0.1} \tag{4}$$

Kashid *et al.* used a Y-junction to mix the two phases. Eq. (4) was developed for the system acid acetic solution-kerosene, similar to the one investigated in this work, which is nitric acid solution – TBP/kerosene (30% v/v). A correction factor, function of the dispersed to continuous phase flow ratio and based on experimental data reported in the literature (Tsaoulidis *et al.*, 2013), has been included in the calculations to take into account the effect of the flow ratio on the interfacial area. Crucial in the model is the estimation of the distribution coefficients, i.e. the ratio between concentrations in the organic and the aqueous phases for each component at equilibrium conditions; the coefficients describe the driving force for the mass transfer. For these calculations, expressions from the literature have been included in the model. Distribution coefficients of U(VI), Pu(IV) and HNO_3 have been calculated according to Richardson and Swanson (1975). Coefficients for U(IV) and Pu(III) have been estimated as suggested by Kumar and Koganti (1998, 2003), whilst for Zr and Ru the correlations developed by Natarajan *et al.* (2012) have been used. Distribution coefficients of Np(IV) and Np(VI) have been estimated according to Kumar *et al.* (2001). Finally, distribution coefficients of HNO_2 and Tc have been calculated as suggested by, respectively, Hongyan *et al.* (2013) and Asakura *et al.* (2005). For Np(V), almost inextractable, a nominal value of 0.01 has been used (Hongyan *et al.*, 2013). In addition to the redox reactions between Np(IV), Np(V) and Np(VI), modelled according to Hongyan *et al.* (2013), many chemical reactions are involved if a reductant is introduced in the system to reduce plutonium from Pu(IV) to Pu(III). U(IV) has been used as reducing agent, and kinetics were estimated according to Tachimori (1994). Pu(III) can reoxidise to Pu(IV) (Tachimori, 1994) hence, to prevent it, $N2H4{\cdot}HNO_3$ is used as nitrite scavenger (Gonda and Matsuda, 1986). $N2H4{\cdot}HNO_3$ also acts as reductant (Tachimori, 1994). Furthermore, U(IV) can oxidise to U(VI) (Gonda and Matsuda, 1986).

The design of a scale out manifold that ensures acceptable flow distribution in the many small channels is essential to process high throughput with small scale devices.

Commenge *et al.* (2011) have proposed a methodology for the design of a comb-like flow network. Pressure drop calculations have been included in the model. Multi-stage counter-current design has been considered, to achieve high extraction efficiency and meet industrial requirements.

The resulting model, posed as an optimisation problem, is a mixed-integer nonlinear problem, implemented in GAMS. The model is large, due to the 50 grid points used to numerically solve Eq. (3) by the 4th order Runge-Kutta method.

3. Case study

To demonstrate the applicability of the model, the COEX process has been investigated. This flowsheet reduced the number of unit operations required, compared to the current PUREX process. In this paper, a further reduction of unit operations has been considered: no second cycle for U purification has been used, significantly reducing the plant cost. This could be achieved if small channels are used, where separations can be improved (in particular, of Np) and further U purification may be not required (U purity of about 99.99% could be achieved, according to the model). Also, no organic recycle stream between steps 8 and 5 is needed. The flowsheet is shown in Figure 1. The aim is to minimise the total annualised cost, which includes operating and capital expenses. Several constraints have been included in the model, to meet industrial requirements in terms of typical decontamination factors (DF): minimum 10,000 for Ru and Zr. DF of Tc is typically around 10, but in this work it has been increased to 50 to avoid further U purification. Recovery of U and Pu in the flowsheet must be higher than 99%, concentration of U in the aqueous phase outgoing step 1 must be ≤ 1.68 mol m^{-3}, ratio Pu/U in the aqueous stream leaving step 8 must be 9.5%.

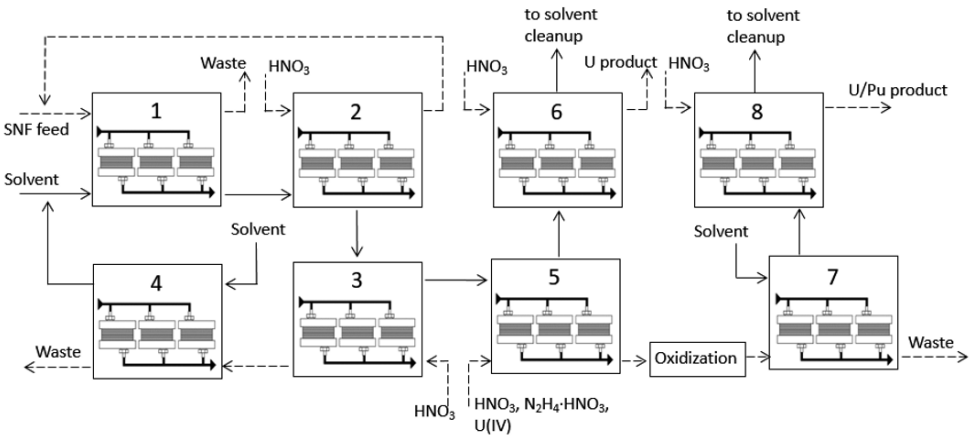


Figure 1: Diagram of COEX process. Continuous line refers to organic stream, dashed line refers to aqueous stream. Flow rate of SNF feed is approximately 200 L h^{-1}.

Design variables are operating variables (flow rates, HNO_3 concentrations), diameter and length of channels (equally sized), number of stages, geometry of manifold (number of elements, width and length of distributors/collectors). Also, concentration of HNO_3, U(IV) and N2H4·HNO_3 in the aqueous streams are free variables.

Table 1: Optimal design. Diameter and length of small channels are, respectively, 2 mm and 100 mm for all steps. No. of channels expressed as: (no. of elements of scale 1) x (no. of elements of scale 2) x (no. of elements of scale 3) x (no. of elements of scale 4).

Step	No. of stage	Aq. to org. flow ratio	No. of channels in manifold
1	4	0.84	67 x 15 x 11 x 1
2	3	0.52	70 x 8 x 8 x 1
3	2	0.24	67 x 12 x 7 x 1
4	1	0.50	70 x 6 x 5 x 1
5	4	0.20	66 x 9 x 6 x 1
6	4	2.17	67 x 14 x 10 x 1
7	4	0.89	70 x 6 x 6 x 1
8	3	3.38	69 x 7 x 6 x 1

3.1. Results

The results of the optimisation are shown in Table 1. Very short residence times, in the order of seconds, are involved using this novel technology. This may represent an important benefit if compared with the residence times of pulsed columns and mixer-settlers, which are in the order of minutes and inevitably cause solvent degradation. High extraction efficiency can be achieved by the high K_La obtained in the small-scale contactors. The values of K_La achieved are, in all steps of the flowsheet, between 0.2 and 0.3 s^{-1}. Roizard *et al.* (1997) listed typical K_La in a stirred tank between 0.03 and 0.8 s^{-1}, and K_La of approximately one order of magnitude lower in columns. For uranium extraction in small channels with same diameter, length and superficial velocity of the mixture, a K_La of approximately 0.1 s^{-1} was reported (Tsaoulidis and Angeli, 2015), using significantly more viscous organic diluents (ionic liquids). The difference between this experimental value and the one used in this work could be explained by two reasons: the lower viscosity, which increases the mass transfer coefficient, and the increased interfacial area, as result of the lower aqueous (dispersed) to organic flow ratio. Another advantage is related to the small size of the channels. In this case study, the average volume of each step is around 120 L. The typical total liquid volume for mixer-settlers is 60,000 L, 800 L for pulsed columns (K. Nash and G. Lumetta, 2011). Smaller volumes lead to safety improvements and a better control of nuclear criticality: neutrons have a greater possibility of escaping with higher surface area to volume ratio.

3.2. Comments on mathematical model

The mathematical model is large, around 700,000 equations, and nonlinear. A sequential modular approach has been first used to solve the flowsheet and provide initialisation. Then, all the unit operations are optimised simultaneously. The mass transfer model, Eq.(4), and its parameters have been derived from experiments in small (< 1 mm) channels. However, Eq.(4) has been used in the literature to fit experimental data of liquid-liquid extraction of uranium in small channels, with diameters up to 2 mm

(Tsaoulidis and Angeli, 2015). The latter value has been assumed as upper bound for the diameter. For larger channels, more data would be required. The current model has been implemented for SNF application, but it may be potentially used for any other LLE applications in small channels.

4. Conclusions

An optimisation model for intensified extraction in small channels has been developed, formulated as MINLP problem. A novel COEX process has been investigated to demonstrate the applicability of the model. Results have confirmed many potential advantages of this novel technology, in terms of performance and size. The mathematical model developed suitable for nuclear application, can be applied to other separation processes involving extractions, by integrating the model with new components and the appropriate equations for thermodynamic calculations.

References

T. Asakura, S. Hotoku, Y. Ban, M. Matsumura, Y. Morita, 2005, J Nucl Radiochem Sci, 6, 3, 271-274.
J. M. Commenge, M. Saber, 2011, Chem Eng J, 173, 541-551.
K. Gonda, T. Matsouda, 1986, J Nucl Sci Technol, 883-895.
C. Hongyan, R. J. Taylor, M. Jobson, A. J. Masters, 2013, Solvent Extr Ion Exc, 34(4), 297-321.
M. N. Kashid, A. Gupta, A. Renken, L. Kiwi-Minsker, 2010, Chem Eng J, 158, 233-240.S. Kumar, S. B. Koganti, 1998, Nucl Technol, 123, 116-119.
S. Kumar, S. B. Koganti, 2001, Indian J Chem Technol 8, 41-43.
S. Kumar, S. B. Koganti, 2003, Solvent Extr Ion Exc, 21 (3), 369-380.
K. Nash, G. Lumetta, 2011, Advanced Separation Techniques for Nuclear Fuel Reprocessing and Radioactive Waste Treatment. Woodhead Publishing, Oxford
R. Natarajan, N. K. Pandey, V. Vijayakumar, R. V. Subbarao, 2012, Procedia Chemistry 7 (2012), 302-308.
G. L. Richardson, J. L. Swanson, 1975, Plutonium partitioning in the PUREX process with hydrazine-stasbilized hydroxylamine nitrate, HEDL-TME.
C. Roizard, G. Wild, J.C. Charpentier, 1997, Absorption avec réaction chimique, Techniques de l'Ingénieur, Traité Génie des Procédés J1079, 1–16.
S. Tachimori, 1994, Extra-m: A computing code system for analysis of the PUREX process with mixer settlers for reprocessing, Japan Atomic Energy Research Institute.
D. Tsaoulidis, P. Angeli, 2015, Chem Eng J, 785-793.
D. Tsaoulidis, V. Dore, P. Angeli, N. V. Pletchkova, K. R. Seddon, 2013, Chem Eng J, 227, 151-157.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50390-6

An MILP-based Operability Approach for Process Intensification and Design of Modular Energy Systems

Vitor Gazzaneo, Juan C. Carrasco, Fernando V. Lima*

Department of Chemical and Biomedical Engineering, West Virginia University, Morgantown, WV 25606, P.O. Box 6102, USA
**Fernando.Lima@mail.wvu.edu*

Abstract

A mixed-integer linear programming (MILP) operability approach is developed for the design of high-dimensional and nonlinear systems. For the approach formulation, classical operability concepts are extended to attain process intensification towards system modularity. Motivated by natural gas utilization processes, a catalytic membrane reactor for the direct methane aromatization (DMA-MR) conversion to hydrogen and benzene is chosen as case study. The DMA-MR is described by a system whose design is challenged by complexity, including nonlinearities and a highly-constrained environment. A DMA-MR subsystem is addressed, for which 80% reduction in membrane area and 78% reduction in reactor volume is obtained when compared to a base case with equivalent performance. The proposed approach presents a computational time reduction greater than 3 orders of magnitude compared with previously introduced nonlinear programming-based operability approaches. These results indicate that the proposed approach is computationally efficient and thus can be extended to address higher dimensional cases.

Keywords: Operability, Process Intensification, Modular Systems, Energy Systems, Optimization.

1. Introduction

Modular systems are promising technologies characterized by versatility. They provide an alternative for the traditional industry by considering small modules of higher efficiency. As these modules can be shipped and assembled, the exploitation of resources in remote and originally untapped areas becomes feasible. A DMA-MR for methane conversion to hydrogen and benzene can be considered as a potential candidate system for modular natural gas utilization that could be transported to stranded gas fields. In this case, the concept of process intensification characterized by the enhanced reaction and separation phenomena in the membrane reactor may enable high efficiency units for modular system applications. Additionally, a substantial decrease in cost and environmental footprint are key outcomes of system modularity that are also associated with process intensification. The design of a modular energy system, however, is challenged by its complexity, including nonlinearities, system dimensionality and a highly-constrained environment. An MILP-based operability approach is proposed here to overcome these challenges. Operability was first described as a mapping between inputs and outputs of a system that was established to measure the ability of a design to achieve specified desired operating states. The aim in the new formulation is to extend

classical operability concepts and tools to attain process intensification targets and ensure system modularity in a computationally efficient manner.

Process operability was initially used for the assessment of low-dimensional nonlinear systems. The expansion to high-dimensional nonlinear systems was successfully addressed by employing optimization algorithms based on nonlinear programming (NLP) [1,2]. However, previously proposed approaches were challenged by the system dimensionality in terms of high computational expense. The novelty in this work consists of the fully utilization of MILP tools to tackle systems of similar complexity, with the advantage of computational time reduction.

2. Process Operability Concepts

In traditional process operability concepts, inputs (u) and outputs (y) are respectively represented by the Available Input Set (AIS) and Achievable Output Set (AOS). The desired operating region for the outputs and the corresponding region for the inputs in order to achieve such outputs are called Desired Output Set (DOS) and Desired Input Set (DIS), respectively. For a given set of inputs, the calculation of the corresponding outputs is performed by direct mapping employing a first-principles process model, M. Conversely, the inverse model, M^{-1}, represents the pathway from the outputs to inputs. For an m x n dimensional input-output system, the operability sets are generically represented by:

$$AIS = \left\{u | u_i^{min} \leq u_i \leq u_i^{max};\ 1 \leq i \leq m\right\};\ DIS = \{u | u = M^{-1}(y); y \in DOS\};$$

$$AOS = \{y | y = M(u);\ u \in AIS\};\qquad DOS = \left\{y | y_i^{min} \leq y_i \leq y_i^{max}; 1 \leq i \leq n\right\};$$

Nonlinear systems require a nonlinear model M. The derivation of M^{-1} may be complex and sometimes not straightforward. One contribution of this work is the numerical computation of M^{-1} employing computational geometry tools, overcoming the challenges associated with the analytical calculation of M^{-1}. The description of such computation can be found in section 3 below, along with the proposed approach.

3. MILP-Based Operability Approach

In the developed framework, the operability mapping accounts for design variables (physical dimensions, feed and operating conditions) as inputs and measures of production (production rates and efficiency) as outputs. Such mapping is established by dividing the original nonlinear space into several linear subspaces described by linearized models. This task is performed with the utilization of computational geometry techniques such as Delaunay-triangulation. The properties of the obtained subsets are then analyzed to find the minimum number of divisions that is needed to adequately represent the original system. Using such divisions, an optimization algorithm based on MILP concepts is formulated, considering all process constraints and specific targets associated with process intensification. The steps of the developed framework can be summarized as follows:

1) The nonlinear system is simulated, generating input-output data points. First-principles models for the application of interest are considered for this task.

2) Using a computational geometry-based technique, e.g., Delaunay-triangulation, sets of triangles are obtained. Each triangle in the AIS has a counterpart in the AOS. This step can be repeated, with gradual reduction in number of considered data points,

until the triangulation is still possible for the minimum number of points. Once executed, this gradual reduction results in the minimum number of triangles needed to adequately represent the original nonlinear system. Alternatively, if needed, this step can be performed using only the points of the operating envelope, i.e., points of the boundary of AIS and AOS.

3) The DOS is selected and its intersection with the AOS is calculated by employing computational geometry tools. The input counterpart of the calculated region is the feasible DIS. This pathway from outputs to inputs is the inverse model calculation performed using linearized models represented by triangles.

4) An MILP minimization problem is formulated to reduce footprint in terms of physical size and ensure the compliance of process constraints and intensification targets. Here, the MILP formulation considers the regions of input/output delimited by the DIS/DOS in step 3. This formulation is mathematically defined as:

$$\Phi = \underset{u}{\text{minimize}}\ (footprint)$$

Subject to: $u \in DIS$, $y \in DOS$
Linearized process model and constraints
Process intensification targets

The mixed-integer nature is necessary due to the existence of the discrete regions represented by triangles. For each triangle, a binary value is assigned and the sum of all assigned values have to be 1. Interpolations are allowed by using weights related to the vertices of existing triangles. The outcome of the optimization is an optimal design inside a unique triangle.

4. Results: DMA-MR Case Study

In this section, each step of the framework described above is applied to a 2x2 DMA-MR subsystem, in which each step is in a separate subsection.

4.1. Simulation of the DMA-MR System

The feed for the shell and tube MR configuration considered is natural gas (methane) in the tube side, and a sweep gas (a laboratory inert gas, e.g., helium) in the shell side. The methane is converted into benzene and hydrogen following a two-step reaction mechanism. Hydrogen is selectively permeated through the membrane, yielding a hydrogen-rich permeate and a benzene-rich retentate. Figure 1 shows the adopted configuration along with reactions.

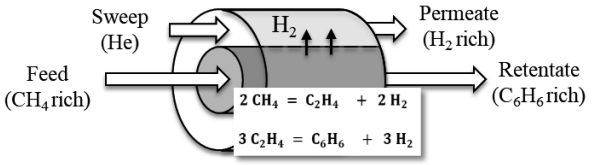


Figure 1. DMA-MR cocurrent configuration scheme.

For the performed operability analysis, the inputs are the reactor dimensions (length and diameter) and the outputs are benzene production and methane conversion. Other variables such as reactor temperature, pressure, feed, etc., are assumed

constant/controlled for this analysis. A nonlinear model is developed, which is characterized by a set of ordinary differential equations (ODEs) obtained from molar balances inside tube and shell (refer to [2] for model details and assumptions). An initial simulation is carried out considering 10,000 combinations of reactor dimensions, with the reactor length range from 10 to 100 cm and tube diameter range from 0.5 to 2.0 cm. Such simulation is performed in MATLAB, employing a stiff ODE solver, "ode15s", due to the nature of the ODE system.

4.2. Triangulation of AIS and AOS

The obtained data points from the input-output simulation are used for the triangulation task. The AIS is constrained to the given rectangular mesh and the AOS is constrained to its original hull. Figures 2 and 3 show the initial triangulation and the one obtained with minimum number of triangles, respectively.

For an evenly spaced AIS grid, the configuration with the minimum number of triangles employing the approach described above is a 38 x 38 grid. To produce a system representation with an even lower number of triangles, an alternative triangulation using only the outside operating envelope has to be performed. In such case, an unconstrained Delaunay-triangulation is applied to the AOS envelope points, resulting in a set of triangles that connect the edges of the original shape. As a lower number of triangles is used to represent the system, an error increase is expected in such representation. The trade-off between computational time and relative error is further addressed in section 5.

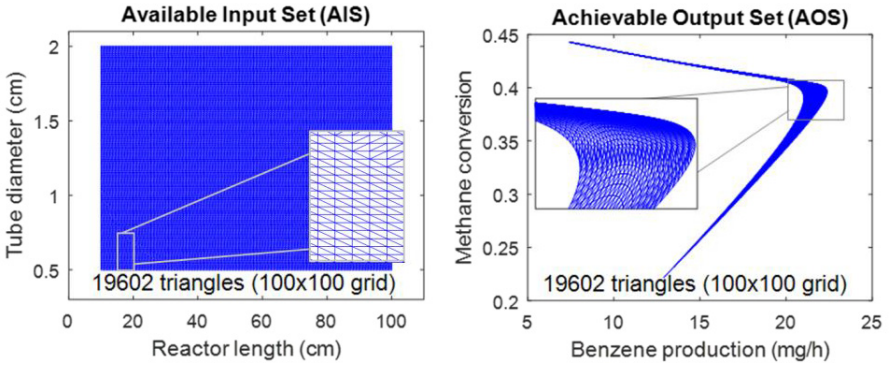


Figure 2. Constrained triangulation of AIS and AOS for initial case.

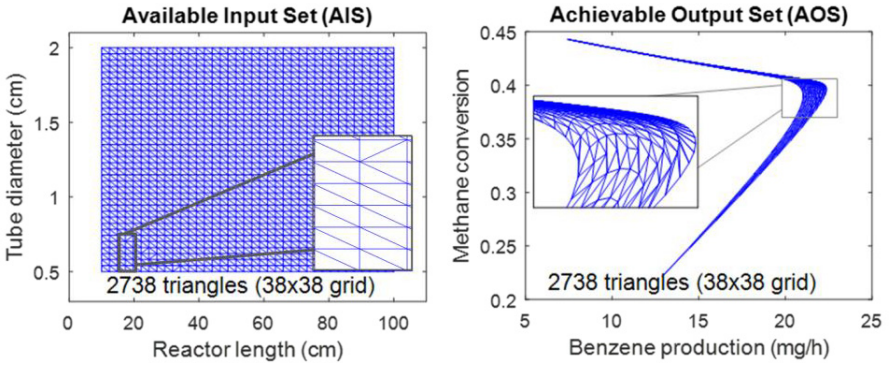


Figure 3. Constrained triangulation of AIS and AOS for case with minimum number of triangles.

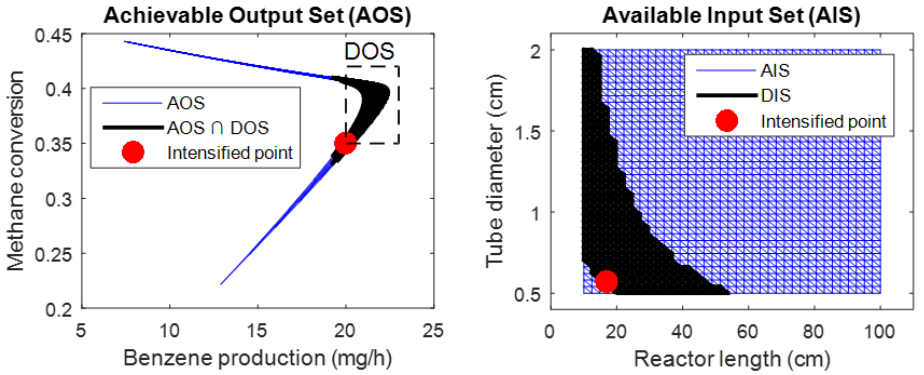


Figure 4. DOS selection, DIS calculation and MILP solution for minimum number of triangles.

4.3. Quantification of Linearized Spaces and DOS Selection

For this application, the DOS region is defined as benzene production ranging from 20 to 23 mg/h and methane conversion from 0.35 to 0.42 as shown in Figure 4. The intersection of this region with the AOS is calculated using the function "isOverlapping" in the MPT toolbox [3] in MATLAB. The inverse mapping step employing the triangles results in the region of desired inputs, the DIS. Figure 4 also illustrates the DIS calculation for the case with minimum number of triangles.

4.4. MILP Formulation and Optimal Design

The reduction of the MR footprint is defined as the MILP objective function. The model for the optimization task corresponds to a set of linearized multi-models. As constraints, the ratio $L/D \geq 30$ is considered here for plug flow, where L is the reactor length and D the tube diameter. An additional performance target is set for a benzene production of at least 20 mg/h. The MATLAB subroutine "intlinprog" is used for this task.

Figure 4 shows the intensified point given by $(L, D) = (17.0, 0.57)\ cm$, obtained using the proposed MILP formulation. A comparison with a base case from literature [2] is established, where $(L, D) = (100, 0.50)\ cm$. The obtained design shows equivalent reactor performance for an 80% reduction in membrane area and 78% reduction in reactor volume.

5. Computational Results

The developed approach is executed for a different number of initial points and the trade-off between computational cost and relative error is analyzed. For such activity, NLP-based approaches [2,4] are also run for similar number of initial points. The NLP result for the case with the largest number of initial points is considered as the benchmark result, given the detailed exploration of the input-output spaces performed in this case using the nonlinear model. Table 1 summarizes the obtained comparison.

From the results in Table 1, note that the proposed approach performs faster in all scenarios for the same number of initial points. When compared to the NLP-based approach benchmark case considering the same number of initial points, it shows a computational time improvement greater than 3 orders of magnitude with a relative error less than 2%. Such improvement indicates the computational efficiency of the developed approach and thus its feasibility to address high-dimensional calculations.

Table 1. Trade-off performed between computational time and relative error for proposed approach when compared to NLP-approaches.

Approach	CPU Time (h:min:s)	# Initial Points	Reactor Length (cm)	Tube Diameter (cm)	Relative Error (w.r.t. membrane area) (%)	Gain in CPU Speed (times) **
NLP-based Approach [2,4]	00: 06: 30.60	25	18.00	0.600	10.23	-
	00: 25: 21.15	100	17.38	0.577	2.31	-
	06: 11: 04.35	1444	17.26	0.570	0.50	-
	10: 35: 50.21	2500	17.14	0.571	0	-
Proposed MILP-based Approach	00: 00: 00.03*	24	18.02	0.601	10.50	12,844
	00: 00: 00.03*	100	17.63	0.588	5.74	44,773
	00: 00: 07.29	1444	17.03	0.568	1.28	3,051
	00: 00: 32.00	2500	17.00	0.567	1.66	1,191

*Rounded values; ** for the same number of initial points

6. Conclusions

The developed MILP-based operability approach was successfully applied to attain process intensification towards footprint reduction and system modularity. When compared to NLP-based operability approaches, it performed faster at any scenario with minimal errors, especially for a higher number of initial points. As computational cost was not a tractability barrier for the developed approach, this approach will be extended next to address higher dimensional cases.

Acknowledgments

The authors gratefully acknowledge the Donors of the American Chemical Society Petroleum Research Fund and the National Science Foundation CAREER Award 1653098 for partial support of this research.

References

J Carrasco and F. Lima, 2017, “An optimization-based operability framework for process design and intensification of modular natural gas utilization systems”, Comput. Chem. Eng., 105, 246–258.

J Carrasco and F. Lima, 2017, “Novel operability-based approach for process design and intensification: application to a membrane reactor for direct methane aromatization”, AIChE J., 63, 3, 975-983.

J Carrasco and F. Lima, 2017, “Operability-based approach for process design, intensification, and control: application to high-dimensional and nonlinear membrane reactors”, In Proceedings of the FOCAPO/CPC.

Herceg M., Kvasnica M., Jones C.N., and Morari M., 2013, “Multi-Parametric Toolbox 3.0”. In Proceedings of the European Control Conference, 502-510.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50391-8

Optimal synthesis of rotating packed bed and packed bed: a case illustrating the integration of PI and PSE

Zhi Qian[a], Qi Chen[b], Ignacio E. Grossmann[b*]

[a]*University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China*
[b]*Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, United States*
grossmann@cmu.edu

Abstract

The rotating packed bed (RPB) that is a novel process intensification (PI) reactor has been successfully adopted in the chemical industry in recent decades. Unlike sheer process optimization, PI aims to break the bottleneck of transfer processes at the micro level, like mixing and reaction kinetics, such that the process shifts to be governed by its innate reaction rate rather than by mass transfer. This work demonstrates a superstructure synthesis approach for evaluating PI device through a case study involving an RPB and a packed bed (PB). A multi-scale model links the micro-mechanisms of intensification in the RPB with the macro-level decisions of equipment selection and interconnection. The proposed superstructure considers configurations of an RPB and PB in series, in parallel, and in between, allowing quantification of their synergy. The overall model is formulated as a nonlinear programming problem and is demonstrated with the case of H_2S removal from refinery tail gas. The computational results show that the configuration of both RPB and PB in parallel is optimal under the assumption of linear investment and operating costs. Over a wide range of possible electricity prices, the hybrid system of both units is shown to be more cost-effective than a single RPB or PB. This synergistic effect is uncovered with the integration of process intensification and superstructure optimization, illustrating a new way for both PI and PSE to incorporate.

Keywords: Process intensification, Process synthesis, Rotating packed Bed, Nonlinear programming

1. Introduction

Process intensification, a promising pathway in the development of sustainable and cost-effective chemical process systems(Van Gerven et al., 2009),is defined by Ramshaw(1995) as a methodology for making remarkable reduction in equipment size, energy consumption, or waste generation while achieving a given production goal. PI has become a topic of increased interest(Stankiewicz, Lutze et al. 2000,2010) with its main goal to coordinate different unit operations/processes effectively and to activate synergistic effects during process development(Huang, 2007). Process synthesis—the assembly and interconnection of units for efficient performance and low energy consumption(Weserberg, Grossmann, 1975, 2000) —is an excellent complement to process intensification(Moulijin, 2008). However, both of these areas are distinct to

some extent. Process intensification places greater emphasis on the micro-mechanisms of intensification and experimental validation of transfer phenomena(Moulijin, 2008). In contrast, process synthesis obtains an optimal configuration given a set of potential alternatives topologies for a process system. By reducing the bottleneck of transfer processes at the micro level, like mixing and reaction kinetics, such that process reactions are governed by their inherent rates rather than by transport phenomena(Baldea, 2015), PI can further improve solutions from process synthesis. Therefore, it is meaningful to obtain insight into the connection between process intensification and process synthesis.

2. Outline of integrated optimization approach

The rotating packed bed is a novel and compact PI operating unit(Ramshaw, Bucklin, et al., 1981, 1987) that utilizes centrifugal acceleration to intensify mass transfer. Its low investment cost leads to growing adoption in gas purification, distillation and nano-material preparation applications(Munjal, Kelleher, Chen, el at., 1989, 1996, 2002). Fig.1 shows the difference in size between the RPB and traditional PB for an industrial desulfurization process. Due to the 1-2 orders of magnitude increase in the volumetric mass transfer coefficient for the RPB, its volume is only 9% as large as the PB. This work examines the industrial desulfurization of cracking dry gas at a refinery in China(see Fig.1) as the specific case study to discuss the integration of process intensification and process synthesis. The dry gas contains about 1vol% H_2S, which is not only a hazard to human health and the environment, but also a burden to pipeline upkeep due to its corrosiveness. The RPB stands beside the packed bed. The dry gas enters the RPB and contacts with an absorbing solution of methyl-diethanolamine (MDEA) counter-currently. The sweet gas after desulfurization leaves the RPB, and goes into a clear dry gas pipeline with a maximum H_2S content specification of 20ppm.Most refineries employ a packed bed for desulfurization. However, conventional PBs require large volumes, leading to high investment costs and large absorbing solution circulation flows. The RPB makes up for these deficiencies of the PB, especially increases with limited space, such as the desulfurization of natural gas on an off shore drilling platform, where the RPB realizes its full potential.

Fig. 1 The comparison between RPB and PB within an industry case

On the other hand, the RPB is rotating equipment that requires significant energy input. As shown in Table 1, the RPB has a distinct advantage over the PB on the space it occupies. However, the RPB requires a motor to provide rotation. Therefore, there are several tradeoffs between RPB and PB. Process intensification and process synthesis

can be coupled to search for novel configurations that combine both technologies as shown in Fig. 2.

In this paper, nonlinear programming models are first presented to optimize the RPB and the PB separately. Next, a superstructure optimization model is proposed to synthesize configurations that integrate the advantages of the packed bed, which requires less energy input, and the rotating packed bed, which has high mass transfer efficiency leading to smaller sizes.

In order to synthesize a design involving RPB and PB units, the superstructure shown in Fig. 3(a) is proposed, in which the liquid and gas streams are distributed among potential streams for countercurrent flow. Gas streams and liquid streams are supplied to the two treatment units. After the gas streams are contacted counter-currently with liquid streams in the units, both gas and liquid are recycled to the treatment units or sent to the outlet stream, subject to a limit on the H_2S concentration of the sweet gas.

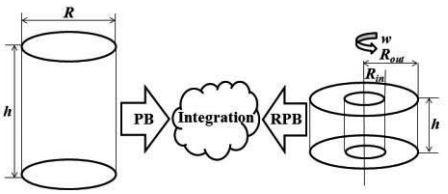


Fig. 2 Processes combining process intensification and synthesis

Table 1 Rotating packed bed vs packed bed

	Rotating Packed Bed	**Packed Bed**
Height	2.89 m	32 m
Equipment volume	3.4 m^3	36 m^3
Motor	37kW	--
H_2S content in sweet gas	<20ppm	<20ppm

The configuration between the RPB and the PB could be in series, as shown in Fig. 3(B), or in parallel, as shown in Fig. 3(C), and even in-between. This is decided by parameters determined from the optimization results. The goal is to minimize the total cost of the integrated system. The synthesis problem is formulated as a nonlinear programming (NLP) model that involves constraints corresponding to the performance of the RPB and the PB, the intensification mechanism of mass transfer, the end-effect quantification of the RPB, and mass balances in the process.

3. Problem statement

The sizes and design parameters of RPB and PB are variables to be determined in the superstructure of Fig 3a so as to meet the gas purifying specification and treatment capacity. The flow rates and concentration of the gas and the liquid in each stream of the superstructure are also variables to be determined. Using mass transfer models for the RPB and the PB, fairly accurate design and operating conditions can be obtained.

The objective is to select the optimal configuration involving the RPB and the PB so as to minimize the total investment and operating costs.

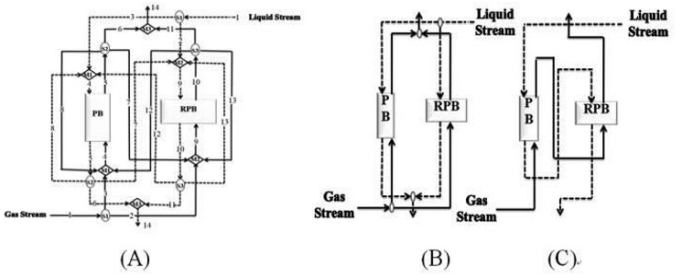


Fig. 3 Superstructure of the integrated RPB and PB processes: (A) overall structure; (B) in parallel configuration; (C) in series configuration. S denotes a splitter and M denotes a mixer, respectively. Gas streams are denoted by solid lines; liquid streams, by dashed lines.

The mathematical model of integrated RPB and PB processes in the superstructure of Fig 3a is NLP model, which consist of mass transfer process intensification mechanisms, fluid flow expressions and mass balance equations.

4.H_2S purification case study

In order to apply the NLP models in the previous section, we consider a case study on a desulfurization process for crack gas at a refinery in China which is shown in Fig.1. The gas contains about 1vol% H_2S and the solvent used is MDEA. This quantity of H_2S is not only a hazard to human health and the environment, but also exacts a high cost for pipeline repair and replacement. The wire mesh packing porosity, ε, is 0.97, and the total specific surface area, a, is 220 m^2/m^3. The liquid jet initial velocity is 6m/s upon entering the rotating bed. Other physical parameters for the H_2S purification case study are presented in Table2. We solve the NLP models for RPB (P1), for PB(P2) and superstructure (P3) using GAMS IPOPT 3.12 and CONOPT 3. The results are presented in Tables3.

Table 2 Parameters for H_2S purification case study

	Rotating Packed Bed	Packed Bed
Content of H_2S	10,000 (ppm)	10,000 (ppm)
Gas flow	10,000 (m^3/hr)	10,000 (m^3/hr)
Liquid film renewal times	90	--
Operating cost scale factor for Rotation,$c_{RPB,1}$	2.2(10,000 CNY/kw)	--
Operating cost factor for RPB volume, $c_{RPB,2}$	121(10,000CNY/m^3)	--
Investment scale factor for PB volume, $c_{PB,1}$	--	5(10,000CNY/m^3)

Operating cost scale factor for RPBliquid pumping, $c_{RPB,3}$	1 (10,000CNY·hr/m^3)	--
Operating cost scale factor for PB liquid pumping, $c_{PB,2}$	--	1.6 (10,000 CNY·hr/m^3)
Diffusion coefficient, D	1.2×10^{-9} (m^2/s)	1.2×10^{-9}(m^2/s)
Reaction rate constant of $H_2S$with MDEA, k_{ov}	1.5×10^{8}(1/s)	1.5×10^{8}(1/s)
H_2S content in exit gas	<100 (ppm)	< 100 (ppm)

Table 3.Comparison between separate reactors and solution from superstructure.

	RPB (Rotating packed bed)	**PB (Packed bed)**	**RPB-PB hybrid**	
			RPB	PB
Height/*m*	0.347	15.275	0.273	7.449
Radius/*m*	0.575 (Out) 0.192 (In)	0.544	0.452(Out) 0.151(In)	0.477
Volume/*m^3*	0.321	14.2	0.156	5.312
Rotating speed/*r/min*	600	--	660	--
Capacity for gas /*m^3/hr*	10008	10008	4845.6	5162.4
Liquid /*m^3/hr*	50	100	25.2	50.4
Equivalent annual cost for equipment /10,000 RMB	38.799	71.002	18.872	26.562
Annual electricity cost for rotor /10,000 RMB	25.915	--	6.318	--
Annual electricity cost for liquid pump /10,000 CNY	5.112	15.926	2.427	8.271
Structure	Individual	Individual	In parallel (b in Fig.3)	
Gross cost /10,000 CNY	69.826	86.928	62.449	

5. Conclusions

This paper investigates the conceptual synthesis of a process involving both conventional and intensified technologies. A superstructure optimization formulation is proposed to consider parallel, series, and in between configurations of the RPB and the PB, with recycle, with the overall goal to minimize the total annual cost while meeting gas capacity and effluent contaminant specifications. To incorporate the micro-mechanisms of intensification into the macro-level superstructure decisions, a multi-scale model was developed that captures the high-level effects of mass transfer in the RPB and the PB, allowing its use in synthesis. The model was applied to a case study involving H_2S removal from refinery tail gas. The computational results shows that the configuration of both RPB and PB in parallel is optimal under the assumption of linear investment and operating costs. Over a wide range of possible electricity prices, the hybrid system of both units is shown to be more cost-effective than a single RPB or PB. This synergistic effect is uncovered with the integration of process intensification and superstructure optimization, illustrating a new way for both PI and PSE to work together.

References

A.L. Stankiewicz, et al. Process intensification: transforming chemical engineering. *Chem. Eng. Prog*. 2000; 22-34.

A.W. Weserberg, et al. Studies in process synthesis - i. branch and bound strategy with list techniques for the synthesis of separation schemes. *Chem. Eng. Sci.* 1975; 30(8): 963-972.

BW Bucklin, et al. Higee contactors for selective H_2S removal and super dehydration. Laurance Reid Gas Conditioning Conference, University of Oklahoma, March 1987.

C Ramshaw. The Incentive for Process Intensification. *Proceedings of 1st International Conference of Process Intensification for Chemical Industry* 1995; London, 18:1.

C. Ramshaw, et al. Mass Transfer Process. U.S. Patent 1981;4283255.

I.E. Grossmann, et al. Research challenges in process systems engineering. *AIChE J.*, 2000; 46: 1700-1703.

JA Moulijin, et al. Process intensification and process systems engineering: a friendly symbiosis. *Comput. Chem. Eng.* 2008; 32(1-2): 3-11.

JF Chen. High Gravity Technology and Application Chemical Industry Press of China, 2002.

K. Huang, et al. Seeking synergistic effect—A key principle in process intensification, *Sep. Purif. Technol.*, 2007;57, 111-120.

M. Baldea et al. From process integration to process intensification. *Comput. Chem. Eng.* 2015; 81: 104-114.

P Lutze, et al. Process intensification: a perspective on process synthesis. *Chem. Eng. Pro. Process intensification* 2010; 49(6): 3-11.

S. Munjal, et al. Mass transfer in rotating packed beds. II. Development of gas−liquid and liquid−solid mass transfer coefficients. *Chem. Eng. Sci.* 1989; 44 (10): 2245−2256.

T. Kelleher, et al. Distillation studies in a high-gravity contactor. *Ind. Eng. Chem. Res.* 1996; 35(12): 4646-4655.

T. Van Gerven, et al. Structure, energy, synergy, time—the fundamentals ofprocess intensification. *Ind. Eng. Chem. Res.* 2009;48:2465–2474.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50392-X

Towards a systematic framework for the synthesis of operable process intensification systems

Yuhe Tian[a], M. Sam Mannan[a,b], Efstratios N. Pistikopoulos[a,c*]

[a]*Artie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX 77843, United States*
[b] *Mary Kay O'Connor Process Safety Center, Artie McFerrin Department of Chemical Engineering Texas A&M University, College Station, TX 77843, United States*
[c] *Texas A&M Energy Institute, Texas A&M University, College Station, TX 77843, United States*
stratos@tamu.edu

Abstract

In this work we present an integrated framework to derive intensified designs with guaranteed safety and operability performance. A step-by-step procedure is described featuring: (i) superstructure-based process synthesis using Generalized Modular Representation Framework, (ii) development of advanced multi-parametric control strategies following the PAROC framework and software platform, and (iii) inherent safety analysis at early design stage utilizing quantitative risk assessment. Two case studies with increasing intensification complexity, namely heat exchanger network synthesis and reactive distillation column design, are investigated to demonstrate the capabilities of the proposed framework.

Keywords: Process intensification, Generalized Modular Representation Framework, multi-parametric Model Predictive Control, risk assessment

1. Introduction

Modular process intensification (PI) offers the potential to drastically reduce the energy consumption and processing cost by involving multifunctional phenomena at different time and spatial scales (Van Gerven and Stankiewicz, 2009). In this context, phenomena-based synthesis tools, comparing to their conventional unit-operation-based counterparts, have attracted much interest as a potential means for automated generation of intensified options from a lower aggregation level without a pre-postulation of plausible flowsheets. Recent progress in this field includes, but not limited to, computer-aided software-tool for phenomena-based multiscale and multistage sustainable process synthesis – intensification by Tula et al. (2017), Infinite DimEnsionAl State-space (IDEAS) framework application on energetic intensification of hydrogen production (Pichardo and Manousiouthakis, 2017), and systematic building block approach by Demirel et al. (2017). However, it is worth noting that the afore-derived intensified designs often reduce the degrees of freedom for safe operation and process control due to high integrity, thus arousing concerns from process safety and operability perspective (Baldea, 2015). In this way, further attention is required to simultaneously synthesize intensified but safely operable systems.

The objective of this paper is to introduce a systematic framework to deliver intensified designs through the phenomenological Generalized Modular Representation Framework

(GMF), and at the same time, to guarantee their safety and operability performance by conducting risk assessment and model predictive controller design. The full framework (Figure 1a) and its detailed methodologies (Figure 1b) are described in section 2. In section 3, the framework is applied to heat exchanger network (HEN) synthesis for thermal intensification. A further application on hybrid reaction/separation processes is demonstrated through a reactive distillation case study in section 4.

2. Framework Description

As shown in Figure 1a, the proposed framework consists of three interactive toolboxes (i.e., process intensification/synthesis toolbox, process simulation/optimization toolbox, and process operability/control/safety toolbox) to link steady-state synthesis, dynamic analysis, and operability assessment at the conceptual design stage. Each toolbox can accommodate different techniques implemented in multiple software environments, thus rendering this framework a desirable flexibility. Verifiable and operable process intensification designs are delivered as the outcome. The procedure followed for this purpose and specific techniques for each toolbox in the current work are presented below on a step-by-step basis (Figure 1b).

Step 1: Process Intensification / Synthesis Representation – Generalized Modular Representation Framework (GMF), presented originally in Papalexandri and Pistikopoulos (1996), is utilized for process synthesis from phenomena level. The GMF uses aggregated multifunctional mass/heat exchange modules and pure heat exchange modules to represent chemical processes, conventional or not, by detecting mass and heat transfer feasibilities based on Gibbs free energy (Ismail et al., 2001). Thus, the discovery of intensified pathways can in principle be achieved by optimizing fundamental performance to overcome process bottlenecks, such as improving mass/heat transfer and/or shifting reaction equilibrium.

Step 2: Process Optimization – A superstructure formulation is developed based on GMF to account for the large search space of intensification possibilities. The overall problem is formulated as a mixed integer nonlinear programming (MINLP) problem and solved using GAMS.

Step 3: Process Operability / Safety Assessment – After obtaining the initial optimal design from the aforementioned steps, feasibility test is performed to evaluate its functionality under varying operating conditions. A multiperiod MINLP model is then formulated considering the detected critical operation points. Moreover, risk assessment is incorporated as a constraint into the synthesis model to indicate the inherent safety performance of the resulting design, accounting for consequence severity and equipment failure frequency (Nemet et al., 2017).

Step 4: Optimal Intensified steady-state designs – Up to this point, optimal intensified designs satisfying flexibility and safety targets are derived from steady-state analysis.

Step 5: Dynamic Simulation and Control – "High fidelity" dynamic model of the resulting PI designs determines the validity of dynamic analysis. This modeling task takes place in gPROMS® Process Systems Enterprise (1997-2017). The development of multi-parametric model-based predictive controller (mp-MPC) is performed through the application of the PAROC framework and software platform (Pistikopoulos et al., 2015). Its simultaneous design and control step gives optimal design via the formulation of a (mixed-integer) dynamic optimization problem (Diangelakis et al., 2017).

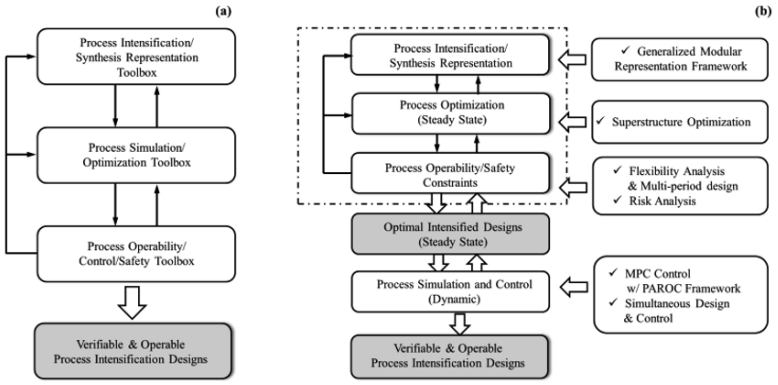


Figure 1. Proposed framework for the synthesis of operable intensification systems (a) The systematic framework, (b) Methodology flow

Step 6: Verifiable and Operable Process Intensification Designs – Closed-loop validation is performed to ensure the consistency throughout the framework, after which verifiable and operable intensification designs will be delivered as the outcome.

3. Heat Exchanger Network Synthesis Example

The afore-described framework is first applied to investigate a heat exchanger network (HEN) synthesis problem for thermal intensification, aiming to demonstrate the steps of the proposed framework as an example from a broader set of intensified systems.

3.1. Problem Description

This case study consists of two hot streams (H1, H2), two cold streams (C1, C2), and one hot utility (HU). Given are: (i) stream flowrate data (Kotjabasakis and Linnhoff, 1987); (ii) uncertain heat transfer coefficient ($U_{H1\text{-}C1}$), for which the flexibility of the network is desired if this stream match is selected; (iii) disturbance and control objective, for which controller design is essential; and (iv) stream toxicity, represented by LC_{50} (i.e., lethal concentration, 50%) of the substance, and equipment data for four types of heat exchanger (HE), namely double pipe HE (DP), plate and frame HE (PF), fixed plate shell and tube HE (SF), and U-tube shell and tube HE (UT), which necessitates inherent safety evaluation (Nemet et al., 2017). The objective is to synthesize a HEN with minimized total annual cost as desired operability performance.

3.2. Steady-state Synthesis with Flexibility and Safety Considerations

The pure heat exchange module of GMF is used here for HEN representation. The superstructure representation accounts for all potential stream matches and their corresponding bypass options. A nominal HEN design without flexibility or safety considerations is first performed to obtain a reference network (Figure 2a). Its risk value is calculated as per introduced in Section 2 Step 3. Two periods of operation are identified by the feasibility test as the extreme values of $U_{H1\text{-}C1}$. To obtain an inherently safer design, risk tolerance is decreased by constraining: (i) overall HEN toxicity risk less than 75% of the nominal risk value, and (ii) individual HE toxicity risk less than 50% of overall risk. Setting the first risk tolerance results in the change of H1-C2 exchanger type from SF to UT (Figure 2b), as the latter one has much higher area density, leading to significant reduction in HE volumes and in the amount of intrinsic

hazards. The risk level is further decreased by confining individual equipment risk so that none of the HEs is particularly risky. In this case, a different network is synthesized to render H1-C2 HE a lower risk level by relieving its heat exchange burden (Figure 2c). These two safely operable HENs separately serve as Candidate 1 and 2 for the next step dynamic analysis.

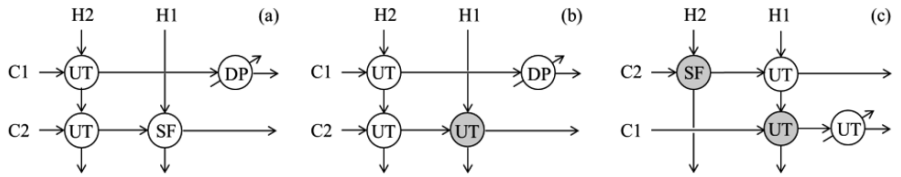


Figure 2. Topology of HEN at different operability level:
(a) Nominal design, (b) Candidate 1, (c) Candidate 2

3.3. Dynamic Modelling and mp-MPC Controller Design

The dynamic HEN is described by a Partial Differential Algebraic Equation (PDAE) model based on gPROMS® Process Model Library for Heat Exchange (PSE, 1997-2017). The network configuration is fixed for each candidate case study as per Figure 3b and 3c, and heat exchanger areas are used as design variables. To ensure the consistency going from steady-state synthesis to dynamic simulation, the dynamic model is validated to match its steady-state synthesis analogue.

In this system, the bypass flowrate and the heat utility duties are the degrees of freedom, and the outlet temperature of stream H1 and C1 are the outputs. The inlet temperature of stream H2 is treated as a disturbance to the operation. The mp-MPC controller design, following PAROC framework (Pistikopoulos et al., 2015), takes place for the two candidates individually. Each mp-MPC problem is formulated using corresponding linear state-space model approximated by the System Identification Toolbox of MATLAB®. Via POP® toolbox in MATLAB®, the problem of Candidate 1 is solved for an output horizon of 2 and a control horizon of 2 resulting in 118 critical regions in solution map, while that of candidate 2 is solved for an output horizon of 2 and a control horizon of 1 resulting in 89 critical regions. Given random disturbances deviating within ±10 K per second, the designed controllers are tested against the "high fidelity" model for closed-loop validation, indicating the agreement between the outputs and setpoints.

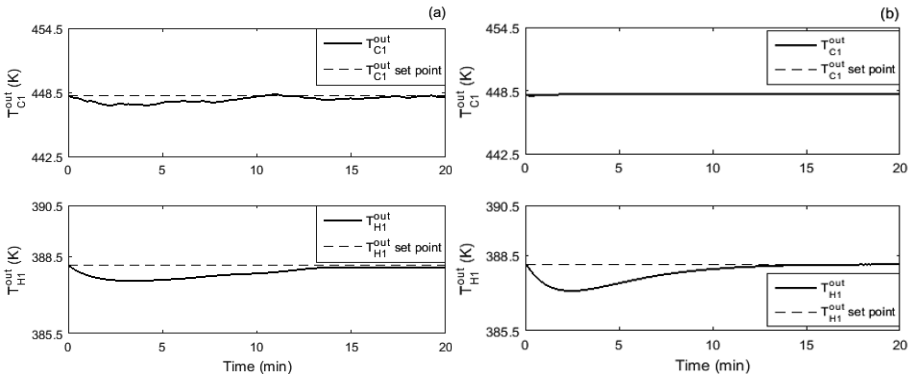


Figure 3. Closed-loop validation of the controller against the high fidelity model:
(a) Candidate 1, (b) Candidate 2

The dynamic optimization problem is then formulated and solved for the minimal total annual cost under each HEN configuration. The obtained results are shown in Table 1. Up to this point, two HENs are designed with different levels of operability, control, and safety. While the final construction decision depends on the trade-off between desired operability behavior and economic performance, this framework demonstrates the potential for comparison of various operable design alternatives.

Table 1. Optimization results for HEN example

	HE Area (m^2)				Investment	Operating Cost
	H1 – C1	H1 – C2	H2 – C1	H2 – C2	Cost (k$)	(k$/yr)
Candidate 1	/	604.4	125.6	99.9	528.4	7357.0
Candidate 2	221.4	522.2	/	165.7	557.4	8593.9

4. Reactive Distillation Example

In this section, the proposed framework is applied on hybrid separation/reaction systems to exploit the synergetic potential of this representative intensification system. The synthesis problem of classical methyl-tertiary-butyl-ether (MTBE) reactive distillation, originally presented in Ismail et al. (2001), is revisited herein for the purpose of demonstration, and to be further investigated for its operability performance.

The design problem aims to produce MTBE bottom product at a rate of 197 mol/s and with the purity higher than 98 mol%. Two types of multifunctional mass/heat exchange modules (i.e., reaction/separation module and pure separation module) and the pure heat exchange modules in GMF are utilized to detect intensification possibilities with respect to the minimal operating and pseudo-capital cost. The optimal design is illustrated in Figure 4. The structure features two hybrid modules for reactive section, one separation module at the bottom separating unreacted methanol back to reactive section, and another separation module at the top transferring n-butene to distillate. This result is consistent while slightly improved comparing to the literature results.

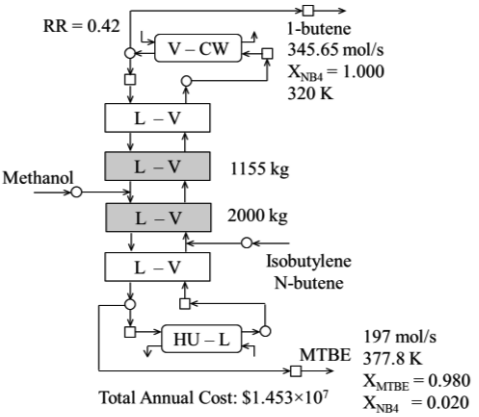


Figure 4. Optimal GMF design of MTBE reactive distillation

In the dynamic analysis, the "high fidelity" dynamic rate-based model of MTBE reactive distillation column is adapted from Schenk et al. (1999), represented by sets of differential algebraic equations (DAE) of index 1. In this system, the design variables include column diameter, reboiler and condenser heat exchange areas. Reflux ratio, steam flow rate, and cooling water flow rate are the three degrees of freedom. Sinusoidal disturbance in the isobutylene inlet composition is introduced while product specification is to be satisfied via the design of mp-MPC controller.

4. Conclusions

In this paper, we present a systematic framework for the design of intensified system with guaranteed operability and safety performance. The framework is applied to two representative intensification systems (i.e., heat integration and reactive distillation). The results indicate that significant improvement can be achieved to exploit intensification opportunities by utilizing phenomena-based Generalized Modular Framework, and to enhance safety and control by performing corresponding assessment at the early design stage. The full framework enables to complete the whole story of synthesizing operable intensification systems. Future work will focus on applying the framework to more intensification systems, as well as zooming in the techniques employed in each toolbox.

5. Acknowledgements

Financial supports from Texas A&M Energy Institute, Shell, and RAPID SYNOPSIS Project are gratefully acknowledged.

References

M. Baldea, 2015. From process integration to process intensification. *Computers & Chemical Engineering*, *81*, 104-114.

S.E Demirel, J. Li, and M.F. Hasan, 2017. Systematic process intensification using building blocks. *Computers & Chemical Engineering, 105*, 2-38.

N.A. Diangelakis, B. Burnak, J. Katz, and E.N. Pistikopoulos, 2017. Process Design and Control optimization: A simultaneous approach by multi-parametric programming. *AIChE Journal*.

S.R. Ismail, P. Proios, and E.N. Pistikopoulos, 2001. Modular synthesis framework for combined separation/reaction systems. *AIChE Journal*, *47*(3), 629-649.

E. Kotjabasakis, and B. Linnhoff, 1987. Flexible Heat Exchanger Network Design: Comments on the Problem Definition and on Suitable Solution Techniques. In *IChemE Symposium" Innovation In Process Energy Utilization*, 1-17.

K.P. Papalexandri, and E.N. Pistikopoulos, 1996. Generalized modular representation framework for process synthesis. *AIChE Journal*, *42*(4), 1010-1032.

P. Pichardo, and V.I. Manousiouthakis, 2017. Infinite DimEnsionAl State-space as a systematic process intensification tool: Energetic intensification of hydrogen production. *Chemical Engineering Research and Design*, *120*, 372-395.

E.N. Pistikopoulos, N.A. Diangelakis, R. Oberdieck, M.M. Papathanasiou, I. Nascu, and M. Sun, 2015. PAROC—An integrated framework and software platform for the optimisation and advanced model-based control of process systems. *Chemical Engineering Science*, *136*, 115-138.

Process Systems Enterprise, 1997-2017. gPROMS. URL http://www.psenterprise.com/gproms/

A. Nemet, J.J. Klemeš, I. Moon, and Z. Kravanja, 2017. Safety Analysis Embedded in Heat Exchanger Network Synthesis. *Computers & Chemical Engineering*. https://doi.org/10.1016/j.compchemeng.2017.04.009

M. Schenk, R. Gani, D. Bogle, and E.N. Pistikopoulos, 1999. A hybrid modelling approach for separation systems involving distillation. Chemical Engineering Research and Design, 77(6), 519-534.

A.K. Tula, D. Babi, J. Bottlaender, M.R. Eden, and R. Gani, 2017. A computer-aided software-tool for sustainable process synthesis-intensification. *Computers & Chemical Engineering, 105*, 74-95.

T. Van Gerven, and A. Stankiewicz, 2009. Structure, energy, synergy, time - The fundamentals of process intensification. Industrial & engineering chemistry research, 48(5), 2465-2474.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50393-1

Efficiency Analysis of Chemical-looping Fixed Bed Reactors integrated in Combined Cycle Power Plants

Chen Chen, George M. Bollas*

Department of Chemical & Biomolecular Engineering, University of Connecticut, 191 Auditorium Road, Unit 3222, Storrs, CT 06269, USA.
george.bollas@uconn.edu

Abstract

In this work, dynamic modeling is used as a tool to analyze the performance of chemical-looping combustion (CLC) integrated with combined cycle (CC) power plants. Fixed bed CLC reactors are incorporated into CC power plants. Efficiency improvements are achieved by optimizing the process design and managing CLC operating strategies. Specifically, this work presents the effect of operating variables on the performance of CLC-CC. The optimal CLC-CC power plant model has a time-averaged efficiency of 51.84%, power output of 223.6 MW, and CO_2 capture efficiency of 96%. The main factor that limits plant efficiency is the maximum reactor temperature, which is constrained by the materials of oxygen carriers. Moreover, it is shown that the impact of the temperature of air fed to the CLC heat removal stage and the pressure ratio are more significant than other process variables.

Keywords: Chemical-looping combustion, CO_2 capture, power generation.

1. Introduction

Greenhouse gas emissions have been identified as a main cause for detrimental global climate change. Fossil fuel combustion mainly contributes to the increasing emissions of CO_2, the key constituent of greenhouse gases. However, due to the intermittent nature of renewable energy sources, the use of fossil fuel remains necessary to balance the electricity grid and avoid other cascading problems. Greenhouse gas emission mitigation can be also accomplished by reducing fossil fuel consumption or by improving the efficiency of energy generation and utilization. Switching to less carbon intensive fuels also reduces CO_2 emissions. Compared with coal, the combustion of natural gas has a smaller carbon footprint and higher thermal efficiency. Combined cycles (CC), using natural gas, can reach higher efficiencies of up to ~60%, which is by far the highest efficiency in thermal power processes. Moreover, carbon capture and storage (CCS) technologies have the potential to accomplish very low levels of CO_2 emissions. Among CCS technologies, chemical-looping combustion (CLC) is a promising approach to capturing CO_2 with a low energy penalty. In CLC, the fuel and air remain unmixed by using a metal oxide as oxygen carrier (OC) to transport oxygen from the air reactor (where OC is oxidized by air) to the fuel reactor (where OC is reduced by fuel). The produced H_2O and CO_2 are inherently separated from other gases in the process avoiding the need for additional equipment or energy consumption for the separation of CO_2. Therefore, process options should include CC and CLC to reduce CO_2 emissions per unit of energy. Fluidized bed CLC reactors, which were largely the focus of prior work, are not able to

maintain stable solids circulation at the high pressure required for CC power plants. Therefore, fixed bed reactor designs were proposed to circumvent the solids separation and loop sealing problems under high pressure. Moreover, CLC fixed bed reactors are easy to scale to meet requirements for intensification and modularization, pressurize, and can handle relatively large range of particle sizes. Due to their compact design, lower capital cost, better utilization of OC, and smaller process footprint, fixed bed reactors are excellent candidates as process intensification options for CLC-CC plants.

System-level dynamic simulation and optimization play a significant role in exploring the feasibility of integrating CC with CLC. For example, Spallina et al. (2014) simulated a series of CLC-CC power plants showing average efficiencies of 41% at CO_2 capture efficiency of 97%. Chen et al. (2016) presented a dynamic model of CC with CLC fixed bed reactors, and showed that the CC performance is only slightly affected by the intrinsic dynamics of the semi-batch operation of fixed beds. Prior research presented the integration CLC with CC, and the static and dynamic estimation of a CC power plant with semi-batch fixed bed CLC reactor (Chen et al., 2016). The power plant model with a simple control architecture was simulated and validated against full load data reported in the literature. The conventional combustor was replaced by a high-pressure CLC island including fixed bed reactors operating in parallel. This analysis showed that the CC power plant generates relatively stable electricity output despite the intrinsic dynamics of fixed beds. This was accomplished by optimizing the CLC operational strategy for stability and thermal efficiency. The estimated efficiency of CLC fixed beds with CC was estimated at ~48% with a small fluctuation of ~2% around the desired steady state.

The objective of this work is to further improve the overall efficiency of the integrated CLC-CC power plant, wherein the conventional combustor is replaced by a CLC island. In the following, the CLC-CC process is described, and the configuration and arrangement of the power plant studied is presented. Thereafter, the optimal control formulation for CLC operation is discussed. The last section communicates efficiency evaluations of the integrated CLC-CC power plant. The efficiency analyses present the impact of certain variables on the power plant operations and potential efficiency improvements by adjusting selected manipulated variables and plant degrees of freedom.

2. Description of the CLC-CC Process

In a typical CC process, fuel is mixed with compressed air in the combustor to generate high-temperature flue gas, which is then expanded in the gas turbine to generate electricity. The gas turbine exhaust feeds a steam (Rankine-type) cycle to evaporate and superheat water, which is expanded in steam turbines to generate additional electricity. In this work, a natural-gas-fueled CC power plant was used as the reference case. This CC power plant has an efficiency of 57.9% and power output of 250 MW. The pressure ratio used for the gas turbine is 30:1. The temperature of flue gas fed to the gas turbine is 1440 °C. The superheated steam is at 565 ℃ and 160 bar. The model development and validation of this commercial system were presented in prior work (Chen et al., 2016).

Figure 1 presents the key design aspects of a CLC-CC plant, where the conventional combustor is replaced by a CLC island. Air feeds the oxidization reactor, and fuel is fed to the reduction reactor. A heat removal stage is used to extract the heat generated by the oxidation of the metallic oxygen carrier, sending high-pressure/high-temperature gas to be expanded in the gas turbine of the CC plant. The exhaust of the gas turbine is fed to a steam cycle. Two new HRSGs (HRSG3 and HRSG4) are incorporated, as compared to

previous designs presented in Chen et al. (2016). The heat of the purge exhaust is utilized to superheat water in the HRSG3. The oxidation exhaust firstly superheats the steam in the HRSG4, and then the HRSG4 exhaust preheats the compressed air.

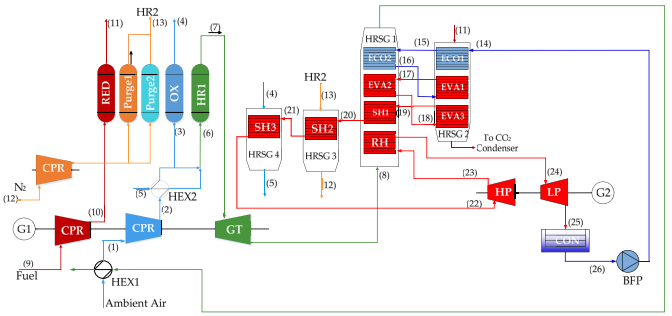


Figure 1. Diagram of the optimal CC power plant with fixed bed CLC reactors.

3. CLC Optimal Control Formulation

The fixed bed CLC reactor was represented with a heterogeneous reactor model (Han et al., 2016a, 2016b) This model includes conservation equations of mass, energy and momentum. Reaction kinetics models were developed on the basis of experimentation and model-based optimal design of experiments. These kinetic models were modified to account for pressure effects with empirical correlations adjusting the atmospheric pressure kinetics, which was verified for accuracy in high-pressure CLC experiments. Overall, the reactor models used in this work are sufficiently accurate to describe the operation of large-scale fixed bed systems, as shown in Zhou et al (2015a, 2015b). The kinetic model is extrapolated in term of temperature range from the developed earlier, to illustrate the benefits of high temperature on CLC-CC efficiency. Using these reactor models an optimal CLC-CC configuration was designed, which included several semi-batch CLC reactors operating in parallel (in different operating mode – reduction, purge, oxidation, heat removal, purge). The integrated CLC island was optimized for thermal efficiency by manipulating the time delay between reactors and operating variables, as discussed in the following section. For a given plant arrangement and oxygen carrier, the optimal control strategy searches for decision variable values that maximize CLC efficiency, while always satisfying plant steady state and dynamic constraints. The system model and optimization were implemented and solved in the commercial software package gPROMS (Process Systems Enterprise, 2014).

Table 1. Control variables in the optimal control problem

Control variables	Notation
Air feed rate Air temperature Fuel feed rate Fuel temperature	$\mathbf{u}_i$
Reduction time interval Oxidation time interval Heat removal time interval	τ_i

As shown in Table 1, the control profile for switching gas streams to the CLC reactor employs piecewise constant functions, represented as $\mathbf{u}(\tau_i)=\mathbf{u}_i$, where $\mathbf{u}$ is the vector of feed rate and temperature of the gas streams and τ_i is the time duration of the i-th CLC

step, i.e., reduction, purge, and oxidation. The set of control variables is summarized in the design vector, represented as $\boldsymbol{\phi}=[\mathbf{u}_i,\boldsymbol{\tau}, \boldsymbol{\omega}]$, which is constrained by upper and lower limits permitted in the design space, $\boldsymbol{\Phi}$.

3.1. Optimization Formulation

Typical targets for CLC-based power plants are ≥ 96% CO_2 capture efficiency and ≥ 98% fuel conversion, as written in Eqs.(1)-(2). Additional constraints defined by power plant considerations included $T_{out}(t)$ (heat removal exhaust gas temperature) stability (within a set tolerance, δ), limits to the pressure drop across the reactor, and maximum allowable internal temperature to ensure material stability, as shown in Eqs.(3)-(5). The overall objective of this optimization was to maximize the CLC power generation capability over the cyclic operation of the fixed bed by varying the control variables shown in Table 1. For the CLC island of the CLC-CC plant, efficiency was represented as the fraction of enthalpy of gas fed to gas turbine. In Eq.(6), T_{out}, $\dot{m}_{out}$, h_{out} are the temperature, mass flow rate, and enthalpy of exhaust streams, τ_{HR} is the heat removal duration, τ_{cycle} is the interval of one complete redox cycle. **f** is the set of differential algebraic equations representing the fixed bed reactor for states **x**, admissible inputs $\mathbf{u}_i$, and cycle duration $\boldsymbol{\tau}_i$. The time horizon for the optimization is at least twice that of the τ_{cycle}, because it takes more than one redox cycle to achieve the cyclic steady state.

$$S_{CO_2}(t)= \frac{\int_{t_0}^{\tau RED} F_{out}(CO_2,t)dt}{\int_{t_0}^{\tau RED} F_{in}(CH_4,t)dt} \geq 96\%, \tag{1}$$

$$X_{fuel}(t)= 1-\frac{\int_{t_0}^{\tau RED} F_{out}(CH_4,t)dt}{\int_{t_0}^{\tau RED} F_{in}(CH_4,t)dt} \geq 98\%, \tag{2}$$

$$TIT-\delta \leq T\_out\ (t) \leq TIT+\delta, \tag{3}$$

$$\left(\frac{\Delta P}{P_{in(t)}}\right) \leq \Delta P^{max}, \tag{4}$$

$$T(t,z)+\delta \leq T^{max}, \tag{5}$$

The control optimization problem was formulated as follows:

$$\max \eta_{HR} = \max \frac{\int_{t_0}^{\tau_{HR}} (\dot{m}_{out}(t)\cdot h_{out}(t))dt}{\int_{t_0}^{\tau_{cycle}} (\dot{m}_{out}(t)\cdot h_{out}(t))dt}$$

subject to: Eq. (1)-(5)

$$\begin{aligned}
&\mathbf{f}(\dot{\mathbf{x}}(t),\mathbf{x}(t),\mathbf{u}(t),\boldsymbol{\theta},t)=0,\\
&\mathbf{f}(\dot{\mathbf{x}}(t_0),\mathbf{x}(t_0),\mathbf{u}(t_0),\boldsymbol{\theta},t_0)=0,\\
&\mathbf{y}(t)=\mathbf{h}(\mathbf{x}(t)),\\
&\mathbf{x}^{\min}\leq\mathbf{x}(t)\leq\mathbf{x}^{\max}\\
&\mathbf{u}_i^{\min}\leq\mathbf{u}_i\leq\mathbf{u}_i^{\max} \qquad \forall \boldsymbol{i}\in[\mathbf{1},\boldsymbol{N_u}]\\
&\tau_i^{\min}\leq\tau_i\leq\tau_i^{\max} \qquad \forall \boldsymbol{i}\in[\mathbf{1},\boldsymbol{N_u}]
\end{aligned} \tag{6}$$

3.2. Case study: Fixed bed CLC with NiO and CH_4

The case study discussed in this section optimizes a Ni-loaded fixed bed CLC reactor with CH_4 as the fuel and the corresponding constraints for optimal control of the CLC-CC power plant. The CLC reactors are designed so that they can handle the same heat input as that of the reference CC power plant. Table 2 shows the specific parameters for a single fixed bed reactor. As shown in Figure 2(a), the oxidation exhaust achieves the

desired set-point with minimal deviations. The heat removal step initiates when the exhaust temperature is within 50 °C from *TIT*. After most of the heat has exited the bed and $T_{out}(t) < TIT\text{-}50$, the reactor is briefly purged and the feed switches to CH_4. The fuel flow rate is much lower than the air feed rate, which explains the sudden drop in exhaust gas enthalpy. In Figure 2(c), there is a sudden drop in O_2 and rise in N_2 when the cycle switches from oxidation (OX) to heat removal (HR), because the OX exhaust was mixed with inlet air during HR, to improve overall energy efficiency.

Table 2. Parameters of the fixed bed CLC reactor

Parameter	Value	Parameter	Value
Weight content of metal oxide	40 wt.% NiO	CH_4 flow rate	8.626 kg/s
Particle diameter	5 mm	Max. reactor temperature	1300 °C
Reactor length	9.4 m	Max. reactor pressure	30 bar
Reactor diameter	4.7 m	Turbine inlet temperature, *TIT*	1250 °C
Max. pressure drop	0.03 bar	Temperature tolerance, δ	50 °C

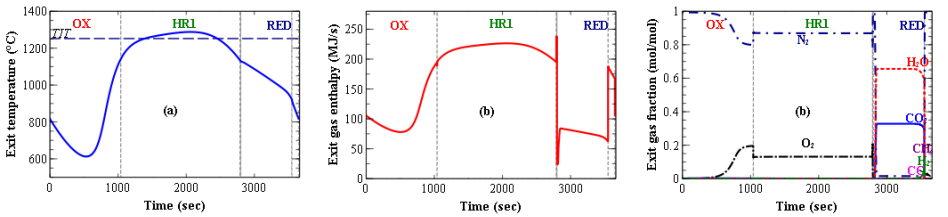


Figure 2. Dynamic profiles of exit gas (a) temperature, (b) enthalpy, and (c) composition.

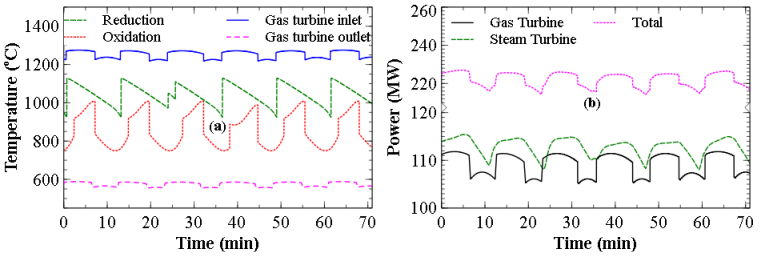


Figure 3. (a) Temperature profiles of the gas streams; (b) dynamic power performance.

3.3. Efficiency Evaluation of CLC-CC

The exhaust gas from each CLC reactor was mixed according to its originating cycle, thus the temperature profiles are less fluctuating than those of the gases from a single reactor, as shown in Figure 3 (a). The exhaust gases can be grouped as: CO_2/H_2O from RED, N_2/O_2 from OX, and air from HR. Here, the CC model was developed in a commercial software Dymola (Elmqvist et al., 1996) and integrated with the CLC model developed in gPROMS. The gPROMS CLC model outputs were brought into the gas turbine in the Dymola CC model as boundary conditions. The stream data of gas fed to the CLC reactor island was matched to those generated by the compressor in the Dymola plant model. The overall plant efficiency was improved by manipulating variables, such as the pressure ratio (*PR*), temperature of the gas compressor feed (T_{CPR}), temperature of the combustor feed (T_{air}), and mass flow of feed water ($\dot{m}_{FW}$). T_{air} is the temperature of air fed to the OX

and HR stages, T_{OX} andT_{HR}, respectively. Table 3 shows the impact of variables on the gas power output (P_{GT}), steam power out (P_{ST}), and total power output (P_{TOT}), and efficiency improvements ($\Delta\eta$).The optimal CLC-CC power plant has a time-averaged efficiency of 51.84%, time-averaged power output of 223.6 MW, and CO_2 capture efficiency of 96%. The efficiency is ~6% points lower than the reference CC power plant, which mainly results from the lower *TIT* reached in CLC (Naqvi and Bolland, 2007).

Table 3. Impact of variables on the power plant performance. *

Variables	Range	P_{GT}	P_{ST}	P_{TOT}	$\Delta\eta$
T_{CPR} [°C]	[23,50]	-	+	-	0.39%
T_{OX} [°C]	[480,550]	+	-	+	0.51%
T_{HR} [°C]	[480,520]	+	-	+	3.83%
PR [-]	[10,30]	-	+	max**	0.7%
$\dot{m}_{FW}$ [kg/s]	[57,76]	/	-	-	0.19%

* "+" means that increasing the variable improves the power generation, while "-" means the opposite; ** PR=20:1, the total power generation reaches the maximum value.

4. Conclusions

The efficiency of an integrated CLC-CC power plant was improved with an optimal CLC control strategy, modification of power plant arrangement and design, and manipulation of the remaining plant degrees of freedom. It was shown that an optimal CLC-CC power plat can reach efficiencies of up to 51.84% at CO_2 capture efficiency of 96%. *PR* and T_{HR} were shown to be more significant than other plant admissible inputs. Higher T_{HR} values originate from higher reactor temperatures, which are only feasible with oxygen carriers that can tolerate these high temperatures. Future oxygen carrier materials research should focus on enabling higher reactor temperatures, which leads to higher CLC-CC power plant efficiency.

References

Chen, C., Han, L., Bollas, G.M., 2016. Dynamic Simulation of Fixed-Bed Chemical-Looping Combustion Reactors Integrated in Combined Cycle Power Plants. Energy Technol. 4, 1209–1220.

Elmqvist, H., Brück, D., Otter, M., 1996. Dymola User's Manual.

Han, L., Zhou, Z., Bollas, G.M., 2016a. Model-based analysis of chemical-looping combustion experiments. Part I: Structural identifiability of kinetic models for NiO reduction. AIChE J. 62, 2419–2431.

Han, L., Zhou, Z., Bollas, G.M., 2016b. Model-based analysis of chemical-looping combustion experiments. Part II: Optimal design of CH4 -NiO reduction experiments. AIChE J. 62, 2432–2446.

Naqvi, R., Bolland, O., 2007. Multi-stage chemical looping combustion (CLC) for combined cycles with CO2 capture. Int. J. Greenh. Gas Control 1, 19–30. Process Systems Enterprise, 2014. The gPROMS platform, URL: http://www.psenterprise.com/gproms/platform.html

Spallina, V., Romano, M.C.C., Chiesa, P., Gallucci, F., van Sint Annaland, M., Lozza, G., 2014. Integration of coal gasification and packed bed CLC for high efficiency and near-zero emission power generation. Int. J. Greenh. Gas Control 27, 28–41.

Zhou, Z., Han, L., Bollas, G.M., 2015a. Model-assisted analysis of fluidized bed chemical-looping reactors. Chem. Eng. Sci. 134, 619–631.

Zhou, Z., Han, L., Nordness, O., Bollas, G.M., 2015b. Continuous regime of chemical looping combustion (CLC) and chemical-looping with oxygen uncoupling (CLOU) reactivity of CuO oxygen carriers. Appl. Catal. B Environ. 166–167, 132–144.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50394-3

Modeling Multistream Heat Exchangers (MHEXs) Using Operational Data

Harsha Nagesh Rao[a] and Iftekhar A. Karimi[a*]

[a]*Dept. of Chemical & Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117585*
cheiak@nus.edu.sg

Abstract

Multi stream heat exchangers (MHEXs) allow simultaneous heat exchange among multiple streams. These exchangers are usually used in energy-intensive applications such as natural gas liquefaction, air separation, etc. Minimizing the energy consumption of processes with existing MHEXs is crucial, but reliable system-wide operational optimization requires the knowledge of flow distribution and individual heat-transfer areas of the exchanging streams in MHEXs. However, MHEXs are virtually black boxes due to their proprietary designs and complex geometry. In this work, we present a generic non-linear programming (NLP) model to estimate the flow distribution and heat transfer areas for each pair of exchanging streams in an MHEX based on its historic operational data. We apply the developed geometry-independent model on the main cryogenic heat exchanger (MCHE) from an existing C3MR LNG plant and estimate its heat transfer parameters successfully. These parameters can then be used for the operational optimization of the process.

Keywords: MHEX, NLP, Optimization, LNG

1. Introduction

Multi stream heat exchangers (MHEXs) are individual process units that allow simultaneous exchange of heat among multiple streams. Mainly used in energy-intensive cryogenic processes like natural gas liquefaction, air separation, etc., MHEXs are complex, proprietary and often associated with phase changes of mixtures. An MHEX normally has several sections in series, called bundles (Hasan et al., 2009; Nagesh Rao & Karimi, 2017). These bundles can be viewed as independent MHEXs which may not necessarily have the same exchanging streams.

While numerous studies have addressed the design of processes employing MHEXs (Nagesh Rao & Karimi, 2017; Tsay et al., 2017), very few studies have attempted the operational optimization of processes employing MHEXs. Operational optimization of a process, in the context of this study, primarily involves obtaining optimal operating conditions for various process units (E.g. stream temperatures, pressure ratios for compressors, etc.) given MHEXs with known flow distribution and individual heat-transfer areas between each pair (hot and cold) of exchanging streams.

However, as MHEXs have proprietary designs and complex fluid flow patterns (Hasan et al., 2009; Nagesh Rao & Karimi, 2017), it is difficult to characterize their flow distribution and heat transfer areas. Hence, we need operational models estimate the flow distribution and the individual heat transfer areas for the streams exchanging heat in an MHEX. Furthermore, the estimated parameters can be used to predict the response of the MHEX for process variations in feed conditions, ambient temperature, etc. Without an

accurate predictive model, operators must depend on their qualitative experience and employ heuristics during process variations (Hasan et al., 2009).

When there is limited knowledge about the internals of the MHEX, simpler models which are applicable to all classes of MHEXs are preferable. However, very few generic operational models exist in literature. Hasan et al. (Hasan et al., 2009) modeled MHEX operation using a hypothetical network of two stream heat exchangers. They proposed a novel mixed-integer nonlinear programming (MINLP) formulation using operational data for finding the best equivalent network representing an MHEX. Their formulation incorporated phase change of streams but allowed only one cold stream. Also, they neglected heat leaks and pressure drop prevalent in MHEX operation. Khan et al. (Khan et al., 2012) modeled MHEXs using artificial neural networks. Jiang et al. (Jiang et al., 2015) developed an equation oriented predictive model for an MHEX in air separation units. They used a regression model to correlate the outlet temperatures with the inlet temperatures and flowrates. However, such data-trained models may not predict the performance accurately when the operation is away from training conditions (Hasan et al., 2009). As they have no physical significance, data-trained may not even ensure thermodynamic feasibility. Clearly, there is a need to develop a reliable and simple operational model for MHEXs to address the above shortcomings.

In this work, we propose a geometry-independent model for MHEXs using operational data. MHEX operation is described using a network of two-stream exchangers following Hasan et al. (Hasan et al., 2009). The model presented can handle multiple hot and cold streams, include ambient heat leaks, and consider pressure drops. Our approach enables us to deduce a series of exchangers that best represents the operation of the MHEX without the need for integer programming, making our formulation NLP rather than MINLP.

2. Problem Statement

Consider a bundle in an existing MHEX. $\boldsymbol{I}$ hot streams ($\boldsymbol{i = 1, 2, \dots, I}$) and $\boldsymbol{J}$ cold streams ($\boldsymbol{j = 1, 2, \dots, J}$) of known physical properties pass through this bundle. The properties include compositions, flowrates, inlet and outlet temperatures, and inlet and outlet pressures. These operational data are available for $\boldsymbol{N}$ distinct steady states ($\boldsymbol{n = 1, 2, \dots, N}$) of the bundle. Given these, we aim to estimate the flow distribution and heat transfer areas for each pair of heat-exchanging streams in the bundle that best describes its operation over all $\boldsymbol{N}$ states. We assume that the hot streams supply heat to cold streams only.

3. Model Formulation

The primary idea is to represent the operation of an MHEX bundle by a network of simple two-stream heat exchangers with known flow distribution and heat transfer areas for each pair of exchanging streams. Given the inlet temperatures of the exchanging streams, this network of exchangers must predict the outlet temperatures close to the actual temperatures. We use the single-stage superstructure (Figure 1) presented in our previous work (Nagesh Rao & Karimi, 2017) to derive this network.

We have N sets ($n = 1, 2, \dots, N$) of operational data. We split hot stream i into J substreams and cold stream j into I substreams to allow heat exchange in $I \times J$ parallel series (S_{ij}, $i = 1, 2, \dots, I$; $j = 1, 2, \dots, J$) of two-stream exchangers. Let S_{ijn} denote the series for data set n. Each S_{ijn} can have at most five of the exchangers ($k = 1, 2, \dots, 5$) to allow dedicated phase change intervals in the series (Nagesh Rao & Karimi, 2017). We

use s to denote any stream entering/leaving a bundle, hot (i) or cold (j), and define the following parameters. $F_{sn}, TI_{sn}, TO_{sn}, PI_{sn}, HI_{sn},$ and HO_{sn} be the molar flow, inlet temperature, outlet temperature, inlet pressure, outlet pressure, inlet molar enthalpy (specific), and outlet molar enthalpy (specific) of stream s in data set n. DP_{in} and BP_{in} be the specific molar enthalpy of hot stream i at its dew point and bubble point in data set n respectively. Similarly, DP_{jn} and BP_{jn} be the specific molar enthalpy of cold stream j at its dew point and bubble point in data set n respectively.

3.1. Estimation of flow distribution

Let f_{ij} denote the fraction of hot substream i flowing through S_{ij}, and g_{ij} denote the fraction of cold substream j flowing through S_{ij}. Therefore,

$$\sum_j f_{ij} = 1 \tag{1}$$

$$\sum_i g_{ij} = 1 \tag{2}$$

Consider an exchanger E_k in S_{ijn}, and call it E_{ijkn}. Let us define the following for E_{ijkn} as shown in Figure 1.

$TIN_{ijkn}, TOUT_{ijkn}$ = Inlet and outlet temperatures of the hot stream
$tin_{ijkn}, tout_{ijkn}$= Inlet and outlet temperatures of the cold stream
$PIN_{ijkn}, POUT_{ijkn}$ = Inlet and outlet pressures of the hot stream
$pin_{ijkn}, pout_{ijkn}$= Inlet and outlet pressures of the cold stream
$HIN_{ijkn}, HOUT_{ijkn}$ = Inlet and outlet specific molar enthalpies of the hot stream
$hin_{ijkn}, hout_{ijkn}$= Inlet and outlet specific molar enthalpies of the cold stream
Q_{ijkn}= Heat duty of E_{ijkn}

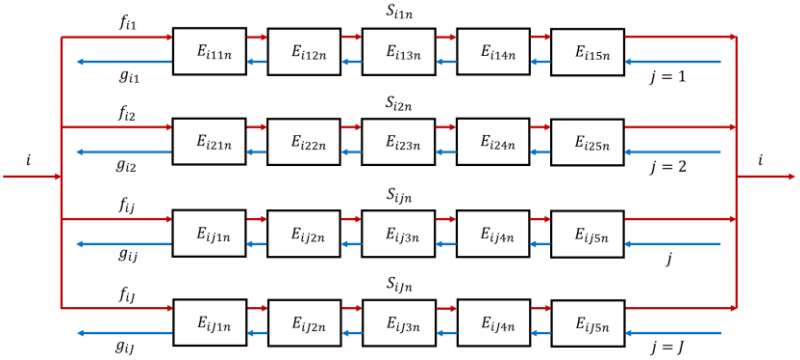


Figure 1. Schematic of our single-stage superstructure with $I \times J$ potential exchanges.

The inlet pressures and molar enthalpies of each series S_{ijk} are specified from the operational data, $HIN_{ij1n} = HI_{in}$, $hin_{ij5n} = HI_{jn}$, $PIN_{ij1n} = PI_{in}$ and $pin_{ij5n} = PI_{jn}$.

Then, in each series S_{ijk} (Figure 1), we have $HIN_{ijkn} = HOUT_{ij(k-1)n}$ and $PIN_{ijkn} = POUT_{ij(k-1)n}$ for $2 \le k \le 5$. Similarly, $hin_{ijkn} = hout_{ij(k+1)n}$ and $pin_{ijkn} = pout_{ij(k+1)n}$ for $1 \le k \le 4$. Next, we write the following related to the energy balance for E_{ijkn}.

$$HIN_{ijkn} \ge HOUT_{ijkn},\ hout_{ijkn} \ge hin_{ijkn} \tag{3}$$

$$Q_{ijkn} = f_{ij}F_{in}\left(HIN_{ijkn} - HOUT_{ijkn}\right) = g_{ij}F_{jn}(hout_{ijkn} - hin_{ijkn}) \tag{4}$$

We distribute the pressure drops across E_{ijkn} in proportions to their heat duties using the following equations.

$$Q_{ijn} = \sum_{k=1}^{5} Q_{ijkn} \quad (5)$$

$$(PIN_{ijkn} - POUT_{ijkn})Q_{ijn} = Q_{ijkn}(PI_{in} - PO_{in}) \quad (6)$$

$$(pin_{ijkn} - pout_{ijkn})Q_{ijn} = Q_{ijkn}(PI_{jn} - PO_{jn}) \quad (7)$$

Next, we prevent phase change in E_{ijkn} to using the following equations.

$$\left[HIN_{ijkn} - DP_{in}\right]\left[HOUT_{ijkn} - DP_{in}\right] \geq 0 \qquad 1 \leq k \leq 5 \quad (8)$$

$$\left[HIN_{ijkn} - BP_{in}\right]\left[HOUT_{ijkn} - BP_{in}\right] \geq 0 \qquad 1 \leq k \leq 5 \quad (9)$$

$$\left[hin_{ijkn} - DP_{jn}\right]\left[hout_{ijkn} - DP_{jn}\right] \geq 0 \qquad 1 \leq k \leq 5 \quad (10)$$

$$\left[hin_{ijkn} - BP_{jn}\right]\left[hout_{ijkn} - BP_{jn}\right] \geq 0 \qquad 1 \leq k \leq 5 \quad (11)$$

Then, the substreams of hot stream i from each series S_{ijn} mix to form the parent stream. The outlet enthalpy of stream i exiting the bundle for data set n is given by

$$\overline{\overline{HOUT}}_{in} = \sum_j f_{ij}.HOUT_{ij5n} \quad (12)$$

As heat leak/gain from the ambient may be prominent in some processes, outlet enthalpy of the cold stream j exiting the bundle for data set n is given by

$$\overline{\overline{hout}}_{jn} = \sum_i\left[g_{ij}.hout_{ij1n} \pm qc_j\left(g_{ij}.hout_{ij1n} - hin_{ij5n}\right)\right] \quad (13)$$

$$0 \leq qc_j \leq 1 \quad (14)$$

where, qc_j is the fraction of the total change in cold stream enthalpy due to heat leak/gain from the ambient. To ensure the minimum approach temperature (MTA_{ij}) in series S_{ij}, we have

$$TIN_{ijkn} \geq tout_{ijkn} + MTA_{ij},\ TOUT_{ij5n} \geq tin_{ij5n} + MTA_{ij} \quad (15)$$

We use the formulation provided in our previous work (Nagesh Rao & Karimi, 2017) to estimate the temperature of the exchanging streams as a function of their specific enthalpy and pressure. Finally, we define the objective function to estimate optimal flow fractions f_{ij} and g_{ij}. Our objective is to minimize the difference between the actual outlet temperatures and the predicted outlet temperature for hot and the cold streams. Hence, we minimize the squared error between the observed and predicted outlet enthalpies summed over the n data sets

$$Min.\ \sqrt{\sum_n\left[\ \sum_i\left[\left(HO_{in} - \overline{\overline{HOUT}}_{in}\right)\right]^2 + \sum_j\left[\left(HO_{jn} - \overline{\overline{hout}}_{jn}\right)\right]^2\ \right]} \quad (16)$$

Let us refer to the above equations as model 1 (M1). The result of M1 provides the optimal flow distribution between each hot stream and cold stream (f_{ij} and g_{ij}), and the duty of each segment in S_{ijn}. In the next section, we develop a separate NLP model M2 to estimate the heat transfer area A_{ij} of the series S_{ij}.

3.2. Areas for two-stream exchanges

Given the duty of each exchanger E_{ijk} in S_{ijn} from the solution of M1, we predict the

heat transfer areas A_{ij} for each series that best describes the operation of the bundle. The log mean temperature difference ($LMTD$) is of the form

$$LMTD_{ijkn} = \begin{cases} \frac{\left(\Delta T^1_{ijkn} - \Delta T^2_{ijkn}\right)}{\left[\ln\left(\Delta T^1_{ijkn}\right) - \ln\left(\Delta T^2_{ijkn}\right)\right]}, & \Delta T^1_{ijkn} \neq \Delta T^2_{ijkn} \\ \frac{\Delta T^1_{ijkn} + \Delta T^2_{ijkn}}{2}, & \Delta T^1_{ijkn} = \Delta T^2_{ijkn} \end{cases} \quad (17)$$

where, $\Delta T^1_{ijknm} = TIN_{ijkn} - tout_{ijkn}$ and $\Delta T^2_{ijkn} = TOUT_{ijkn} - tin_{ijkn}$ are the two temperature approaches. We follow Tsay et al. (Tsay et al., 2017) to estimate the tube side heat transfer coefficient α^{tube}_{ijknm} and the shell side heat transfer coefficient α^{Shell}_{ijknm} of the MHEX. Then, we have the overall heat transfer coefficient U_{ijkn} for E_{ijkn} given by

$$U_{ijkn} = \frac{\alpha^{tube}_{ijkn}.\alpha^{Shell}_{ijkn}}{\alpha^{tube}_{ijkn} + \alpha^{Shell}_{ijkn}} \quad (18)$$

Then, the heat transfer area for E_{ijkn} is given by

$$A_{ijkn} = \frac{Q_{ijkn}}{U_{ijkn} LMTD_{ijkn}} \quad (19)$$

The area A_{ij} of the heat exchangers in series S_{ij} can be predicted by minimizing the following objective function

$$Min. \sqrt{\sum_i \sum_j \sum_n \left(A_{ij} - \sum_k A_{ijkn}\right)^2} \quad (20)$$

4. LNG Case Study

We applied our models M1 and M2 on an existing main cryogenic heat exchanger (MCHE) of an LNG plant in Qatar employing the C3MR liquefaction process. We implemented these M1-M2 in BARON/GAMS 24.6.1 in a 64-bit Windows 7 desktop with a 3.6 GHz Intel Core i7 processor and 16 GB of RAM. As shown in the Figure 2, the MCHE consists of three bundles, viz. hot, mid and cold bundles. The hot and mid bundles have four hot streams viz. Natural gas (NG), mixed refrigerant liquid (MRL), mixed refrigerant vapour (MRV) and liquefied petroleum gas (LPG) being cooled by the mixed refrigerant (MR). LPG mixes with NG post its exit from the mid bundle, and enters the cold bundle along with MRV. MRV exiting the cold bundle returns to the bundle via a Joule-Thomson (J-T) valve as a cold stream.

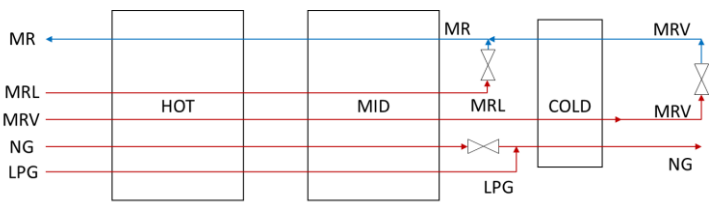


Figure 2. MCHE configuration in the LNG case study

We use the operational data presented in Hasan et al. (Hasan et al., 2009). Furthermore, we pre-processed data to include the J-T valve expansions of the streams exchanging heat in the MCHE. The flow distribution and the heat exchange areas are shown in Figure 3. The areas are given in a scaled unit au to ensure confidentiality.

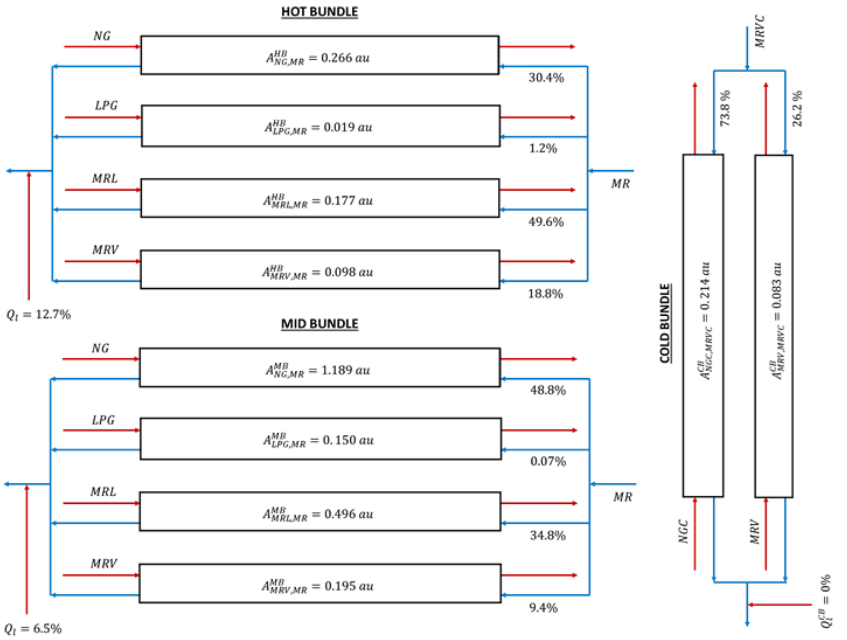


Figure 3. Estimated flow distribution and individual heat transfer areas for the MCHE

5. Conclusions

We presented an operational model for MHEXs to estimate the flow distribution and the individual heat transfer areas of the exchanging streams. This data-driven NLP model is geometry-independent and can be applied to any class of MHEXs. We used a simpler single-stage superstructure to develop the model which can handle multiple hot and cold streams, consider pressure drops and include ambient heat leaks. We then applied our model on a 3-bundle MCHE from an existing LNG plant in Qatar, and estimated the flow fractions and the heat exchange areas between the hot streams and the refrigerant in each bundle. Along with enabling the operational optimization of the process, the estimated parameters can also be used for predicting the response of MHEXs under variations in feed conditions, ambient temperature, etc.

This work was funded under Energy Innovation Research Programme (EIRP) Award NRF2014EWTEIRP003-008, administrated by the Energy Market Authority (EMA).

References

Hasan, M. M. F., Karimi, I. A., Alfadala, H. E., & Grootjans, H. (2009). Operational modeling of multistream heat exchangers with phase changes. AIChE Journal, 55, 150-171.

Jiang, L., Zhou, K., & Zhu, L. (2015). Equation-oriented Modeling of Multi-stream Heat Exchanger in Air Separation Units (Vol. 37).

Khan, M. S., Husnil, Y. A., Getu, M., & Lee, M. (2012). Modeling and Simulation of Multi-stream Heat Exchanger Using Artificial Neural Network. Proceedings of the 11th International Symposium on Process Systems Engineering, 31, 1196-1200.

Nagesh Rao, H., & Karimi, I. A. (2017). A superstructure-based model for multistream heat exchanger design within flow sheet optimization. AIChE Journal, 63, 3764-3777.

Tsay, C., Pattison, R. C., & Baldea, M. (2017). Equation-oriented simulation and optimization of process flowsheets incorporating detailed spiral-wound multistream heat exchanger models. AIChE Journal, 63, 3778-3789.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50395-5

Simulation and optimization of a combined cycle gas turbine power plant under part-load operation

Zuming Liu, Iftekhar A. Karimi*

Department of Chemical & Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117585
cheiak@nus.edu.sg

Abstract

As the electricity demand fluctuates frequently, combined cycle gas turbine (CCGT) power plants often run at part-load conditions. In order to improve the plant performance, a simulation-based optimization approach is proposed to determine an optimal operating strategy that maximizes the plant efficiency. The influence of the optimal strategy is compared with that of fuel flow control (FFC) and inlet guide vane control (IGVC). It is shown that the plant efficiency is increased by as much as 2.63% over FFC and 0.93% over IGVC. Moreover, it seems that FFC tends to prioritize the gas turbine and IGVC tends to prioritize the steam cycle, while the optimal strategy strikes a balance between the two.

Keywords: Simulation, Optimization, Gas turbine, Steam cycle, Part-load operation

1. Introduction

Combined cycle gas turbine (CCGT) power plants fired by natural gas have higher thermal efficiency and lower CO_2 emissions than their coal-fired counterparts. With the rapid increase of energy demand and stricter environmental regulation, CCGT plants are now undergoing widespread installation. However, as the energy demand fluctuates frequently with time, CCGT plants often run at off-design conditions. Under this circumstance, the plant operating strategies have significant influence on the plant performance. Kim and Hwang (2006) analyzed the performance of recuperative gas turbines during off-design operation. Various operating strategies, including fuel flow control (FFC), variable speed control (VSC), and inlet guide vane control (IGVC) for single-shaft configuration, as well as FFC and variable area nozzle control (VANC) for two-shaft configuration, were investigated. Haglind (2011) studied the effects of variable geometry gas turbine on the off-design performance of combined cycles, and found that the plant performance was improved by using IGV and VAN. Jimenez-Espadafor Auilar et al. (2014) considered eight operating strategies for a combined heat and power (CHP) plant, and showed that IGVC offered the best regulation capacity. The above studies simulate the operation of CCGT plants for a given policy, but they cannot optimize the policy. In other words, one must manually test pre-formulated operating strategies to improve part-load operations. Therefore, a comprehensive multi-variable simulation-based optimization approach is necessary to find the best operation strategy for any given part-load.

2. System description

Figure 1 shows the schematic of a typical CCGT power plant. The pressurized ambient air from the air compressor enters the combustor, where a gaseous fuel (natural gas) is burnt. The combustion gas expands in the turbine to produce power. Some air from one or more stages of the compressor is supplied to the turbine blades to keep them cool. The exhaust gas from the turbine then enters a heat recovery steam generator (HRSG) to generate high pressure (HP), intermediate pressure (IP), and low pressure (LP) steam, which drives which drives HP, IP and LP steam turbines to generate extra power. The exhaust steam from LP steam turbine goes to a condenser, and the condensate is finally pumped back to the HRSG.

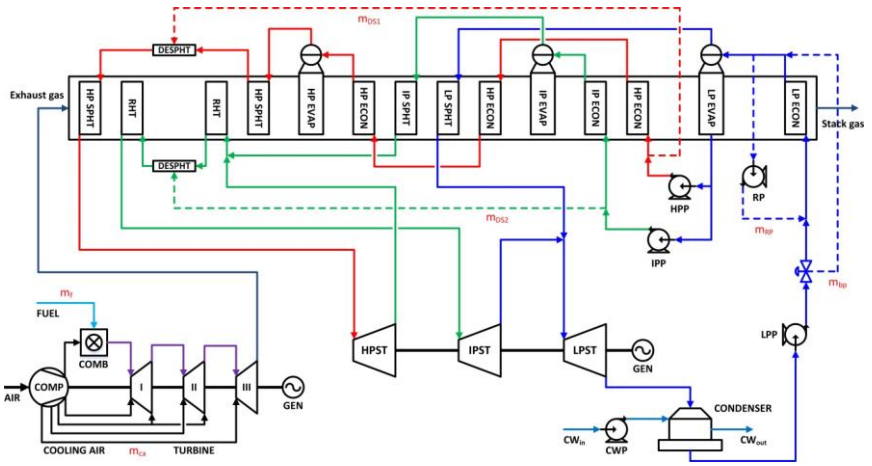


Figure 1 Schematic of a triple-pressure reheat combined cycle power plant.

3. Methodology

3.1 CCGT modeling

The CCGT plant is modeled using a commercial simulator GateCycle. The air compressor is modeled by the compressor map. The turbine is represented by the choking condition. The cooling air from the inter-stages of the compressor to cool the turbine blades is determined using a differential pressure correlation. The turbine efficiency is corrected using a semi-empirical formula. The HRSG is modeled as a series of heat exchangers with off-design correction of overall heat transfer coefficients. The steam turbine is modeled via the modified Stodola's law, and the water pump is modeled by the pump curves. All of these are implemented in GateCycle to form a holistic model for the CCGT plant (Liu and Karimi (2017)).

3.2 Simulation-based optimization

We use IGV angle ($\Delta\alpha$), fuel flow (m_f), cooling airflow to turbine stage t ($m_{ca,t}$), and four water flows [desuperheaters (m_{DS1}, m_{DS2}), recirculation (m_{RP}), and bypass (m_{bp}) as in Fig. 1] in HRSG as the optimization variables to maximize combined cycle efficiency (CCE) at a given part-load in order to find an optimal

operating strategy. Maximizing CCE is same as minimizing the fuel flow for a given part-load.

$$\min m_f\left(\Delta\alpha, m_{ca,t}, m_{DS1}, m_{DS2}, m_{RP}, m_{bp}\right) \tag{1}$$

We enforce the following operational constraints during optimization:

- Each turbine blade metal temperature must not exceed its design metal temperature:

$$T_b \le T_{b,d} \tag{2}$$

- HP steam and reheat steam temperatures must not exceed their design values:

$$T_{\text{HP steam}} \le 565^{\circ}C, T_{\text{Reheat steam}} \le 565^{\circ}C \tag{3}$$

- Stack gas temperature must exceed a lower limit to prevent corrosion:

$$T_{\text{Stack gas}} \ge 99.7^{\circ}C \tag{4}$$

A simulation-based optimization approach (see Fig. 2) is used to optimize the plant performance for any given part-load. The plant is simulated in GateCycle, and the particle swarm optimization (PSO) algorithm is implemented in Matlab for optimization. GateCycle is interfaced with Matlab via Cyclelink.

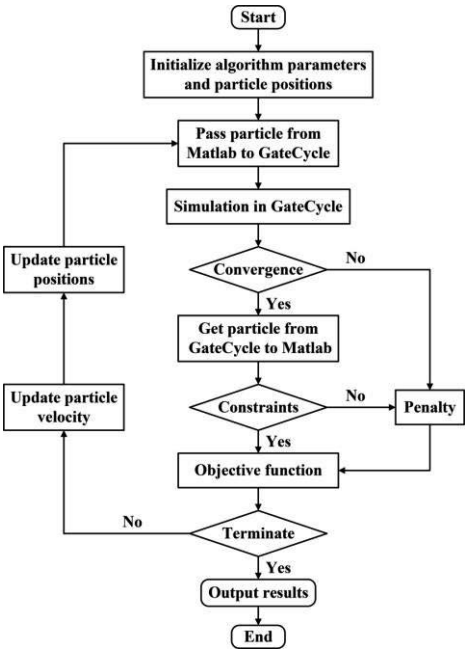


Figure 2 Simulation-based optimization using PSO algorithm.

4. Results and discussion

In CCGT plants, the following two operating strategies are widely used:

- Fuel flow control (FFC): the fuel flow is reduced to achieve the desired part-load.
- Inlet guide vane control (IGVC): IGVs and fuel flow are changed simultaneously to achieve the desired part-load, while maintaining the turbine exhaust temperature (TET) at its design value.

Figure 3 shows the results for FFC. The turbine inlet temperature (TIT) reduces by 26.9% from 100-40%, as FFC manipulates only the fuel flow to change the power output. Because the compressor operating point moves downward along the constant speed line, the pressure ratio reduces by 13.3%. However, the airflow increases by about 1.0%, causing that TIT drops quickly. As the cooling airflows are affected by pressure ratio, the total cooling airflow (summed over all GT stages) reduces by 10.0%. Moreover, both TET and blade metal temperature drop quickly with TIT. Because of these changes, the gas turbine (GT) power and gas turbine efficiency reduce by 62.8% and 11.3%, respectively. As TET drops quickly, the waste heat available for the steam cycle (SC) reduces. As a result, the ST power and steam cycle efficiency (SCE) by decrease by 54.6% and 8.8%. Overall, CCE decreases by 14.3% from 56.3% at 100% load to 42.0% at 40% load. This indicates part-load operation has a huge impact on the plant performance under FFC.

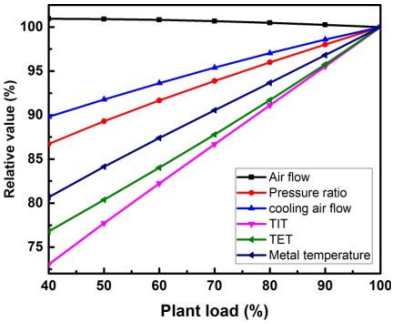


Figure 3 Plant operating variables versus plant load under FFC.

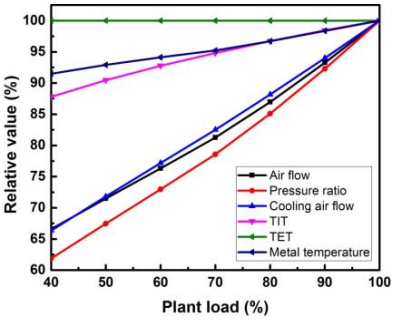


Figure 4 Plant operating variables versus plant load under IGVC.

Figure 4 presents the plant performance under IGVC. The airflow and pressure ratio are reduced by 33.4% and 38.1%, since the IGVs are closed gradually. As a result, the total cooling airflow reduces by 33.6%. Moreover, the drop in TIT is retarded due to the simultaneous reductions in the fuel flow and airflow. Thus, TIT reduces by 12.2% as compared to 26.9% for FFC. As TET remains at its design value, GT and HRSG run much hotter in IGVC than FFC. The drop in TIT reduces the GT power and GTE by 72.9% and 17.5%, respectively. However, the SC performance is improves compared to FFC due to the higher TET. Hence, the ST power decreases by 35.6% versus 54.6% for FFC, and SCE reduces only slightly by 0.6%. Although GTE for IGVC is lower than that for FFC, the higher SC performance improves CCE for IGVC, which decreases by only 12.6% from 56.3% to 43.7%. This clearly suggests that IGVC is more efficient than FFC.

In contrast to FFC and IGVC, the optimal strategy (Figure 5) gives the best CCE via manipulating several variables. The optimizer minimizes the cooling airflows to

keep GT as hot as allowable. As a result, the pressure ratio and total cooling airflow reduce by 30.2% and 81.4%. Moreover, the optimal strategy uses 47.8-71.2% lower total cooling airflow than FCC and IGVC. The optimizer also tends to maximize the blade metal temperature. This leads to a drop in TIT by 17.2%, more than IGVC, but less than FFC. As TET only reduces by 4.0%, the optimal strategy maintains GT at lower average temperatures than IGVC. This results in nil desuperheater flows. Furthermore, the stack gas temperature is kept at its minimum to ensure maximum heat recovery by manipulating the recirculation and bypass flows. Under the optimal strategy, the GT power and GTE reduce by 70.0% and 15.0% respectively, while the SC power and SCE decrease by 41.0% and 1.6% respectively. However, CCE is higher than FFC and IGVC, and reduces by 11.7% from 56.3% to 44.6%, which is the lowest drop among the three operating policies, as shown in Figure 7.

For implementing the optimal strategy for the CCGT plant, IGVs is manipulated to maintain TET variation tendancy. Moreover, cooling air flows are changed to keep blade metal temperatures constant. The recirculation flow is regulated to maintain LP economizer inlet temperature constant, while the by-pass flow is varied to keep the stack gas temperature at its minimum value. Finally, fuel flow is reduced to decrease the plant load.

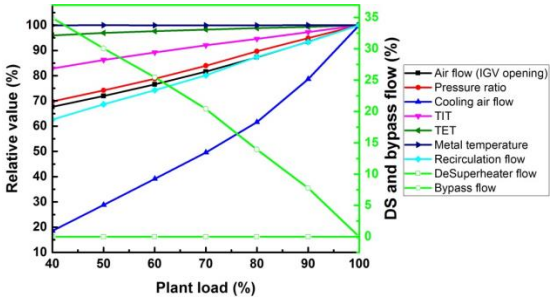


Figure 5 Plant operating variables versus plant load for the optimal strategy.

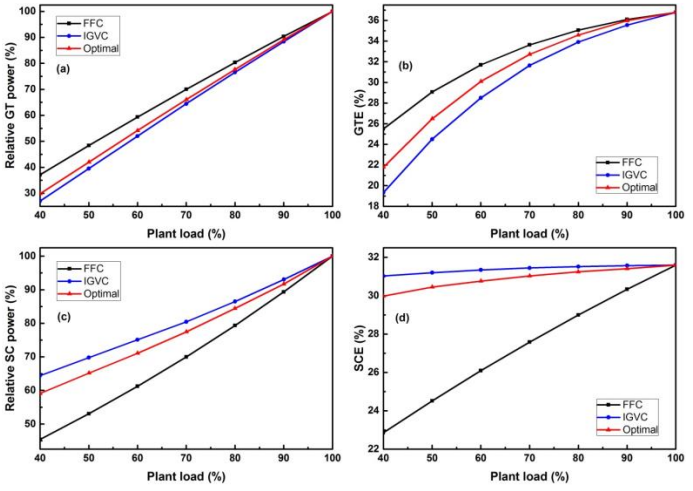


Figure 6 GT power, GTE, SC power, and SCE versus plant load under FFC, IGVC, and optimal strategy.

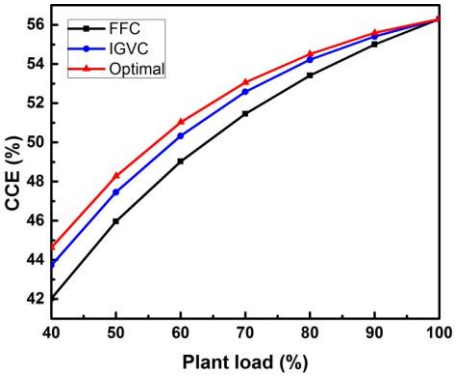


Figure 7 CCE versus plant load under FFC, IGVC, and optimal strategy.

5. Conclusions

In this work, a rigorous multi-variable simulation-based optimization approach is proposed to obtain an operating strategy that maximizes the overall plant efficiency. It is shown that FFC prioritizes GTE, and gives the worst plant performance. IGVC prioritizes SCE, and gives much better performance than FFC. However, it is not the best operation strategy for a CCGT plant. Our optimal strategy achieves the best trade-off between the GT and SC cycles, and gives the best plant performance. It can increase CCE by as much as by 2.63% and 0.93% compared to FFC and IGVC, respectively. Our optimization approach thus provides a basis for evaluating various operating policies.

This work was funded in part under Energy Innovation Research Programme (EIRP) Award NRF2014EWTEIRP003-008, administrated by the Energy Market Authority (EMA).

Reference

F. Haglind, 2011. Variable geometry gas turbines for improving the part-load performance of marine combined cycles – Combined cycle performance. Applied Thermal Engineering, 31, 467-476.

F. Jimenez-Espadafor Aguilar, R. R. Quintero, E. C. Trujillo, M. T. García, 2014. Analysis of regulation methods of a combined heat and power plant based on gas turbines. Energy, 72, 574-589.

J. H. Kim, T. S. Kim, J. L. Sohn, S. T. Ro, 2003. Comparative Analysis of Off-Design Performance Characteristics of Single and Two-Shaft Industrial Gas Turbines. Journal of Engineering for Gas Turbines and Power, 125, 954.

T. Kim, S. Hwang, 2006. Part load performance analysis of recuperated gas turbines considering engine configuration and operation strategy. Energy, 31, 260-277.

Z. Liu, I.A. Karimi. 2017. Simulation and optimization of a combined cycle gas turbine power plant for part-load operation. Chemical Engineering Research Design. https://doi.org/10.1016/j.cherd.2017.12.009.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50396-7

Optimal Design of Boil-Off Gas Liquefaction in LNG Regasification Terminals

Harsha Nagesh Rao[a], Iftekhar A. Karimi[a*], and Shamsuzzaman Farooq[a]

[a]*Dept. of Chemical & Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117585*
cheiak@nus.edu.sg

Abstract

Boil-off gas (BOG) generation is inevitable in LNG regasification terminals, which import and store huge amounts of LNG. The energy and costs required to reliquefy the generated BOG is substantial for a regasification terminal. Consequently, it is important to optimize BOG reliquefaction in regasification terminals. In this work, we develop a novel superstructure for BOG reliquefaction comprising of all the process options discussed in our previous work (Rao et al., 2016) and use a nonlinear programming (NLP) formulation to minimize the total annualized cost (TAC) of the process. We then use a derivative-free optimization approach to solve this simulation-based optimization problem. For a case study terminal (Park et al., 2012), we discuss the optimal configurations for BOG reliquefaction corresponding to varying amounts of BOG in the terminal.

Keywords: LNG, Regasification Terminal, Design.

1. Introduction

Natural gas (NG) is the fastest growing fossil energy resource with its trade poised to increase significantly in the foreseeable future. Liquefied Natural Gas (LNG) is the most common and economical mode of transporting NG over long distances. As LNG is a cryogenic liquid (around –160°C) at atmospheric pressure, heat ingress into LNG throughout its storage and transport is inevitable regardless of heavy insulation. This vaporizes a portion of the LNG to generate boil-off gas (BOG). LNG regasification terminals which import and store huge amounts of LNG are concentrated sources of BOG. Unless reliquefied and/or reused, BOG is an economic loss, and alters LNG composition and quality over time. The power consumption and the costs associated with the reliquefaction of BOG are significant for a regasification terminal, and hence, reducing them is important. Therefore, it is crucial to optimize the configuration of BOG reliquefaction in regasification terminals.

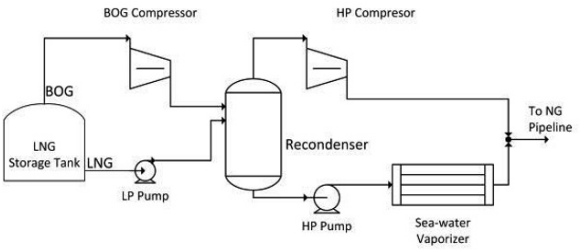


Figure 1: A typical LNG regasification terminal

Consider a typical LNG regasification terminal shown in Figure 1. A regasification terminal stores the incoming LNG in atmospheric tanks, and regasifies it continuously to supply NG to gas distribution grids or customers at high pressures (40-80 bar). BOG from the LNG storage tanks is compressed to a sufficiently high pressure, and recondensed using low-pressure LNG (LP-LNG) pumped from the LNG tanks. Subsequently, the recondensed liquid is pressurized to the send-out pressure of the terminal to form high-pressure LNG (HP-LNG). HP-LNG is then regasified in a seawater vaporizer to form the send-out NG.

Most literature on LNG regasification concerns the utilization of cryogenic energy of LNG to produce power by integrating various cycles such as Organic Rankine, Brayton, etc. Very few studied the retrofit design of a BOG reliquefaction system (Chaowei Liu, 2010; Park et al., 2012). In our previous study (Rao et al., 2016), we evolved reliquefaction schemes to minimize the terminal power consumption based on physical arguments. However, it is not straightforward to synthesize schemes to lower total annualized cost (TAC) due to the trade-offs involved between capital and operating costs. Therefore, we develop a superstructure comprising of several process options and optimize the process systematically.

2. Problem Statement

Consider the seawater-based LNG regasification terminal in Figure 1. On a long-term basis, it receives and gasifies F_{OUT} tonne/h (send-out rate) of LNG on an average, and supplies NG to a local gas distribution grid at specified conditions. Let βF_{OUT} $(0 < \beta < 1)$ tph of BOG exit the tanks at (P_{G1}, T_{G1}), and $(1-\beta)F_{OUT}$ tph of LNG be available for send-out at (P_{L1}, T_{L1}). We allow the user to specify BOG and LNG conditions independently based on real terminal data. With this, we address the following problem in this work.

Given:

1. LNG throughput or gas send-out rate F_{OUT} tonne/h (tph). Composition and conditions (P_{L1}, T_{L1}) of the LNG available for send-out on an average.
2. BOG composition, generation fraction β, and conditions (P_{G1}, T_{G1}) on an average.
3. Desired send-out temperature (T_{OUT}) and pressure (P_{OUT}).
4. Ratio (p_{sw}) of the power required to pump seawater to the rate of heating required for regasification.

Obtain: Configuration, design targets (e.g., heat-transfer areas), and optimal operating conditions for various process units that correspond to the minimum total annualized cost (TAC) of the reliquefaction process for a given β.

3. Novel Superstructure

We concluded from our previous study (Rao et al., 2016) that the indirect cooling of BOG prior to compression using high pressure (HP)- and medium pressure (MP)-LNG are effective process options to reduce the power consumption of the terminal. However, we did not include the impact of direct cooling BOG using LNG at low pressure, called desuperheating. This process option involves spraying low pressure (LP)-LNG at the inlet of a knockout drum to reduce the temperature of BOG. The residue liquid from the knockout drum can be used for BOG recondensation by elevating its pressure. The sprayed LP-LNG vaporizes upon contact with the superheated BOG, resulting in an increase in the BOG flowrate. Consequently, the

decrease in the compression work due to the reduction in BOG temperature is slightly offset by the rise in the compression work due to the increased BOG flowrate. Therefore, desuperheating may not be the best option to precool BOG from an energy perspective. However, desuperheating requires low capital investment as it mainly consists of a drum operated at low pressure. Hence, in this study, we incorporate all plausible combinations (series and/or parallel) to cool BOG before compression using indirect heat exchange with HP- and MP-LNG, and direct heat exchange by LP-LNG (Figure 2). These coolers are placed in the order of increasing potency of the coolants (HP-LNG < MP-LNG < LP-LNG). Let us refer to these components collectively as the cooling unit (Figure 2) for further use.

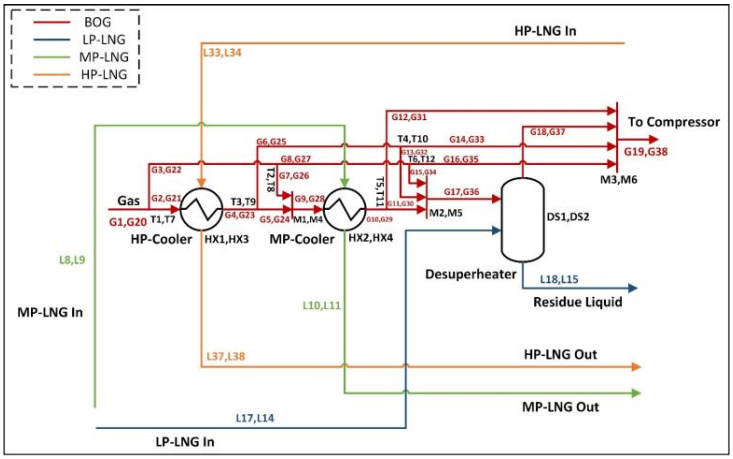


Figure 2: The cooling unit

The overall superstructure for BOG reliquefaction is shown in Figure 3. Derived from schemes presented Rao et al. (Rao et al., 2016), this superstructure can synthesize schemes with either one stage or two stages of recondensation. BOG from storage tanks, shown in red colour, enters the cooling unit-1 (CU-1) prior to compression by C1. If BOG undergoes recondensation in one stage, BOG from C1 enters the cooling unit-2 (CU-2) followed by the second stage of compression (C2). The compressed BOG can be partially or fully cooled by the HP cooler HX5 or even bypass it before entering the recondenser R1, where it is fully reliquefied. If BOG undergoes recondensation in two-stages, BOG from C1 enters the cooling unit-2 (CU-2) that consists only of the HP-LNG cooler HX3 along with the option to partially or fully bypass it. This is because BOG enters the recondenser R1 subsequently. We learnt from the previous work (Rao et al., 2016) that direct heat exchange with LP-LNG is superior to indirect cooling by HP- or MP-LNG. Next, the uncondensed BOG from R1 is fed to the second compressor C2. The output of C2 can be partially/fully cooled by the HP-LNG cooler HX5 before entering the second recondenser R2, where BOG is fully reliquefied. To ensure feasibility, splitters T13 and T15 cannot allow partial splits. In other words, only one of their two outputs can exist. Similarly, M7 and M9 are binary mixers with one of the inlets present.

LNG send-out from the storage tanks, indicated in blue colour, is divided for desuperheating and recondensation purposes. A part of the LNG to be used for desuperheating (at tank pressure) is sent to the desuperheater DS1 of cooling unit-1 (CU-1) at BOG pressure via valve V1. Remaining LNG to be used for desuperheating is

pumped using pump P4 to the BOG pressure at the inlet of the second desuperheater DS2 of CU-2. The residual liquids from DS1 and DS2 are pressurized the recondenser R1 pressure using pumps P3 and P5, and mixed with LP-LNG in M10. LNG to be used for recondensation is pumped to the recondenser R1 pressure using the LP-Pump P1. Subsequently, while most of the LP-LNG is sent to mixer M10 followed by R1, a part of LP-LNG is pumped using MP-pump P2 to form MP-LNG (shown in green colour). MP-LNG is used for cooling BOG in the MP-LNG coolers HX2 of CU-1 and HX4 and CU-2. The MP-LNG streams exiting HX2 and HX4 are mixed in M12. Further, the recondensed liquid exiting R1 is pressurized by P6 and fed to the second recondenser R2, if present. However, a portion of this liquid has an option to partially bypass R2 via pump P7 even in the case of two-stage recondensation to limit the size of R2. In this scenario, P7 is used to elevate its pressure to that of R2. The recondensed liquid from R1 and R2 are pressurized by P8 to be mixed with MP-LNG stream exiting mixer M12, if MP-LNG is at a higher pressure. If MP-LNG (at M12) is at a lower pressure than the recondensed liquid (at M11), pump P9 is used to match pressures. This mixture from M13 is pressurized to the send-out pressure by the high-pressure pump (P10) to form HP-LNG (shown in orange colour). The seawater vaporizer is generally over-designed, and has a complex design with special considerations to avoid ice formation when seawater contacts HP-LNG. Therefore, it has a high capital cost almost insensitive to the scheme of BOG reliquefaction. Hence, we do not include the costs associated with the SW regasifier in this study.

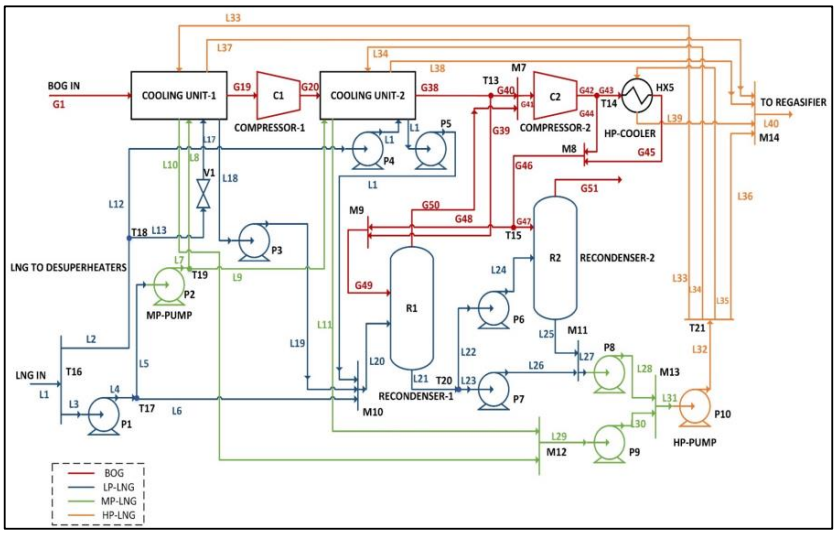


Figure 3: Superstructure for BOG reliquefaction

4. Model Formulation and Solution Strategy

We develop independent modules for the process units of our superstructure such as heat exchangers, pumps, compressors, recondensers, desuperheater, mixers and splitters in a suitable programming platform. We do not use unit modules in commercial simulators as they generally have fixed input-output structures, and offer limited flexibility. This is a medium-scale simulation-based NLP model with 35 independent variables. We implemented the model in MATLAB R2015a while accessing the Peng

Robinson fluid package from Aspen Hysys using the hysyslib toolbar, and used the Nonlinear Optimization by Mesh Adaptive Direct Search (NOMAD) algorithm to solve the problem.

5. Case study

We consider the Pyeongtaek LNG receiving terminal case study discussed in Park et al. (Park et al., 2012). We solved the model for $0.03 \leq \beta \leq 0.15$ to study the effect of varying BOG concentration in the terminal. Figure 4 present the variation of the minimum TAC of the terminal with β.

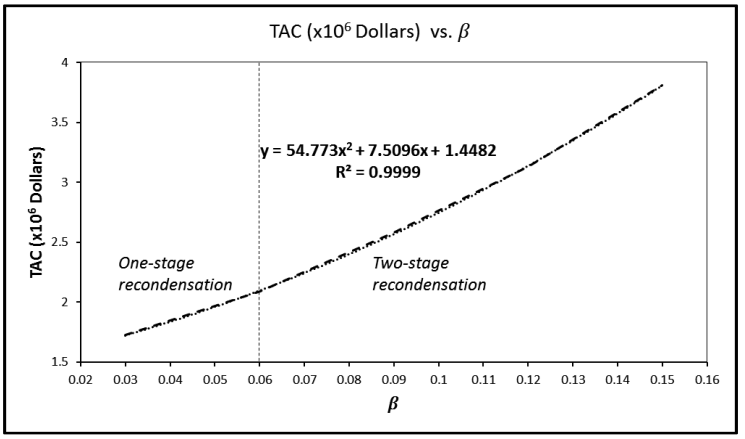


Figure 4: Minimum TAC vs β

We inferred from the Rao et al. (Rao et al., 2016) that recondensation in two-stages is much more energy efficient than recondensation in one stage. This is because BOG flow to the second compressor is reduced due to the partial recondensation in stage I. Besides, BOG entering the second compressor is saturated. Therefore, the feed temperature of BOG and hence the compression work are low. However, the capital investment on the second stage of recondensation is justified only if the there is considerable reduction in the power consumption (or OPEX) of the terminal. As the savings in power consumption of the terminal increases with β, we find that recondensation in two stages is optimal only for $\beta \geq 0.06$.

The optimal terminal configuration for $\beta = 0.03$ with one stage of recondensation is shown in Figure 5a. BOG is slightly precooled in the desuperheater by LP-LNG prior to compression, and subsequently recondensed in R1 completely. Next, the savings in the OPEX is just sufficient to invest on another stage of recondensation for $\beta = 0.06$ (Figure 5b). BOG is compressed to a low pressure and sent to the first recondenser R1 at a low temperature, wherein it is partially reliquefied. The uncondensed gas from R1 is compressed further. As the compressor output is sufficiently cold, it is directly fed to the second recondenser R2 for complete reliquefaction. Then, Figure 5c shows the optimal configuration for the range of $0.09 \leq \beta \leq 0.12$. A desuperheater prior to C1 is added to the configuration from the previous case. Lastly, consider the case of highest BOG concentration ($\beta = 0.15$). The optimal configuration shown in Figure 5d does not have any coolers from cooling unit-1 (CU-2). In this case, HP-LNG has very less precooling potential. However, there is a HP cooler after CU-1.

Overall, we find that two-stage recondensation is optimal for terminals with $\beta \geq 0.06$. Desuperheating provides direct heat exchange with a superior coolant LP-LNG at a low capital cost. Hence, it is often employed to precool BOG in the optimal configurations discussed above. Though HP-LNG and MP-LNG coolers minimize the power consumption of the reliquefaction process, capital investments on them is not justified for low-moderate concentration of BOG in the terminal.

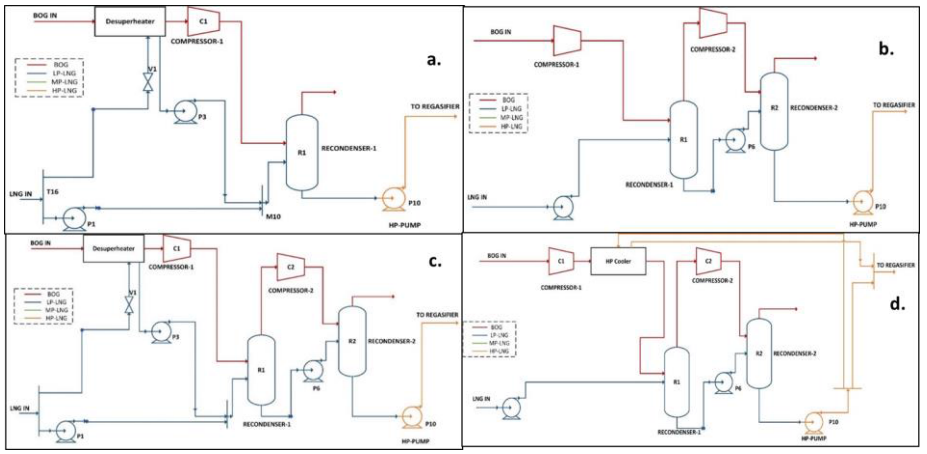


Figure 5a-d: Minimum TAC configurations

6. Conclusions

We developed a novel comprehensive superstructure for BOG reliquefaction. The superstructure allowed up to two stages of recondensation, incorporated three pressure levels of LNG (LP-, MP-, and HP-LNG) and all of their combinations(series and parallel) to cool BOG via direct and indirect heat exchange. We used a modular NLP formulation to minimize the total annualized costs of the reliquefaction process that is suitable for derivative free optimization by NOMAD. For the case study terminal from Park et al. (Park et al., 2012), the optimization results suggest that two-stage recondensation is optimal beyond 6% BOG concentration in the terminal and desuperheating using LP-LNG is a favourable precooling option.

7. Acknowledgements

This work was funded under Energy Innovation Research Programme (EIRP) Award NRF2014EWTEIRP003-008, administrated by the Energy Market Authority (EMA).

References

Chaowei Liu, J. Z., Qiang Xu, John L. Gossage. (2010). Thermodynamic-Analysis-Based Design and Operation for Boil-Off Gas Flare Minimization at LNG Receiving Terminals. Industrial & Engineering Chemistry Research, 49, 7412–7420.

Park, C., Song, K., Lee, S., Lim, Y., & Han, C. (2012). Retrofit design of a boil-off gas handling process in liquefied natural gas receiving terminals. Energy, 44, 69-78.

Rao, H. N., Wong, K. H., & Karimi, I. A. (2016). Minimizing Power Consumption Related to BOG Reliquefaction in an LNG Regasification Terminal. Industrial & Engineering Chemistry Research, 55, 7431-7445.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50397-9

Monitoring of two-phase slug flow in stacked multi-channel reactors based on analysis of feed pressure

Osamu Tonomura[a*], Ryuichi Kobori[a], Satoshi Taniguchi[a], Shinji Hasebe[a], Akira Matsuoka[b]
[a]*Kyoto University, Nishikyo, Kyoto 615-8510, Japan*
[b]*KOBE STEEL, LTD., Onoecho Ikeda, Kakogawa 675-0023, Japan*
tonomura@cheme.kyoto-u.ac.jp

Abstract

Microreactors with two-phase slug flow have an advantage over the conventional batch reactors in mass transfer because of the internal circulation flow within the slugs. The slug flow in the microreactors have to be monitored, because the slug lengths influence mass transfer performance. In this work, an experimental investigation is carried out to analyze the formation of gas-liquid slug flow in a single T-shaped channel. The result shows that the feed pressures oscillate periodically along the formation of a pair of gas and liquid slugs. Based on this result, a feed pressure-based method is developed to monitor whether a slug having the desirable length is generated. By the developed method, the slug flows can be monitored without complicated processing, such as the placement of the windows or electrodes on channels. In addition, the applicability of the developed method to a multi-channel reactor connected to manifold-type flow distributors is investigated. As a result of the investigation, it is proposed that, by using Fourier transform analysis of pressure signals obtained at the inlet of either flow distributor, the number of pressure sensors required for estimating all the pressure oscillation cycles is reduced to just one. The effectiveness of the proposed method is experimentally demonstrated under both normal and abnormal conditions of the process.

Keywords: Process monitoring, Microreactors, Multi-channel reactors, Gas-liquid slug flow, Fourier transform analysis.

1. Introduction

Microreactors can realize short residence time as well as high mass and heat transfer rates. Recently, microreactors with multi-phase flow have attracted researchers' and engineers' attention by their unique flow characteristics. For example, when two mutually immiscible fluids, such as gas and liquid, are simultaneously fed into a microchannel, various flow patterns such as annular, dispersed and slug flows are generated as to the design and operating conditions (Shao et al., 2009). Among these, slug flow, also called Taylor flow or segmented flow, is especially attractive for its small residence time distribution and large interface area per fluid volume (Kreutzer et al., 2008). In addition, rapid mass transfer between two phases is achieved because the liquid interface is constantly renewed by the circulation flow inside the slug (Kashid et al., 2005). As an application example, Takebayashi et al. (2012) reported that the direct carbonylation of nitrobenzene to phenylisocynate was conducted by using a T-shaped microchannel with gas-liquid slug flow, and the isocyanate yield in the microchannel was three to six times higher than that in the batch reactor.

A process monitoring method needs to be developed to realize stable long-term operation of microreactors with gas-liquid slug flow, because the slug lengths, which influence mass transfer performance, are shifted to undesired values under abnormal conditions such as the deposition of solid on the surface of channels. Optical methods (Ide et al., 2009) and electrical methods (Ye et al., 2011) have been proposed as the measuring methods of void fraction in channels. Though these methods are applicable to the monitoring of slug flows, there are some drawbacks. As the optical method requires the placement of the windows on the channels to visualize the slug flow, it is not applicable to microreactors made of opaque materials. The electrical method uses the electrical resistance and capacitance of fluid in the channels to estimate the flow condition in the channel. If the device is made of electrically-conductive materials such as stainless steel, complicated electrically-insulated placement of electrodes is required. Hence, the objective of this research is to develop alternatives to the conventional monitoring methods for slug flows in channels.

The outline of this paper is as follows: In Section 2, a monitoring method for a single T-shaped channel with two-phase slug flow is proposed on the basis of the experimental observation about the relationship between the flow behavior and feed pressure changes of gas and liquid. In Section 3, an experimental study is carried out to investigate the applicability of the developed monitoring method to a multi-channel reactor connected to manifold-type flow distributors. In addition, a method of solving the problem that the number of sensors increases as the number of channels increases is proposed.

2. Slug flow monitoring for a single channel reactor

In this section, the characteristic points of feed pressure profiles that correspond to the formation of gas and liquid slugs in a single T-shaped channel are made clear. In addition, a slug flow monitoring method for the single T-shaped channel is proposed.

2.1. Experimental setup

Figure 1 shows an experimental system used in this study. Pure nitrogen (gas) and distillated water (liquid) are fed into a T-shaped channel from a gas cylinder and a pressure tank, respectively. The gas flow rate is controlled by a mass flow controller (FCST-1005LC, Fujikin). The liquid flow rate is controlled by a mass flow controller (LV-F50PO, Horiba STEC). After gas and liquid are contacted in the mixer, slug flow is formed and goes through a succeeding tube. The end of the tube is immersed in a beaker filled with water. The beaker is open to the atmosphere, and the water level of the beaker is kept constant. The sizes and materials of tubes and mixers are summarized in Table 1. The length between the pressure sensor and the mixer is defined as the length of feeding tube. The mixers are made of polymethyl methacrylate (PMMA). The tubes (JR-T-6806-M10, VICI AG International) are made of fluorinated ethylene propylene (FEP). All of the channels and tubes have circular cross-sections. The slug formation behavior in the mixer is observed by high-speed video camera (VW-6000, KEYENCE), and the gas and liquid feed pressures are measured by pressure sensors (PA-830-101G-10, NIDEC COPAL ELECTRONICS). The digital fiber-optic sensor unit (FU-L51Z, KEYENCE) is installed at near the inlet of FEP tube to measure superficial velocity of slug flow and slug lengths of gas and liquid. The measurements acquired from pressure sensors and digital fiber-optic sensor unit are collected in a computer through a data logger (NR-600, KEYENCE).

2.2. Slug formation and pressure oscillation

The experimental result is explained. Figure 2 shows a photo of gas-liquid slug flow in the succedent FEP tube, where gas and liquid feed flow rates are 1.0 mL/min and 1.0 mL/min, respectively. All of gas slugs in the FEP tube have no liquid film, and the shapes are almost cylindrical. This is because FEP has hydrophobic property. Figure 3 shows the pressures at gas and liquid feed flows. Both pressures oscillate periodically and the oscillation phases of gas and liquid pressures have a half cycle difference. For clearly observing the profile of gas and liquid pressure measurements, smoothed lines obtained by the Savitzky–Golay filter are shown in black lines.

Figure 4 shows the gas-liquid slug formation behavior in the mixer. Photos in Fig. 4 are taken at time A to time E in Fig. 3. The gas and liquid discharge alternately from the confluence part of the mixer into the succeeding tube. The relationship between the periodic pressure fluctuation and the formation of gas and liquid slugs is explained as follows: During the period between Fig. 4(A) and just before Fig. 4(C), the gas penetrates to the confluence part of the mixer, but it is still blocked by liquid flow. Thus, the gas pressure continuously increases because gas is constantly supplied to the confluence part. At the time shown in Fig. 4(C), the gas occupies the confluence part, and the gas starts flowing from the confluence part to the succeeding tube. By the flow, the pressure of the gas feeding section gradually decreases. The flow behavior of liquid during the periods from Fig. 4(C) to (E) is explained in a similar manner as that of gas during the periods from Fig. 4(A) to (C). These phenomena are repeated as far as gas-liquid slugs are formed. The magnitude of liquid pressure oscillation is larger than that of gas pressure oscillation, and the profile of liquid pressure change is clearer than that of gas pressure change. This is due to the difference of the fluid properties between gas and liquid, such as viscosity, compressibility and wettability. Thus, the liquid pressure profile is used in the slug flow monitoring method explained in the next subsection.

2.3. Slug flow monitoring method

Based on the result of Sec. 2.2, a monitoring method for gas and liquid slug lengths in a single T-shaped channel reactor is developed. When the T-shaped channel reactor with gas-liquid slug flow is operated at Q_G and Q_L, feed pressures of gas and liquid are fluctuated periodically with the cycle time t_{GL}, which is equal to the generation cycle time of a pair of gas and liquid slugs. Hence, the estimates of gas and liquid slug lengths

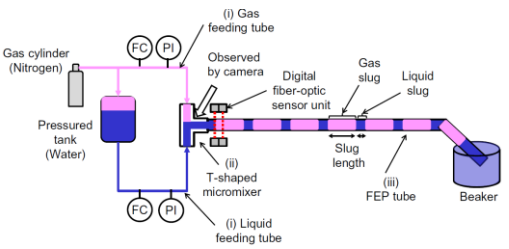


Figure 1 Experimental system of a single T-shaped channel reactor

Table 1 Sizes and materials of tubes and mixers

Location in Fig. 1	I.D. (mm)	Length (mm)	Material
(i) Feeding tube	1.00	100	FEP
(ii) Micromixer	1.30	40 for all branches	PMMA
(iii) Reactor tube	1.00	500	FEP

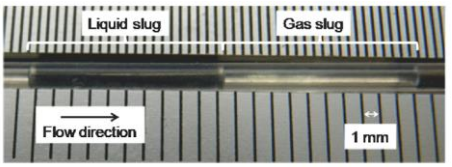


Figure 2 Slug flow in the FEP tube

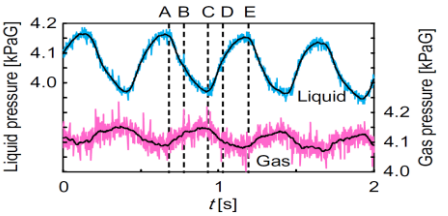


Figure 3 Gas and liquid feed pressures

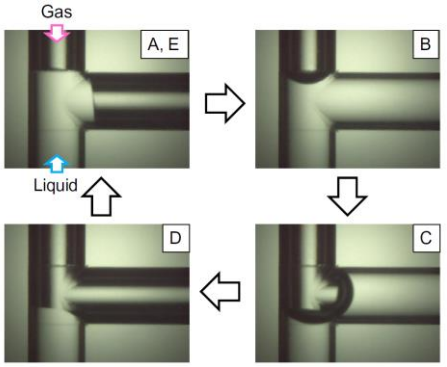


Figure 4 Photos of flow behavior in the mixer

Table 2 Estimates and measurements of gas and liquid slug lengths

	Q_G/Q_L	L_G^*: L_G [mm]	L_L^*: L_L [mm]
Case 1	1.0	11.1:11.0	11.1:11.1
Case 2	2.0	15.2:15.3	7.6:7.6
Case 3	0.5	7.6:7.5	15.1:15.2

are derived by the following equations:

$$L_G^* = Q_G t_{GL} / A \tag{1}$$

$$L_L^* = Q_L t_{GL} / A \tag{2}$$

where A is the cross sectional area of gas and liquid slug flow, and L_G^* and L_L^* are estimated values of gas slug length L_G and liquid slug length L_L, respectively. Accuracy of the developed method was evaluated. The results are summarized in Table 2. It is concluded that slug lengths can be estimated accurately by using the developed method.

3. Slug flow monitoring for a multi-channel reactor

The applicability of the developed method to a multi-channel reactor connected to manifold-type flow distributors is investigated. In addition, a method for reducing the number of installed sensors is proposed.

3.1. Experimental setup

Figure 5 shows a schematic diagram of an experimental system developed in this study. Gas (nitrogen) and liquid (distilled water) are fed by mass flow controller (MFC) at Q_{GT} and Q_{LT}, respectively. The outlets of gas and liquid MFCs are connected to a gas manifold (MG) and a liquid manifold (ML), respectively. The MG and ML distribute gas and liquid to parallelized channels (Ch.1 to Ch.4), respectively. Each channel is composed of a gas barrier channel (BG), a liquid barrier channel (BL), a gas feed channel (FG), a liquid feed channel (FL), a T-shaped union (T) and a reaction channel

(R). The barrier channels, which have high hydraulic resistance, are used to realize equal flow distribution. Pressure sensors (PI1 to PI4) are installed between BLs and FLs of Ch.1 to Ch.4, and a pressure sensor (PI0) is installed at the inlet of the ML. The flow rates of gas and liquid in Ch.1 to Ch.4 (Q_{Gi}, Q_{Li}; i=1, 2, 3, 4), the slug lengths of gas and liquid in Ch.1 to Ch.4 (L_{Gi}, L_{Li}; i=1, 2, 3, 4) and the generation cycle time of a pair of gas and liquid slugs in Ch.1 to Ch.4 (t_{GLMi}; i=1, 2, 3, 4) are measured by using the digital fiber-optic sensors installed on the Rs of Ch.1 to Ch.4.

3.2. Applicability of the developed method to multi-channel reactor

To verify the applicability of the monitoring method developed in the previous section to a multi-channel reactor, the relationship between the generation cycle of a pair of gas and liquid slug in each channel and the fluctuation cycle of feed pressure in each channel is investigated through an experimental case study. Figure 6 shows the profiles of liquid pressure changes acquired from PI1 to PI4 under a normal condition where $Q_{GT} = Q_{LT} = 8.0$ mL/min and gas and liquid are equally distributed to Ch.1 to Ch.4. By comparing one-minute average values of the pressure fluctuation cycles in Ch.1 to Ch.4 (t_{GLi} [s] ; i=1, 2, 3, 4) and those of t_{GLMi} (i=1, 2, 3, 4), it is confirmed that, in every channel, the fluctuation cycle of feed pressure is equal to the generation cycle of a pair of gas and liquid slugs. In addition, the same result is confirmed under an abnormal condition where blockage occurs in Ch.1. Therefore, it is shown that the proposed method for a single T-shaped channel is applicable to the multi-channel reactor.

3.3. Slug flow monitoring method

Figure 7 (a) and (b) show the results of Fourier transform analysis of the pressure data acquired from PI0 to PI4 under the normal and abnormal conditions, respectively. It is shown that, under both conditions, peak frequencies of PI0 are almost equal to those of PI1 to PI4: namely, the pressure fluctuation cycles at PI1 to PI4 can be estimated from PI0. This result opens up the possibility of reducing the number of pressure sensors. In this study, just three measurements, Q_{GT}, Q_{LT} and the pressure at PI0, are used to monitor the gas and liquid slug lengths in all parallelized microreactors. The verification result shows that the estimation error is within 5% of real value.

4. Conclusions

In this study, from the experimental investigation of a single T-shaped channel, it was made clear that the pressures in gas and liquid feeding tubes oscillate periodically along the formation of a pair of slugs. By using this fact, a slug flow monitoring method based on the feed pressure measurements was developed, whose effectiveness was verified through several case studies. In addition, the developed method was applied to the multi-channel reactor connected to manifold-type flow distributors, and it was proposed that, by using Fourier transform analysis of pressure signals obtained at the inlet of the flow distributor, the number of pressure sensors required for estimating the pressure fluctuation periods of every microchannel is reduced to just one. The effectiveness of the proposed method was experimentally demonstrated under both normal and abnormal conditions of the process.

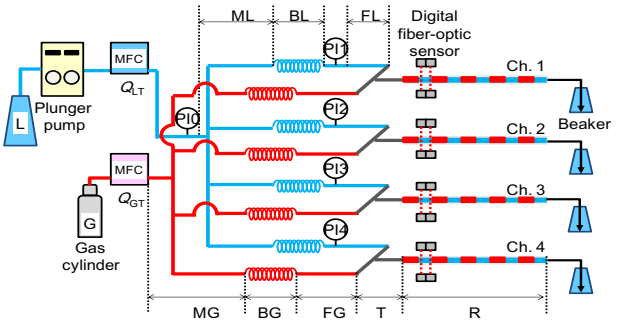


Figure 5 Experimental system of a multi-channel reactor

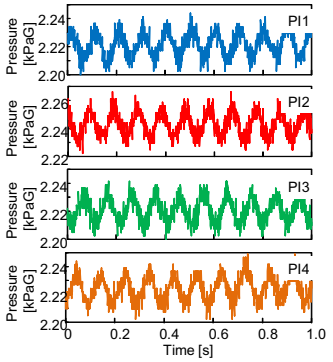


Figure 6 Pressure data acquired from PI1 to PI4

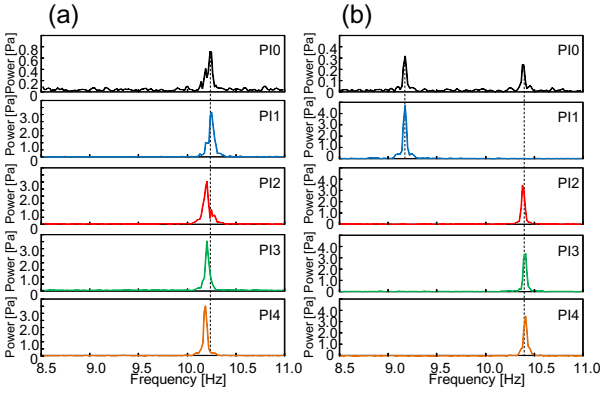


Figure 7 Fourier transform of data acquired from PI1 to PI4: (a) normal and (b) abnormal conditions

References

N. Shao, A. Gavriilidis, P. Angeli, Flow Regimes for Adiabatic Gas-Liquid Flow in Microchannels, Chemical Engineering Science, 64, 2749-2761 (2009)

M.T. Kreutzer, A. Gunther, K.F. Jensen, Sample Dispersion for Segmented Flow in Micro-channels with Rectangular Cross Section, Analytical Chemistry, 80 (5), 1558-1567 (2008)

M.N. Kashid, I. Gerlach, S. Goetz, J. Franzke, J.F. Acker, F. Platte, D.W. Agar, S. Turek, Internal Circulation within the Liquid Slugs of a Liquid-Liquid Slug-Flows Capillary Microreactor, Industrial & Engineering Chemistry Research, 44, 5003-5010 (2005)

Y. Takebayashi, K. Sue, S. Yoda, T. Furuya, K. Mae, Direct Carbonylation of Nitrobenzene to Phenylisocyanate using Gas-Liquid Slug Flow in Microchannel, Chemical Engineering Journal, 180, 250-254 (2012)

H. Ide, R. Kimura, M. Kawaji, Optical Measurement of Void Fraction and Bubble Size Distributions in a Microchannel, Heat Transfer Engineering, 28, 713–719 (2009)

J. Ye, L. Peng, W. Wang, W. Zhou, Optimization of Helical Capacitance Sensor for Void Fraction Measurement of Gas-liquid Two-phase Flow in a Small Diameter Tube, IEEE Sensors Journal, 11 (10), 2189-2196 (2011)

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50398-0

Modular Concept Inspired by Microchemical Systems and Application to Distillation

Ken-Ichiro Sotowa*, Takahiro Aoyama, Ryo Takagi, Kohei Ito, Jesús Rafael Alcántara Avila and Toshihide Horikawa

Department of Science and Technology, Tokushima University, 2-1 Minamjosanjimacho, Tokushima 770-8506 Japan
**sotowa.ken-ichiro@tokushima-u.ac.jp*

Abstract

Microreactors and micro heat exchangers have very unique characteristics due to the small size of their channels. Our close examination of the properties of microchemical systems has resulted in the finding that microchannel reactor systems have modular structures, a feature which may be exploited to create novel systems. In microreactors, reagents are mixed in a dedicated device while another microchannel is used to exchange heat. For this reason, the chemical reactor, which is conventionally treated as a unit operation, may be decomposed into more elementary operations, thereby enabling the development of new chemical apparatuses by reconfiguring the individual modules. This presentation introduces a conceptual framework for the elementary operation of these chemical apparatuses. An example of the application of these concepts to the development of a new distillation device will be discussed.

Keywords: microreactors, unit operations, elementary operations, distillation.

1. Introduction

Microreaction technology has attracted the attention of chemical engineers for decades. Whereas classical chemical engineering focuses on the use of large vessels to increase processing rates, in micro-scale reactions, small channels are exploited to enhance performance. The challenge of decreased throughput is addressed by using a so-called numbering up in which large scale production is achieved by designing many parallel microchannel reactors. The combination of the remarkable heat transfer and flow controlling properties of microchannels with the development of a unique mass production strategy has enabled microreaction technology to become a promising technology that will be exploited in chemical industries in the near future.

The concept of modularization has existed in the field of microreaction technology for many years, and a set of microchannel modules have been commercialized by several hardware manufacturers. This approach has much in common with that of the process design approach adopted in conventional chemical engineering, where chemical processes are treated as a set of unit operations.

During the early stages of microreactor development, researchers proposed to use the technology to develop small chemical processes for the on-site on-demand production of chemicals. This application required the development of a variety of small unit operations for separation and purification. To meet this need, we began to develop a microchannel distillation device. However, we discovered that it is extremely difficult to achieve countercurrent contact of vapor and liquid in microchannels. After an

extensive examination of the fundamental properties of microchannels and distillation, we came to the conclusion that there are a set of fundamental operations which govern many chemical processes and apparatuses. These elementary principles underpin the established set of unit operations. This concept can be exploited to create new chemical apparatuses.

In this report, we summarize the foundation of the new set of fundamental operations and show an example application of these concepts to the development of a new distillation apparatus.

2. Elementary operation in microchemical processes

In microchemical plants, microchannels are exploited to achieve a variety of unit operations such as reactions, heat exchange, or extraction. With their small characteristic diameters and high surface-to-volume ratios, microchannels are suited to intensify the mixing and heat transfer of fluids. In addition, the nature of flow apparatuses allows for the precise control of residence time. It is even possible to sequentially inject reagents at an interval as short as a few milliseconds. This is extremely useful in applications which include active reaction intermediates that deactivate within milliseconds. These features are rooted to the properties of the microchannels.

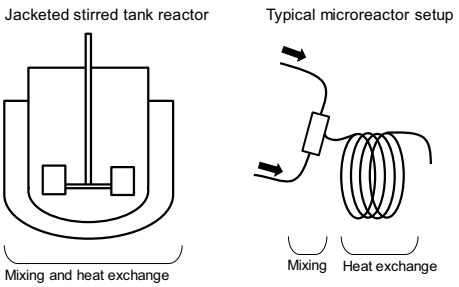


Figure 1. Functions of stirred tank reactor and microreactor setup

Our study of the properties of microreactors revealed that in addition to the utility of fine channels, another aspect of the geometry of microreactors plays an important role. That is, microreactors have intrinsically modular structures. Figure 1 compares a conventional stirred tank reactor and a typical microreactor setup. For a reaction to take place, usually two or more reagents must be mixed well, and the heat of reaction must be added or removed at an appropriate rate. In conventional chemical engineering, all of these operations are carried out in a single reactor unit. In the microreactor setup, although both mixing and heat exchange are necessary for the reaction, they are achieved in a different manner. A dedicated mixing device is employed, and heat exchange is accomplished using a downstream tube, which also ensures an appropriate residence time. This observation tells us that it is necessary to split the function of the reactor into mixing and heat exchange before converting reactions to be performed on microchannel devices. In addition, it is important to note that the reactor is divisible, and mixing and heat exchange are the elementary operations which are common across all chemical reactors.

It is expected that a variety of chemical apparatuses can be decomposed into elementary operations in a similar way. Examination of the configuration of various devices used in

microchemical processes reveal that each is realized by the combination of elementary operations such as liquid–liquid contact, gas–liquid contact, mixing, and heat exchange (Figure 2). This demonstrates that every conventional unit operation can be decomposed into a set of more elementary operations. Once a unit operation is treated as a set of elementary operations, a new device can be created by intensifying and reconfiguring each operation.

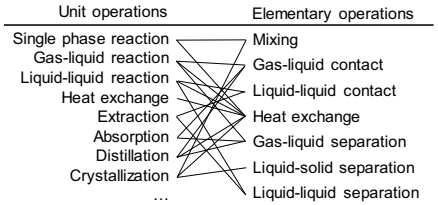


Figure 2. Unit and elementary operations

3. Application to distillation

3.1. Elementary operation of distillation

Distillation is a conventional unit operation. For this reason, it is usually considered one of the smallest units which constitutes a chemical process. In our study, we discussed further the decomposition of distillation from the viewpoint of elementary operations. As a first step, distillation can be seen as a collection of stages. A stage is not an elementary operation: its function can be further split into gas–liquid contact, equilibration, and gas–liquid separation. Any chemical apparatus with these three functions can play the role of a distillation stage. Figure 3 shows a distillation device that was designed using this idea (Sotowa et al., 2003). Experiments showed that it can separate a water–methanol mixture with a tray efficiency of 70%.

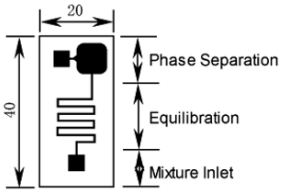


Figure 3. Distillation device consisting of phase separation and equilibration sections (Sotowa et al., 2003)

3.2. Multistage distillation

In theory, multistage distillation can be achieved using interconnecting devices as shown in Figure 3. However, it would be undesirable to simply replace the trays in a conventional distillation column with these devices. Because the tray temperatures vary in conventional distillation, the temperature of each device should be independently controlled. This calls for the development of an alternative strategy to achieve multistage distillation.

Thus, we proposed a new multi-stage distillation configuration in which the entire system is kept at a uniform temperature and the mixture is separated by forming a pressure distribution. Temperature control is easily achieved by using a thermostatic bath or circulating heat medium.

Conventional distillation columns have excellent separation performance but usually consume a high volume of energy. Although a large amount of heat is required to generate a vapor stream at the reboiler, a large portion of it is released from the condenser. Researchers have developed methods of thermal integration within a column, leading to the invention of distillation techniques such as VRC and HIDiC. In the proposed method, the temperature of the entire distillation column is controlled using a single heat medium so that the heat released from the condensing unit is automatically transported to the evaporation unit. Thus, it is expected that the proposed method will enable separation at a much smaller energy cost.

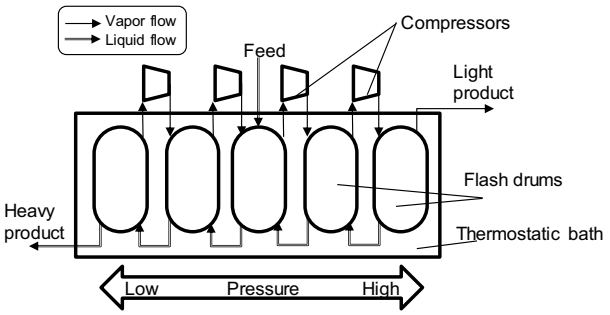


Figure 4. Pressure-driven distillation system (PDDS)

3.3. Scaling up

A distillation apparatus consisting of microchannels can process only small volume of chemicals. To increase the throughput of microchemical devices, parallelization or numbering up is usually employed. It should be noted that the chief advantage of the distillation method proposed in this study is its high energy efficiency, which is brought about by isothermal operation rather than the properties of microchannels. Thus, conventional scale-up (rather than numbering up) is suited for increasing the throughput of the proposed distillation device. Finally, we conceived the distillation system which is shown in Figure 4 and called a pressure-driven distillation system (PDDS) (Asada et al., 2010). The major components of PDDSs are flash drums and compressors. The compressors are necessary to form the pressure distribution required to achieve separation. The drum with the lowest pressure corresponds to the reboiler, from which the high boiling product is obtained. The distillate flows out from the drum with the highest pressure. The reflux ratio is determined by the pressure of the drum of highest pressure, which plays a role similar to the partial condenser in traditional distillation technology.

The highest and lowest pressures in this system are determined by the vapor pressures of the components in the mixture at the operating temperature. For example, when this was applied to a water–methanol separation, the lowest and highest pressures were 1.0 and 3.5 atm, respectively, when the operating temperature was 100 °C. When this mixture was separated using a 20-stage PDDS, the average compression ratio per unit was $10^{(1/20)} = 1.06$. A compression ratio of 1.06 is small; indeed, it is so low that low power compressors or blowers would be sufficient. The distillation system could be designed using the well-known McCabe–Thiele method using the *x-y* relationship in the *P-xy* VLE data. Once the composition of each stage was obtained using the McCabe–Thiele method, the pressure of each stage was determined from *P-xy* data. The resulting

set of pressures could then be directly used in the final design or fine-tuned to minimize the total rate of energy consumption.

The energy efficiency of PDDS was evaluated by conducting a process simulation for a methanol–water system. The feed was assumed to be an equimolar mixture, and the number of stages was 13. The specifications for both the top and bottom products were both 99% and the feed rate was 10 kmol/h. A PDDS was designed using the McCabe–Thiele method. The result of this design was input as the initial condition of an optimization calculation to minimize the total rate of energy consumption. The sum of the electric energy required to operate the compressors was 14.5 kW. The energy consumption rate of the comparable conventional reaction was also calculated. The ratio of the overall reflux ratio to the minimum reflux ratio was 1.01. The reboiler duty was calculated as 67.7 kW. To enable comparison, the electric energy was converted to thermal energy by assuming that a generation efficiency was 0.35. The results of this study suggest that the PDDS could be used to cut the consumption of thermal energy by 39% (Ito et al., 2014)

3.4. Experimental evaluation of separation performance

We are currently conducting an experimental study to validate the PDDS design concept. Our present focus is on the understanding its separation performance because a PDDS large enough to evaluate the rate of energy consumption may not be practical for a laboratory setting.

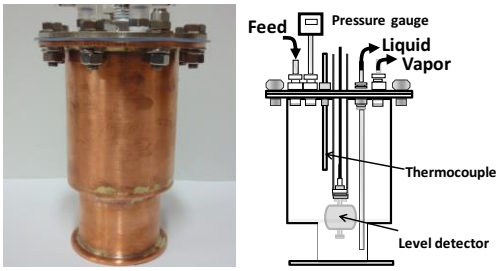


Figure 5. A device for a stage in PDDS

Figure 5 is a small-stage device developed in our laboratory. The body is made of copper to facilitate heat exchange with an external fluid. It is equipped with a pressure sensor, thermocouple, and liquid level sensor. The liquid level was controlled by manipulating a solenoid valve connected downstream of the liquid extraction line. A diaphragm pump within the vapor stream is activated when the pressure exceeds a set point.

A four-stage PDDS (Figure 6 (left)) was constructed in a thermostatic bath. A separation experiment was carried out using a methanol–water mixture. The four devices were connected to achieve a countercurrent flow of vapor and liquid and placed in a thermostatic bath at 80 °C. The pressure set points were 1.15, 1.3, 1.45, and 1.6 atm, respectively. The methanol composition of the feed was 66 mol%. The feed mixture was fed to the second stage from the low-pressure side at a rate of 2.0 mL/min. Samples of the liquid and vapor phases were collected after the system reached a steady state. The composition of each sample was determined by densitometry.

The measured compositions were plotted against *P-xy* data as shown in Figure 6 (right). The compositions of the samples were comparable to those given by the VLE data. A

slight deviation in the sample composition from the equilibrium can be observed at the highest-pressure stage. This is due to the large pressure fluctuation caused by the on–off operation of the diaphragm pump for pressure transport. This fluctuation is expected to be resolved by constructing a large apparatus and employing a more sophisticated control algorithm. Taken together, these observations suggest that the separation performance of PDDS is expected to be comparable to that of a conventional distillation column.

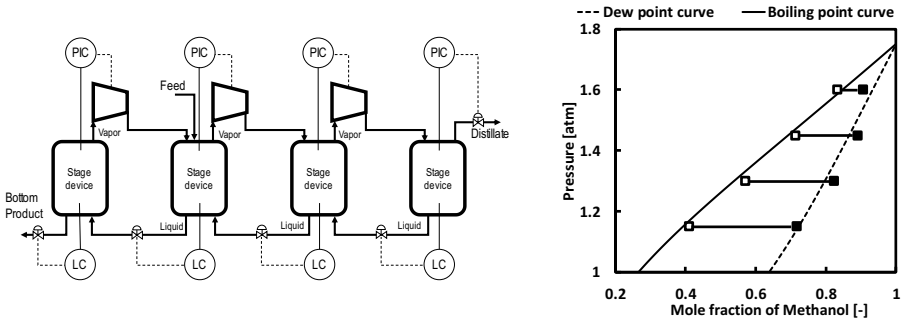


Figure 6. Configuration of 4-stage PDDS (left) and composition distribution (right)

4. Conclusions

Through close examination of the properties of microchemical process technology, the elementary operations in chemical apparatuses were identified. Conventional unit operations can be decomposed into a set of elementary operations, as shown in Figure 2. This framework allows us to create new chemical devices with distinct properties from their conventional analogues.

The development of PDDS was presented as an application of this idea. The proposed PDDS is unique in that a mixture is separated using a pressure distribution. It is expected be highly energy-efficient. Experimental evidence confirms that the PDDS can achieve satisfactory separation performance. Larger scale experimental work is being undertaken to evaluate the energy efficiency of PDDS.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers 17H01339, 17K19015, 21656246 and 15760560.

References

K.-I. Sotowa and K. Kusakabe, 2003, *Development of Micro Distillation Device*, 4th European Conference on Chemical Engineering (ECCE-4), Granada, Spain.

K. Asada, K.-I. Sotowa, K. Nakagawa and S. Sugiyama, 2010, *Study on the Energy Consumption Rate of Pressure Driven Distillation Systems*, 23rd Symposium on Chemical Engineering, Fukuoka, Japan.

K. Ito, K.-I. Sotowa, T. Horikawa and J. R. Alcantara Avila, 2014, *Energy Efficiency of Pressure Driven Distillation System for Separation of Bioethanol*, 10th International Conference on Separation Science and Technology (ICSST14), Nara, Japan.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50399-2

Dynamic Bounds on Stochastic Chemical Kinetic Systems through Semidefinite Programming

Garrett R. Dowdy and Paul I. Barton*

Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts, 02139, United States of America
pib@mit.edu

Abstract

The method of moments has been proposed as a potential means to reduce the dimensionality of the chemical master equation (CME) appearing in stochastic chemical kinetics. However, attempts to apply the method of moments to the CME usually result in the so-called closure problem. Several authors have proposed moment closure schemes, which allow them to obtain approximations of quantities of interest, such as the mean count of molecules of each species. However, these approximations have the dissatisfying feature that they come with no error bounds. This paper presents a fundamentally different approach to the closure problem in stochastic chemical kinetics. Instead of making an approximation to compute a single number for the quantity of interest, we calculate mathematically rigorous bounds on this quantity by solving semidefinite programs (SDPs). These bounds provide a check on the validity of the moment closure approximations, and are in some cases so tight that they effectively provide the desired quantity. This paper extends a previously published method for calculating steady-state bounds to the problem of calculating time-varying bounds.

Keywords: semidefinite programming, chemical master equation, closure problem, stochastic chemical kinetics, moments

1. Introduction

Chemical reactions are often modelled using deterministic rate laws. However, when the number of molecules in a reaction system is relatively small, as in a single biological cell, these deterministic rate laws become unreliable, and the system's time evolution is best viewed as a stochastic process. This stochastic process is governed by the chemical master equation (CME), which describes the time evolution of a probability distribution describing the system. In principle, one could solve the CME to obtain the time evolution of the distribution. However, in any system of practical importance, the number of states is very large, rendering direct solution of the CME impractical (see Higham, 2008).

In an effort to circumvent this problem, instead of modelling the time evolution of the full probability distribution, some have proposed modelling only its moments up to some finite order m. The moments could then be used to calculate quantities of interest such as the mean count of molecules of each chemical species and the variances in these counts. The difficulty with this approach is that the system of ordinary differential equations describing how the moments change over time usually suffers from the so-called "closure problem", in which the time evolution of the moments up to order m depends on the values of higher-order moments.

To deal with this problem, various authors have proposed "closure scheme" approximations. For example, C. S. Gillespie (2009) assumes that the unknown distribution is normal; Smadbeck and Kaznessis (2013) assume it is maximum entropy. With such assumptions, the low-order moments determine the entire distribution (including the high-order moments), and the closure problem is bypassed. One can then approximate both dynamic trajectories and steady-state values for the moments. Unfortunately, these are just approximations, with no bounds on error. The approximation could be good; on the other hand, it could be quite bad. These two scenarios are indistinguishable, unless some second, independent method is applied to the problem.

In this paper, we propose an alternative strategy. Instead of selecting one of the many distributions satisfying prescribed conditions on the low-order moments and calculating some approximation of the quantities of interest, we acknowledge that many distributions satisfying the prescribed conditions exist and calculate rigorous bounds on the quantities of interest. These bounds are computed by solving semidefinite programs (SDPs), convex optimization problems which can be solved efficiently, both in theory and practice (Vandenberghe and Boyd, 1996).

This paper extends a previously published method for calculating steady-state bounds to the problem of calculating time-varying bounds (see Dowdy and Barton, 2018 and the references therein).

2. Mathematical Background

2.1. Notation of Stochastic Chemical Kinetics

Consider a stochastic chemical kinetic system with N distinct chemical species and R reactions. At any point in time, the state of the system is described by a vector $\boldsymbol{x} = (x_1, \ldots, x_N) \in \mathbb{Z}_+^N$, where $x_i \geq 0$ is the count of molecules of species i present.

The state changes with the occurrence of each reaction. For example, if $\boldsymbol{s}_r \in \mathbb{Z}^N$ is the vector of stoichiometric coefficients of reaction r, then an occurrence of reaction r takes the system from state $\boldsymbol{x}$ to $\boldsymbol{x} + \boldsymbol{s}_r$.

2.2. Invariants

In general, given an initial state $\boldsymbol{x}_0 \in \mathbb{Z}_+^N$ and a stoichiometry matrix $\boldsymbol{S} = [\boldsymbol{s}_1 \ \cdots \ \boldsymbol{s}_R] \in \mathbb{Z}^{N \times R}$ not every state $\boldsymbol{x} \in \mathbb{Z}_+^N$ is accessible to the system. The set of possible states X must be contained in the affine subspace $\{\boldsymbol{x}_0 + \boldsymbol{S}\boldsymbol{v} : \boldsymbol{v} \in \mathbb{R}^R\}$. If $\boldsymbol{S}$ has a nontrivial left null space, spanned by the linearly independent basis vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_L \in \mathbb{R}^N$, and we define the matrix $\boldsymbol{B} \equiv [\boldsymbol{b}_1 \ldots \boldsymbol{b}_L]^{\mathrm{T}} \in \mathbb{R}^{L \times N}$, then $\boldsymbol{B}\boldsymbol{x} = \boldsymbol{B}\boldsymbol{x}_0$ for all $\boldsymbol{x} \in X$. That is, certain linear combinations of the components of $\boldsymbol{x}$ are conserved by the reaction system. These linear combinations are the "invariants" of the reaction system.

2.3. The Chemical Master Equation

Because of the stochastic nature of the system, there is some uncertainty as to the state at time t, and we express this uncertainty by assigning a probability $P(\boldsymbol{x}, t)$ to each of the possible states $\boldsymbol{x} \in X$. This probability distribution $P(\cdot, t)$ changes over time according to the chemical master equation (CME):

$$\frac{dP}{dt}(\boldsymbol{x},t) = \sum_{r=1}^{R}[P(\boldsymbol{x}-\boldsymbol{s}_r,t)a_r(\boldsymbol{x}-\boldsymbol{s}_r) - P(\boldsymbol{x},t)a_r(\boldsymbol{x})], \quad \forall \boldsymbol{x} \in X, \tag{1}$$

where a_r is the "propensity function" of reaction r, described by Higham (2008), which is proportional to the rate constant c_r. Note that this equation holds for all possible states $\boldsymbol{x} \in X$.

2.4. Moments in Stochastic Chemical Kinetics

The probability distribution $P(\cdot,t)$ can be characterized by its moments. In particular, for any multi-index $\boldsymbol{j} = (j_1,\ldots,j_N) \in \mathbb{Z}_+^N$ we have a moment $\mu_{\boldsymbol{j}}(t)$ defined as

$$\mu_{\boldsymbol{j}}(t) \equiv \sum_{\boldsymbol{x}\in X} P(\boldsymbol{x},t)\,\boldsymbol{x}^{\boldsymbol{j}}, \tag{2}$$

where the sum is over the set X of all possible states, and $\boldsymbol{x}^{\boldsymbol{j}} = \prod_{k=1}^{N} x_k^{j_k}$ is a monomial. The order of the moment $\mu_{\boldsymbol{j}}$ is defined as the sum $\sum_{k=1}^{N} j_k$.

Some moments have intuitive meanings. For example, the zeroth-order moment $\mu_{\boldsymbol{0}}$ indexed by $\boldsymbol{0} = (0,\ldots,0)$ is simply the sum of probabilities across all possible states, so that $\mu_{\boldsymbol{0}}(t) = 1$ for all times t. More important for our purposes, the first-order moment $\mu_{\boldsymbol{e}_i}$ indexed by $\boldsymbol{e}_i = (0,\ldots,1,\ldots,0)$ is the mean number of molecules of species i.

2.5. The Closure Problem

As described by Smadbeck and Kaznessis (2012) and Gillespie (2009), the CME can be used to derive a system of linear ordinary differential equations describing how the moments of the distribution change over time. In particular, we can write

$$\frac{\mathrm{d}\boldsymbol{\mu}_L}{\mathrm{d}t}(t) = \boldsymbol{A}_L\boldsymbol{\mu}_L(t) + \boldsymbol{A}_H\boldsymbol{\mu}_H(t), \tag{3}$$

where $\boldsymbol{\mu}_L(t)$ is a vector of moments up to arbitrary order m, $\boldsymbol{\mu}_H(t)$ is a vector of finitely many moments with order greater than m, and $\boldsymbol{A}_L$ and $\boldsymbol{A}_H$ are constant matrices, whose sizes depend on m. The fact that the time derivatives of the moments with order up to m depend on higher-order moments is the essence of the closure problem. It is unclear how to solve such a system.

One could apply a closure scheme, assuming that the underlying distribution is normal (C.S. Gillespie, 2009) or maximum entropy (Smadbeck and Kaznessis, 2013). Doing so allows us to express the high-order moments in terms of the low-order moments, and thus "close" the system of equations. However, as noted in the introduction, these closure schemes provide no bound on the error introduced in making their assumptions. In the following section, we propose a different approach.

3. The Bounding Method

3.1. The Paradigm

Instead of trying to calculate the quantity of interest directly, let us calculate rigorous bounds on that quantity. For example, for the mean molecular count of species i at time t, $\langle x_i(t)\rangle$, let us calculate two numbers, $\langle x_i(t)\rangle^L$ and $\langle x_i(t)\rangle^U$, such that

$$\langle x_i(t)\rangle^L \leq \langle x_i(t)\rangle \leq \langle x_i(t)\rangle^U \tag{4}$$

is guaranteed.

How might we calculate these bounds? Suppose that we write down several mathematical constraints that the moment vector $\boldsymbol{\mu}(t)$ at time t must necessarily satisfy. Suppose that we then optimize over the set of vectors $\boldsymbol{\mu}$ satisfying these necessary conditions, searching for the vector $\boldsymbol{\mu}$ with the greatest component $\mu_{\boldsymbol{e}_i}$:

$$\begin{aligned} \mu^*_{\mathbf{e}_i} \equiv \max_{\boldsymbol{\mu}} \; & \mu_{\mathbf{e}_i} \\ \text{s.t.} \; & \boldsymbol{\mu} \text{ satisfies necessary conditions} \end{aligned} \tag{5}$$

Then, because the vector $\boldsymbol{\mu}(t)$ corresponding to true moments at time t satisfies the necessary conditions, it is a feasible point for this optimization problem. It follows that the optimal value $\mu^*_{\boldsymbol{e}_i}$ is an upper bound on $\mu_{\boldsymbol{e}_i}(t) = \langle x_i(t)\rangle$, and we can set $\mu^*_{\boldsymbol{e}_i} \equiv \langle x_i(t)\rangle^U$.

Following the same reasoning, we can calculate at lower bound on $\langle x_i(t)\rangle$ by solving the minimization version of Problem (5).

3.2. Necessary Conditions

What exactly are the necessary conditions appearing in Problem (5)? First, we have some linear equalities implied by the invariants of the reaction system:

$$\sum_{\boldsymbol{x}\in X} P(\boldsymbol{x},\tau)\, \boldsymbol{x}^{\boldsymbol{j}} \boldsymbol{B}(\boldsymbol{x}-\boldsymbol{x}_0) = \mathbf{0}, \tag{6}$$

for finitely many multi-indices $\boldsymbol{j} \in \mathbb{Z}^N_+$. These equations can be written concisely as

$$\widetilde{\boldsymbol{B}}\boldsymbol{\mu}(t) = \mathbf{0}. \tag{7}$$

Next, we have several linear matrix inequalities which reflect the support of the underlying probability distribution $P(\cdot, t)$:

$$\boldsymbol{M}^{\mathbf{0}}_n\big(\boldsymbol{\mu}(t)\big) \succcurlyeq \mathbf{0}, \tag{8}$$

$$\boldsymbol{M}^{\boldsymbol{e}_j}_{n-1}\big(\boldsymbol{\mu}(t)\big) \succcurlyeq \mathbf{0}, \;\; \forall j \in \{1,\dots,N\}. \tag{9}$$

The exact definitions of these matrices are given in our related publication on stochastic chemical systems at steady state (Dowdy and Barton, 2018). However, for now, all that matters is that the matrices are affine functions of the vector $\boldsymbol{\mu}(t)$.

Next, we have a family of equalities which can be derived from Equation (3), which reflect the dynamics of the system, and which tie $\boldsymbol{\mu}(t)$ to the initial moment sequence $\boldsymbol{\mu}(0)$:

$$\boldsymbol{\mu}_L(t) - e^{\rho t}\boldsymbol{\mu}_L(0) = (\boldsymbol{A}_L - \rho \boldsymbol{I})\boldsymbol{z}^{(\rho)}_L + \boldsymbol{A}_H \boldsymbol{z}^{(\rho)}_H, \quad \forall \rho \in \{\rho_1,\dots,\rho_d\}. \tag{10}$$

These equations are valid necessary conditions for any choice of the values of $\rho_1,\dots,\rho_d$. However, roughly speaking, we will try to pick these values to match the eigenvalues of the dominant dynamic modes of the chemical master equation (1).

Equation (10) introduced a set of decision vectors $\boldsymbol{z}^{(\rho)}, \forall \rho \in \{\rho_1,\dots,\rho_d\}$, which are defined by

$$z_{\boldsymbol{j}}(t) \equiv \int_0^t e^{\rho(t-\tau)} \sum_{\boldsymbol{x}\in X} P(\boldsymbol{x},\tau)\, \boldsymbol{x}^{\boldsymbol{j}}\, \mathrm{d}\tau, \qquad \forall \rho \in \{\rho_1, \ldots, \rho_d\},\ \ \forall \boldsymbol{j} \in \mathbb{Z}_+^N, \tag{11}$$

and can be thought of as "generalized moments". As moments, these vectors $\boldsymbol{z}^{(\rho)}$ must satisfy several constraints analogous to Conditions (7)-(9):

$$\widehat{\boldsymbol{B}}\boldsymbol{z}^{(\rho)} = \boldsymbol{0}, \qquad \boldsymbol{M}_n^{(\boldsymbol{0},\boldsymbol{0})}\big(\boldsymbol{z}^{(\rho)}\big) \succcurlyeq \boldsymbol{0}, \tag{12}$$

$$\boldsymbol{M}_{n-1}^{(\boldsymbol{0},\boldsymbol{e}_j)}\big(\boldsymbol{z}^{(\rho)}\big) \succcurlyeq \boldsymbol{0}, \qquad \forall j \in \{1, \ldots, N\}, \tag{13}$$

for all $\rho \in \{\rho_1, \ldots, \rho_d\}$.

3.3. A Semidefinite Program

Substituting Necessary Conditions (7)-(10) and (12)-(13) into Problem (5), we obtain a semidefinite program:

$$\begin{aligned}
\langle x_i(t)\rangle^U \equiv \max_{\boldsymbol{\mu}(t),\boldsymbol{z}^{(\rho_1)},\ldots,\boldsymbol{z}^{(\rho_d)}} \ & \mu_{\boldsymbol{e}_i}(t) \\
\text{s.t.}\ \ & \boldsymbol{\mu}_L(t) - e^{\rho t}\boldsymbol{\mu}_L(0) = (\mathbf{A}_L - \rho\mathbf{I})\mathbf{z}_L^{(\rho)} + \mathbf{A}_H\mathbf{z}_H^{(\rho)}, \\
& \qquad \forall \rho \in \{\rho_1, \ldots, \rho_d\}, \\
& \widehat{\mathbf{B}}\mathbf{z}^{(\rho)} = \mathbf{0},\ \mathbf{M}_n^{(\mathbf{0},\mathbf{0})}(\mathbf{z}^{(\rho)}) \succcurlyeq \mathbf{0},\ \forall \rho \in \{\rho_1, \ldots, \rho_d\}, \\
& \mathbf{M}_n^{(\mathbf{0},\mathbf{e}_j)}(\mathbf{z}^{(\rho)}) \succcurlyeq \mathbf{0},\ \ \forall j \in \{1, \ldots, N\},\ \forall \rho \in \{\rho_1, \ldots, \rho_d\}, \\
& \widehat{\mathbf{B}}\boldsymbol{\mu}(t) = \mathbf{0},\ \mathbf{M}_n^{\mathbf{0}}(\boldsymbol{\mu}(t)) \succcurlyeq \mathbf{0}, \\
& \mathbf{M}_n^{\mathbf{e}_i}(\boldsymbol{\mu}(t)) \succcurlyeq \mathbf{0},\ \ \forall j \in \{1, \ldots, N\},
\end{aligned} \tag{14}$$

Solving this SDP gives us our upper bound on $\langle x_i(t)\rangle$. Solving the corresponding minimization bounding problem gives us our lower bound.

4. Example

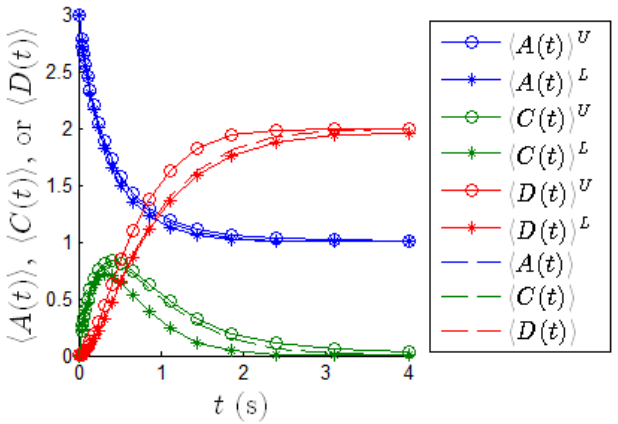


Figure 1 Time-varying bounds on the mean molecular counts of several species appearing in Reaction System (15).

Consider the simple sequential reaction system

$$A + B \xrightarrow{c_1} C \xrightarrow{c_2} D \tag{15}$$

with $c_1 = 1$ s^{-1}, $c_2 = 2$ s^{-1}, and an initial state of $x_A = 3,\ x_B = 2,\ x_C = 0$, and $x_D = 0$. Repeatedly solving SDP (14) and its lower-bounding analogue for this system, using $\{\rho_1, \ldots, \rho_d\} = \{0, -c_1, -c_2\}$ and a range of t values, gives the time-varying bounds on $\langle A \rangle, \langle C \rangle$, and $\langle D \rangle$ shown in Figure 1. All SDPs were solved using SeDuMi (Sturm, 1999).

This system is small enough that we can calculate the exact time-varying means by solving the CME (dashed lines in Figure 1). These exact means lie within the bounds, as predicted by the theory. Also, in this case, the bounds are tight around the exact means. We have chosen an example for which the exact means are available because it supports the theory. However, in general, the exact solution will not be available.

5. Conclusions

The proposed bounding method is a fundamentally different approach to the problem of characterizing time-varying distributions in stochastic chemical kinetic systems. In this paper, it has been applied to calculate bounds on the mean molecular counts. These bounds are theoretically guaranteed to be accurate, in contrast with the results of moment closure schemes. Of course, bounds on a quantity are less useful than the quantity itself, but we've seen that the bounds are sometimes so tight that they effectively give us the desired quantity.

There are several possible extensions and variations of the bounding method. First, one can formulate an SDP for calculating an upper bound on the variance in the count of each molecular species. Second, one can formulate an SDP to calculate bounds on the probability that the count of molecules of each species lies within a particular range. This leads to bounds on a histogram describing the underlying unknown distribution. Finally, there are computational efficiencies to be gained by considering reduced descriptions of the state of the system in which only the "independent components" are considered. These extensions and variations of the bounding method will be detailed in a forthcoming publication.

References

C. Gillespie, 2009, "Moment-closure approximations for mass-action models", IET Systems Biology, 3, 1, 52-58.

G. Dowdy, P. I. Barton, 2018 "Bounds on Stochastic Chemical Kinetic Systems at Steady State." In Review.

D. Higham, 2008, "Modeling and simulating chemical reactions", SIAM Review, 50, 2, 347-368.

P. Smadbeck, Y. Kaznessis, 2012, "Efficient moment matrix generation for arbitrary chemical networks", Chemical Engineering Science, 84, 612-618.

P. Smadbeck, Y. Kaznessis, 2013, "A Closure Scheme for Chemical Master Equations", Proceedings of the National Academy of Sciences, 110, 35, 14261-14265.

L. Vandenberghe and S. Boyd, 1996, "Semidefinite programming", SIAM Review, 38, 1, 49-95.

J. Sturm, 1999, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones." Optimization methods and software, 11, 1, 625-653.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50400-6

Multiscale Three-Dimensional CFD Modeling for PECVD of Amorphous Silicon Thin Films

Marquis Crose[a], Anh Tran[a], Gerassimos Orkoulas[b], and Panagiotis D. Christofides[a,c,*]

[a]*Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA, 90095-1592, USA*
[b]*Department of Chemical Engineering, Widener University, Chester, PA 19013, USA*
[c]*Department of Electrical Engineering, University of California, Los Angeles, CA 90095-1592, USA*
pdc@seas.ucla.edu

Abstract
A three-dimensional, multiscale computational fluid dynamics (CFD) model is presented here which aims to capture the deposition of amorphous silicon thin films via plasma-enhanced chemical vapor deposition (PECVD). The two domains which define the multiscale model are linked using a dynamic boundary which is updated at the completion of each time step. A novel parallel processing scheme built around a message passing interface (MPI) structure is applied in order to allow for transient simulations to be conducted using a mesh with greater than 1.5 million cells. Due to thin film thickness non-uniformity present in the *a*-Si:H product, an improved PECVD reactor geometry is proposed. The resulting geometry is shown to reduce the product offset from ~8% to less than 4% using identical deposition parameters.

Keywords: Multiscale modeling, computational fluid dynamics, kinetic Monte-Carlo, thin film growth.

1. Introduction
The past two decades have seen continual development in the multiscale modeling of plasma-enhanced chemical vapor deposition (PECVD) with specific applications to the manufacturing of silicon thin films for use in the photovoltaic and microelectronics industries (da Silva and Morimoto, 2002, and Collins et al., 1994). Accurate modeling of thin film deposition remains a key element in the effort to improve product quality and to cut down on manufacturing costs due to the difficulties associated with continuous and/or in situ measurements during chambered deposition processes. Recently, Crose et al. (Crose et al., 2017) demonstrated a novel multiscale computational fluid dynamics (CFD) model, which combined a macroscopic CFD domain with a microscopic surface domain through a common boundary which lies on the surface of the silicon wafer. Despite the success of this model in capturing the behavior of the PECVD reactor with regard to the non-uniform deposition of a-Si:H films and in providing a basis for future multiscale CFD modeling endeavors, the two-dimensional axisymmetric nature of the model limits the exploration of some phenomena which exist in the three-dimensional in space process. Specifically, two-dimensional (2D) axisymmetric models cannot represent the full geometry of showerhead holes that provide process gas to the plasma region -- a key feature when considering the uniformity of thin film products.

Given the aforementioned motivations, the framework previously developed for use in 2D applications is extended to the three dimensional domain. Using a three-dimensional (3D) render which captures the typical geometry of chambered, parallel-plate PECVD reactors, a CFD model is proposed in the present work which is capable of reproducing both accurate plasma chemistry and fluid flow into the reaction zone through the showerhead region. With regard to the microscopic domain (i.e., the surface of the silicon wafer), a detailed kinetic Monte Carlo (kMC) algorithm developed previously by Crose et al. (Crose et al., 2017) is applied in order to capture both the exchange of mass and energy, as well as the microstructure of the a-Si:H thin-film. Given that the startup and operation of PECVD reactors are inherently dynamic, the proposed simulations cannot derive accurate rector behavior from steady-state solutions; however, the computationally demanding nature of the transient simulations necessitates the use of a parallel computation strategy as well as taking advantage of multiple-time-scale phenomena occurring in the process. In this work, a Message Passing Interface (MPI) structure is adopted which allows for the discretization of both the macroscopic CFD volume and the microscopic kMC algorithm. The outlined multiscale model is applied to the deposition of 300 nm thick a-Si:H films revealing significant non-uniformities in the thickness of the thin-film product. An improved reactor geometry is proposed which utilizes a polar showerhead arrangement and a radially-adjusted showerhead hole diameter. This geometry is shown to reduce thickness non-uniformity in the *a*-Si:H thin film product from ~8% to less than 4%, representing significant product quality improvement and financial savings.

2. Three-dimensional Modeling

Recently, Crose et al. (Crose et al., 2017) demonstrated the need for multiscale modeling in the PECVD process due to the interconnection between the macroscopic, reactor scale and the microscopic, thin film growth domains. Specifically, a two-dimensional (2D) axisymmetric geometry was used in the creation of the reactor mesh and a hybrid kinetic Monte Carlo (kMC) algorithm was applied to capture the growth of amorphous silicon (*a*-Si:H) thin film layers. Although this model has proved useful in designing optimized reactor operational strategies, some features of the PECVD reactor are lost when using 2D representations. In looking at Fig. 1, the difference between showerhead geometries becomes clear: the cylindrical showerhead holes and their associated spatial arrangement of the 3D system cannot be directly translated to 2D models. Given the primary motivation of addressing *a*-Si:H thin film product quality through improved PECVD reactor design, in this work the development of a 3D multiscale CFD model will be discussed including both the macroscopic and microscopic domains, as well as the associated dynamic boundary conditions and parallel computing implementation.

2.1. Macroscopic domain

As discussed previously, we utilize a 3D, cylindrical PECVD reactor geometry (see Fig. 1, right) with dimensions typical of those used in industry. The showerhead holes, visible in Figs. 1 and 2, have a diameter of 1 cm and are evenly spaced in a rectangular array throughout the inlet region. In order to solve the partial differential equations which capture the gas phase mass, momentum and energy balances, the reactor geometry is discretized using an unstructured mesh containing ~1.5 million tetrahedral cells. Although the choice to use an unstructured mesh rather than a simplified

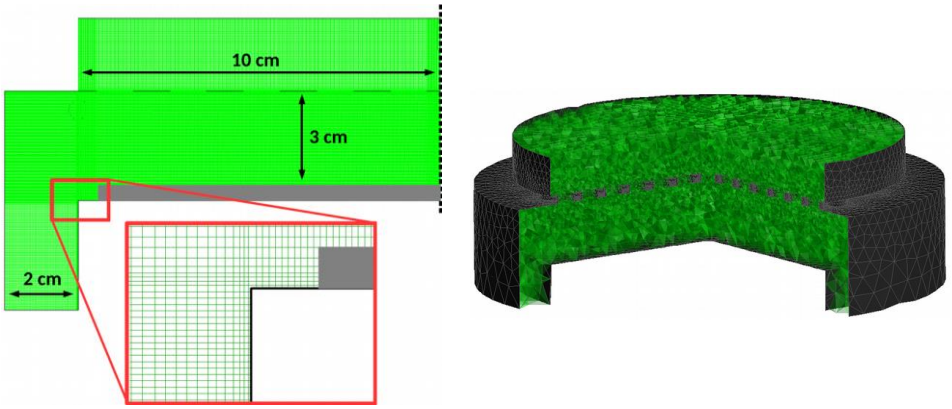


Figure 1. (left) 2D axisymmetric geometry after discretization using a structured mesh containing 120,000 cells. (right) Collection of 1.5 million polygons which define the unstructured, 3D mesh.

structured mesh comes at the cost of computational efficiency, the curvature of the cylindrical reactor shell and showerhead holes favors its use. It is important to note that throughout this work ANSYS software is applied to the creation of the geometric mesh (specifically, ICEM meshing) and as a solver for the PDEs mentioned above.

The ANSYS Fluent software alone cannot yield a multiscale model for the PECVD process of interest; consequently, three user defined functions (UDFs) have been developed to tailor the solver to the deposition of *a*-Si:H thin films. More specifically, the thirty four most dominant gas phase reactions have been accounted for via a volumetric reaction scheme which includes terms for the nine primary plasma reactions which lead to thin film growth. The necessary electron density within the plasma region is calculated using a product of the zero-order Bessel function and a sine function which account for the spatial dependencies within the cylindrical reaction zone. The third UDF necessary to the multiscale nature of the model simulates the growth of *a*-Si:H thin films along the surface of the wafer substrate, and will be touched upon briefly in the following section. The details of these UDFs, in particular the kinetic Monte Carlo (kMC) algorithm which defines the microscopic domain, will not be provided here due to space limitations; however, an in-depth discussion can be found in Crose et al. 2017.

2.2. Microscopic domain

As radicals diffuse from the gas phase down to the substrate surface, they enter the microscopic domain. While tracking of each individual particle remains a computationally infeasible task, growth of *a*-Si:H thin film layers may still be achieved using kinetic Monte Carlo (kMC) algorithms. Specifically, the substrate surface is broken down into discrete regions along the *x*-*y* plane as shown in the right-hand side of Fig. 2, and within each region an independent kMC simulation is executed. Each simulation consists of a lattice of width 1200 particles which captures the growth of a representative *a*-Si:H layer within the associated region. The tetrahedral cells bordering each region are assigned dynamic boundary conditions corresponding to the exchange of mass and energy due to thin film growth at each time step (see Fig. 2, left). Details concerning the development of the kMC algorithm used here may be found in Crose et al. (Crose et al., 2017) and will not be discussed further at this time.

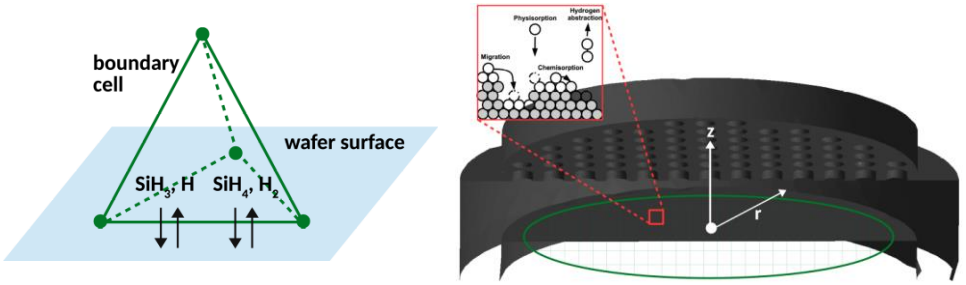


Figure 2. (left) Finite element adjacent to the substrate surface with dynamic boundary condition calculated via microscopic simulation domain. (right) Kinetic Monte Carlo setup within overall multiscale simulation. Wafer substrate discretized in both the *x* and *y* directions forming a 'grid' structure.

2.3. Multiscale workflow

The methodology for connecting the macro- and microscopic domains is of particular importance to the function of the multiscale model; consequently, the simulation workflow must be examined here. At the start of each transient batch simulation, $t = 0$, the first time step for the macroscopic CFD domain is calculated within every cell of the unstructured mesh using ANSYS Fluent as discussed previously. Additionally, the first two UDFs (e.g., the volumetric reaction and electron density scripts) are executed and the results of which are fed into Fluent's PDEs. Once a time step has been completed, $t = t_1$, the species concentration, the temperature and the pressure along the boundary of the wafer substrate are transferred to the microscopic domain (i.e., Fig. 2, left). The discrete kMC simulations discussed above are then initialized using this information allowing for growth of *a*-Si:H thin film layers to begin. Once all of the microscopic simulations have reached t_1, the dynamic boundary conditions for cells bordering the wafer surface are updated based on the mass and energy transfer within the associated region. Again, the macroscopic PDEs are solved such that the time moves forward to t_2 and the cycle continues. In this way, the multiscale model progresses until the end of the batch deposition process is reached (i.e., until $t = t_{batch}$).

2.3.1. Parallel computation

The move from 2D to 3D PECVD reactor models comes at the cost of computational efficiency. In the recent work of Crose et al. (Crose et al., 2017), the 2D axisymmetric CFD simulations required the use of a message passing interface (MPI) structure in order to parallelize the domain and to achieve feasible computation times (e.g., <1 day per batch simulation). In this work, the computational demands are further increased due to the mesh containing ~1.5 million cells (as opposed to 120,000 for the 2D model), tetrahedral cell shape and far more discrete kMC simulations required to span the substrate surface. As a result, parallelization of the multiscale domain remains crucial to operation of transient simulations. Decomposition of the reactor mesh will not be detailed here but may be found in Crose et al. (Crose et al., 2017); however, orchestration of the independent kMC simulations is unique to this work and will be provided below.

As mentioned in previous sections, the substrate surface is broken down into the grid structure shown in the right-hand side of Fig. 2. Given that kMC simulations within this structure are independent (i.e., do not share mass and energy directly), they are potential candidates for distribution amongst the computation nodes. At the start of each batch

simulation, the N_{kMC} wafer regions are assigned evenly among the 128 compute nodes available. The compute nodes then calculate the growth of *a*-Si:H thin film layers for their assigned regions during each microscopic time step. Before the next macroscopic step can begin, growth information from each node is communicated to the host process and the cell boundary conditions are updated.

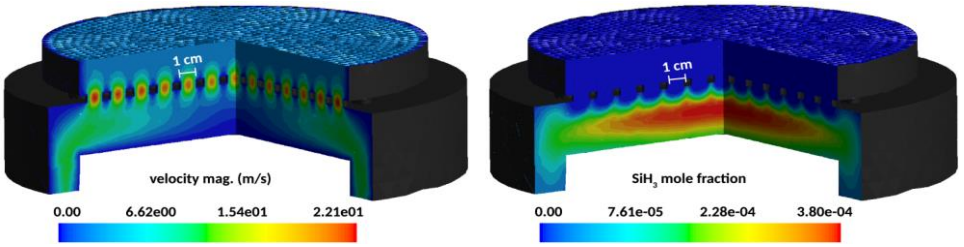


Figure 3. (left) Velocity magnitude within 3D PECVD reactor showing dead-zone near substrate center. (right) Non-uniform, steady-state SiH$_3$ concentration.

3. Results

3.1. Non-uniform deposition

The PECVD reactor cross sections shown in Fig. 3 have been obtained after the brief startup period to give an idea of the gas phase throughout most of the thin film growth (note: the startup period constitutes less than 2% of the deposition time). In looking at the flow field (Fig. 3, left), three distinct regions become apparent: (1) rapid acceleration as the precursor gas passes through the showerhead, (2) increased flow velocity as the unused gas is forced through the narrow outlet regions and (3) a 'dead zone' just above the center of the substrate. The effect of the flow field on product quality can be readily seen in Figs. 3 and 4; the concentration of the primary deposition species, SiH_3, is non-uniform. Not only is there a sharp drop-off in the mole fraction of SiH_3 in the radial, r, direction, but along the azimuthal, θ, direction significant variation exists. The triangular data points in Fig. 5 further highlight this phenomena. In moving radially outward from r = 0 cm to r = 8 cm (the edge of the wafer substrate), the mole fraction of SiH_3 steadily decreases. Additionally, for a fixed radial position (e.g., r = 4 cm) azimuthal variation in x_{SiH3} is worthy of note. In recent years, spatial non-uniformity of deposition species has been shown to contribute to thin film products with thickness offset from the desired set-point (da Silva and Morimoto, 2002, Crose et al., 2016 and Crose et al., 2017), causing device quality issues and poor solar conversion. As such, elimination of 'hot-spots,' or regions defined by significantly different concentrations, is of great interest from a manufacturing perspective.

3.2. Adjusted reactor geometry

By capturing the detailed showerhead geometry, the 3D multiscale CFD model presented here is capable of evaluating alternative PECVD reactor designs. To this end, an adjusted showerhead design is presented here with the goal of reducing spatial non-uniformity in the concentration of SiH_3 such that the *a*-Si:H product is of improved quality. Two key changes have been implemented in the design of the showerhead holes: (1) the rectangular showerhead hole array (e.g., Fig. 4, left) has been altered in favor of the polar array shown in the right-hand side of Fig. 4, and (2) the diameter of the holes are radially dependent. By maintaining larger showerhead holes near the edges of the inlet region with smaller holes at the center, the distribution of SiH_3 is significantly more uniform. In looking at the right of Fig. 4, the hot-spots present in the

original reactor geometry have been eliminated in both the radial and azimuthal directions. In terms of product quality, the circular data points in Fig. 5 demonstrate that the *a*-Si:H thickness offset has been reduced from ~8% to less than 4%.

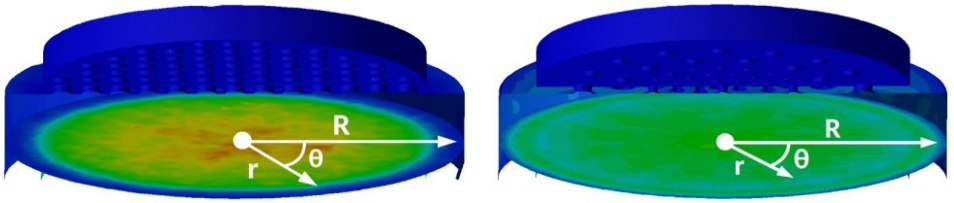


Figure 4. (left) Cross section of SiH_3 concentration taken just above the surface of the wafer substrate (right) SiH_3concentration above the wafer surface for the adjusted reactor geometry.

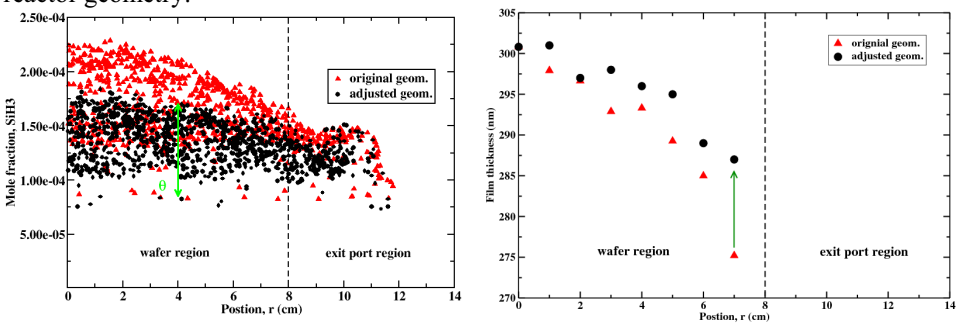


Figure 5. (left) SiH_3 mole fraction as a function of radial position, r, and azimuthal position, θ. (right) Radial *a*-Si:H product thicknesses.

4. Conclusions

An alternative PECVD reactor design has been evaluated using a three-dimensional multiscale CFD model which coordinates communication between the macroscopic reactor scale and the microscopic thin film growth domain. Application of this model to two representative PECVD reactor geometries has shown that thickness non-uniformity in the *a*-Si:H product can be minimized by adjusting the positions and size of the showerhead holes. As a result, the 3D CFD model presented here holds promise for not only improving product quality in PECVD processing, but also for significant savings in time and resources otherwise spent on testing and manufacture of reaction chambers.

References

1. ANSYS Inc. ANSYS Fluent Theory Guide 15.0 (November), November 2013.
2. D. Collins, A. Strojwas, and D. White. A CFD model for the PECVD of silicon nitride. IEEE transactions on semiconductor manufacturing, 7:176–183, 1994.
3. M. Crose, Kwon, JSI., A. Tran, and P. D. Christofides. Multiscale modeling and run-to-run control of PECVD of thin film solar cells. Renewable Energy, 100:129–140, 2017.
4. M. Crose, A. Tran, and P. D. Christofides. Multiscale computational fluid dynamics: Methodology and application to PECVD of thin film solar cells. Coatings, 7:23, 2017.
5. A. da Silva and N. Morimoto. Gas flow simulation in a PECVD reactor. In 2002 International Conference on Computational Nanoscience and Nanotechnology, pages 22–25, 2002.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50401-8

Multi-Objective Computer-Aided Solvent Design for Selectivity and Rate in Reactions

Eliana Grant[a,b], Yueyuan Pan[a], Jeffery Richardson[c], Joseph R. Martinelli[d], Alan Armstrong[b], Amparo Galindo[a], Claire S. Adjiman[a*]

[a]*Department of Chemical Engineering, Centre for Process Systems Engineering, Imperial College London, London, SW7 2AZ, UK*
[b]*Department of Chemistry, Imperial College London, London, SW7 2AZ, UK*
[c]*Discovery Chemistry Synthesis Group, Lilly UK, Erl Wood Manor, Windlesham, Surrey GU20 6PH, UK*
[d]*Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, Indiana 46285, United States*
c.adjiman@imperial.ac.uk

Abstract

A hybrid empirical computer-aided methodology to design the solvent for a reaction, incorporating both selectivity and rate, is presented. A small initial set of diverse solvents is used, for which experimental, *in situ* kinetic data are obtained. A surrogate model is utilized to correlate the reaction kinetics with solvent properties and a computer-aided molecular design (CAMD) multi-objective optimization problem is then formulated to identify solvents with improved performance compared with the initial solvent set. This methodology is applied to an S_NAr reaction of 2,4-difluoroacetophenone with pyrrolidine, which demonstrates an interesting effect of solvent on both the selectivity of the *ortho*-:*para*-substitution ratio and the overall rate of the reaction. A set of Pareto optimal solutions is identified, highlighting the trade-off between reaction rate and selectivity.

Keywords: solvent effects, selectivity, optimization, computer-aided molecular design.

1. Introduction

Solvents can have significant effects on liquid-phase chemical reactions, such as on the rate, selectivity and solubility, yet there is no general model for predicting the solvent effects on reactions (Reichardt and Welton, 2010). In the production of pharmaceuticals and fine chemicals, solvents characteristically constitute between 80 and 90 % of the materials used by mass (Constable et al., 2007). Additionally, given that many traditional solvents are volatile organic compounds and account for around half of the greenhouse gas emissions in the production of pharmaceuticals (Jimenez-Gonzalez et al., 2005), the industry is facing increasing pressure to employ more environmentally friendly solvents, ideally without compromising on reaction efficiency and performance.

Consequently, many methods have been developed to aid in the selection of solvents for reactions. When a vast array of experimental data is available chemometrics can be implemented in order to identify promising solvents (Havet et al., 2002). However, this method is not applicable when experimental data is lacking or laborious to acquire. Computer-aided molecular design (CAMD) has thus evolved as a predictive and systematic method for solvent design. Early CAMD work involved using generate-and-test methods whereby target solvent properties were identified and all the possible

solvents that fit these properties were designed using group contribution (GC) methods (Gani and Brignole., 1983). Following from this, optimization-based CAMD was employed and process performance criteria were targeted (Odele and Machietto, 1993). Whilst generate-and-test methods may be efficient for problems with a small design space, as they involve generating all possible combinations, including many that may be infeasible solutions to the problem, for large design space problems, optimization methods are preferable (Austin et al., 2016).

More recently reaction kinetics have been incorporated into a CAMD framework, both experimentally (Folic et al., 2004) and theoretically (Struebing et al., 2013; Austin et al., 2017), where maximizing the rate constant for the reaction was set as the solvent design target. This concept has been extended to reaction selectivity whereby hypothetical experimental data was generated and a surrogate model was derived based on the solvatochromic equation (Folic et al., 2008). The optimization objective was to maximise the difference in concentration between the desired product and a by-product. In our work an empirical-CAMD methodology is developed to maximise both reaction rate and selectivity in the case of competing reaction pathways. Kinetic data are measured experimentally in order to derive the surrogate model, and multi-objective optimization is employed to incorporate both selectivity and rate of the desired reaction as design objectives. The consideration of both selectivity and rate is of great importance given that they respectively affect the amount of waste and the scale of a typical industrial process. Thus, these objectives are important from a green chemistry and an economic perspective.

2. Empirical-CAMD methodology

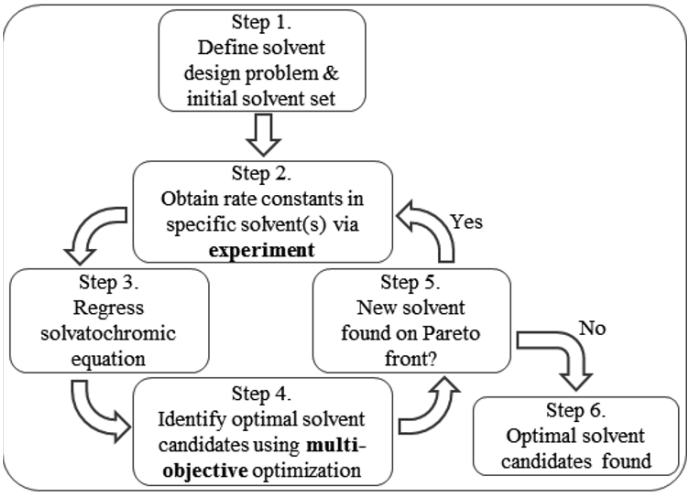


Figure 1: Empirical-CAMD algorithm for identifying optimal solvents for reaction productivity where the solvent influences multiple reaction parameters of interest.

In Step 1 the design objective and constraints are determined as well as the initial solvent set, which should include a diverse range of functional groups and polarity. In Step 2 the rate constants in these solvents are obtained experimentally. A number of possible experimental techniques can be employed, the choice of which is dependent on the specific reaction being studied. In Step 3 a surrogate model is derived from the experimental rate constants. The surrogate model used here is the solvatochromic

equation, which correlates linearly the properties of solvent *i* with the rate constant of a reaction in that solvent (log k_i) using 6 reaction specific coefficients:

$$\log k_i = c_0 + c_A A_i + c_B B_i + c_S S_i + c_\delta \delta_i + c_H(\delta^2_{Hi}/100) \quad (1)$$

where A_i and B_i are hydrogen bonding acidity and basicity of solvent *i* respectively and S_i is its polarisability/polarity (Folic et. al., 2007), δ_i is determined by the number of halogen and aromatic functional groups (Sheldon, et al., 2005) and δ^2_{Hi} is the square of its Hildebrand solubility parameter (Hukkerikar, et al., 2012) all of which are obtained using GC methods. The reaction specific coefficients c_i, i = 0, A, B, S, δ, H are obtained by linear regression of Eq.(1). In Step 4 the solvent design problem is translated into a mathematical optimization problem. A multi-objective optimization (MOO) method is utilised. The ϵ-constraint method (Haimes, et al., 1971) is employed to pose and solve the problem, which transforms one or more of the objectives into a constraint. The general formulation of the problem tackled here is written as follows:

$$\max \quad f_1(\boldsymbol{x}) \quad (2)$$

$$s.t. \quad f_2(\boldsymbol{x}) \geq \epsilon \quad (3)$$

$$\boldsymbol{h}_1(\boldsymbol{x}) \leq Z\boldsymbol{y} \quad (4)$$

$$C\boldsymbol{y} \leq \boldsymbol{d} \quad (5)$$

$$\boldsymbol{x}_i \in \mathbb{R}^m \quad (6)$$

$$\boldsymbol{y}_i \in \{0, 1\}^u \quad i = 1,\ldots,q, \quad (7)$$

where $f_1(\boldsymbol{x})$ is one objective function and $f_2(\boldsymbol{x})$ is the other that is transformed into a constraint using ϵ which is set by the user, $\boldsymbol{h}_1(\boldsymbol{x}) \leq Z\boldsymbol{y}$ is a set of molecular complexity, chemical feasibility and structure-property constraints, $C\boldsymbol{y} \leq \boldsymbol{d}$ is a set of constraints on the binary variables, $\boldsymbol{x}$ is an m-dimensional vector of variables denoting the number of groups as well as the physical properties of the solvent, $\boldsymbol{y}$ is a $q \times u$ matrix of binary variables used to constrain continuous variables such as the number of groups, to integer values and to activate functional groups. The problem is formulated as a mixed-integer linear problem and is solved using the CPLEX solver with the GAMS software, version 24.9.1.

In Step 5 the Pareto-optimal solvents generated in the design step are crosschecked against the solvents used to derive the surrogate model. If at least one of the Pareto-optimal solvents had not been included, the algorithm returns to Step 2; if all the Pareto-optimal solvents were included, the algorithm continues to Step 6, reaching convergence.

3. Case Study - S_NAr Reaction of 2,4-Difluoroacetophenone with Pyrrolidine

Nucleophilic aromatic substitution (S_NAr) reactions are used widely in medicinal chemistry (Brown and Boström, 2016). In addition, where there is the possibility for substitution at more than one site on the aromatic ring, improving the selectivity is important to maximize the reaction productivity and to reduce waste (Wang et al., 2009). Our empirical-CAMD methodology is applied to an S_NAr reaction of 2,4-difluoroacetophenone (A) with pyrrolidine (B) (Fig. 2) where the solvent influences both the rate of the reaction and the selectivity. As shown in Table 1, reactions in solvents with high selectivity values (k_o/k_p), where k_o and k_p are the rate constants for *ortho*- and *para*-substitution respectively, generally result in lower reaction rate constants (k_o). Consequently, both log k_o as a measure of rate and log k_o – log k_p as a measure of selectivity are used as key performance objectives, resulting in a MOO problem.

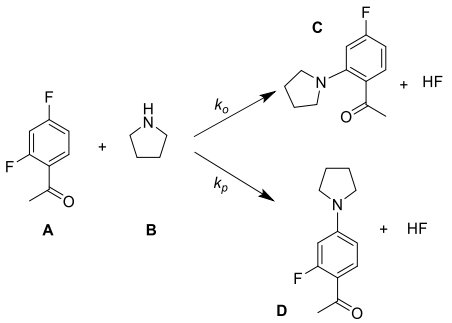


Figure 2: S_NAr reaction that exhibits an effect of solvent on the k_o:k_p selectivity.

In Step 1 the initial solvent set is chosen, as listed in Table 1, with the solvent design objective set to maximize both the rate (f_1=log k_o) and selectivity (f_2=log k_o – log k_p) for *ortho*-substitution. In Step 2 the concentrations of the reaction species in these solvents are monitored using *in situ* ^{1}H NMR experiments at 323 K and ambient pressure. The data is fitted to a kinetic model (one per solvent) to obtain the experimental rate constants k_o and k_p. In Step 3 the solvatochromic equation is regressed with the experimental rate constants k_o and k_p to obtain the surrogate model, the reaction specific coefficients.

Table 1: Experimental *Ortho* to *Para* selectivity of the initial solvent set.

Solvent number	Solvent	k_o x 10^3 (M^{-1} s^{-1})*	Selectivity (k_o/k_p)*	Dielectric constant†
1	1,4-Dioxane	0.44	9.68	2.2
2	Triethylamine	0.20	7.87	2.4
3	Chloroform	0.24	10.80	4.7
4	Chlorobenzene	0.58	14.53	5.7
5	1-butanol	0.13	1.05	17.3
6	Acetonitrile	0.90	1.50	35.7
7	Dimethylformamide	1.12	0.53	37.2
8	Dimethylacetamide	1.14	0.41	37.7
9	Dimethyl sulfoxide	1.97	0.32	46.8

*Experimental rate constants obtained from ^{1}H NMR *in situ* experiments at 323 K and ambient pressure.
†Experimental dielectric constant at 298K (Winget et al., 1999).

A range of algorithms have been proposed to solve MOO problems, specifically the sandwiching algorithm has been applied to mixed-integer problems that have a discontinuous Pareto front (Burger et al., 2015). Here we explore the benefits of MOO without attempting to control computational cost. Instead, we solve problem (2)-(7) repeatedly using integer cuts to generate 50 solutions for each value of ϵ and deduce the Pareto front from this. Constraints are placed on the boiling and melting points to insure that the solvents generated are liquid within the range 298 K to 323 K. Additional environmental and safety constraints are also imposed using GC methods, specifically on

toxicity (LD50) (Hukkerikar, et al., 2012a), the octanol-water partition coefficient at infinite dilution at 298 K (log K_{ow}) and the flash point (Hukkerikar, et al., 2012b).

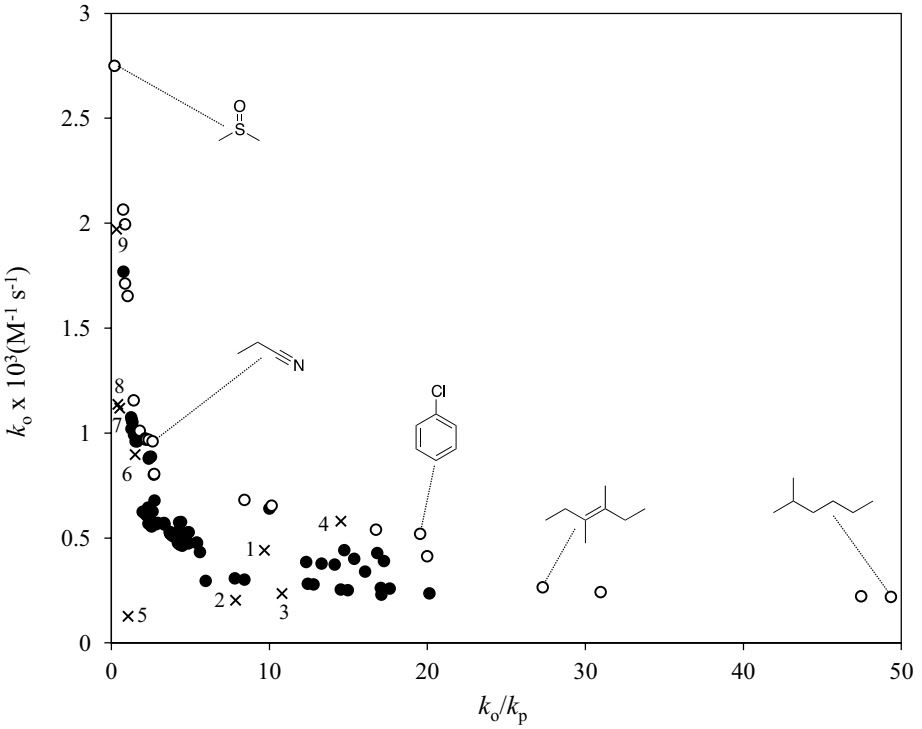


Figure 3: Experimental and CAMD solutions to the MOO problem. Numbered crosses correspond to the experimental values for the solvents in the initial set, closed circles correspond to dominated solutions to the optimization problem, and open circles indicate the Pareto-optimal solutions with example molecules highlighted.

In Fig. 3 solutions to the problem for maximising both rate and selectivity are presented, compared with the experimental values for the initial solvent set and the trade-off between rate and selectivity is evaluated. Propionitrile is generated as a Pareto-optimal solution with a k_o value 65 % lower than for dimethyl sulfoxide, the Pareto-optimal solution with the maximum rate, however an order of magnitude improvement in the selectivity is obtained. Considering the high selectivity region in Fig. 3, the selectivity increases by 81 % from 3,4-dimethyl-3-hexene to 2-methylhexane, the Pareto-optimal solution with maximum selectivity, without a significant negative impact of 17 % on the rate. MOO allows for the trade-off between rate and selectivity to be assessed, enabling a more informed solvent choice for the user, which will ultimately be dictated by considering ease of product separation compared with reactor running costs and logistics.

At this point the solvent design methodology has been implemented for one iteration. However, some of the Pareto-optimal solutions were not included in the initial solvent set and the algorithm will therefore return to Step 2, where in future work the reaction kinetics in these solvents, that are commercially available, will be measured to update the surrogate model. The methodology will then be implemented iteratively until convergence, when all Pareto-optimal solutions are included in the initial solvent set.

4. Conclusions

We have proposed an empirical-CAMD methodology that utilises multi-objective optimization and applied this to identify optimal solvents for an S_NAr reaction. A trade-off between the rate and selectivity is evaluated and a set of Pareto-optimal solutions is identified. Future work involves running the algorithm to convergence in order to obtain the most accurate set of optimal solutions by iteratively improving the surrogate model.

5. Acknowledgements

The authors acknowledge funding from the Pharmacat II Consortium and funding from the Engineering and Physical Sciences Research Council (EPSRC) (EP/M507878/1).

References

N. Austin, N. V. Sanhinidis, D. W. Trahan, 2016, Chemical Engineering Research and Design, 116, pp. 2-26
N. Austin, N. V. Sanhinidis, D. W. Trahan, 2017, Chemical Engineering Science, 159, pp. 93-105
J. Burger, V. Papaioannou, S. Gopinath, G. Jackson, A. Galindo, C. S.Adjiman, 2015, AIChE Journal, 61, pp. 3249-3269
D. G. Brown, J. Boström, 2016, Journal of Medicinal Chemistry, 59, pp. 4443-4458
M. Folić, C. S. Adjiman, 2004, European Symposium on Computer-Aided Process Engineering, 14, pp. 175-180
M. Folić, C. S. Adjiman, E. N. Pistikopoulos, 2007, AIChE Journal, 53, pp. 1240-1256
M. Folić, C. S. Adjiman, E. N. Pistikopoulos, 2008, Industrial & Engineering Chemisry Research, 47, pp. 5190-5202
R. Gani, E. A. Brignole, 1983, Fluid Phase Equilibria, 13, pp. 331-340
Y. Y. Haimes, L. S. Lasdon, D. A. Wismer, 1971, IEEE Transactions on Systems Man and Cybernetics, SMC-1, pp. 296-297
J. Havet, M. Billiau-Loreau, C. Porte, A. Delacroix, 2002 Journal of Automated Methods & Management in Chemistry, 24, pp. 111-119
A. S. Hukkerikar, S. Kalakul, B. Sarup, D. M. Young, G. Sin, R. Gani, 2012, Journal of Chemical Information Modeling, 2012, 52, pp. 2823-2839
A. S. Hukkerikar, B. Sarup, A. T. Kate, J. Abildskov, G. Sin, R. Gani, 2012, Fluid Phase Equilibria, 321, pp. 25-43
O. Odele, S. Macchietto, 1993, Fluid Phase Equilibria, 82, pp. 47-54
C. Reichardt, T. Welton, 2010, Solvent and Solvent Effects in Organic Chemistry, 4th Ed., WILEY-VCH
T. J. Sheldon, C. S. Adjiman, J. L. Cordiner, 2005, Fluid Phase Equilibria, 231, pp. 27-37
X. Wang, E. J. Salaski, D. M. Berger, D. Powell, 2009, Organic Letters, 11, pp. 5662-5664
P. Winget, D. M. Dolney, D. J. Giesen, C. J. Cramer, 1999, Minnesota Solvent Descriptor Database
H. Struebing, Z. Ganase, P. G. Karamertzanis, E. Siougkrou, P. Haycock, P. M. Piccione, A. Armstrong, A. Galindo, C. S. Adjiman, 2013, Nature Chemistry, 5, pp. 952-957

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50402-X

Integrated design of ORC process and working fluid for transient waste-heat recovery from heavy-duty vehicles

Johannes Schilling[a], Katharina Eichler[b], Stefan Pischinger[b] and André Bardow[a,c*]

[a]*Institute of Technical Thermodynamics, RWTH Aachen University, Aachen, Germany*
[b]*Institute for Combustion Engines, RWTH Aachen University, Aachen, Germany*
[c]*Institute of Energy and Climate Research - Energy Systems Engineering (IEK-10), Forschungszentrum Jülich GmbH, Jülich, Germany*
andre.bardow@ltt.rwth-aachen.de

Abstract

Organic Rankine Cycles (ORC) transform low- and medium-temperature heat into mechanical power. One promising application of ORCs is the recovery of exhaust gas heat from heavy-duty vehicles. To utilize the full potential of the transient exhaust gas heat, both, the ORC process and working fluid have to be designed. To integrate the working fluid design into the process design, we developed the so-called 1-stage CoMT-CAMD approach, which allows us to identify the optimal combination of ORC process and working fluid. However, 1-stage CoMT-CAMD is limited to steady-state heat input preventing the consideration of the transient exhaust gas behavior. In this work, we propose an iterative algorithm combining 1-stage CoMT-CAMD with time-series aggregation to tackle the challenge of transient exhaust gas behavior, so-called time-resolved 1-stage CoMT-CAMD. By using time-series aggregation, the transient exhaust gas behavior can be represented with sufficient accuracy by a few time steps serving as quasi-steady-state input for 1-stage CoMT-CAMD. The presented algorithm efficiently identifies the optimal working fluid and ORC process while capturing the transient exhaust gas behavior.

Keywords: 1-stage CoMT-CAMD, PC-SAFT, transient input, time-series aggregation.

1. Introduction

Nearly 30 % of the CO_2 emission in 2014 result from the transportation sector (Muncrief and Sharpe, 2015). The CO_2 emission can be reduced by reducing the fuel consumption of heavy-duty vehicles (HDV). This reduction can be obtained by transforming the exhaust gas waste heat of the engine into mechanical power using Organic Rankine Cycles (ORC). ORCs on HDVs reduce the total cost of ownership and improve the compliance with future CO_2-regulations (Eichler et al., 2015). However, ORCs have high investment and low efficiencies. To ensure an efficient process, tailor-made ORCs are required.

For tailoring an ORC, the working fluid selection has to be integrated into the process design leading to a mixed-integer nonlinear program (MINLP) (Zhang et al., 2016). Solution approaches to this MINLP have been recently reviewed by Linke et al. (2015). In our previous work, we proposed the Continuous-Molecular Targeting−Computer-Aided Molecular Design (CoMT-CAMD) approach for the integrated design of molecules and processes (Bardow et al., 2010; Lampe et al., 2015). Herein, the PC-SAFT equation of state (Gross and Sadowski, 2001) is used as thermodynamic model. In the CoMT stage, a hypothetical optimal fluid is identified by relaxation of the discrete pure component parameters of PC-SAFT. In a second stage, CAMD and a Taylor approximation around

the hypothetical optimum are used to identify real fluids. Recently, the CAMD problem has been directly linked to the CoMT problem in the so-called 1-stage CoMT-CAMD approach enabling us to solve the resulting MINLP in one stage (Schilling et al., 2017a). So far, CoMT-CAMD is limited to a steady-state input. However, for ORCs on HDVs, the exhaust gas heat source shows a transient behavior. Since the optimal working fluid depends on the temperature and mass flow of the heat source, a single design point is not sufficient to capture the transient behavior. However, the consideration of the full transient behavior within the integrated design generally results in large optimization problems, which requires dynamic models and computationally efficient solution methods.

In this work, we therefore propose an iterative algorithm to capture the transient heat input within the 1-stage CoMT-CAMD approach. For this purpose, time-series aggregation techniques are used to transform the transient exhaust gas behavior into quasi-steady-state parameters serving as heat input in 1-stage CoMT-CAMD. An iterative algorithm determines the aggregated time steps required to accurately represent the full time series. The presented algorithm results in an optimal combination of ORC process and working fluid while capturing the full range of the transient exhaust gas behavior.

2. Transient heat source for ORCs on heavy-duty vehicles

For the design of ORCs on heavy-duty vehicles, the full transient driving cycle has to be considered. In this work, the driving cycle is described by the VECTO Long Haul Cycle (LHC) (Dünnebeil et al., 2015). The LHC time series consists of 10656 time steps t for the vehicle velocity and the road slope during $5328\ s$ (Figure 1, left). The LHC time series of the vehicle velocity and road slope are inputs to models for the vehicle and engine of a $40\ t$ HDV with a 6-cylinder $11\ l$ on-road diesel engine with $340\ kW$ at rated power (Bergmann et al., 2017). Thereby, we obtain the time series of the exhaust gas temperature T_t and mass flow rate $\dot{m}_t$ used as input of the ORC (Figure 1, right).

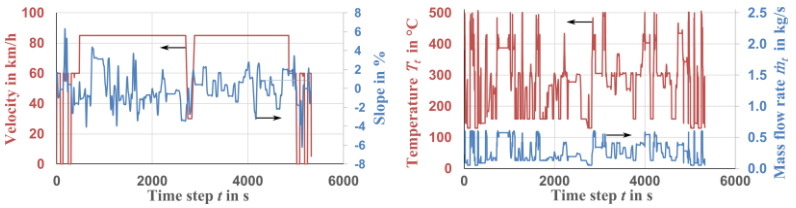


Figure 1: Time series of (left): the velocity and road slope of the VECTO LHC (Dünnebeil et al., 2015) and (right): the temperature and mass flow rate of the exhaust gas (Bergmann et al., 2017).

3. Time-resolved 1-stage CoMT-CAMD

1-stage CoMT-CAMD can be used for the integrated design of ORC process and working fluid considering steady-state heat input (Schilling et al., 2017a). In general, a transient heat source would require a dynamic model leading to optimization problems of prohibitive complexity. Thus, we employ a discrete-time approximation assuming quasi-steady-state behavior. Quasi-steady-state approximation of dynamic ORC models have been successfully used, e.g., by Preißinger et al. (2017). Within the present design context, a quasi-steady-state model allows to capture the full range of the transient exhaust gas, which can be regarded as conservative compared to a dynamic model, which might smoothen peaks of the transient behavior. The resulting time-resolved 1-stage CoMT-CAMD algorithm combines 1-stage CoMT-CAMD with time-series aggregation (Bahl et al., 2017) to efficiently solve the integrated design problem based on quasi-steady-state input.

3.1. Time-resolved 1-stage CoMT-CAMD algorithm

The time-resolved 1-stage CoMT-CAMD algorithm is shown in Figure 2. The input of the process is defined by a time series T. In step 1, we use the k-means algorithm (MacQueen, 1967) to aggregate $t \in T$ time steps to N_k aggregated time steps $k \in K$, where N_k is initialized by the user for the first iteration. Here, we start with a single time step. Each aggregated time step k represents a subset T_k of time steps by one characteristic mass flow rate $\dot{m}_k$ and temperature T_k of the exhaust gas as well as a duration Δt_k. The aggregated time steps serve as heat input to the ORC process. In step 2, we use the aggregated time steps in 1-stage CoMT-CAMD to identify a ranking of working fluids (see section 3.2). In step 3, an individual process optimization of the full time series T is performed for each identified fluid, to validate the quality of the time-series aggregation. The mean relative deviation $\overline{\Delta r}$ is calculated between the objective function with N_k aggregated time steps and the full time series T (step 4). We iteratively increase the number of aggregated time steps N_k, until $\overline{\Delta r}$ satisfies a maximum allowed error ε. The final ranking of working fluids is the result of the time-resolved 1-stage CoMT-CAMD algorithm.

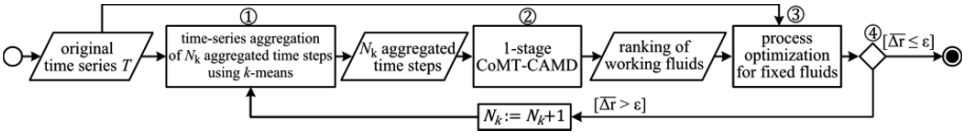


Figure 2: Workflow of the time-resolved 1-stage CoMT-CAMD algorithm.

3.2. Problem definition of time-resolved 1-stage CoMT-CAMD

In time-resolved 1-stage CoMT-CAMD, the N_k aggregated time steps are combined in one overall objective function $\tilde{f}$. The resulting MINLP problem is defined as follows:

$$
\begin{aligned}
\min_{x_k, y^S} \quad & \tilde{f} = \sum_{k=1}^{N_k} \Delta t_k \cdot f_k(x_k, \Theta_k) \\
s.t. \quad & g_{1,k}(x_k, \Theta_k) = 0 && \forall\, k \in K && \text{process model} \\
& g_{2,k}(x_k, \Theta_k) \leq 0 && \forall\, k \in K && \text{process model} \\
& \Theta_k = h(x_k, z, y^S) && \forall\, k \in K && \text{PC-SAFT} \\
& z = GC \cdot y^S && && \text{CAMD and GC method} \\
& F_1 \cdot y^S = 0 && && \text{CAMD and GC method} \\
& F_2 \cdot y^S \leq 0 && && \text{CAMD and GC method} \\
& x_{lb} \leq x_k \leq x_{ub} \in \mathbb{R}^n && \forall\, k \in K \\
& y^S_{lb} \leq y^S \leq y^S_{ub} \in \mathbb{Z}^I.
\end{aligned}
\tag{1}
$$

In this problem, a process-based objective function $\tilde{f}$ (e.g., net power output) is calculated as the sum of N_k objective functions f_k multiplied by the relative duration Δt_k of the aggregated time step k. The objective functions f_k compute the objective for an aggregated time step k and depend on degrees of freedom of the process x_k (e.g., pressure levels) and properties of the working fluid Θ_k (e.g., enthalpies). For each aggregated time step k, equality constraints $g_{1,k}$ (e.g., energy balances) and inequality constraints $g_{2,k}$ (e.g., pressure limits) of the process are considered. The PC-SAFT equation of state (Gross and Sadowski, 2001) is used as consistent thermodynamic model to calculate equilibrium properties Θ_k. In PC-SAFT, a fluid is described by its pure component parameters z. A CAMD formulation enables the design of novel working fluids. Here, a working fluid is represented by the vector y^S, which contains the number of functional groups constituting the molecular structure. In this work, we employ the 20 functional groups presented in Schilling et al. (2017a). Additionally, functional groups for 1-alcohols and

1-amines as well as 16 individual groups for small molecules are considered which cannot be combined with other functional groups (e.g., methane). The molecular structure y^S is linked to the pure component parameters of PC-SAFT using the homosegmented group contribution (GC) method of PC-SAFT (Sauer et al., 2014). Additional constraints F_1 and F_2 ensure the structural feasibility of the molecule taken from Struebing et al. (2011).

The MINLP results in the optimal working fluid y^S and the corresponding optimal process x_k of all aggregated time steps. A ranking of working fluid candidates is calculated using integer-cuts. The MINLP is solved using the local MINLP solver DICOPT (Grossmann et al., 2002); for details see Schilling et al. (2017a).

4. Case study

The time-resolved 1-stage CoMT-CAMD algorithm is applied to the integrated design of a subcritical ORC for waste heat recovery from heavy-duty vehicles (Figure 3). The heat source is characterized by the exhaust gas temperature T_k and mass flow rate $\dot{m}_k$ of aggregated time step k considering the LHC driving cycle as input.

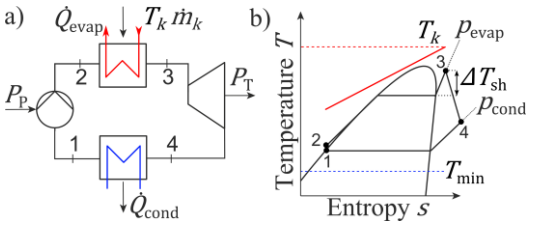


Figure 3: ORC process in (a) flowsheet and (b) qualitative temperature-entropy diagram.

4.1. General specifications of the process

The mean net power output $\bar{P}_{net}$ is considered as objective function $\tilde{f}$ in problem (1):

$$\tilde{f} = \bar{P}_{net} = \sum_{k=1}^{N_k} \Delta t_k \cdot P_{net,k}(x_k, \Theta_k) = \sum_{k=1}^{N_k} \Delta t_k \cdot \left[P_{T,k}(x_k, \Theta_k) - P_{P,k}(x_k, \Theta_k)\right] \quad (2)$$

The degrees of freedom of the process x_k are the mass flowrate of the working fluid $\dot{m}_{wf,k}$, the reduced pressure levels in the condenser p_{cond}^{red} and evaporator $p_{evap,k}^{red}$ and the degree of superheating $\Delta T_{sh,k}$. $\dot{m}_{wf,k}$, $p_{evap,k}^{red}$ and $\Delta T_{sh,k}$ are optimized for aggregated time step k, while p_{cond}^{red} is defined as equal for all aggregated time steps since it is not affected by the transient heat source. In practice, the degree of superheating ΔT_{sh} is not a degree of freedom, but results from the heat transfer in the evaporator, which is not considered in this study. The inequality constraints $g_{2,k}$ of problem (1) specify the minimal approach temperature in the evaporator ($\Delta T_{min} = 20\ K$), the minimal temperature in the condenser ($T_{min} = 80\ °C$), the minimal and maximal pressure levels ($1\ bar \leq p \leq 35\ bar$). We limit the maximal number of functional groups of the molecular structure to $n_{max} = 25$ leading to 126 binary degrees of freedom describing the molecular structure.

4.2. Results and Discussion

In this case study, a ranking of 25 working fluids is calculated for each iteration with N_k aggregated time steps using 1-stage CoMT-CAMD. A maximum allowed error of $\varepsilon = 1\ \%$ is assumed and the proposed algorithm is initialized with $N_k = 1$ aggregated time steps. For a better illustration of the time-resolved 1-stage CoMT-CAMD algorithm, the algorithm is not stopped when $\overline{\Delta r} < \varepsilon$, but continued to $N_k = 8$ and the standard deviation σ of the relative deviation Δr over all working fluids of each ranking is calculated.

For $N_k = 1$ aggregated time step, the mean relative deviation is $\overline{\Delta r} = 37.7\ \%$ with a standard deviation of $\sigma = 5\ ppts$ (percentage points) (Figure 4). The mean relative deviation $\overline{\Delta r}$ decreases quickly and for $N_k = 3$ the time-resolved 1-stage CoMT-CAMD algorithm satisfies the required accuracy of $\varepsilon = 1\ \%$ ($\overline{\Delta r} = 0.98\ \%$ with $\sigma = 0.8\ ppts$). Thus, for the integrated design of process and working fluid, the 10656 time steps of the LHC driving cycle are sufficiently reflected by 3 aggregated time steps. However, the monotony of the algorithm cannot be guaranteed ($N_k = 4$; $\overline{\Delta r} = 2.6\ \%$; $\sigma = 1.4\ ppts$), but 5 to 8 aggregated time steps show again a mean relative deviation of $\overline{\Delta r} < \varepsilon = 1\ \%$ with $\sigma < 0.6\ ppts$. Thus, the mean relative deviation $\overline{\Delta r}$ and standard deviation σ does not steadily decrease with increasing number of aggregated time steps and a systematic validation using the full time series is important (Bahl et al., 2017). The presented time-resolved 1-stage CoMT-CAMD algorithm ensures such a systematic validation.

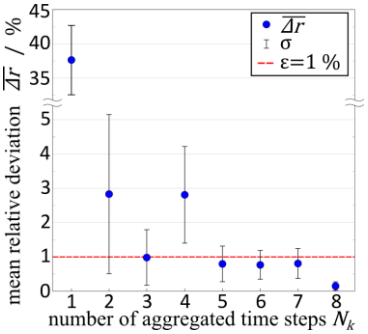


Figure 4: Mean relative deviation $\overline{\Delta r}$ of the objective function value of the 25 working fluids identified with 1-stage CoMT-CAMD using N_k aggregated time steps and the full time series as well as the corresponding standard deviation σ over all working fluids.

Table 1: The top 5 identified working fluids as well as the benchmark ethanol and the corresponding objective function value considering the full time series of the LHC driving cycle.

Rank	Name	P_{net} / kW
1	Butanone	8.15
2	Benzene	8.15
3	n-Butylamine	7.94
4	Butyraldehyde	7.92
5	Acetone	7.91
-	Ethanol	7.45

Due to the local nature of the optimization solver and different inputs, the identified rankings are not necessarily identical between different numbers of aggregated time steps. Only 5 of the 25 working fluids identified with 1 aggregated time step are also identified in the rankings of 2 aggregated time steps, whereas more than 20 of the 25 identified fluids are identical for 2-8 aggregated time steps. The optimal identified working fluid is butanone with $P_{\text{net}} = 8.15\ kW$ (Table 1). Compared to the benchmark ethanol with $P_{\text{net}} = 7.5\ kW$, which is commonly used as working fluid for HDVs, butanone shows a 9.5 % higher objective. Thus, the time-resolved 1-stage CoMT-CAMD algorithm efficiently identifies promising working fluids for ORCs on HDVs while capturing the full transient exhaust gas behavior. In future work, the ranking has to be assessed regarding further criteria, which are not captured within the thermodynamic quasi-steady-state model of the presented case study (e.g., flammability, toxicity, economics or dynamics). Furthermore, the impact of model uncertainties and the local nature of the MINLP solver has to be investigated for the time-resolved 1-stage CoMT-CAMD algorithm (for details on the impact on 1-stage CoMT-CAMD see Schilling et al. (2017a; 2017b)).

5. Conclusions

In this work, we present the time-resolved 1-stage CoMT-CAMD algorithm for the integrated design of process and working fluid for ORCs on heavy-duty vehicles considering the full transient exhaust gas behavior as input of the ORC. For this purpose, we combine

1-stage CoMT-CAMD with time-series aggregation techniques to capture the transient time series of the exhaust gas by aggregated time steps. An accuracy measure is introduced to identify the number of aggregated time steps that are required to reflect the full time series accurately. The resulting algorithm is applied to the design of an ORC on a HDV considering the LHC as driving cycle with 10656 time steps. The results show that the proposed algorithm accurately identifies promising working fluids using 3 aggregated time steps. We identify butanone as the most promising working fluid, which increases the performance by 9.5 % compared to the commonly used working fluid ethanol.

Acknowledgements

We thank the Deutsche Forschungsgemeinschaft (DFG) for funding this work (BA2884/4-2).

References

B., Bahl, A. Kümpel, H. Seele, M. Lampe, and A. Bardow, 2017, Time-series aggregation for synthesis problems by bounding error in the objective function. Energy 135, 900–912.

A., Bardow, K. Steur, and J. Gross, 2010, Continuous-Molecular Targeting for Integrated Solvent and Process Design. Ind. Eng. Chem. Res. 49 (6), 2834–2840.

D. Bergmann, S. Lauer, S. Ghetti, K. Eichler, and S. Petri, 2017, Holistic simulation tool for the commercial vehicle powertrain development, in: 14. International Conference Commercial Vehicles 2017: truck, bus, van, trailer: Friedrichshafen, May 16 -17, 2017 / VDI Automotive and Traffic Systems Technology. VDI Verlag GmbH, Düsseldorf, pp. 209–222.

F. Dünnebeil, C. Reinhard, U. Lambrecht, A. Kies, S. Hausberger, and M. Rexeis, 2015, Future measures for fuel savings and GHG reduction of heavy-duty vehicles. Federal Environment Agency (Germany), Nr. 32.

K. Eichler, Y. Jeihouni, C. Ritterskamp, 2015. Fuel Economy Benefits for Commercial Diesel Engines with Waste Heat Recovery. SAE Int. J. Commer. Veh. 8 (2), 491–505.

J. Gross, and G. Sadowski, 2001, Perturbed-Chain SAFT: An Equation of State Based on a Perturbation Theory for Chain Molecules. Ind. Eng. Chem. Res. 40 (4), 1244–1260.

I.E. Grossmann, J. Viswanathan, A. Vecchietti, R. Raman, E. Kalvelagen, 2002, GAMS/DICOPT: A discrete continuous optimization package. GAMS Corporation Inc.

M. Lampe, M. Stavrou, J. Schilling, E. Sauer, J. Gross, and A. Bardow, 2015, Computer-aided molecular design in the continuous-molecular targeting framework using group-contribution PC-SAFT. Comput. Chem. Eng. 81, 278–287.

P. Linke, A. Papadopoulos, and P. Seferlis, 2015, Systematic Methods for Working Fluid Selection and the Design, Integration and Control of Organic Rankine Cycles-A Review. Energies 8 (6), 4755-4801.

J. MacQueen, 1967, Some methods for classification and analysis of multivariate observations, in: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics. University of California Press, Berkeley, Calif., pp. 281–297.

R. Muncrief, and B. Sharpe, 2015. Overview of the heavy-duty vehicle market and CO2 emissions in the European Union. Washington, DC: The International Council on Clean Transportation.

M. Preißinger, J.A.H. Schwöbel, A. Klamt, and D. Brüggemann, 2017. Multi-criteria evaluation of several million working fluids for waste heat recovery by means of Organic Rankine Cycle in passenger cars and heavy-duty trucks. Applied Energy 206, 887–899.

E. Sauer, M. Stavrou, and J. Gross, 2014. Comparison between a Homo- and a Heterosegmented Group Contribution Approach Based on the Perturbed-Chain Polar Statistical Associating Fluid Theory Equation of State. Ind. Eng. Chem. Res. 53 (38), 14854–14864.

J. Schilling, M. Lampe, J. Gross, and A. Bardow, 2017a. 1-stage CoMT-CAMD: An approach for integrated design of ORC process and working fluid using PC-SAFT. Chem. Eng. Sci. 159, 217–230.

J. Schilling, D. Tillmanns, M. Lampe, M. Hopp, J. Gross, and A. Bardow, 2017b. From molecules to dollars: Integrating molecular design into thermo-economic process design using consistent thermodynamic modeling. Mol. Syst. Des. Eng. 2 (3), 301–320.

H. Struebing, C.S. Adjiman, and A. Galindo, 2011. Optimal solvent design for reactions using computer-aided molecular design. Imperial College London. http://www.minlp.org/library/problem/mod/index.php?lib=MINLP&i=180&pi=-137. Accessed 9 November 2015.

L. Zhang, D.K. Babi, and R. Gani, 2016. New Vistas in Chemical Product and Process Design. Annu. Rev. Chem. Biomol. Eng. 7, 557–582.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50403-1

Study on DFT Based Computation for Quantum Chemical Descriptors of Neodymium Carboxylate Molecules

Shuguang Xiang *, Jiye Wang

Institute of Process Systems Engineering，Qingdao University of Science & Technology，Qingdao 266042，China
xsg@qust.edu.cn

Abstract

As a relatively new research domain of Process Systems Engineering (PSE), Computer-aided Molecular Design (CAMD) explores the micro-scale systems, such as compound molecules, to discover the quantitative structure property relationship (QSPR) or the quantitative structure activity relationship (QSAR) of a molecule. One of the approaches to create QSPR or QSAR is to compute quantum chemical descriptors by some commercial quantum chemistry software. In this study, ADF2014, a piece of quantum chemistry computational software based on density functional theory (DFT), was chosen to compute the titled compound molecules and obtained well ballanced accuracy and efficiency of computation. Detailed computation methods were developed, including how to build an initial molecular geometry, how to set the computing parameters, how to simplify the initial geometry, how to deal with the the conformation search, etc. Four descriptors including the HOMO energy, the LUMO energy, the dipole moment and the polarizability of ligand were selected according to the variation percentage and the correlation matrix. The four descriptors of 17 neodymium carboxylates were computed. Results indicate that the repeatability of the computation methods is well and the variation among computed values of a quantum chemical descriptor of different neodymium carboxylates is outstanding, which meets the requirement for QSAR modeling.

Keywords: computation method, quantum chemical descriptor, neodymium carboxylate

1. Introduction

Process Systems Engineering (PSE) is nowadays developing from the traditional meso scale into the macro and the micro scale. In the direction of micro scale, Product Engineering or Computer-aided Molecular Design (CAMD) focuses on studying the quantitative structure property relationship (QSPR) of a molecule (Alexandre et al., 2014). For a catalyst, the quantitative structure activity relationship (QSAR) between its molecular structure and catalytic activity can be studied by computing its quantum chemical descriptors, which is an approach well developed in the pharmaceutical field (Boone et al., 2016; Gubskaya, 2010) for molecular design of medicines or pesticides. However, it is not an easy thing to compute quantum chemical descriptors of a compound molecule, especially for such compound as neodymium carboxylate with a neodymium atom in its molecule, due to the many-body effect, the relativistic effect and the multi-configuration effect (Martelli et al., 2014; Xiang et al., 2016).

The neodymium carboxylate is usually used as a key component of catalyst for coordination polymerization of dienes in the rubber manufacturing industry to produce synthetic rubbers. Impoving its catalytic activity has been an everlasting subject. While the traditional research method is uasually by experiments in the laboratry, the QSAR method that combines quantum chemical computation and experiments has not been reported for the titled compound. But we believe it is a hopeful method to greatly increase the catalytic activity if a proper approach to compute the quantum chemical descriptors is to be created, which attracted us to carry out this study.

The density functional theory (DFT) approach is known as the most popular one in quantum chemical computation. It allows for correlation effects to be included in a computation that takes roughly the same time as a Hartree-Fock computation that does not include correlation effects (Levine, 2013; Yu et al., 2016). The ADF module in ADF2014 used in this study is based on DFT (Baerends et al., 2014). Evenmore, as a semiempirical approach, the MOPAC imbedded in ADF has advantages in geometry preoptimization of neodymium complexes (Stewart, 2012). With all above, the computation efficiency and accuracy are anticipated to be well ballanced.

2. Computations

2.1. Hardware and Software

Three DIY desktop PCs each with a 1T hard disk, a 16G RAM and an Intel Core i7 CPU (3.4 GHz); The ADF module of ADF2014, version ADF2014.07 201505162227 r46825.

2.2. Computation Methods

2.2.1. Development of computation methods

(1) Determination of initial geometry

According to Wang and Xiang (2015), and Zhang et al. (1990), and for computation efficiency, the initial coordination structure for neodymium carboxylate complexes was simplified as a six-coordinated monomeric one and pre-optimized with UFF (Rappe et al., 1992) and MOPAC into a trigonal antiprism or an octahedron, shown in Fig. 1.

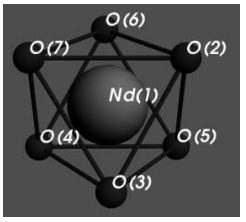


Figure 1. Simplified initial coordination structure of carboxylate (Other atoms are omitted)

(2) About conformation search and the most stable conformer

Geometry optimization usually converges to a local minimum other than a global one. In order to arrive at the global minimum, one should do conformation search and geometry optimization of each conformer, which is a trivial and time-consuming job (Xiang et al., 2016). For simplification, these were omitted in this work, which made the computation efficiency increased by at least 10 times. It is believed that the geometry optimization of neodymium carboxylate complexes initially optimized by MOPAC2012

may reach the global minimum or a nearby local minimum (Dutra et al., 2013). Moreover, with the same meothods, the optimized geometries, though not the global minimums, ought to be comparable, which is enough for computation of descriptors to be used in QSAR modeling.

(3) Computation of quantum chemical descriptors

For QSAR modeling, eight quantum chemical descriptors related to catalytic activity were computated, including the frontier orbital energies, ε_{HOMO} and ε_{LUMO}, the Mullikin charge of Nd, q, the dipole moment, μ, the oil-water partition coefficient, logP, the ligand molecular polarizability, α, the electrophilicity index, ω, and the frontier orbital energy gap, ε_{HOMO}-ε_{LUMO}.

2.2.2. Detailed computation methods

(1) Method to build the initial geometry

In the INPUT module of ADF2014, build an initial geometry as shown in Fig. 1, and optimize with the MOPAC module: Method=PM6， Total charge=0.0， Spin polarization=3，Unrestricted=No，Sparkles=yes.

(2) Precise optimization of geometry

Precise optimization was performed in the ADF module: Total charge =0.0，Spin polarization = 3，Unrestricted = yes，XC potential in SCF = GGA: BLYP，Relativity (ZORA) = Scalar，Basis set = TZP，Frozen core = Large，Numerical Quality = Normal/STO fit; Coordinates used for optimization = Cartesian, Optimization method = AUTO, Number of geometry iterations = 200, Maximum number of SCF cycles = 300, SCF convergence method = A-DIIS, Molecular symmetry = NOSYM.

(3) Stability verification of optimized geometry

Parameter settings for frequency calculation included those in the precise optimization of geometry and, Analytical frequencies = yes, Max number of CPKS iterations = 50.

(4) Computation of quantum chemical descriptors

The HOMO energy, the LUMO energy, the Mulliken charge, the dipole moment and the polarizability of ligand can be obtained directly from the out file of geometry optimization of complex or ligand. The electrophilicity index, ω, was calculated as Eq. (1). The logP was calculated using the solvation model, COSMO-RS, with concentration in mole fraction, solvents of n-hexane and water, and temperature of 298.15 K.

$$\omega = \frac{(\varepsilon_{LUMO}+\varepsilon_{HOMO})^2}{4(\varepsilon_{LUMO}-\varepsilon_{HOMO})} \quad (1)$$

3. Results and discussions

3.1. Verification of computation methods

First, repeated the computation of neodymium n-pentanoate for six times to obtain the eight descriptors related to the catalitic activity. Results are shown in Table 1, in which CV is the coefficient of variation and defined as the ratio of the standard deviation, s, to the absolute average value of a descriptor, |av|, in percentage, as Eq. (2). Results in Table 1 show that the CV values are small, indicating good repetitiveness.

$$CV = s/|av| \times 100 \quad (2)$$

Second, the descriptors and their CV values for five different neodymium carboxylate complexes were computed and listed in Table 2. The descriptors with CV values more than ten times larger as CV_0 include α, logP, μ, ε_{HOMO} and ε_{LUMO}. That is to say, these descriptors differ significantly among different neodymium carboxylate complexes and are thus suitable for building QSAR. Therefore, the computation methods are proved to be suitable for QSAR study.

Table 1. Repeated computation results of descriptors for neodymium n-pentanoate

No.	ε_{HOMO} a.u.	ε_{LUMO} a.u.	q /	M D	logP /	α a.u.	ω a.u.	ε_{HOMO}-ε_{LUMO} a.u.
1	-0.23840	-0.10658	1.5862	0.6615	4.557	68.37	7.9596	-0.13182
2	-0.23858	-0.10626	1.5854	0.5752	4.524	68.62	8.2277	-0.13232
3	-0.23840	-0.10646	1.5851	0.6496	4.618	68.75	7.9537	-0.13194
4	-0.23835	-0.10642	1.5864	0.6621	4.548	68.75	8.1058	-0.13193
5	-0.23845	-0.10641	1.5856	0.6331	4.454	68.75	7.9728	-0.13204
6	-0.23851	-0.10642	1.5865	0.6104	4.547	68.75	8.1439	-0.13209
CV_0, %	0.04	0.10	0.04	5.38	1.17	0.22	1.43	0.131

Table 2. Computation results of descriptors for different neodymium carboxylates

	ε_{HOMO} a.u.	ε_{LUMO} a.u.	Q /	M D	logP /	α a.u.	ω a.u.	ε_{HOMO}-ε_{LUMO} a.u.
Acetate	-0.24129	-0.10972	1.5898	0.0767	-1.530	38.25	9.0124	-0.13157
Propionate	-0.23936	-0.10720	1.5905	0.4175	0.770	48.16	8.0700	-0.13216
N-butyrate	-0.23908	-0.10659	1.5885	0.2567	2.650	58.40	7.9346	-0.13249
N-pentanoate	-0.23835	-0.10642	1.5864	0.6621	4.548	68.75	8.1058	-0.13193
N-hexanoate	-0.23821	-0.10588	1.5827	1.0494	6.456	78.88	7.6353	-0.13233
CV, %	0.46	1.26	0.18	68.74	108.40	24.63	5.66	0.24
CV/CV_0	11.5	12.6	4.4	12.8	92.6	111.9	4.0	1.9

3.2. Selection of proper quantum chemical descriptors

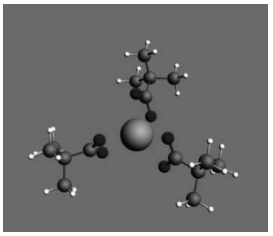
As stated in 3.1, some descriptors are suitable for QSAR modelling. But, not all are necessary because they may be correlated to one another and correlation with the molecular structure may not be significant. Therefore, the linear correlation was studied with the correlation matrix and the results show that four quantum chemical descriptors, i.e. ε_{HOMO}, ε_{LUMO}, μ and α, can be chosen as the representative.

3.3. Computation results of descriptors

Seventeen neodymium carboxylate complexes with different ligands from the acetic acid to the n-tetradecanoic acid were selected to compute their quantum chemical descriptors. One sample of the optimized molecular geometries and the computation results of all the seventeen are shown in Fig. 2 and Table 3.

Results in Table 3 show that CV/CV_0 are all larger than 10.0, which means descriptors ε_{HOMO}, ε_{LUMO}, μ and α all have significant variations among the seventeen complexes

compared with the accidental computation error and are thus suitable for QSAR modeling.



Big grey ball-Nd
Small white ball-H
Middle black ball-O
Middle grey ball-C

Figure 2. Optimized molecular geometries of neodymium neopentanoate

Table 3. Computed descriptors of neodymium carboxylate complexes

No.	Name	Formula	ε_{HOMO} a.u.	ε_{LUMO} a.u.	μ D	α a.u.
1	Acetate	$Nd(CH_3COO)_3$	-0.24129	-0.10972	0.0767	38.25
2	Propionate	$Nd(C_2H_5COO)_3$	-0.23936	-0.1072	0.4175	48.16
3	N-butyrate	$Nd(C_3H_7COO)_3$	-0.23908	-0.10659	0.2567	58.4
4	N-pentanoate	$Nd(C_4H_9COO)_3$	-0.23835	-0.10642	0.6621	68.75
5	Neopentanoate	$Nd((CH_3)_3CCOO)_3$	-0.23539	-0.10276	0.2626	73.61
6	N-hexanoate	$Nd(C_5H_{11}COO)_3$	-0.23800	-0.10545	0.9882	78.32
7	2-Ethylbutyrate	$Nd(C_2H_5C_2H_5CHCOO)_3$	-0.23582	-0.10344	0.0748	77.06
8	2-Methylpentanoate	$Nd(C_3H_7CH_3CHCOO)_3$	-0.23553	-0.10322	0.2688	82.82
9	Neohexanoate	$Nd(C_2H_5(CH_3)_2CCOO)_3$	-0.23495	-0.10182	0.3513	84.29
10	N-heptanoate	$Nd(C_6H_{13}COO)_3$	-0.23797	-0.10602	0.726	88.33
11	N-octanoate	$Nd(C_7H_{15}COO)_3$	-0.2379	-0.10553	0.8624	98.58
12	2-Ethylhexanoate	$Nd(C_4H_9C_2H_5CHCOO)_3$	-0.23537	-0.10536	0.3552	115.44
13	N -nonanoate	$Nd(C_8H_{17}COO)_3$	-0.23779	-0.10574	0.3348	108.06
14	N -decanoate	$Nd(C_9H_{19}COO)_3$	-0.23793	-0.10588	0.9055	118.11
15	Neodecanoate	$Nd(C_6H_{13}(CH_3)_2CCOO)_3$	-0.23416	-0.10117	0.6093	123.36
16	N -dodecnoate	$Nd(C_{11}H_{23}COO)_3$	-0.23816	-0.10682	0.3563	138.58
17	N -tetradecanoate	$Nd(C_{13}H_{27}COO)_3$	-0.23821	-0.10671	0.4720	160.16
		CV，%	0.79	2.05	59.19	35.22
		CV/CV_0	19.7	20.5	11.0	160.1

4. Conclusions

(1) The ADF module of ADF2014 was chosen to calculate descriptors of neodymium carboxylate molecules, which well ballanced the computation accuracy and efficiency.

(2) Detailed computation methods were developed. Results proved that the developed methods have good repeatability and the selected descriptors have significant enough variation among different neodymium carboxylate complexes, which is required by QSAR modeling.

(3) With the help of the variation coefficient and the correlation matrix, four suitable quantum chemical descriptors were selected, including the energy of the highest occupied molecular orbital, the energy of the lowest unoccupied molecular orbital, the dipole moment and the polarizability of ligand.

(4) The values of the four selected descriptors of seventeen neodymium carboxylate complexes were computed and obtained. The results are suitable for QSAR modeling and further for CAMD.

Acknowledgement

Thanks to National Natural Science Foundation of China for its financial support (project No. 21476119).

References

A.K. Rappe, C.J. Casewit, K.S. Colwell, W.A. Goddard Ⅲ, et al., 1992, UFF, a full periodic table force field for molecular dynamics simulations, Journal of the American Chemical Society, 114, 25, 10024-10035.

A.V. Gubskaya, 2010, Quantum-chemical descriptors in QSAR/QSPR modeling: Achievements, perspectives and trends, Wiley-VCH Verlag GmbH & Co. KGaA, 693-721.

C.D. Alexandre, S.B. Costin, A.K. Anton, 2014, Chapter 12 - Chemical product design, Computer Aided Chemical Engineering, 35, 489-523.

D. Zhang, C. Huang, G. Xu, 1990, Structural studies on rare earth carboxylates I. Studies on the crystal structure of tris (neopentanoato) – tris (neopentanoic acid) neodymium (III) dimer, [$NdA_3(HA)_3]_2$, Chinese Journal of Chemistry, 8, 6, 529-535.

E.J. Baerends, T. Ziegler, J. Autschbach, et al., 2014, ADF2014, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, http://www.scm.com.

F.Martelli, Y. Jeanvoine, T. Vercouter, C. Beuchat, R. Vuilleumier, R. Spezia, 2014, Hydration properties of lanthanoid (III) carbonate complexes in liquid water determined by polarizable molecular dynamics simulations, Physical Chemistry Chemical Physics, 16, 8, 3693-3705.

H.S. Yu, S.L. Li, D.G. Truhlar, 2016, Perspective: Kohn-Sham density functional theory descending a staircase, The Journal of Chemical Physics, 145, 130901.

I. N. Levine, 2013, Quantum Chemistry, Seventh Edition, Pearson Press.

J.D. Dutra, M.A. Filho, G.B. Rocha, R.O. Freire , A.M. Simas , 2013, Sparkle/PM7 lanthanide parameters for the modeling of complexes and materials, Journal of Chemical Theory & Computation, 9, 8, 3333–3341.

J.J.P. Stewart, 2012, MOPAC2012, Molecular Orbital PACkage, Stewart Computational Chemistry, Colorado Springs, Co, USA.

J. Wang, S. Xiang, 2015, Research on catalytic activity rules of neodymium carboxylate complexes for stereospecific polymerization of 1,3-butadiene, Computers and Applied Chemistry, 32, 4, 413- 418.

K. Boone, F. Abedin, M. R. Anwar, K.V.Camarda, 2016, Chapter 8 - Molecular design in the pharmaceutical industries, Computer Aided Chemical Engineering, 39, 221-238.

S. Xiang, J. Wang, X. Sun, 2016, Quantum chemistry based computational study on the conformational population of a neodymium neodecanoate complex, Chinese Journal of Structural Chemistry, 35, 6, 833-838.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50404-3

Experimental validation of scheduled PID control for directed self-assembly of colloidal particles in microfluidic devices

Yu Gao, Richard Lakerveld*

Department of Chemical and Biological Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong S.A.R
kelakerveld@ust.hk

Abstract

Directed self-assembly provides a promising route to assemble small-scale structures for engineering applications. Typically, a local particle density has to be controlled to assemble specific structural features defect-free. The strong non-linearity of self-assembly systems complicates such control. Therefore, in this work, a gain-scheduling approach for automated control of directed self-assembly is developed and experimentally tested. An empirical input-output model is developed from experimental data, which is used to design a scheduling approach for a feedback controller. The controller is tested in a microfluidic cell with patterned electrodes to direct the self-assembly process with electric fields. The experimental results demonstrate a significantly improved dynamic performance over a broad range of operating parameters when gain-scheduling is considered compared to a conventional feedback controller.

Keywords: directed self-assembly, small-scale systems engineering, feedback control, gain scheduling

1. Introduction

Self-assembly is the spontaneous association of small particles into structures, which has great potential for the precise manufacture of advanced materials. Electric fields can be used to direct the self-assembly into a desired direction flexibly due to their tunable strength and direction. Once in-situ observation of the system is available, automated control can be implemented to dynamically shape the electric field to direct the self-assembly process, which is an example of small-scale process control (Ulissi, 2013; Paulson, 2015). Feedback control has been demonstrated to be effective to self-assemble colloidal crystals with low defects (Juárez and Bevan, 2012; Tang et al., 2016) or to position particles along predefined lines to fabricate non-periodic structures with single-particle resolution (Gao and Lakerveld, 2017). Tight control of the local particle density is important for both crystalline structures (Juárez et al., 2012) and individual particle alignment (Gao and Lakerveld, 2017). Furthermore, local particle density control is an important basis for the implementation of hierarchical multi-resolution approaches to self-assemble more complicated non-periodic structures (Lakerveld et al., 2012; Solis et al., 2010). However, such automated control of a local particle density is challenged by the stochastic and highly nonlinear behavior of directed self-assembly systems when external electric fields are used as manipulated variables. The abundance of engineering systems with non-linear behaviour has led to much progress on control

strategies for nonlinear systems such as Linear Parameter-Varying (LPV) control (Apkarian et al., 1995), which includes gain-scheduling (Shamma and Athans, 1992). However, to the best of our knowledge, gain scheduling methods have not been applied for controlling directed self-assembly systems despite their great potential for this new type of application area.

In this work, a gain-scheduling feedback control strategy for directed self-assembly of colloidal particles is developed and validated experimentally. The control objective is to control a local particle density using patterned electrodes in a microfluidic cell over a broad range of operating conditions despite the strong non-linear behaviour of such system. An input-output model describing the particle density as function of both electric field frequency and voltage between electrodes is developed from experimental data. Subsequently, an optimal value for voltage to maximize the range of attainable particle density is identified, which is used to develop an automated control strategy based on scheduled PID control with field frequency as manipulated variable. Finally, the novel particle-density controller is implemented experimentally in the microfluidic device for performance assessment.

2. Input-output model development and controller design

A microfluidic cell with patterned electrodes was used for model development and controller implementation. The details of this experimental setup are described elsewhere (Gao and Lakerveld, 2017). In summary, a suspension with microspheres is placed between two transparent ITO-coated glass slides that act as electrodes, of which one has patterns of photoresist to create a non-uniform field. The glass slides are separated by a thin rubber spacer. A low-voltage AC electric field is induced between both electrodes using a function generator (33500B, Keysight Technologies) to direct the self-assembly process. An optical microscope (Nikon, Ni-U) equipped with a research-grade digital camera (Nikon Digital Slight Qi2) and automated image analysis (Matlab) are used to observe the positions of all particles in real time with a typical time lag of 1s.

The particle density in a specific area of the cell was measured for various voltages in the range of 2-10V and frequencies in the range of 60-300kHz to develop the empirical input-output model. Five sequential images were used to determine the stable particle density for a given pair of voltage and frequency (Figure 1A) once there were no more significant changes in particle density. Illustrative microscopy images corresponding to the various measured particle densities are shown in Figure 2A. It can be seen that the particle density depends approximately linearly on the voltage for the applied frequencies. More data would be needed to reveal any higher-order dependency. Depending on the frequency, particle density will either go up or down when increasing the voltage. This behaviour is likely caused by a switch in dominating physical phenomena of the self-assembly system, i.e., the system behaviour is driven by positive dielectrophoresis (pDEP) at the lower frequencies and negative dielectrophoresis (nDEP) at higher frequencies (Oh et al., 2009). The measured particle densities clearly demonstrate the non-linear behaviour of the system, which motivates the use for a scheduling approach.

To construct the empirical model, linear interpolation between two logarithmic functions that describe the particle density ρ (μm^{-2}) as function of frequency f (kHz) at the lowest and highest voltage tested is used as follows:

$$\rho(f,v) = a_1 \cdot \log_{10}\left(\frac{f}{f_0}\right) + b_1 + \frac{a_2 \cdot \log_{10}\left(\frac{f}{f_0}\right) + b_2 - a_1 \cdot \log_{10}\left(\frac{f}{f_0}\right) - b_1}{v_{max} - v_{min}}\left(v - v_{min}\right) \quad \forall v = \left[v_{min}, v_{max}\right] \quad (1)$$

where a_1, a_2, b_1, b_2 are fitting parameters obtained from the experimental data and v_{max} and v_{min} are the maximum and minimum voltages tested, respectively. Numerical values for the model parameters are given in Table 1. Figure 1B and 1C show that the measured particle density at the lowest and highest voltage can be well described using a logarithmic function with two fitting parameters each.

Figure 2B and Figure 2C show a plot of the density as function of frequency and voltage as described by Eq.(1) with parameters as given in Table 1. The largest range of attainable particle density can be achieved for a voltage of 10V when varying frequency. Therefore, a constant voltage of 10V is used for the design of a PI controller with frequency as the manipulated variable.

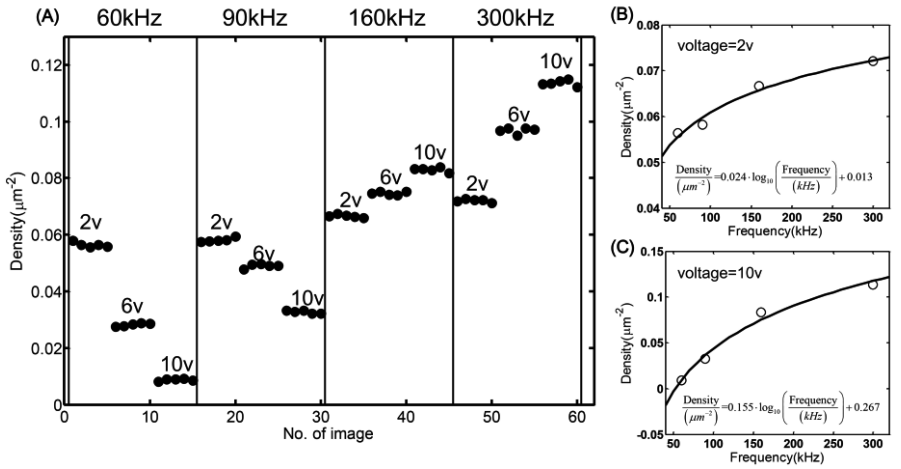


Figure 1. (A) Particle density of self-assembly system as function of frequency and voltage; (B) Empirical model for particle density as function of frequency at a constant voltage of 2V; (C) Empirical model for particle density as function of frequency at a constant voltage of 10V.

Table 1. List of numerical values for the model parameters

a_1/μm^{-2}	b_1/μm^{-2}	a_2/μm^{-2}	b_2/μm^{-2}	v_{min}/V	v_{max}/V	f_0/kHz
0.024	0.013	0.155	0.267	2	10	1

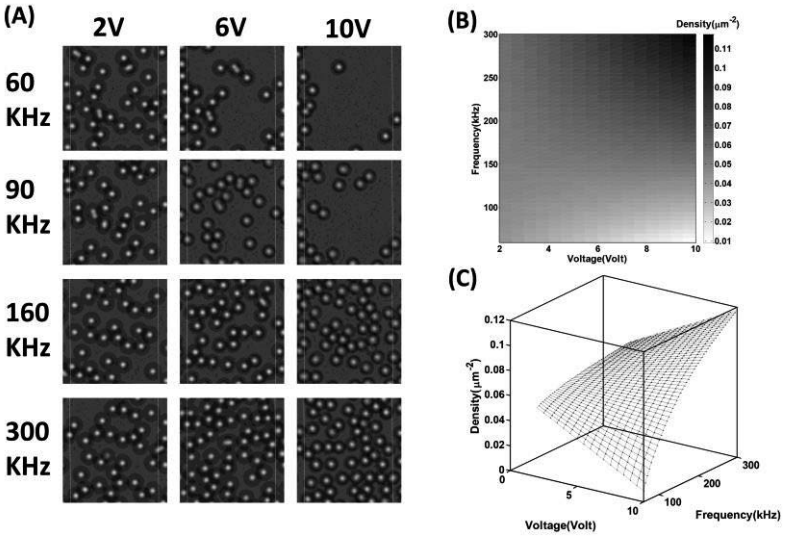


Figure 2. (A)Microscopy images of particle densities at different operating frequencies and voltages. Each image shows a part of the area that is used for measuring the density with image analysis for illustrative purposes; (B) Grayscale mapping of the density with respect to frequency and voltage; (C) Surface plot of the particle density as function of frequency and voltage.

Derivative action is not included in the controller design as it is complicated by measurement noise, which is common for self-assembly systems due to Brownian motion. A gain scheduling procedure is applied to account for the nonlinear dependency of the manipulated variable (frequency) on controlled variable (local particle density). The process gain K can be obtained directly from the empirical model:

$$K=\left.\frac{\partial\rho(f,v)}{\partial f}\right|_{v=10}=\frac{a_2}{\ln 10}\cdot\frac{1}{f}\ , \tag{2}$$

which is used to calculate the tuning parameters for a gain-scheduled PI controller using the Internal Model Control tuning method as follows:

$$K_p=\frac{2\tau+\theta}{2K\lambda},\ \tau_i=\tau+\frac{\theta}{2},\ \lambda\geq 1.7\theta\ , \tag{3}$$

where θ is the system dead time, which is equal to 1s due to the delay from image analysis. τ is the process time constant, which is approximated as 30s by measuring the dynamic response of particle density with respect to a step change in frequency. Finally, λ is obtained as 25s by trial and error.

3. Controller implementation and performance assessment

The designed controller is implemented in MATLAB using the Keysight Command Expert interface between Matlab and the function generator within the experimental setup. Image analysis is used to measure the local particle density with a delay of approximately 1s. The dynamic development of the local particle density with the proposed gain-scheduled controller in closed-loop mode for a set-point tracking experiment is illustrated in Figure 3A. The solid step-wise line represents the set point and the solid and dashed continuous lines represent the measured particle density with and without gain-scheduling, respectively. It can be seen that the density is well controlled around each set point despite the large range over which the set points are varied. Noise is present in the measured density due to the stochastic nature of Brownian motion. Furthermore, it can be seen that the gain-scheduled controller reaches the set point faster compared to a conventional PI controller with essentially no overshoot for both cases. The gain-scheduled controller particularly outperforms the conventional controller for higher set points, which is due to the diverging proportional gain compared to the base case as shown in Figure 3B.

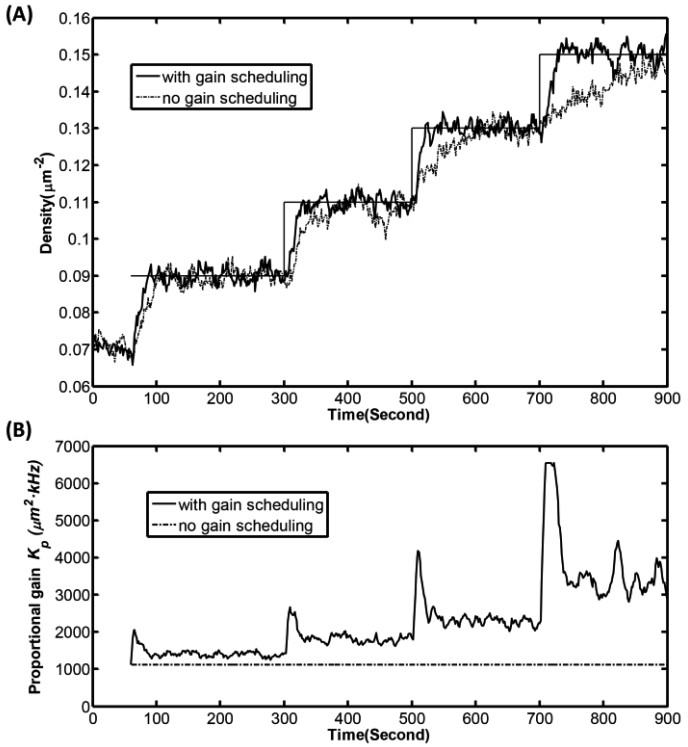


Figure 3. (A) Density and (B) Proportional gain as function of time during a set-point tracking experiment for cases with and without gain scheduling.

4. Conclusions and outlook

A gain-scheduled feedback controller for directed self-assembly has been developed and experimentally tested in this work. An empirical model is used to estimate the process gain online for different operating conditions such that a local particle density can be controlled effectively over a broad range of set points with electric field frequency as the manipulated variable, despite the strong nonlinear behaviour of such system. A conventional controller without gain-scheduling shows a more sloppy behaviour when the set point moves away from the base-case conditions. Future work may involve the extension of the feedback controller for particle density with controllers for particle alignment such that complicated non-periodic structures can be self-assembled, possibly in the framework of a multi-resolution approach (Solis et al. 2010). A more robust model-based control system such as MPC is of interest for future work, which may need a dynamic model rather than the current model. Finally, specific applications such as self-assembly of biological cells are of interest to be investigated.

Acknowledgement

This work was financially supported by the Research Grant Council of Hong Kong under Grant No. 16214617.

References

P. Apkarian, P. Gahinet, G. Becker, 1995, Self-scheduled H∞ control of linear parameter-varying systems: a design example, Automatica, 31, 9, 1251-1261

Y. Gao, R. Lakerveld, 2017, An automated two-step control strategy to self-assembly colloidal particles into non-periodic structures with single-particle resolution, Submitted

J. Juárez, B. Bevan, 2012, Feedback Controlled Colloidal Self-Assembly, Advanced Functional Materials, 22, 8, 3833-3839

J. Juárez, P. Mathai, J. Liddle, B. Bevan, 2012, Multiple electrokinetic actuators for feedback control of colloidal size, Lab on a Chip, 12, 20, 4063-4070R. Lakerveld, G. Stephanopoulos, P. Barton, 2012, A master-equation approach to simulate kinetic traps during directed self-assembly, The Journal of chemical physics, 136, 18, 184109

J. Oh, R. Hart, J. Capurro, H.M. Noh, 2009, Comprehensive analysis of particle motion under non-uniform AC electric fields in a microchannel, Lab on a Chip, 9, 1, 62-78

J. Paulson, A. Mesbah, X. Zhu, M. Molaro, R. Braatz, 2015, Control of self-assembly in micro- and nano-scale systems, Journal of Process Control, 27, 38-39

E. Solis, P. Barton, G. Stephanopoulos, 2010, Controlled Formation of Nanostructures with Desired Geometries. 2. Robust dynamic paths, Industrial&Engineering Chemistry Research, 49, 17, 7746-7757

J. Shamma, M. Athans, 1992, Gain scheduling: Potential hazards and possible remedies, IEEE Control Systems, 12, 3, 101-107

X. Tang, B. Rupp, Y. Yang, T. Edwards, M. Grover, M. Bevan, 2016, Optimial Feedback Controlled Assembly of Perfect Crystals, ACS nano, 10, 7, 6791-6798

Z. Ulissi, M. Strano, R. Braatz, 2013, Control of nano and microchemical systems, Computers and Chemical Engineering, 51, 149-156

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50405-5

Design of Doped Perovskite Oxygen Carriers Using Mathematical Optimization

Christopher L. Hanselman[a], De Nyago Tafen[bc], Dominic R. Alfonso[b], Jonathan W. Lekse[b], Christopher Matranga[b], David C. Miller[b], Chrysanthos E. Gounaris[a*]

[a]*Dept. of Chemical Enginering, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA*
[b]*National Energy Technology Laboratory, 626 Cochrans Mill Road, Pittsburgh, PA 15236, USA*
[c]*AECOM, P.O. Box 1959, Albany, OR 97321 USA*
gounaris@cmu.edu

Abstract

Mathematical optimization can readily support the design of nanostructured materials in areas that are typically explored via design of experiments. The combinatorial nature of possible configurations of nanomaterials leads to a high-dimensional space, which can be efficiently explored by algorithms traditionally developed in process systems engineering. In this work, we demonstrate a mathematical optimization based framework for the design of doped perovskite oxygen carriers. Specifically, we developed an approach for incorporating evaluations from density functional theory into simplified structure-function relationships that can be embedded directly into mathematical models. The material design question was formulated as a mathematical optimization model and solved to identify patterns of dopant that are expected to yield a highly reducible material. The results from the dopant placement optimization model can inform the theoretical bounds of material performance as well as identify structures that can serve as targets for experimental synthesis.

Keywords: mixed-integer linear programming, nanotechnology, oxygen carriers

1. Introduction

Ongoing developments in the synthesis of nanostructured materials are enabling the fabrication of increasingly complex materials. This results in combinatorially-complex design spaces, which in turn require the use process systems engineering techniques to efficiently explore and find optimal structures. We have previously demonstrated a materials design framework that uses mathematical optimization to design nanostructured materials bottom-up from the building blocks of matter (Hanselman and Gounaris, 2016). This optimization framework utilizes a generic modeling approach that can be applied to other material systems to identify novel, high-performance materials. In this work, we adapt the aforementioned materials design framework for the design of doped perovskites. Previous work by Lekse et al. (2014) shows that doped perovskites can achieve rapid and reversible oxygen release, making them promising materials for oxygen carriers for chemical looping. In particular, we consider a $BaFeO_3$ perovskite doped with indium that exhibits tunable oxygen release with respect to dopant content. The addition of indium replaces iron in the lattice at specific sites ("B-sites") in the perovskite material and is expected to strengthen bonds to neighboring oxygen atoms, while weakening bonding to oxygen further away in the lattice. This tradeoff of bond

strengths gives rise to the design question of how to best place the available amount of dopant throughout the crystal lattice. Such an optimal doping pattern could then be the aim of a controlled synthesis effort based on growing the material layer-by-layer or via some other nanostructuring technique. In order to quantitatively study the material properties as a function of specific dopant placements, density functional theory (DFT) can be used to predict the reducibility of the doped perovskite lattice. The relevant quantity in this context is the "oxygen excess energy," defined as the energy of the complete lattice minus the energy of the lattice with a particular oxygen removed. Since Motohashi et al. (2010) showed the perovskite oxygen release process was generally not kinetically limited, we assume that the oxygen excess energy serves as a good indicator of the reducibility of the carrier. For our purposes, a carrier containing a large number of oxygen sites that are easily removed will correspond to a highly reducible material. While DFT is useful for predicting the reducibility of oxygen sites in the perovskite material, it is not tractable to employ directly in conjunction with a mathematical optimization process, where the evaluation of material performance will be frequently required. Instead, we seek to develop a simplified structure-function relationship that uses algebraic constraints to predict the oxygen excess energy as a function of dopant placement in the lattice. In this work, we have identified a set of relevant motifs to characterize the impact of dopant on the oxygen excess energy of a nearby oxygen site. This set of motifs was evaluated using DFT, and then a simplified structure-function relationship was identified using standard model identification tools. This simplified structure-function relationship was then embedded in an optimization model to design the dopant placement in a perovskite lattice.

2. Computational Methods

In order to enumerate relevant motifs, we first proposed a motif definition that was expected to explain the variation in oxygen reducibility with respect to nearby dopant placement around a central oxygen atom. Given the computational cost of DFT calculations, it was deemed important to not define a motif that included too many B-sites which would result in the enumeration of a large list of conformations. It was expected that only the closest neighbours contribute significantly to bonding in the lattice; therefore, we chose to define motifs according to the doping pattern of the 10 closest B-sites to an oxygen on the perovskite lattice. Indium-doped $BaFeO_3$ perovskites are experimentally observed to be stable in the cubic phase with between 10 % and 30 % of the B-sites doped, so we limited our enumeration to only include motifs near that range, with between 0 % and 50 % dopant locally around each oxygen site. More specifically, we identified 74 motifs that included all possible ways to place between 0 and 5 indium atoms in the total 10 B-sites of the motif in a rotationally-unique way.

We performed spin-polarized DFT GGA+U calculations using the Perdew-Burke-Enzerhoff functional and the rotationally invariant formulation as implemented in the VASP code (Kresse and Furthmüller, 1996). Solid solutions with different structural motifs were modelled using a 4x4x4 supercell model containing 320 atoms. This size was chosen to make the calculations tractable without compromising the accuracy of the description of the electronic structure. We found that truncating the plane wave basis set at 520 eV kinetic energy cut-off and using a 4x4x4 k-point mesh yielded converged total energies. For one structure, performing the DFT calculations on an Intel-based High Performance Computing (HPC) system with 128 cores typically took 200 cpu hours.

Given the set of possible motifs, we can propose several geometric descriptors of dopant placement to explain the variation between motifs. These descriptors can include counts of dopant at different distances away from the central oxygen, counts of pairs of dopant in various configurations, as well as imbalances in dopant across the central oxygen. Importantly, these values can all be encoded in an algebraic model using propositional logic with binary variables; therefore, it is possible to express the problem as a mixed-integer linear programming mathematical optimization model.

Together, the DFT evaluations and the set of geometric descriptors for dopant placement can be used to propose simplified structure-function relationships for oxygen excess energy. This process is readily supported by standard model identification tools such as ALAMO, which build simple and predictive regression models by maximizing an information criterion (Cozad et al., 2014). In this way, we were able to identify which geometric descriptors had a significant impact on the evaluation of oxygen excess energy.

This simplified structure-function relationship was later embedded in a mathematical optimization model, which was solved using CPLEX 12.7.1, primarily using default settings (IBM Corp., 2017). While effort was put towards tailoring the formulation of the model to improve its numerical performance and be in position to address larger problem instances, the main goal was to demonstrate the potential of expressing the material design problem in a suitable optimization model.

3. Results

The set of rotationally-unique conformations featuring between 0 and 5 indium atoms in the nearest 10 B-sites was enumerated exhaustively, resulting in 74 motifs. Some example motifs are presented in Figure 1.

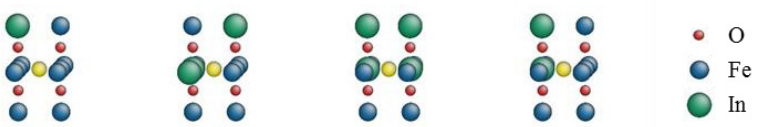


Figure 1: Example motifs and 4x4x4 supercells for DFT evaluation

Given these motifs, geometric descriptors can be defined based on rotationally-unique ways of counting dopant placed around the central oxygen atom. Some example geometric descriptors are presented in Figure 2, including counts of combinations of dopant up to two atoms. The procedure for identifying geometric descriptors can in principle be continued to identify counts of triplets or quadruplets of dopant atoms, but these additional descriptors were expected to have less physical significance on the evaluation of oxygen excess energy, and where hence omitted from consideration.

The DFT data were then paired with the tabulated geometric descriptors and presented to ALAMO for model identification. The model that minimized the Bayesian information criterion is presented below in Eq.(1). It included 6 of the 10 descriptors, exhibiting an average absolute deviation of 0.21 eV. The parity between DFT ("true") evaluations and the simplified structure-function relationship is given below in Figure 3.

$$E = 6.801A_0 - 0.0462B_0 - 0.121B_1 - 0.270B_2 - 0.143B_3 + 0.103B_4 + 4.535 \quad (1)$$

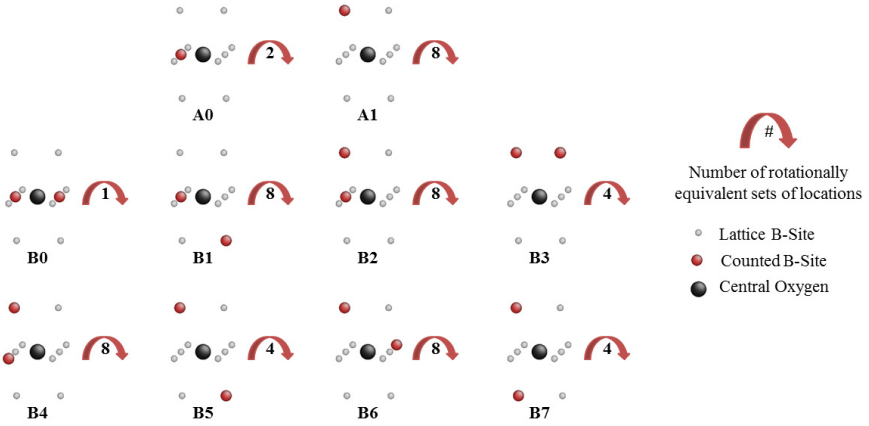


Figure 2: Examples of possible geometric descriptors for placement of dopant on B-sites

Once the oxygen excess energy could be predicted via an algebraic equation, it was possible to formulate a mathematical optimization model that identifies the dopant patterns that are expected to lead to highly reducible oxygen carriers, Eq. (2)-(11). More specifically, the model identifies the placement of dopant that maximizes the density (number) of "easily removable" oxygen sites, which are defined as those sites with a predicted oxygen excess energy that falls below a given threshold. The geometric descriptors for counts of individual dopant atoms (A_{im}) are encoded in Eq.(3). Similarly, the geometric descriptors for counts of pairs of dopant (B_{in}) are encoded in Eq.(4) utilizing variables, X_{ij}, representing the presence of a pair of dopant atoms. The X_{ij} variables are uniquely determined by Eq.(5)-(7). Eq.(8) encodes the simplified structure-function relationship for oxygen excess energy as a function of geometric descriptors. Finally, Eq.(9) indicates the presence of a site with oxygen excess energy below the target threshold, E^*, using indicator variable Z_i. The total number of such sites is maximized in the objective, Eq.(2). Additional constraints on the local concentration of dopant (i.e., per motif) and the global concentration of dopant (i.e., across the entire cell) are also included in Eq.(10) and Eq.(11), respectively. The formulation presented in Equations 2-11 results in a very challenging optimization problem. In order to improve tractability, the authors took advantage of symmetry-breaking constraints presented previously by Hanselman and Gounaris (2016). Furthermore, concentrations of 0 to 30 % dopant locally and 10 to 30 % dopant globally were enforced, representative of materials that can be successfully synthesized experimentally. Even with these constraints, the design space is still highly combinatorial. Models instantiated for supercells of size 3x3x3 can be solved in under 30 seconds, while models for supercells of size 4x4x4 typically require in excess of 24 hours. Further work will attempt to improve the numerical performance by investigating decomposition techniques, which may open the way to obtain solutions that exhibit periodicity over an even larger range. Some observations can be made from the results of this optimization model solved at various levels of target oxygen excess energy, E^*, illustrated in Figure 4. Since there is only one motif with oxygen excess energy below 4.4 eV that satisfies the local dopant concentration limits, the optimal dopant pattern is a single, periodic row of dopant with approximately 15 % of the oxygen sites achieving the target. At a target of 4.5 eV, there are two such motifs, which causes additional packings to be optimal, though they achieve the same density of ideal sites. As the target is raised to

4.55 eV, the majority of oxygen sites become ideal, utilizing the same simple design with a periodic row of dopant.

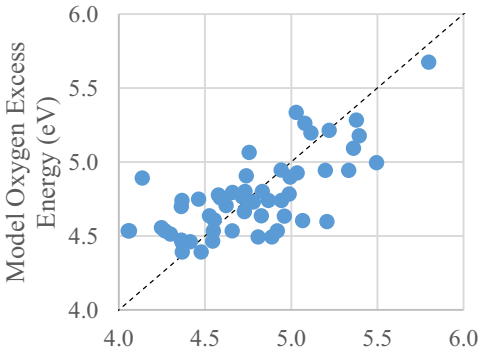


Figure 3: Parity between DFT evaluations and simplified structure-function model evaluations

$$\max_{\substack{Y_j, X_{jk}, \\ A_{im}, B_{in}, Z_i}} \quad \sum_{i \in I} Z_i \tag{2}$$

s.t.

$$A_{im} = \sum_{j \in J_{im}} Y_j \qquad \forall i \in I \quad \forall m \in M \tag{3}$$

$$B_{in} = \sum_{(j,k) \in K_{in}} X_{ij} \qquad \forall i \in I \quad \forall n \in N \tag{4}$$

$$X_{jk} \leq Y_j \qquad \forall (j,k) \in K_{in}\ \forall n \in N\ \forall i \in I \tag{5}$$

$$X_{jk} \leq Y_k \qquad \forall (j,k) \in K_{in}\ \forall n \in N\ \forall i \in I \tag{6}$$

$$X_{jk} \geq Y_j + Y_k - 1 \qquad \forall (j,k) \in K_{in}\ \forall n \in N\ \forall i \in I \tag{7}$$

$$E_i = \sum_{m \in M} \alpha_m A_{im} + \sum_{n \in N} \beta_n B_{in} \qquad \forall i \in I \tag{8}$$

$$E_i - E^* \leq M(1 - Z_i) \qquad \forall i \in I \tag{9}$$

$$L^{\mathrm{LB}} \leq \sum_{j \in \mathrm{Neigh}_i} Y_j \leq L^{\mathrm{UB}} \qquad \forall i \in I \tag{10}$$

$$G^{\mathrm{LB}} \leq \sum_{j \in J} Y_j \leq G^{\mathrm{UB}} \tag{11}$$

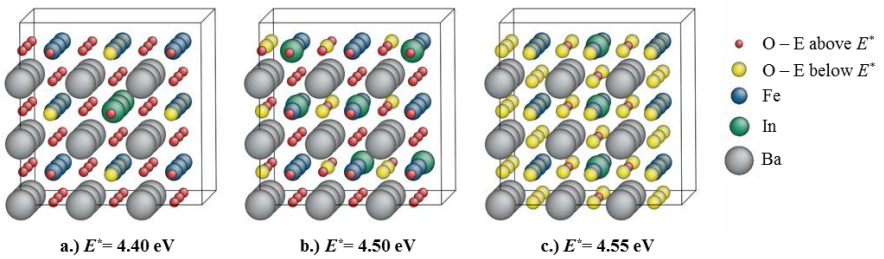


Figure 4: Example results on a 3x3x3 supercell, at various levels of E^*

4. Conclusions

In this work, we demonstrated the application of mathematical optimization to answer an important materials design question, namely how to identify the patterns of dopant that maximize the potential for oxygen release from a perovskite oxygen carrier. This involved using DFT calculations to elucidate the form of a simplified structure-function relationship that could predict perovskite oxygen excess energy in a way that was amenable to be incorporated in a formal optimization model. Optimal patterns for dopant placement were thus identified as a function of target oxygen excess energy, and these patterns can now serve as targets for experimental synthesis. Furthermore, this overall approach built upon a previously developed materials design framework by demonstrating the necessary steps to set up the optimization model from scratch, without a pre-existing structure-function relationship. Importantly, our work illustrated the benefit nanostructuring can have on the performance of oxygen carrier materials.

Acknowledgements and Disclaimer

This work was performed with support from the DOE Office of Fossil Energy via the Institute for the Design of Advanced Energy Systems, and in support of the National Energy Technology Laboratory's research under RES contract DE-FE0004000. This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

References

C. L. Hanselman, C. E. Gounaris, 2016, A mathematical optimization framework for the design of nanopatterned surfaces, AIChE Journal 62, 9, 3250-3263.

J. W. Lekse, S. Natesakhawat, D. R. Alfonso, C. Matranga, 2014, An experimental and computational investigation of the oxygen storage properties of $BaLnFe_2O_{5+d}$ and $BaLnCo_2O_{5+d}$ (Ln = La, Y) perovskites, Journal of Materials Chemistry A, 2, 7, 2397–2404.

T. Motohashi, T. Ueda, Y. Masubuchi, M. Takiguchi, T. Setoyama, K. Oshima, S. Kikkawa, 2010, Remarkable oxygen intake/release capability of $BaYMn_2O_{5+\delta}$: Applications to oxygen storage technologies, Chemistry of Materials, 22, 10, 3192-3196.

G. Kresse, J. Furthmüller, 1996, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, Computational Materials Science, 6, 1, 15-50.

A. Cozad, N. V. Sahinidis, D. C. Miller, 2014, Learning surrogate models for simulation-based optimization, AIChE Journal, 60, 6, 2211–2227.

IBM Corp., 2017, IBM ILOG CPLEX Optimization Studio 12.7.1

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50406-7

Process-Structure-Property Relationships for Design of Polymer Organic Electronics Manufacturing

Michael McBride[a], Carlex Morales[a], Elsa Reichmanis[a,b,c], Martha A. Grover[a,*],

[a]*School of Chemical & Biomolecular Engineering, Georgia Institute of Technology, 311 Ferst Dr, Atlanta, GA, 30332*
[b]*School of Chemistry & Biochemistry, Georgia Institute of Technology, 901 Atlantic Drive, Atlanta, GA, 30332*
[c]*School of Materials Science and Engineering, Georgia Institute of Technology, 771 Ferst Dr NW, Atlanta, GA, 30332*
[*]*martha.grover@chbe.gatech.edu*

Abstract

Conjugated semiconducting polymers present a new paradigm in electronics manufacturing. Examination of structure-property relationships in poly(3-hexylthiophene) (P3HT), a workhorse semiconducting polymer, has indicated that highly ordered crystals with long range interconnectivity exhibit the highest charge carrier mobility. Here we demonstrated the utility of using structural characterization to probe for tie chains, chains that tie together crystalline domains. P3HT nucleated via UV-irradiation was mixed with non-nucleated P3HT in a continuous flow system in which the flow rate and recollection tube length were varied. The setting with the lowest flow rate and shear rate (LL:LF) displayed the highest charge carrier mobility. This was associated with a trade-off between the quantity of fibril bundles formed and the interconnectivity. Comparable batch studies with solution aging exhibited higher mobilities than the flow system. The lack of correlations between the orientational order parameter and fiber packing density to charge mobility demonstrated that the structural design space is multidimensional and requires numerous structural metrics.

Keywords: poly(3-hexylthiophene), OFET, solution processing, structure quantification

1. Introduction

The solution processability of organic, conjugated polymeric semiconductors promises a new paradigm in electronic device fabrication. The ability to utilize roll-to-roll processing or other existing printing technologies to produce large area, flexible, and stretchable devices has the potential to drastically reduce manufacturing costs. While organic electronics are not expected to replace high-end, high-performance devices currently dominated by silicon, they are expected to be major competitors in markets that are less demanding computationally but require flexibility and large area coverage. Examples include organic field-effect transistors (OFETs), organic photovoltaics (OPVs), organic light emitting diodes (OLEDs), and biosensors (Facchetti, 2011). The realization of organic semiconducting polymer-based devices is dependent upon maintaining reproducible and optimized electronic properties. A plethora of research has been conducted to relate the processing conditions to charge carrier mobility, a key electronic metric describing how quickly electrons (or holes) can flow through a

material given an applied voltage. With hundreds of process-property data points in the literature each with a unique microstructure, a concerted effort to quantify and identify key process-structure-property relationships is beginning to emerge (Persson et al., 2016a). However, these relationships span multiple length scales, are multidimensional, highly nonlinear, and subject to a high degree of uncertainty. A rigorous process systems approach to elucidate predictive relationships could vastly enhance the field of organic electronics. To this end, poly(3-hexylthiophene) (P3HT) has been utilized as a workhorse polymer to study the role of processing on microstructural order and electronic properties. Advancements in structural characterization techniques have suggested two main features are crucial to high-mobility OFET devices: i) highly ordered crystals on the micro- to mesoscales providing for electron delocalization, and ii) long-range interconnectivity between crystalline domains to provide a percolative network for charge transport. Exposure of P3HT solutions to low-dose UV-irradiation followed by subsequent solution aging has been demonstrated to be a controllable technique to form highly crystalline nanofibers (Chu et al., 2016). However, methods specifically targeting the formation of "tie chains", defined as polymer chains that tie crystalline domains together, are generally lacking despite their perceived importance in macroscale charge transport. In this work, we present a case study to demonstrate the complexity of developing process-structure-property relationships in P3HT-based OFETs. Experimentally, the objective is to utilize structural characterization to probe for tie chains through the design of a continuous flow processing technique. It was found that batch processing of P3HT using UV-irradiation can provide higher charge carrier mobilities compared to the continuous flow process but requires two days of solution aging. Furthermore, the mobility of P3HT can be manipulated in a continuous flow system by controlling the flow rate and tube recollection length. Analysis of the UV-vis absorption spectra in solution indicated that an optimal aggregate percent is necessary for charge transport above 0.1 cm^2/V-s. Finally, analysis of the surface morphology indicates that metrics such as orientational order or fiber length density do not globally predict charge mobility. Instead, the structural design space is multidimensional and cannot be understood using a simple structural measurement yielding a single structural metric.

2. Experimental Methods

P3HT with molecular weight of 95 kDa (PDI 2.2, regioregularity 97 %) was purchased from Rieke Metals. Chloroform was purchased from Sigma-Aldrich. Tubes with an inner diameter of 0.762 mm were purchased from Cole-Parmer Instrument Company. In the continuous flow setup, two tubes of length 3.35 m each were connected to two syringe pumps containing P3HT dissolved in chloroform at a concentration of 5 mg/ml. One of these tubes was exposed to UV light over a residence time of 15 seconds to form nucleated P3HT; the other tube was not exposed to UV light. The streams from these two tubes were then combined in a static mixing tee and flowed through a recollection tube to allow the nucleated and non-nucleated P3HT chains to form an interconnected network. A two-level, two-factor experimental design scheme was used to investigate the impact of flow rate and recollection tube length on P3HT self-assembly into ordered, interconnected fibrillar domains. The high and low levels for the flow rates were 0.025 ml/s and 0.0125 ml/s while the recollection tube length was either 1.5 m or 4.5 m. These settings are summarized in Table 1. A batch processed sample comprising UV-irradiated P3HT mixed with a P3HT solution that had not been exposed to UV

served as the control. This sample was characterized on Day 0 and after 2 days of solution aging.

Table 1. Factor names with corresponding recollection length and flow rate settings

Sample Name	Recollection Tube Length (m)	Flow Rate (ml/s)
LL:HF	1.5	0.0250
LL:LF	1.5	0.0125
HL:HF	4.5	0.0250
HL:LF	4.5	0.0125

*Sample names refer to factor settings in format of recollection length:flow rate.

Electrical characterization of the polymer semiconductors was conducted using a bottom-gate, bottom-contact OFET. Heavily n-doped silicon wafer substrates acted as the gate electrode, and thermally grown SiO_2 (300 nm) served as the dielectric layer (capacitance per unit area: 1.15×10^{-18} cm^{-2}). Source and drain electrodes, 30 nm of Cr as an adhesion layer and 50 nm of Au, were deposited via standard lift-off photolithography techniques. The electrodes had a channel width of 50 μm and length of 2000 μm. The semiconducting films were then deposited using a blade coating apparatus. Drops of solution (4 μL) were placed on the substrate, and blade speed and height were 2 mm/s and 10 μm, respectively. Additional experimental details can be found in previous work (Chu et al., 2016). UV-vis solution spectra were recorded on an Agilent HP 8510 UV-vis spectrophotometer. The surface morphology of the thin films was characterized using a Bruker Dimension Icon AFM system in tapping mode.

3. Results and Discussion

3.1. Electrical Properties

Optimal devices are characterized by high charge carrier mobility with minimal run-to-run variance. The extracted charge carrier mobilities for both the batch experiments and the continuous flow designs are presented in Figure 1.

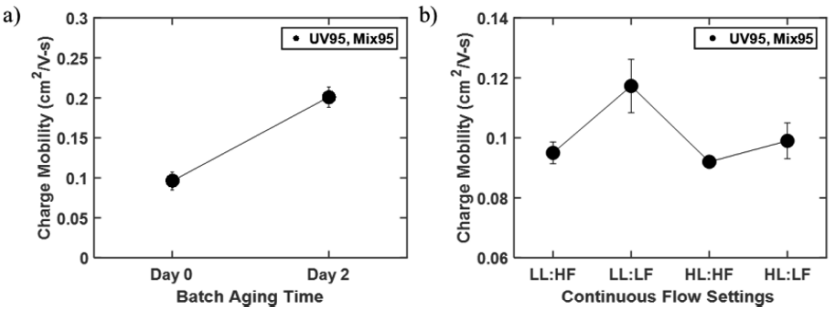


Figure 1. (a) Average charge carrier mobilities of OFETs for batch processing conditions. (b) Average charge carrier mobilities of OFETs for continuous flow processing conditions. Error bars denote standard errors over multiple replicates.

Inspection of Figure 1a shows the impact of batch processed solution aging on mobility, where mobility increases with aging over 2 days. The results are consistent with previous studies and demonstrate the nucleation and growth mechanism associated with UV-irradiation and aging in solution processing (Chu et al., 2016). In the continuous flow study, the low recollection length, low flow rate (LL:LF) setting displayed the highest charge carrier mobility with a value of 0.117 cm^2/V-s. However, this processing

condition also afforded the largest standard error. The mobilities for all other settings were statistically equivalent according to a t-test, with unequal variance for a 95 % confidence level and were only slightly lower than the LL:LF setting. These results suggest a trade-off between processing time and charge carrier mobility. While the batch processing method provided a thin film with the highest overall mobility, this came at the expense of two days of solution aging. In contrast, the Day 0 and continuous flow setup exhibited lower average mobilities but were obtained with no process delays. Further optimization of the continuous system may lead to increased mobility, enabling high performance devices without solution aging.

3.2. Solution UV-Vis Absorption

UV-vis absorbance spectroscopy was used to probe the degree of polymer conformational order and aggregate formation in solution. For all P3HT solutions, a broad absorption peak centred at ca. 455 nm accounts for the wide distribution of isotropic, amorphous chains that do not form aggregates. Shoulder peaks with vibronic bands at ca. 570 nm and 615 nm are associated with the formation of P3HT aggregates. The relative degree of polymer aggregation in solution can be estimated by deconvoluting the absorption spectrum into amorphous and aggregated contributions through the fitting of a Frank-Condon progression shown in Figure 2a (Panzer et al., 2015).

The aggregate fraction values for both the continuous flow and batch results are given in Figure 2b. Unlike the mobility results, there are significant differences in the solution characteristics prior to thin film deposition. Batch processing resulted in 10 % and 19 % solution aggregates for Day 0 and Day 2, respectively. The results for continuous flow processing were markedly different. The LL:LF setting provided an aggregate fraction of 22 %, which was statistically different from the other continuous flow settings according to a t-test with unequal variance for a 95% confidence level. All other settings formed larger aggregate domains with the percentage of aggregates exceeding 35 %. This can be attributed to both the effect of shear on the planarization of polymer chains and the time in which confinement effects force chains to interact and self-assemble.

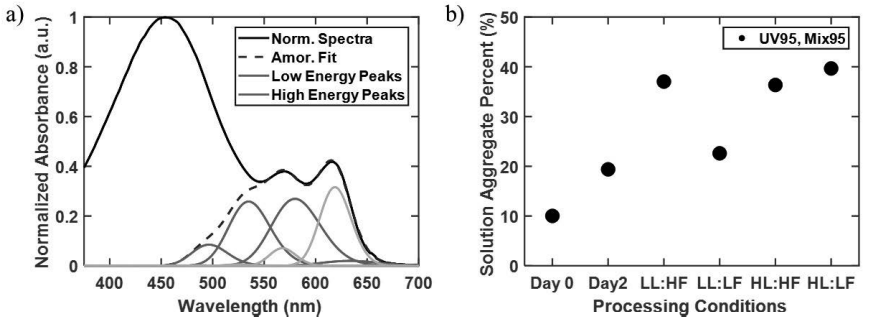


Figure 2. (a) Normalized solution UV-vis absorbance spectra of LL:LF sample deconvoluted into amorphous and aggregate contributions. (b) Solution aggregate percent for both batch and continuous flow processing experiments. The standard errors over multiple replicates were all less than 1.15 %.

Unlike the conventional wisdom where increased levels of crystallization are expected to afford higher mobility, the results here indicate that larger aggregate fractions do not correlate with high mobility. Rather, the data suggests an optimal range of solution aggregate fraction (15 % to 30 %) to obtain mobilities exceeding 0.1 cm^2/V-s. From a

mechanistic perspective, the existence of an optimal range can be readily justified. When the aggregate percent falls below the lower limit, the number of aggregates is insufficient to form the percolated network necessary for efficient charge transport; whereas when the degree of aggregation increases above the upper limit, the resultant films possess too many aggregates resulting in grain boundaries that limit transport pathways.

3.3. Surface Morphology

The nanofibrillar surface morphology of the semiconductor-air interface was imaged using AFM. Typically, AFM image analysis has been used to qualitatively relate the density, orientation, and dimensions of fibers to measured mobility values. However, recent advances in image processing techniques have demonstrated that crucial insights into the mechanisms of aggregate formation can be extracted through quantification of the surface morphology. Persson et al. (2016b) developed a general, open-source computational approach to skeletonize and vectorize individual fibers. This enabled calculation of the orientational order parameter, S_{2D}, the fiber length density (FLD), andapparent fiber dimensions. This process is demonstrated in Figure 3a. Examination of the extracted S_{2D} order parameter and FLD for both the batch and the continuous flow studies suggests different self-assembly mechanisms. The Day 0 batch AFM images appear featureless despite exhibiting a mobility just under 0.1 cm^2/V-s. This lack of nanofibers is vastly different from the continuous flow cases that exhibit similar mobilities (LL:HF, HL:LF, HL:HF), where the surface morphology exhibits a high degree of orientational order and dense fibrillar packing. Furthermore, neither S_{2D} nor FLD appear globally predictive of charge carrier mobility as devices with mobility

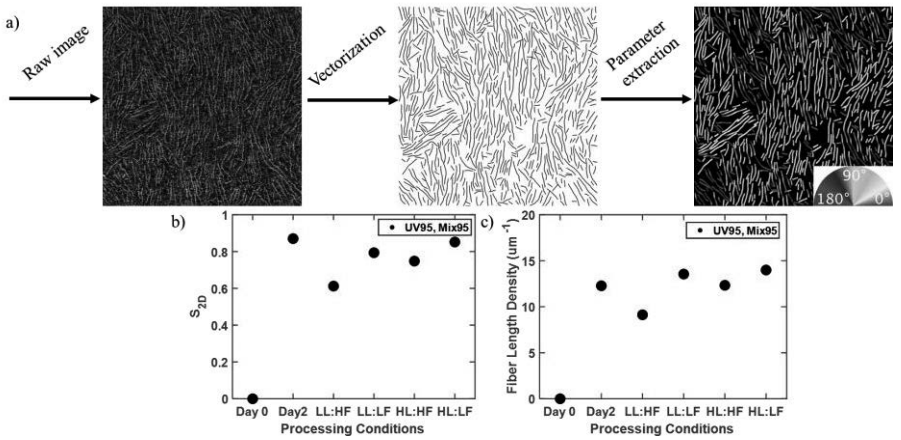


Figure 3. (a) Example of vectorization of P3HT fibers into quantifiable format using LL:LF sample. The image on the far right is referred to as an orientation map (b) S_{2D} orientational order parameter and (c) fiber length density values for batch and continuous flow processing.

both above and below 0.1 cm^2/V-s display similar values. Past studies have correlated these metrics with mobility, but this study demonstrates that the structural design space is multidimensional and a single metric such as S_{2D} cannot fully describe the structure in process-structure-property relations. However, the quantification of AFM facilitates the elucidation of mechanistically different self-assembly mechanisms and their relation to charge carrier mobility.

The orientation map of the LL:LF setting in Figure 3a shows local aligned bundles of nanofibers intermixed with other bundles showing a different directional preference. These fiber bundles are referred to as "shish-kebabs" and have been identified in other flow processing studies (Wang et al., 2015). These structures vastly differ from the well-aligned network structure that forms during batch self-assembly. Notably, the LL:LF setting has the lowest shear rate across the shortest distance, resulting in the self-assembly of fewer shish-kebab bundles. Mechanistically, this implies that more free chains are available to form tie chains during thin film deposition, providing the higher mobility for this flow condition. Understanding the driving force of shish-kebab formation in terms of applied solution shear, tube dimensions and residence time is essential to transfer processing lessons learnt from batch processing to continuous systems for industrial scale up.

4. Conclusions

The requirement of highly crystalline nanofibers with a high degree of interconnectivity to form a percolative network for charge transport has dominated the design of polymer semiconductor solution process techniques. However, limitations in structural characterization techniques have slowed the advancement of robust process-structure-property relationships that can enable the large-scale solution processing techniques. Here, we have proposed a continuous flow method that mixes nucleated and non-nucleated P3HT with the aim of improving charge transport mobility. It was found that while signatures of similar degrees of crystallinity can be found in the solution state, batch and continuous flow processing provide vastly different thin film morphologies. Previously discovered correlations between crystallinity, orientational order, and charge carrier mobility for batch studies do not as readily apply in a continuous flow system. As the organic electronics field progresses, it is imperative to provide as much structural characterization as possible so that multidimensional process-structure-property relationships enabling advancements in charge carrier mobility can be discovered.

References

A. Facchetti, 2011, π-conjugated polymers for organic electronics and photovoltaics cell applications, Chemistry of Materials, 23, 3, 733-758.

F. Panzer, M. Sommers, H. Bassler, M. Thelakkat, A. Kohler, 2015, Spectroscopic signature of two distinct H-aggregate species in poly(3-hexylthiophene), Macromolecules, 48, 5, 1543-1553.

G. Wang, N. Persson, P. Chu, N. Kleinhenz, B. Fu, M. Chang, N. Deb, Y. Mao, H. Wang, M. Grover, E. Reichmanis, 2015, Microfluidic crystal engineering of π-conjugated polymers, ACS Nano, 9, 8, 8220-8230.

N. Persson, M. McBride, M. Grover, E. Reichmanis, 2016a, Silicon Valley meets the ivory tower: Searchable data repositories for experimental nanomaterials research, Current Opinion in Solid State and Materials Science, 20, 16, 3380343.

N. Persson, M. McBride, M. Grover, E. Reichmanis, 2016b, Automated analysis of orientational order in images of fibrillar materials, Chemistry of Materials, 29, 1, 3-14.

P. Chu, N. Kleinhenz, N. Persson, M. McBride, J. Hernandez, B. Fu, G. Zhang, E. Reichmanis, 2016, Toward precision control of nanofiber orientation in conjugated polymer thin films: Impact on Charge Transport, Chemistry of Materials, 28, 24, 9099-9109.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50407-9

Developing Non-linear Rate Constant QSPR using Decision Trees and Multi-Gene Genetic Programming

Shounak Datta, Vikrant A. Dev, Mario R. Eden*

Department of Chemical Engineering, Auburn University, Auburn, AL 36849, USA
edenmar@auburn.edu

Abstract

Developing a QSPR model, which not only captures the influence of reactant structures but also the solvent effect on reaction rate, is of significance. Such QSPR models will serve as a prerequisite for the simultaneous computer-aided molecular design (CAMD) of reactants, products and solvents. They will also be useful in predicting the rate constant without entirely relying on experiments. To develop such a QSPR, recently, Datta et al. (2017) used the Diels-Alder reaction as a case study. Their model displayed great promise, but, there is scope for improvement in the model's predictive ability. In our work, we improve upon their model by introducing non-linearity. This is achieved using multi-gene genetic programming (MGGP). In our methodology, a combination of genetic algorithm (GA) and directed trees was used to develop a branched version of chromosomes, allowing additional possibilities in the generated models. In our work, prior to model development through MGGP, principal component analysis (PCA) was conducted. Lastly, models were evaluated based on metrics such as R^2, Q^2, and RMSE.

Keywords: Multi-gene genetic programming, hybrid algorithm, nonlinear regression, machine learning, stochastic optimization

1. Introduction

The search for QSPR models to predict influence of structures of reactants and solvents on reaction rate constant has long been a challenge. According to Roy et al. (2015), QSPR (Quantitative Structure Property Relationship) models are generally linear or non-linear mathematical relationships that correlate a particular property or activity of a chemical species with their structure. Such structures are generally represented numerically by descriptors, which can be determined experimentally or theoretically as per the definition. Early attempts to develop QSPR models for the prediction of rate constant of a reaction have been limited. Either the rate constant was studied as a function of structures of reactants while keeping the solvent structure constant or the solvent structures were varied but the reactants' structures were kept constant. With regards to the study of the influence of reactants' structures, Chaudry and Popelier (2003) developed a property model to predict the rate constant of hydrolysis of esters by utilizing quantum chemical descriptors. Estrada and Matamala (2007) have used generalized topological indices to predict the gas phase reaction rate constants of alkanes and cycloalkanes with OH, Cl and NO_3 radicals. Recently, Struebing et al. (2013) developed a methodology to design solvents by utilizing surrogate models and quantum chemical calculations. There is a need for QSPR models that capture the

combined influence of reactants' structures and solvent structures. Such models will serve two purposes: The first would be to quickly predict the rate constant without relying fully on experiments, while the second purpose will be the simultaneous design of reactants, products and solvents. With respect to QSPRs that capture reactant and solvent influence, Nandi et al. (2013) developed a quantitative structure-activation barrier relationship for the Diels-Alder reaction that utilizes quantum chemical descriptors. Their aim was to construct a relationship between the activation energy and the structures of the utilized reactants and solvent. However, their data set lacked solvent variety. Recently, Zhou et al. (2014) have studied a variety of solvents for the Diels-Alder reaction in their search for new solvent descriptors though they only used one set of reactants. Thus, we have combined the data sets utilized by Nandi et al. (2013) and Zhou et al. (2014) and created a set which offers more diversity in terms of the solvents utilized. We have utilized this more diverse data set to develop a rate constant model in terms of connectivity indices. It is worth noting that Nandi et al. (2013) relied on the data set utilized in the work of Tang et al. (2012). In our previous work, a hybrid GA-DT algorithm was designed to develop a linear model (Datta et al., 2017). In this work, we have proposed an efficient Multi-gene Genetic Programming (MGGP) algorithm using initialization by a modified DT algorithm for model development which utilizes the "divide and conquer" strategy in combination with Principal Component Analysis. This DT algorithm checks if the addition of branched gene improves model fitness. The MGGP algorithms hold promise as they possess the ability of developing models using a wide variety of nonlinear mathematical basis functions. Both internal and external validations were performed separately to determine model confidence. Additionally, model RMSE and R^2 values in case of both external and internal validation were calculated as they describe model fitness of data.

1.1. Diels-Alder Reaction

The Diels-Adler reaction is a well-studied organic chemical reaction involving a conjugate diene and an alkene, which is also termed as dienophile. Evans and Johnson (1999) have discussed this reaction in detail in their work. This reaction involves cycloaddition of two reactants in the presence of a solvent. Earlier, Rideout and Breslow (1980) discussed the hydrophobic acceleration of the Diels-Alder reaction in their work. Specifically, the influence of hydrophobic cavity in organic structures for acceleration of the reaction rate was discussed. In both of the aforementioned works, the impact of the solvent on the rate constant of the reaction has been observed. This feature of this reaction makes it an appealing choice for the aimed study.

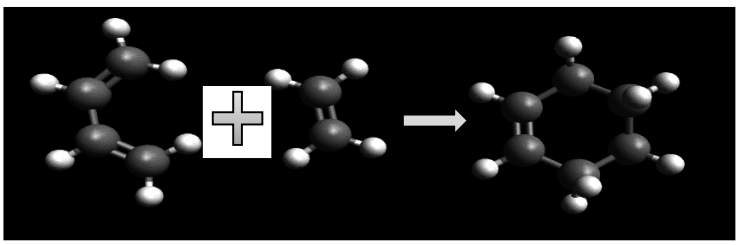


Figure 1: Example of a simple Diels-Alder Reaction drawn in Avogadro Software

2. Methodology

From the work of Tang et al. (2012) and Zhou et al. (2014, 2015), we generated a diverse data set of 72 reactions that consisted of 38 different dienophiles, 19 dienes and 10 solvents. All chemical species were designed using Avogadro software. The structures were optimized using MMFF94s, a built-in geometry optimization algorithm of Avogadro software, as suggested by Datta et al. (2015). The optimized geometries were saved as MOL files. These files were then used as input for Dragon 6 software to calculate descriptors. In our work, one sixth of the reactions were separated for external validation and all other reactions were used to train the model. Keeping reaction design simplicity in mind, only connectivity descriptors were used for model development. Further information about the versions of divide and conquer method and decision tree (DT) method used to reduce number of data points can be found in the work of Datta et al. (2017).

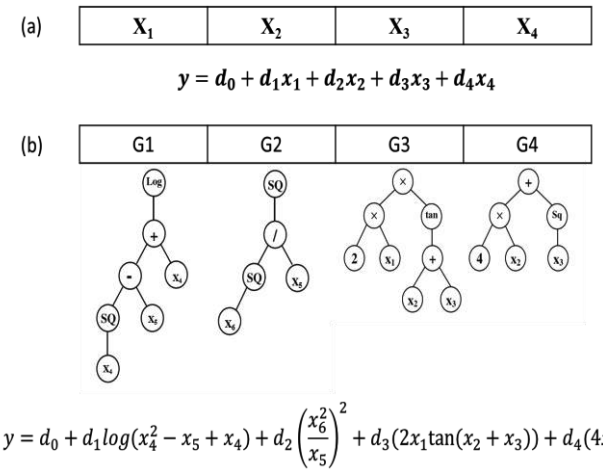
2.1. Multi-Gene Genetic Programming (MGGP)

Genetic Algorithm is widely employed in developing QSPR/QSAR models. As Whitley (1994) describes, such an evolutionary algorithm is efficient as a function optimizer and its applications are very diverse. Houck et al. (1996) has presented a basic idea of GA. In their work, they have proposed these basic steps that can be used as a guideline for designing a GA process for function optimization problems.

- Supply an initial population P_0 of N individuals with respective function values and constraints, if any.
- $i = 1$
- $P_i^{'}$ = selection_function(P_{i-1})
- P_i = reproduction_function($P_i^{'}$)
- evaluate(P_i)
- $i = i+1$
- Repeat 3rd step until termination condition met
- Print best solution achieved

In a feature selection approach using Genetic Algorithm, the chromosomes of the population generally contain position of variables (descriptors) that can be used to develop a property prediction model. The difference in MGGP is, instead of such variable positions, the chromosomes contain genes developed using directed trees (Gandomi and Alami, 2012). In Figure 2, a small example of model development using both (a) GA and (b) MGGP is presented. The genes can be considered branched version of individuals. The figure also shows the development of the final model from the individual developed in MGGP. It is important to note that the initial node of gene developing an MGGP individual is always an arithmetic operator. For maintaining mathematical acceptance of the genes, descriptor nodes are always separated with an arithmetic operator node between them. The tree develops by placing descriptor nodes after operator nodes, and moderators (e.g. sin, e^x, log, root, power) in between as needed. Two operator nodes or two descriptor nodes cannot be connected by a directed edge. However, the moderator nodes can be connected by directed edges. When needed, two or more moderator nodes can be connected with each other. For the crossover operation, the scatter swap method was selected. The mutation operation, however, could choose any node in the genes to alter.

Symbolic regression is used to determine the values of d_0, d_1, d_2, d_3, and d_4. Here, the weight of each gene is calculated and evolutionary algorithms (EA) are used to tune the values (Riolo and Worzel, 2003). For execution of this operation, the GPTIPS 2.0 (Searson et al., 2010) toolbox of MATLAB was used due to its wide flexibility. However, the code was modified to present and improve models based on not only training set fitness but also test set fitness. A great advantage of MGGP over other non-linear model development algorithms is the flexibility of control over choosing numerous arithmetic combinations with high nonlinearity. This is highly useful in cases like this where high level of nonlinearity is a possibility (Gandomi and Alavi, 2011).



$$y = d_0 + d_1 log(x_4^2 - x_5 + x_4) + d_2\left(\frac{x_6^2}{x_5}\right)^2 + d_3(2x_1\tan(x_2 + x_3)) + d_4(4x_2 + x_3^2)$$

Figure 2: Individual development of (a) Genetic Algorithm, and (b) Multi-Gene Genetic Programing

3. Results

As GPTIPS 2.0 gives the opportunity to use various numbers and depths of the genes, the objective was to develop a better model than presented by Datta et al. (2017) using a nonlinear model that contained the minimum number of descriptors possible. Although only connectivity index descriptors were used in this effort, it is beneficial to develop a model with fewer descriptors to increase the interpretability of the model. The population size and number of iterations were kept constant at respectively 50 and 250, respectively. In different settings of number of genes, gene depths, and crossover possibilities iteration numbers, three models possessed better performance metrics than the model presented in the work of Datta et al. (2017). For convenience, they are identified as M1, M2, and M3. It should be noted that, all the selected models represent influence of all the chemical classes involved in the reactions. However, the numbers of descriptors used to develop the models are different. M1 required 13 descriptors, where as M2 and M3 required 10 and 18 descriptors respectively. Here, R^2 and Q^2 represent model fitness for training set and test set data respectively, and RMSE expresses overall root mean squared error of the model. Here, k is reaction rate constant. As presented in Table 1, M3 model shows the best fit compared to the other models. It can also be seen that, in case of M2 and M3, decreasing crossover possibility, and increasing number and depth of gene was able to improve fitness.

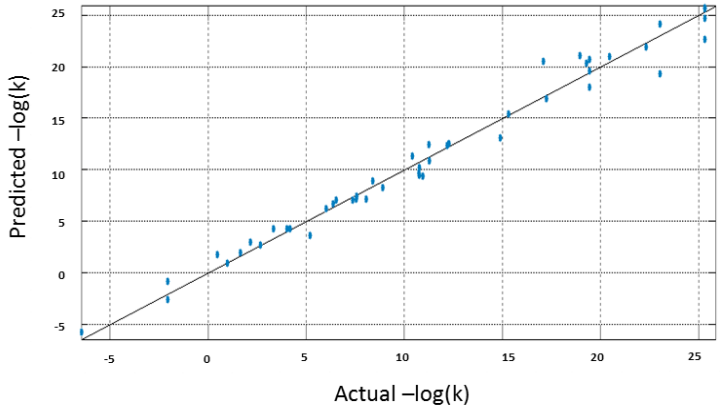


Figure 3: Actual vs. Predicted -log(k) values for model M3

Table 1: Properties of Models developed

Model Number	M1	M2	M3
R^2	0.9825	0.9011	0.9538
Q^2	0.8943	0.8213	0.9316
RMSE	0.5471	0.8410	0.3918
Gene Number	4	5	7
Gene Depth	3	4	5
Crossover	0.8	0.7	0.6
Mutation	0.02	0.02	0.02

4. Conclusions

In this project, the aim was to develop a better model than that previously proposed for the same property using QSPR analysis (Datta et al., 2017). From the work of Datta et al. (2017), it was clear that the hybrid GA-DT method was very efficient in developing linear models of this type. Dev et al. (2017) also concluded that the hybrid GA-DT method provided the best possible model, thus developing a nonlinear model was the only option left to develop a better property model. From the results it can be noticed that M1 and M3 models have similar metrics. However, in case higher level of accuracy is required, use of model M3 is advised. It can be noticed that Model M3 used five more descriptors than M1, but as connectivity index descriptors are very easy to calculate, no severe rise in computational expense is expected. Finally, it can be noticed that the models developed had better quality than the linear model presented for the same property by Datta et al. (2017).

References

U.A. Chaudry, P.L.A. Popelier, 2003, Ester Hydrolysis Rate Constant Prediction from Quantum Topological Molecular Similarity Descriptors, J. Chem. Phys. Chem. A, 107, 4578-4582.

S. Datta, R.H. Herring III, M.R. Eden, 2015, Data Mining and Regression Algorithms for the Development of a QSPR Model Relating Solvent Structure and Ibuprofen Crystal Morphology, Computer-Aided Chemical Engineering, 37, 1439-1444.

S. Datta, V.A. Dev, M.R. Eden, 2017, Hybrid genetic algorithm-decition tree approach for rate constant prediction using structures of reactants and solvents for Diels-Alder reaction, Computers and Chemical Engineering, 106, 690-698.

V.A. Dev, S. Datta, N.G. Chemmangattuvalappil, M.R. Eden, 2017, Comparison of Tree Based Ensemble Machine Learning Methods for Prediction of Rate Constant of Diels-Alder Reaction, Computer Aided Chemical Engineering, 40, 997-1002.

E. Estrada, A.R. Matamala, 2007, Generalized Topological Indices. Modeling Gas-Phase Rate Coefficients of Atmospheric Relevance, J. Chem. Inf. Model., 47, 794-804.

D.A. Evans and J.S. Johnson, Diels-Alder Reactions, 1999, Chapter 33.1, Comprehensive Asymmetric Catalysis I-III, Springer-Verlag, Berlin Heidelberg, 1178-1235.

A.H. Gandomi and A.H. Alavi, 2011, Multi-stage genetic programming: A new strategy to nonlinear system modeling, Information Sciences, 181,23, 5227-5239.

A.H. Gandomi and A.H. Alavi, 2012, A new multi-gene genetic programming approach to nonlinear system modeling. Part I: materials and structural engineering problems, Neural Computation and application, 21, 171-187

C.R Houck, J.A. Joines, M.G. Kay, 1996, A Genetic Algorithm for function Optimization: A MATLAB implementation (accessed 01.05.2017) https://www.researchgate.net/publication/2386612_A_Genetic_Algorithm_for_Function_Optimization_A_MATLAB_implementation.

K. Roy, S. KAr, R.N. Das 2015, A Primer on QSAR/QSPR Modeling, DOI 10.1007/978-3-319-17281-1_2.

S. Nandi, A. Monesi, V. Drgan, F. Merzel, M. Novic, Quantitative structure-activation barrier relationship modeling for Diels-Alder ligations utilizing quantum chemical structural descriptors, Chemistry Central Journal, 7, 171.

D.C. Rideout and R. Breslow, 1980, Hydrophobic acceleration of Diels-Alder reactions, J. Am. Chem. Soc., 102, 7817-7818.

R. Riolo and B. Worzel, 2003,Classification of gene expression data with genetic programming, Genetic Programming Theory and Practice, New Yor, USA.

D.P. Searson, D.E. Leahy, M.J. Willis, 2010, GPTIPS: an open source genetic programming toolbox for multigene symbolic regression, Proceedings of International Multiconference of Enginners and Computer Scientists 2010, Hong Kong.

H. Struebing, Z. Ganase, P.G. Karamertzanis, E. Siougkrou, P. Haycock, P.M. Piccione, A. Armstrong, A. Galindo, C.S. Adjiman, 2013, Computer-aided molecular design of solvents for accelerated reaction kinetics, Nature Chemistry, 5, 952–957

S. Tang, J. Shi, Q. Guo, 2012, Accurate prediction of rate constants of Diels–Alder reactions and application to design of Diels–Alder ligation, Org. Biomol. Chem., 10, 2673

D. Whitley, 1994, A Genetic Algorithm Tutorial, Statistics and Computing, 4,65-85

T. Zhou, K. Mcbride, X. Zhang, Z. Qi, K. Sundmacher, 2014, Integrated solvent and process design exemplified for a Diels–Alder reaction, AIChE Journal, 61, 1, 147-158

T. Zhou, Z. Lyu, Z. Qi, K. Sundmacher, 2015, Robust design of optimal solvents for chemical reactions-A combined experimental and computational strategy, Chemical Engineering Science, http://dx.doi.org/10.1016/j.ces.2015.07.010

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50408-0

Goal-directed Design of Dynamic Experiments for Cybernetic Models of Bioreactors

Martin F. Luna, Ernesto C. Martínez*

INGAR (CONICET-UTN), Avellaneda 3657, Santa Fe S3002 GJC, Argentina
ecmarti@santafe-conicet.gov.ar

Abstract

The distinctive feature of cybernetic models of bioreactors is their capacity to account for regulatory mechanisms in a cell metabolism by modeling the synthesis of enzymes and their activities. From a process engineering viewpoint, to guarantee its predictive capabilities regarding one or more process objective or goals (e.g. optimization, controllability, etc.), experimental data used to fit a cybernetic model parameters should be the most informative bearing in mind the adequacy of the resulting model to describe the specific objective of interest. To excite the most relevant metabolic modes in the cybernetic model, a dynamic experiment is optimally designed by accounting for the sensitivity of the chosen objective to operating conditions. The bioreactor feeding profile and sampling times are selected to maximize a global sensitivity index. As a case study, biomass production or fermentation to ethanol conversion in the fed-batch cultivation of *Saccharomyces Cerevisiae* are considered as alternative objectives.

Keywords: Bioreactor; Cybernetic Model; Design of experiments; Optimization.

1. Introduction

Mathematical models of bioreactors are gaining widespread acceptance in both academic and industrial practice since it is vividly clear that, in order to be competitive and to keep pace with the new developments in the biopharmaceutical industry, bioprocesses need to be designed, optimized, monitored and controlled using process engineering methods (van Daele, et al., 2017). Cell metabolism is well-known for its dynamic complexity in comparison with pure reaction networks. The microorganisms used as catalyst are living beings that shown different metabolic behaviors depending on the operating conditions. Thus, simple kinetic and transport phenomena are usually not enough to describe the system dynamics. Some insight into the regulatory system of the cell must be considered in order to have a useful model that can predict the response of the microorganism to changes in the abiotic environment. The supplied nutrients undergo different metabolic pathways within the cells, and these alternative uptake modes are regulated by the microorganism enzymes. In this regard, cybernetic models (Ramkrishna and Song, 2016) are capable of explicitly modeling the metabolic behavior by resorting to matching and proportional control laws for the production and activities of enzymes.

However, the complex nature of cell metabolism is translated into a complex formulation of the mathematical models. Usually, cybernetic models involve several differential equations with many parameters. Thus, experimental information from dynamic experiments is needed in order to parameterize them. But if a metabolic mode is not excited during an experimental run, it may be impossible to obtain meaningful values of the parameters related to it. This may hamper model prediction capabilities

which may lead the industrial process to suboptimal operation. Thus, a method to optimize the Design of Dynamic Experiments (DoDE) is needed in order to obtain the maximum information from each experimental run (Franceschini and Macchietto, 2008; Martinez et al., 2013). The design will depend on the way the information gain is measured. In this work, the sensitivity of the process objective will be evaluated as a scalar of the Global Sensitivity Matrix (GSM). Thus, the process objective chosen (e.g. biomass production, metabolite expression, etc.) will determine which metabolic modes are excited during the experiment so that all the available resources are used to account for them properly in the cybernetic model. Here, two possible objective functions are considered for a bioreactor involving glucose consumption by yeast in a fed-batch bioreactor. Both goal-directed optimal designs are compared and the adequacy of each is discussed. It is shown that explicitly accounting for a given objective in the DoDE is advantageous for data generation in fitting a cybernetic model that properly describes regulatory mechanisms among different metabolic pathways.

2. Cybernetic modeling

In the cybernetic approach for bioreactors, the regulatory system of the microorganism is modeled by describing the enzymes production rates and activities. In order to do so, it uses the matching and proportional laws to link the uptake rates to the different metabolic modes. Here, the term metabolic mode refers to the preferential uptake of a carbon source and the subsequent metabolic pathway that it undergoes within the cell. As a case study here, the cultivation of baker's yeast (*Saccharomyces Cerevisiae*) with a feed of glucose as a carbon source, there are three main metabolic modes: i) the uptake of glucose by a fermentative pathway (fermentation), ii) the uptake of glucose by an oxidative pathway (respiration) and iii) the uptake of ethanol (produced by fermentation) by an oxidative pathway. Each mode has different rates and yields, and the switch from one mode to another depends on the cybernetic laws (Ramkrishna and Song, 2016). In this work, the model used is presented in Eq. (1) through Eq. (8). The matching and proportional laws are described by Eq. (7) and Eq. (8).

$$\frac{dX}{dt} = \left(\sum_i r_i v_i - \frac{F}{V}\right) X \tag{1}$$

$$\frac{dGl}{dt} = (Gl_F - Gl)\frac{F}{V} + \left(\frac{r_1 v_1}{Y_1} - \frac{r_2 v_2}{Y_2}\right) X \tag{2}$$

$$\frac{dEt}{dt} = -Et\frac{F}{V} + \left(p\frac{r_1 v_1}{Y_1} - \frac{r_3 v_3}{Y_3}\right) X \tag{3}$$

$$\frac{dV}{dt} = F \tag{4}$$

$$r_i = e_i \mu_i \frac{S_i}{S_i + K_i}\frac{K_{EI}}{Et + K_{EI}} \tag{5}$$

$$\frac{de_i}{dt} = \alpha u_i \frac{S_i}{S_i + K_i} - \left(\sum_j r_j v_j - \beta\right) e_i + \alpha^* \tag{6}$$

$$v_i = \frac{fc_i\, r_i}{\max(fc_j r_j)} \tag{7}$$

$$u_i = \frac{fc_i\, r_i}{\sum_j fc_j\, r_j} \tag{8}$$

Here, X stands for biomass concentration, Gl stands for glucose concentration (subscript F indicates that it is the concentration in the feed) and Et stands for ethanol concentration. Eq. (5) through Eq. (8) are repeated for i=1, 2, 3, which correspond to the three modes: fermentation, respiration and ethanol oxidation. In Eq. (5) and Eq. (6) S_i is equal to Gl for i=1,2 and to Et for i=3. V stands for volume, F stands for the feeding profile, e stands for the enzymes concentration and v and u for the cybernetic variables. In Eq. (7) and Eq. (8) fc_i is the number of carbon atoms per molecule of substrate, which is 6 for i=1, 2 and 2 for i=3. The rest of the symbols shown correspond to the model parameters to be fitted by the method.

The bioreactor is operated in fed-batch mode with complete mixing and it is assumed that no diffusion limitations are present (thus oxygen balance is not taken into account). The feeding profile is parameterized as a sequence of steps. The cybernetic model is parameterized by regression of experimental data generated in one or more dynamic experiments. Distributions (histograms) of the model parameters are generated using bootstrapping and the least-square error method is used to fit them.

3. Design of Dynamic Experiments

The complexity of the proposed model and the number of parameters involved to optimize the information content J of each experiment in order to minimize experimental costs and time constraints. Based on an a priori chosen distribution model parameters, the DoDE optimization problem is posed here as follow:

$$\begin{aligned} &\max\ J = det(Q'.Q) \\ &\text{s.t: } \varphi_{LB} \le \varphi \le \varphi_{UB} \end{aligned} \tag{9}$$

In this work, the information content of a dynamic experiment is measured as the determinant of the matrix Q defined below. The design vector φ includes the feeding profile parameters and the sampling schedule. Each entry in the vector φ has lower and upper limits φ_{LB} and φ_{UB}, respectively. Each entry of the matrix Q corresponds to the global sensitivity index SI_{ij} of the process objective function with respect to the model parameters i at sampling time j. Accordingly, the Global Sensitivity Analysis (GSA) is defined as follows:

$$Q = \begin{pmatrix} SI_{11} & \dots & SI_{1j} \\ \vdots & \ddots & \vdots \\ SI_{i1} & \dots & SI_{ij} \end{pmatrix} \tag{10}$$

This definition of the information content J is key in the formulation of the mathematical program (9). Since the cybernetic model can be used for different process objectives (e.g. biomass or ethanol production), different experimental designs can be used in order to gather more relevant information about the influence of model parameters in better predicting mode switching and its influence on the process objective function of interest.

It is worth noting that the evaluation of the information content J is computationally expensive due to GSA. A Bayesian optimization (BO) method is used here to find the global solution to the mathematical program in Eq. (9). The details for the implementation of the BO algorithm used can be found elsewhere and references therein (Shahriari et al., 2016).

4. Results

The proposed method is used to design dynamic experiments to parameterize a cybernetic model that describes the operation of a bench-scale fed-batch bioreactor containing baker's yeast, with glucose as carbon source. The overall duration of the experiment is set to 9.5 h, including a batch phase of 2 h after which the feeding of glucose begins. The feeding profile is made up of 5 successive steps of fixed duration (1.5 h), whose inflow rates are the entries of the design vector φ. The concentration of glucose in the bioreactor feed is 100 g/L. The sampling schedule consists of six sampling times, of which five has to be designed and the remaining one corresponds to the final time (9.5 h). The concentrations of biomass, glucose and ethanol are determined at each sampling time. The initial volume of the bioreactor is 1 L, with an inoculum of yeast of 6 g/L and initial glucose concentration of glucose of 1.2 g/L (no ethanol is present at the beginning of each experiment).

The initial parameterization of the cybernetic model is fitted with data gathered from the experiments performed in the work of Luna and Martinez (2017). In the available experimental data, the ethanol concentrations at sampling times were not available. Furthermore, the process objective function was the production of biomass, thus the glucose oxidative pathway, which has the higher biomass yield, was of special concern. As a result, the distribution of parameters related to the production and uptake of ethanol has a significantly higher variance.

Two different process objective functions J_p are proposed here to compare the influence of these goals on the design of the experiments. First, the process objective function is chosen to be the production of biomass, measured as the mass of yeast in the reactor. Alternatively, the production of ethanol is selected.

4.1. Biomass production

Using the cybernetic model fitted with the aforementioned data, the DoDE is performed using the production of biomass as the process objective function:

$$J_p = X\,V \tag{11}$$

The feeding profile is presented in Fig. 1(a) along with optimal sampling times. In Fig. 1(b), the cybernetic variables v for the three main metabolic modes are shown. Time-varying profiles of each cybernetic variable is an indicator of the prevalence of each mode over the experiment (a value of 1 indicates that the enzyme of the mode has maximum activity). As can be seen, during the first 2 h when no glucose is fed, the initial amount of glucose is consumed by the fermentative pathway (producing ethanol). When the concentration of glucose decreases the importance of the glucose oxidative mode increases and even surpass the fermentation mode, but it is rapidly overtaken by the ethanol oxidative mode. This metabolic mode prevails until fed-batch operation begins and the cell favors the glucose fermentative mode again. However, the glucose

oxidation pathway is not at all absent and gains momentum as the glucose concentration decreases again (due to an increase in biomass). When the inlet flow rate drops after the fifth hour, the glucose oxidative mode becomes prevalent.

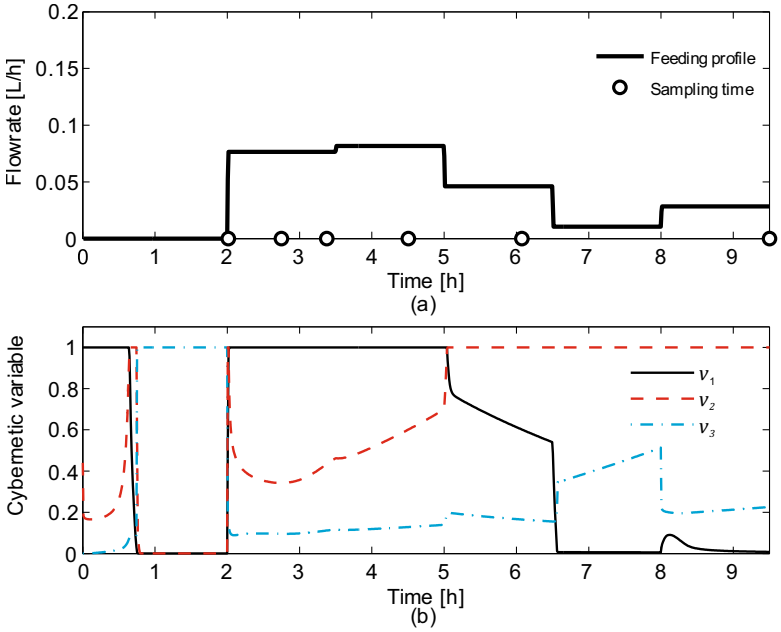


Figure 1. (a) Optimal feeding profile and sampling schedule for the case of biomass production. (b) Values of cybernetic variables for the optimal experimental design.

4.2. Ethanol production

A new experiment is designed using the same model parameters, but using the production of ethanol as the objective function:

$$J_p = Et\ V \tag{12}$$

The feeding profile and the sampling schedule are shown in Fig. 2(a), whereas the cybernetic variables *v* are depicted in Fig. 2(b). As can be seen, all metabolic modes are conveniently excited which generates highly informative data.

It is worth cross-comparing the values of *J* for both optimal designs. The value of *J* for the first design (the one obtained section 4.1), using the biomass production goal is 4.75 10^{-11}, whereas for the second optimal design (the one obtained in this section), using the same process objective the value for the information content *J* is 1.85 10^{-11}. The value for the information content *J* in the first and second DoDE using the ethanol production goal are 1.96 10^{-13} and 1.54 10^{-10}, respectively. The results obtained from cross comparison among optimal designs and information contents are rather expected: the optimal DoDE is dependent on the process objective the model is helping to optimize.

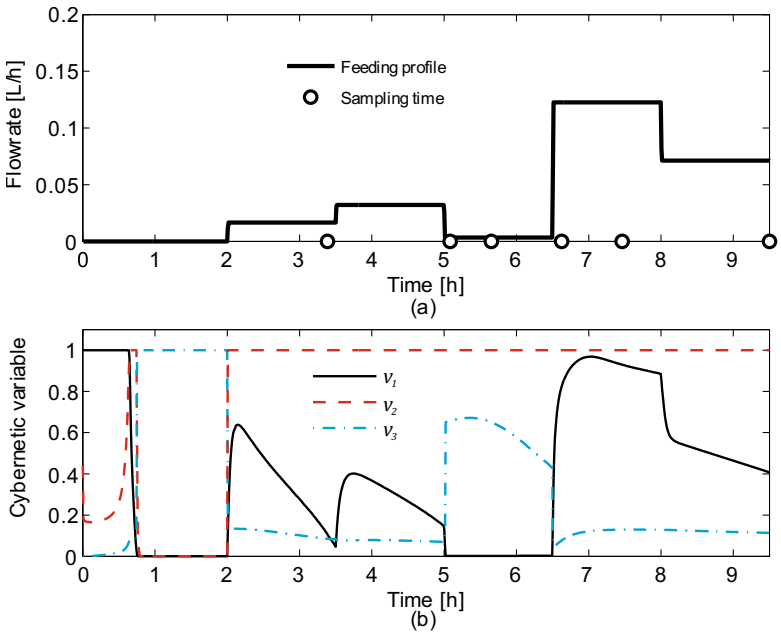


Figure 2. (a) Optimal feeding profile and sampling schedule for the case of ethanol production. (b) Values of cybernetic variables for the optimal experimental design.

5. Conclusions

In this work, a DoDE method for goal-directed parameterization of cybernetic models has been proposed. The importance of the process objective to be modeled is explicitly accounted for in the design of a dynamic experiment, since the same model may need different types of information to better describe the process objective that is being pursued. The method has proven to be useful in the maximization of the information content. However, the high computational burden of the GSA gives rise to the need of speeding up simulation-intensive algorithms used to deal with parameter distributions.

References

T. van Daele et al., 2017, Application of Iterative Robust Model-based Optimal Experimental Design for the Calibration of Biocatalytic Models, Biotechnology Progress, 33, 5, 1278–1293.
G. Franceschini, S. Macchietto, 2008, Model-based design of experiments for parameter precision: State of the art, Chemical Engineering Science, 63, 19, 4846-4872.
M.F. Luna, E.C. Martínez, 2017, Iterative modeling and optimization of biomass production using experimental feedback, Computers & Chemical Engineering, 104, 151-163.
E. C. Martínez, M.D. Cristaldi, R.J. Grau, 2013, Dynamic optimization of bioreactors using probabilistic tendency models and Bayesian active learning, Computers & Chemical Engineering, 49, 37-49.
D. Ramkrishna, H.S. Song, 2016, Analysis of Bioprocesses. Dynamic Modeling is a Must, Materials Today: Proceedings, 3, 10, 3587-3599.
B. Shahriari, K. Swersky, Z. Wang, R.P. Adams, N. de Freitas, 2016, Taking the human out of the loop: A review of Bayesian optimization, Proceedings of the IEEE, 104, 1, 148-175.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50409-2

Transition Phase Approach for Statistical Model in Multiphase Process

Hye Ji Lee[a], Dong Hwi Jeong[a], Jaehan Bae[a], Jong Min Lee[a*]

[a] *School of Chemical and Biological Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea*
jongmin@snu.ac.kr

Abstract

Online monitoring of batch reactor is essential to maintain the performance against batch-to-batch variations. A major challenge in using a statistical model for monitoring is to guarantee the data from different batches have similar dynamic characteristics at the same time points. By detecting singular points, uneven batch data are synchronized and the DPLS model of each singular episode is used for process monitoring. A new phase division method is proposed by introducing transition phase at a phase boundary to improve estimation performance around the transition region.

Keywords: Uneven batch synchronization, DTW, PLS, transition phase

1. Introduction

Batch reactors are used in a number of industries producing small quantities of high-valued materials such as cell cultivation, polymer synthesis and crystallization. To operate the batch reactor with success, the final quality performance has to maintain with minimum cost. Though operated with the same recipe, the batch process shows batch-to-batch variations in its specified trajectories. Therefore, online process monitoring is essential to achieve successful batch operation.

Several techniques have been developed to deal with real-time observation of process variables and detection of abnormal operations. The model-based method is the most traditional way to estimate the operation trajectory. It consists of the first principle equation based on mass balance and empirical equations. Though this method shows considerable prediction performance, it is used in limited conditions because it needs much time and cost due to the complexity. Data-driven methods can be an alternative as they need little knowledge of the process. The stored data is used to extract correlation between process variables and quality indicators.

For batch processes, multiway principal component analysis (MPCA) was developed (Wold *et al.*, 1987). In case both quality indicators and process variables are available, multiway partial least square (MPLS) was proposed to estimate batch performance (Nomikos and MacGregor, 1995). Since these models use samples from one operating point, they could deal with a static behaviour only. To analyze time dependent relations, dynamic PCA (Ku *et al.*, 1995) and dynamic PLS (Kaspar and Ray, 1993) were developed. In this study, dynamic version of the MPLS method, called batch dynamic PLS (BDPLS), is used to construct the dynamic model.

Batch synchronization is an important pre-process because batch operations with a same recipe could have various batch lengths due to the inherent nonlinearity and operational constraints. Given the assumption that the data point at the same time point will be similar

in each batch, several approaches are available to synchronize the sampled data for different batches. For example, indicator variables representing the batch progress for the reactions with well-known kinetics were proposed (Nomikos and MacGregor, 1995). After the reference batch is selected, the other batches are adjusted so that the indicator variables would have same values at the same time point.

Dynamic time warping (DTW) minimizes the distance between the trajectories of different time series (Sakoe and Chiba, 1978), which can be used without any prior knowledge of the process. This study employed a phase-based DTW technique (Doan and Srinivasan, 2008) to synchronize the trajectories of different batches by adjusting each phase segment to be a same length.

Singular point (SP) refers to the point that has more information than others (Srinivasan and Qian, 2005), for example, extrema. SPs can be used to separate the phase as the dynamic behavior changes at that point. Though the SP based phase division method has acceptable performance through the whole trajectory, it shows insufficient accuracy around the SP, when the phase changes. In this study, a new method for singular point based phase division is proposed by separating a transition phase from normal ones defined by the SP, so that the performance of the multiphase statistical model is improved.

2. Uneven batch synchronization

2.1. Dynamic Time Warping (DTW)

DTW synchronizes two time series with different lengths in the way to minimize the distance between them (Doan and Srinivasan, 2008). Asymmetric DTW used in this paper adjusts one time series to the same length of the other's, reference, without a change in the length of the reference time series. The series of mappings from $T(M \times J)$ to $R(K \times J)$ with J number of variables and lengths M and K, respectively, is denoted by F:

$$\mathrm{F} = [\mathrm{c}(1), c(2), \ldots, c(P)] \tag{1}$$

$$\mathrm{c(p)} = [i(p), j(p)], \quad 1 \le \mathrm{p} \le \mathrm{K} \tag{2}$$

DTW maps T onto R by optimizing the sequence F of P points by minimizing the distance between two series, $D(T, R)$:

$$\mathrm{D(T,R)} = \frac{\sum_{p=1}^{P} d(c(p)) \cdot w(p)}{\sum_{p=1}^{P} w(p)} \tag{3}$$

with the local distance between specified points in T and R denoted as $d(i, j)$. The weighting coefficients denoted by $w(p)$ are specified 1 in this work.

$$\mathrm{d(i,j)} = \|R(i) - T(j)\| \tag{4}$$

Then the optimal warping sequence is given by

$$F^* = \underset{F}{\operatorname{argmin}}[D(T, R)] \tag{5}$$

2.2. Singular Point (SP)

There are special points called singular point (SP) containing the majority of features of the signal because the information density in a signal is not uniformly distributed. SPs consist of extrema, sharp change or discontinuity and trend change points. The segment

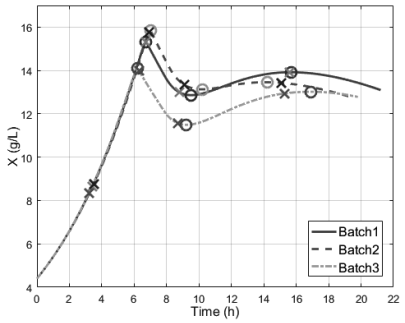


Figure 1. Unsynchronized batch profiles of bioreactor. Each marker indicates the location of SPs in the batch.

Table 1. Position of SP in each batch

Batch 1		Batch 2		Batch 3	
Position[h]	SP Type	Position[h]	SP Type	Position[h]	SP Type
3.4	SP_T	3.5	SP_T	3.2	SP_T
6.7	SP_T	6.9	SP_T	6.2	SP_E
6.7	SP_E	7	SP_E	6.2	SP_S
6.7	SP_S	7	SP_S	6.3	SP_T
8.8	SP_T	9.1	SP_T	8.7	SP_T
9.5	SP_E	10.2	SP_E	9.2	SP_E
15.6	SP_T	14.2	SP_E	15.3	SP_T
15.7	SP_E	15.1	SP_T	16.9	SP_E

separated by adjacent SPs is called a singular episode (Srinivasan and Qian, 2005), treated as a "phase." The phase refers to a distinguishable segment of a batch trajectory (Doan and Srinivasan, 2008).

Figure 1 shows various SP locations among several batches, and Table 1 specifies the types and positions of SPs. SP_T, $\mathrm{SP_E}$, and SP_S denote the type of SPs: trend change SP, extrema and sharp change SP, respectively. As the batch progresses are different, a statistical model can be inaccurate without synchronization because it assumes the data with the same time point in each batch shows similar behaviour. Though normal DTW using whole series at a time could improve the performance as well, phase-based DTW which synchronizes each phase defined by SPs shows higher accuracy (Doan and Srinivasan, 2008).

3. Multiphase PLS

3.1. Partial Least Square (PLS)

Partial least squares (PLS) is used to predict unmeasured properties based on measured process variables (Nomikos and MacGregor, 1995). The 3-way batch data is unfolded and separated into two blocks, independent block X and dependent block Y which both are two dimensional vectors. They are decomposed into the product of the score vector t, and the loading vector p plus residuals E and F, respectively.

$$\mathrm{X} = \sum_{k=1}^{r} t_k p_k^T + E = TP^T + E \tag{6}$$

$$\mathrm{Y} = \sum_{k=1}^{r} t_k q_k^T + F = TQ^T + F \tag{7}$$

where r denotes the number of principal components used to construct a PLS model.

3.2. Dynamic PLS

The data with autocorrelation can be represented by containing previous observations in an augmented form. Dynamic PLS model (Kaspar and Ray, 1993) uses this approach by including previous d observations at each time step:

$$X_d = \begin{bmatrix} x^T(k-d) & x^T(k-d+1) & \cdots & x^T(k) \\ x^T(k-d+1) & x^T(k-d+2) & \cdots & x^T(k+1) \\ \vdots & \vdots & \vdots & \vdots \\ x^T(k+l-d-1) & x^T(k+l-d) & \cdots & x^T(k+l-1) \end{bmatrix} \tag{8}$$

$$\mathrm{x(k)} = \left[x_{1,k}, x_{2,k}, \cdots, x_{J,k}\right]^T \tag{9}$$

where $x(k)$ is the J-dimensional vectors containing the observations of J process variables at time k.

3.3. Multiphase PLS with transition phase

For each phase defined by SPs, DPLS model is constructed assuming a similar behaviour within the phase. The transition region near the SP, however, shows nonlinearity (Ge *et al.*, 2012) having different behaviour from that of the previous phase. In this study, the SP based phase division method is modified by separating the phase transition region as a transition phase. For each SP, the previous d observations and subsequent d observations are included in the transition phase. In case of the SPs with a close position, observations from the point d-step before the former SP to the point d-step after the later SP are considered as a transition phase. The points in a transition phase could be estimated by combining two models of neighboring normal phases. The weighting factors of models to be combined are proportional to the inverse of the distance from each point.

4. Results

4.1. Batch synchronization and phase division

For the 76 batch data, SP based synchronization was carried out so that the positions of SPs become identical throughout the batches. Asymmetric DTW was used for each singular episode and the synchronization results for the batches in Figure 1 are shown in Figure 2. In this study, three methods including normal DTW, SP based phase division and modified SP based phase division, applying transition phase concept, are compared. In normal DTW case, which is a control group, asymmetric DTW is applied to the whole signal all at once without any singular episode considered. After the synchronization, the data is separated into phases defined by SPs to construct DPLS model for each phase. On the other hand, SPs are used to synchronize the data and construct the model in SP based

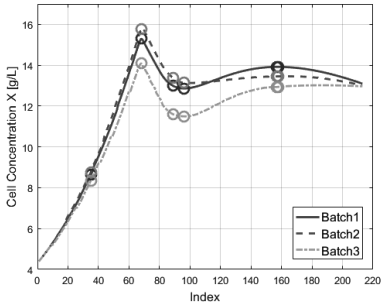


Figure 2. Synchronized batch profiles presented in Figure1. Marker denotes SPs.

phase division method. For the transition phase method, after the SP-based synchronization, the models are constructed only for the normal phases and the transition phases are estimated from them.

4.2. Estimation results

Figure 3 shows estimation results of the three methods around the transition region. The SP based phase division method is more accurate than the normal DTW method because of the improved performance of the synchronization. It has a rapidly changing point where the phase changes occur and lead to discontinuity in estimates as well as increase in error. The transition phase method, on the other hand, shows improvement in estimation accuracy, especially in the transition region. The estimation performances are compared in terms of error in Figure 4 and Table 2.

5. Conclusion

A modified method for estimating batch process using uneven batch data is proposed. The singular points of each uneven batch are detected and synchronized by phase-based DTW. By introducing the transition phase concept, the estimation performance of the DPLS model for each phase is improved, especially at the boundary of each phase.

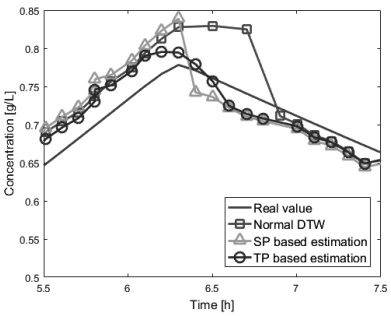


Figure 3. Estimation results of transition region

Table 2. Mean error

	DTW	SP	TP
Total Error [%]	3.16	2.97	2.42
Transition Phase Error [%]	5.09	3.83	2.99

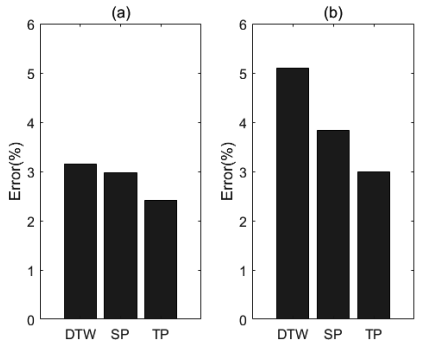


Figure 4. Mean error. (a) Total, (b) Transition phase. DTW, SP and TP means normal DTW method, SP based phase division method and transition phase method, respectively.

Acknowledgments

This research was respectfully supported by Engineering Development Research Center (EDRC) funded by the Ministry of Trade, Industry & Energy (MOTIE). (No. N0000990)

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (MSIP) (NRF-2016R1A5A1009592).

References

J. Chen and K.C. Liu, 2002, On-line batch process monitoring using dynamic PCA and dynamic PLS models, Chemical Engineering Science, 57, 63-75

X.T. Doan and R. Srinivasan, 2008, Online monitoring of multi-phase batch processes using phase-based multivariate statistical process control, Computer and Chemical Engineering, 32, 230-243

Z. Ge, L. Zhao, Y. Yao, Z. Song and F. Gao, 2012, Utilizing transition information in online quality prediction of multiphase batch processes, Journal of Process Control, 22, 599-611

M.H. Kaspar and W.H. Ray, 1993, Dynamic PLS modelling for process control, Chemical Engineering Science, 48, 20, 3447-3461

W. Ku, R.H. Storer and C. Georgakis, 1995, Disturbance detection and isolation by dynamic principal component analysis, Chemometrics and Intelligent Laboratory Systems, 30, 179-196

P. Nomikos and J.F. MacGregor, 1995, Multi-way partial least square in monitoring batch processes, Chemometrics and Intelligent Laboratory Systems, 30, 97-108

H. Sakoe and S. Chiba, 1978, Dynamic-programming algorithm optimization for spoken word recognition, IEEE Transactions on Acoustics, 26, 43-49

R. Srinivasan and M.S. Qian, 2005, Off-line Temporal Signal Comparison Using Singular Points Augmented Time Warping, Industry & Engineering Chemistry Research, 44, 4697-4716

S. Wold, P. Geladi, K. Esbensen and J.Ohman, 1987, Multi-way principal components and PLS analysis, Journal of Chemometrics, 1, 41-56

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50410-9

Integrated Metabolic and Process Modeling of Bubble Column Reactors for Gas Fermentation

Xiangan Li[a], Jin Chen[a], Derek Griffin[b], Xueliang Li[b], Michael A. Henson[a*]

[a]*Department of Chemical Engineering and Institute for Applied Life Sciences, University of Massachusetts, Amherst, MA 01003 USA*
[b]*LanzaTech Inc., Skokie, IL 60077 USA*
mhenson@umass.edu

Abstract

Gas fermentation has emerged as an attractive option for producing renewable fuels and chemicals from carbon monoxide (CO) rich waste streams. Bubble column reactors are being developed for large-scale production, motivating the investigation of multiphase reactor hydrodynamics. We combined two-phase hydrodynamics with a genome-scale reconstruction of *Clostridium autoethanogenum* metabolism and multiphase convection-dispersion equations to compare the performance of bubble column reactors with and without liquid recycle. Hydrodynamics were predicted to diminish column performance with respect to CO conversion, biomass accumulation and ethanol production when compared to column models in which the gas phase was modeled as ideal plug flow plus axial dispersion. Liquid recycle was predicted to be advantageous by increasing CO conversion, biomass production, and ethanol and 2,3-butanediol production compared to the non-recycle reactors. Our results demonstrate the power of combining metabolic models and hydrodynamics for simulating and optimizing gas fermentation reactors.

Keywords: Gas fermentation; metabolic models; hydrodynamics; multiphase reactors

1. Introduction

A promising route to renewable chemicals is the fermentation of carbon rich gas streams (Daniell et al., 2012). Typical feedstocks include industrial waste gases (mainly CO, CO_2 and N_2) and synthesis gas (mainly CO and H_2) produced by gasification of solid waste feedstocks. Compared with alternative technologies, gas fermentation offers several important advantages including operation at ambient temperatures and low pressures, high conversion efficiencies, robustness to gas impurities and high product yields. *Clostridium autoethanogenum* is an anaerobic bacterium that efficiently converts carbon rich gas streams into products such as ethanol and 2,3-butanediol. Wild-type strains of *C. autoethanogenum* produce low yields of ethanol compared to acetate. Researchers at LanzaTech developed an improved *C. autoethanogenum* strain which provides increased CO utilization, ethanol/acetate selectivity and ethanol tolerance (Liew et al., 2017).

Most academic research on gas fermentation has focused on batch and continuous stirred tank reactors due to their operational simplicity. Because mass transfer between the gas and liquid phases typically is the rate limiting step, these reactor configurations are not economically viable at industrial scale due to the high energy cost of agitation. Bubble column reactors are particularly promising for large-scale production due to their low power requirements and low maintenance and operational costs (Munasinghe and Khanal, 2010). We have previously developed computational models of bubble column reactors for converting synthesis gas into ethanol with the bacterium

Clostridium ljungdahlii (Chen et al., 2015). More recently, we experimentally tested a model of a bench-scale bubble column reactor for converting CO into ethanol with *C. autoethanogenum* as the biocatalyst (Chen et al., 2018).

These models were based on highly simplified hydrodynamics, most notably ideal plug flow for the vapor phase and plug flow plus axial dispersion for the liquid phase. Industrial bubble column reactors are complex multiphase processes in which spatial variations in the superficial gas velocity, gas holdup and interfacial area can have profound effects on fermentation performance (Kantarci et al., 2005). In this paper, we demonstrate how two-phase hydrodynamics can be incorporated into our modeling framework and assess the impact of hydrodynamics on the performance of bubble column reactors with and without liquid recycle. The proposed approach allowed the effects of the superficial gas velocity, gas bubble size, gas holdup and gas-liquid mass transfer coefficient on CO conversion and ethanol production to be predicted.

2. Methods

The bubble column models consisted of a genome-scale reconstruction (GEM) of *C. autoethanogenum* metabolism, uptake kinetics for dissolved CO, reaction-convection-dispersion equations for the liquid phase, and reaction-convection equations plus hydrodynamic equations for the gas phase. The GEM developed for LanzaTech's proprietary *C. autoethanogenum* strain consisted of 1,102 metabolites and 1,103 intracellular and exchange reactions. The uptake kinetics for dissolved CO were specified to follow the Michaelis-Menten equation and accounted for CO inhibition. Uptake parameters were estimated from experiments performed at LanzaTech.

The simulated feed gas consisted of CO (50 mole %), CO_2 (20 mole %), and N_2 (30 mole %) to mimic an industrial waste gas stream. Flux balance analysis (FBA) calculations with the *C. autoethanogenum* GEM showed that CO uptake resulted in the production of ethanol, acetate, CO_2 and 2,3-butanediol. Neither CO_2 or N_2 were consumed. Therefore, extracellular mass balances were posed for *C. autoethanogenum* biomass, gas-phase and liquid phase CO and CO_2, and liquid-phase acetate, ethanol and 2,3-butanediol. Transport processes were assumed to occur only in the axial direction of the column such that the spatial variation could be captured by a single variable z.

Two column configurations were investigated, both of which assumed countercurrent flows of the liquid and gas streams. The recycle column (Figure 1) had liquid recycle from the bottom of the column to the top of the column, while liquid recycle was absent in the conventional column. The recycle column was more consistent with industrial practice (Chen et al., 2018), while conventional columns have been the focus of previous UMass research (Chen et al., 2015). Due to the high velocity of the liquid recycle stream (50 m/h) compared to the low velocity of the liquid media stream (0.6 m/h), the recycle column was characterized by low reaction rates relative to convective mass transfer rates. Therefore, the liquid phase was assumed to be well mixed rather than spatially varying. This assumption was validated by comparing predictions to a more complex model in which the liquid phase was spatially varying.

The incorporation of hydrodynamics is illustrated below for the recycle column. The liquid- and gas-phase mass balances have been reported previously (Chen et al., 2018) with the exception that the current model: (1) does not assume constant liquid- and gas-phase volume fractions; and (2) allows CO and CO_2 gas-liquid mass transfer

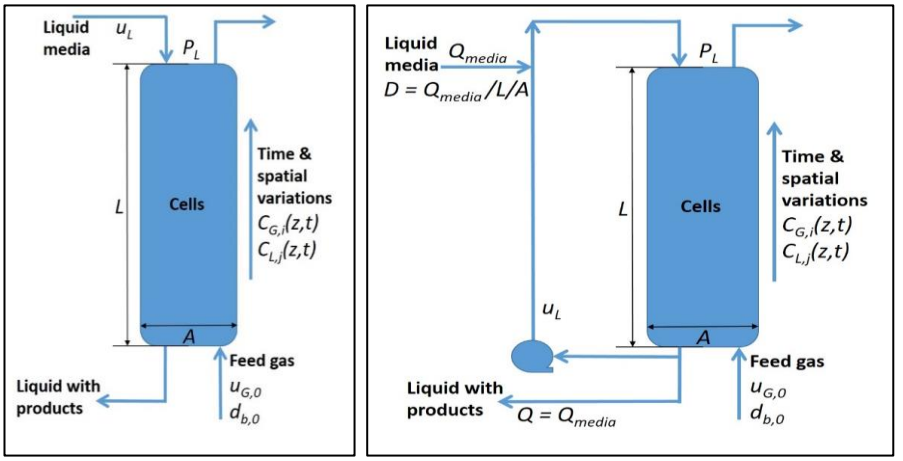


Figure 1. Schematic representation of the countercurrent bubble column reactors. $u_{G,0}$ and $d_{b,0}$ are the feed superficial gas velocity and bubble diameter, u_L is the liquid velocity, L is the reactor length, A is the reactor cross-sectional area, P_L is atmospheric pressure, Q_{media} is the media feeding rate, and D is the dilution rate.

coefficients to vary with time and spatial location according to the local interfacial gas-liquid mass transfer area. These variables were determined from the equations described below. The gas holdup was calculated from the one-dimensional drift-flux model,

$$\frac{u_G}{\varepsilon_G} = C_0(u_G + u_L) + v_b(1 - \varepsilon_G) \tag{1}$$

where ε_G is the liquid holdup, C_0 is the distribution parameter and v_b is the bubble rising velocity. The drift-flux model is commonly used for describing the relative motion of multiphase flows without solving the detailed momentum and energy equations. Eq. 1 is applicable to the bubbly flow regime and gas holdups less than 0.25 (Clark et al., 1990). The equation for bubble rising velocity had the form (Peebles, 1953),

$$v_b = 0.33g^{0.76}\left(\frac{\rho_L}{\mu_L}\right)^{0.52}\left(\frac{d_b}{2}\right)^{1.28} \tag{2}$$

where g is the gravitational constant, d_b is the bubble diameter, and ρ_L and μ_L are the density and viscosity of liquid phase.

The equation for bubble diameter had the form,

$$d_b = d_{b,0}\sqrt[3]{\frac{P_L n_G}{P n_{G,0}}} \tag{3}$$

where $d_{b,0}$ is the bubble diameter of the feed gas, $n_{G,0}$ is the molar flow rate of the feed gas and n_G is the local molar flow rate of the gas stream. This relation accounted for increasing bubble size due to decreasing pressure and decreasing size due to gas consumption. The following equation was used to calculate the superficial gas velocity,

$$u_G = u_{G,0}\frac{P_L}{P}\frac{n_g}{n_{g,0}} \tag{4}$$

where $u_{G,0}$ is the superficial velocity of the feed gas. The interfacial gas-liquid mass transfer area was calculated as,

$$a = \frac{6\varepsilon_G}{(1-\varepsilon_G)d_b} \tag{5}$$

Model parameters were obtained from the literature to the extent possible and are not reported here. The feed superficial gas velocity $u_{G,0}$ and bubble diameter $d_{b,0}$ were specified at atmospheric pressure and were pressure corrected to calculate the actual feed conditions. This approach was necessary because the pressure P_L at the bottom of the column depended on the gas holdup and was unknown until the model was solved. The model consisted of linear programs (LPs) for *C. autoethanogenum* metabolism, algebraic equations (AEs) for hydrodynamics, ODEs in time for liquid-phase mass balances, and partial differential equations (PDEs) in time and space for gas and dissolved gas mass balances. We used our solution strategy (Chen et al., 2015) based on spatial discretization of the PDEs followed by integration of the resulting differential-algebraic equation (DAE) system with embedded LPs with DFBAlab (Gomez et al., 2014). DFBAlab used the MATLAB code ode15s for DAE solution and Gurobi for LP solution.

3. Results and Discussion

To investigate the combined effects of hydrodynamics and liquid recycle, we developed four models: the conventional column (Figure 1A) with and without hydrodynamics and the liquid recycle column (Figure 1B) with and without hydrodynamics. The column models without hydrodynamics used constant values of the superficial gas velocity, gas holdup, gas bubble diameter and interfacial area. To facilitate meaningful comparisons, these values were specified as the corresponding values at the bottom of the column in the models with hydrodynamics. The conventional column models included equations for a spatially varying liquid phase (Chen et al., 2015) and used the same parameter values as the recycle column models except that a much smaller liquid flow rate was used to avoid reactor washout. Dynamic simulations were run for 500 h to obtain steady-state solutions. Because the conventional and recycle reactors had different exiting liquid flow rates, we plotted mass flow rates at the bottom of the column rather than concentrations.

Liquid recycle was predicted to generate high biomass concentrations compared to the conventional reactors without recycle (Fig. 2a) due to increased gas holdup and gas-liquid mass transfer (see below). Correspondingly, the recycle reactors were predicted to have higher CO gas conversion (Fig. 2c) and higher ethanol, acetate and 2,3-butanediol production (Fig. 2d-f). Ethanol and acetate were produced with an approximately equal selectivity for the four cases. By contrast, 2,3-butanediol was produced in appreciable amounts only with the recycle reactors. The conventional reactors produced higher dissolved CO concentrations (Fig. 2b) due to lower CO conversions and higher CO concentrations in the gas phase. The omission of hydrodynamic effects resulted in the overestimation of mass transfer performance, because in reality a significant reduction in gas residence time occurs due to bubble expansion (see below). For both column configurations, hydrodynamics were predicted to reduce biomass production, CO gas conversion, and ethanol, acetate and 2,3-butanediol production. The impact of hydrodynamics was most significant for the recycle reactor.

Differences in reactor performance between the four cases were analyzed with respect to the hydrodynamic variables. For either reactor configuration, the spatially varying hydrostatic pressure was predicted to have a substantial impact on superficial gas velocity (Fig. 3a), gas bubble diameter (Fig. 3b), gas holdup (Fig. 3c) and gas-liquid mass

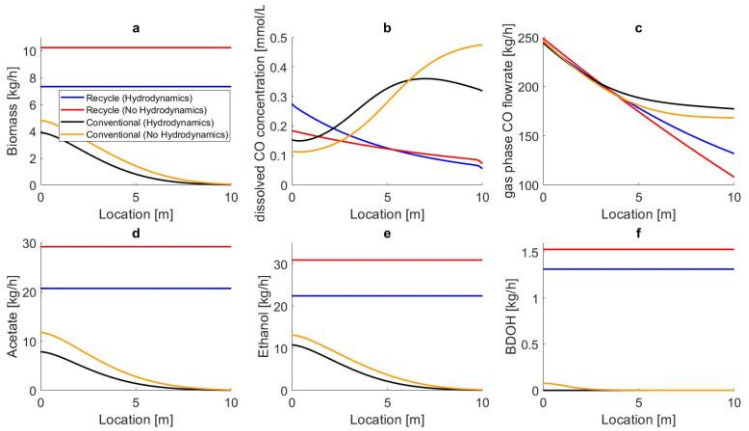


Figure 2. Steady-state production rate profiles predicted for two bubble column reactors (conventional, liquid recycle) with and without hydrodynamics. The liquid phase exited the column at $z = 0$ m and the gas phase exited the column at $z = 10$ m. Liquid-phase concentration profiles for the recycle reactors were flat due to the assumption of homogeneity.

transfer coefficient (Fig. 3d) variations across the column. The gas velocity and bubble diameter were predicted to increase along the length of the column because the effect of reducing hydrodynamic pressure was more significant than the effect of gas consumption (see Eq. 3 and 4). The drift-flux model predicted an increase in gas holdup along the column, primarily due to the increasing volumetric gas flow rate. The gas-liquid mass transfer coefficient was predicted to increase along the column because the effect of increasing gas holdup was more significant than the effect of increasing bubble size (see Eq. 5). The major differences between the conventional and recycle column models without hydrodynamics were the values of the gas holdup and the gas-liquid mass transfer coefficient. Liquid recycle yielded increased values of these two variables as well as a high volumetric flow rate, all of which remained constant along the column in the model where the hydrodynamic effect was neglected. Analogous predictions were obtained for the two column models with hydrodynamics.

Interestingly, liquid recycle was predicted to reduce *C. autoethanogenum* specific growth rates (Fig. 3e) and CO uptake rates (Fig. 3f) over most of the column compared to the conventional columns. The conventional reactors were predicted to have higher average specific CO uptake and specific growth rates due to higher dissolved CO concentrations. However, liquid recycle allowed the use of a media flow rate more than double that possible in the conventional column because recycle substantially increased the range of flow rates that did not result in reactor washout. Thus, the recycle reactors generated higher biomass and byproduct mass flow rates despite lower growth rates.

4. Conclusions

We investigated the impact of two-phase hydrodynamics on bubble column reactor performance by developing spatiotemporal metabolic models for conversion of carbon monoxide (CO) to ethanol using the microbial catalyst *Clostridium autoethanogenum*. The hydrodynamic model allowed the prediction of the superficial gas velocity, gas holdup, gas bubble diameter and interfacial area as a function of time and column location. The incorporation of hydrodynamics was predicted to substantially diminish

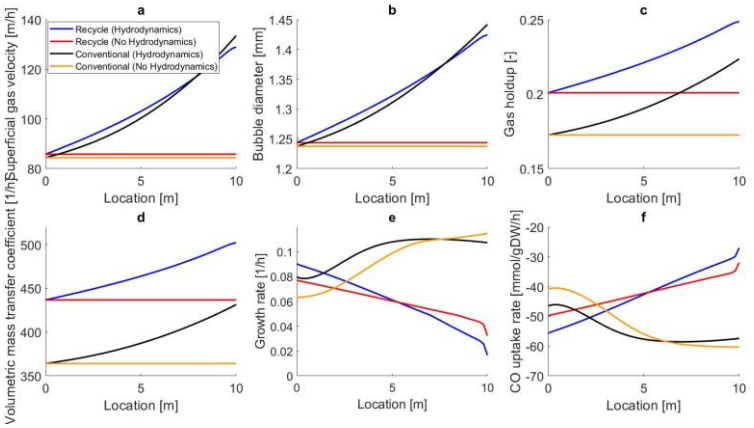


Figure 3. Steady-state profiles of hydrodynamic and GEM variables for two bubble column reactor configurations with and without hydrodynamics.

bubble column performance with respect to CO conversion, biomass accumulation and ethanol production. Therefore, we concluded that the inclusion of gas-phase hydrodynamics is important to generate high fidelity models of these complex multiphase processes. Our models also predicted that liquid recycle would substantially improve performance compared to a conventional reactor without recycle.

References

Chen, J., J.A. Gomez, K. Höffner, P.I. Barton and M.A. Henson, *Metabolic modeling of synthesis gas fermentation in bubble column reactors*. Biotechnology for biofuels, 2015. **8**(1): p.89.

Chen, J., J. Daniel, D. Griffin, X, Li and M.A. Henson, *Experimental testing of a spatiotemporal metabolic model for carbon monoxide fermentation with Clostridium autoethanogenum.* Biochemical engineering journal, 2018. **129**: p. 64-73.

Clark, N., J. Van Egmond, and E. Nebiolo, *The drift-flux model applied to bubble columns and low velocity flows.* International journal of multiphase flow, 1990. **16**(2): p. 261-279.

Daniell, J., M. Köpke, and S.D. Simpson, *Commercial biomass syngas fermentation.* Energies, 2012. **5**(12): p. 5372-5417.

Gomez, J.A., K. Höffner, and P.I. Barton, *DFBAlab: a fast and reliable MATLAB code for dynamic flux balance analysis.* BMC bioinformatics, 2014. **15**(1): p. 409.

Kantarci, N., F. Borak, and K.O. Ulgen, *Bubble column reactors.* Process Biochemistry, 2005. **40**(7): p. 2263-2283.

Liew, F., A.M. Henstra, M. Köpke, K. Winzer, S.D. Simpson and N.P. Minton, *Metabolic engineering of Clostridium autoethanogenum for selective alcohol production*. Metabolic engineering, 2017. **40**: p. 104-114.

Munasinghe, P.C. and S.K. Khanal, *Biomass-derived syngas fermentation into biofuels: opportunities and challenges.* Bioresource technology, 2010. **101**(13): p. 5013-5022.

Peebles, F.N., *Studies on the motion of gas bubbles in liquids.* Chem. Eng. Prog., 1953. **49**: p. 88-97.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50411-0

Process Variability Source Analysis for a Multi-step Bio-process

Yuan Jin[a], S. Joe Qin[a*], Victor Saucedo[b], Zheng Li[b], Angela Meier[c], Siddhartha Kunda[d], Salim Charaniya[e]

[a]*Mork Family Department of Chemical Engineering and Materials Science, University of Southern California, Los Angeles, CA, 90089, USA*
[b]*Process Development Engineering, Genentech, South San Francisco, CA, 94080, USA*
[c]*Late Stage Cell Culture, Genentech, South San Francisco, CA, 94080, USA*
[d]*MSAT, Genentech, Vacaville, CA, 95688, USA*
[e]*Global MSAT, Genentech, South San Francisco, CA, 94080, USA*
sqin@usc.edu

Abstract

Typical manufacturing bio-processes operate as multi-step batch operations. Abnormal variations are often experienced in commercial manufacturing which affects the ability to maintain the product supply to the patients. The low titer is usually highly correlated with high lactate in recombinant Chinese hamster ovary (CHO) cells producing immunoglobulin G, but there are limited means to analyze the vast amount of process data archived from manufacturing sites and understand the causes of the multiple sources of process variability. In this study, we investigate diverse data from a cell culture process, from historical runs of a multi-step manufacturing process for a recombinant antibody. Data includes intra-batch measurements, batch conditions and media preparation measurements. Cell culture process stages from inoculum train to production runs.

To analyze manufacturing variability in multiple stages with various sources of data sets, we first adopt principal component analysis (PCA) based monitoring to find in which space high lactate and low lactate batches can be separated, and apply clustering to identify fault directions for intra-batch data from production runs. We then predict final lactate using data from inoculum stages with a nonlinear support vector machine (SVM) classifier. These models are used to uncover potential causes of high lactate. This data-driven modelling can provide new insights on process characteristics to understand the sources of process variability and thus mitigation procedures can be implemented.

Keywords: process variability, cell culture process, PCA monitoring, SVM.

1. Introduction

Multi-step batch processes are widely used in pharmaceutical manufacturing and process variability may be experienced in these complex bioprocesses that utilize mammalian cells for production of recombinant protein therapeutics (Charaniya 2010). Data of enormous volume and various kinds are generated and collected from different cell culture process stages. Therefore, mining the historical data is pivotal to uncover the latent reasons that cause the process variability, and potentially provide means to detect them in early stages and improve process performance.

Low titer in CHO monoclonal antibodies (mAb) production is often correlated to lower cell growth and with high lactate production (Charaniya 2010). Several studies have employed lactate metabolism model to study this phenomena (Mulukutla 2015, Zagari 2013). For example, Mulukutla (2015) proposed a steady state topological metabolic model, indicating the mechanism of lactate consumption and accumulation can explain the cause of high lactate. Nevertheless, these knowledge-based models are built with lactate metabolic pathway simulation and can hardly be applied to the large-scale manufacturing process since concentrations of additional metabolites are needed, which is not done routinely in commercial manufacturing.

Multivariate analysis (MVA) is also a useful way to explore the bioprocess data (Nomikos 1995, 1996). Methods such as Multi-way principal component analysis (mPCA) and Multi-way partial least square (mPLS) have been extensively used to monitor batch process data. For example, Nomikos et al. (1995) applied mPCA and mPLS to extract the information in the trajectory data and developed simple monitoring charts for process fault detection.

This paper aims to investigate the cell culture process variability in multiple manufacturing stages with diverse bio-process data sets. Following introduction, Section 2 presents the large-scale cell culture process and characteristics of the data sets. Section 3 describes the detailed modelling procedure for process anomaly source analysis in different stages and their results. Conclusions and future work are discussed in Section 4.

2. Cell culture process and data set characteristics

In this study, we investigate 241 historical runs of large-scale cell culture process data during year 2008- 2015 from Genentech's facility. The recombinant CHO cells were expanded from inoculum train to production runs. For each stage, data includes offline measurements, initial batch conditions and media preparation measurements.

For each run, the lactate concentration at the end of the production process is measured, scaled to range 0-1 based on the min and max reading and used as the criterion to distinguish high performance run and low performance run. The distribution of the final lactate concentration is shown in Fig.1, runs are recognized as low performance (high lactate) if the normalized value is larger than 0.4, otherwise, are recognized as high performance (low lactate). There are 75 high lactate runs and 166 low lactate runs in total.

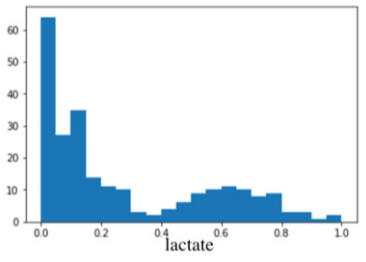


Figure 1 Final lactate concentration distribution

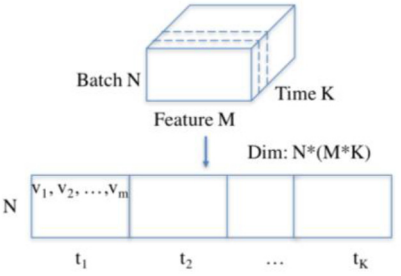


Figure 2 Process data unfolding

3. Multi-step batch data analysis

In this section, we describe how to find process anomaly source in multi-step bio-process. After unfolding the production run batch data, we first find a latent space, and apply clustering in this space to find different hidden patterns, we then adopt classification on inoculum run in order to diagnose the anomaly source in an early stage.

3.1. Intra-batch data preprocessing and unfolding

The features in the intra-batch data set are measured twice a day. After removing duplicates and filling missing data by local interpolation, each feature has 23 sampling points during the production runs. We focus on the first 13 sampling points before the irreversible lactate accumulation for high lactate runs is observed.

The intra-batch data is organized in a 3D tensor X (N * M * K) as shown in Fig.2. It contains n = 1, 2, …, N batch runs, where each of them has m = 1,2,…,M features with k = 1,2, …,K samples. To analyze it at batch level, we decompose it into a 2D matrix with dimension (N * (M * K)) (Nomikos 1995). The 2D matrix has dimension 241* 130, where each column represents a feature at a certain sampling timestamp.

3.2. Production run intra-batch data analysis

Fig.1 indicates that around 30% of the batches end with high lactate concentration, possibly there exist multiple reasons that are correlated to the abnormal lactate accumulation. In this section, we want to figure out how many potential patterns exist by applying clustering methods. The data sets used in this section are the intra-batch data in the production run. By applying unsupervised learning, we want to discover different hidden patterns of process data that are correlated to the final high lactate concentration.

3.2.1. PCA based monitoring

The high and low lactate batches cannot be separated in the original space, so we need to find a latent space in which they are separable. We first do PCA based monitoring on intra-batch data for all 166 low lactate batches, the number of PCs is selected as 48 since it can explain 95% of the total variance. We then calculate corresponding T2 and SPE control limit, which are 65.17 and 9.55 respectively (Qin 2003, 2012). The T2 index and SPE index for both high lactate and low lactate batches are plotted in Fig.3.

It can be observed that in residual space (Fig.3b), the high lactate and low lactate runs can be easily differentiated, while in the principle space, they cannot be easily differentiated. Therefore, further analysis will be performed on the residual space obtained by PCA monitoring.

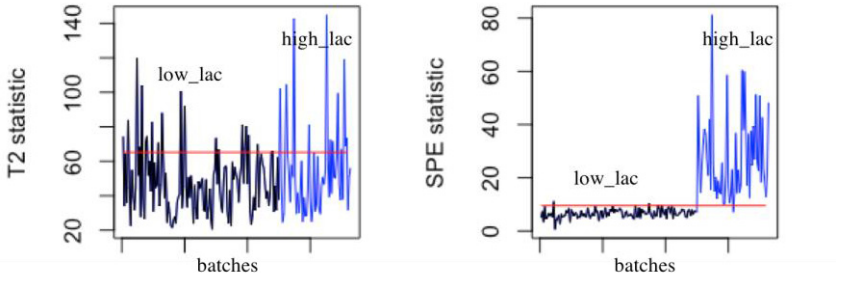


Figure 3. PCA monitoring results. a) in principal space b) in residual space

3.2.2. Hierarchical clustering

It is important to find hidden patterns in the intra-batch data; to achieve it, we perform hierarchical clustering on the residual space obtained by PCA monitoring for all high lactate batches. The number of clusters is 4, which is determined by the elbow method (Kodinariya 2013). Fig.4 shows the hierarchical clustering tree and Fig.5 shows the curve of within-cluster sum of square (WSS) corresponding to the number of cluster k.

To visualize the special distribution of the four high lactate clusters and the low lactate cluster, linear discriminant analysis (LDA) is performed. As shown in Fig.5, in the LD1 and LD2 plot, clusters 2 and 3 are separated and in the LD1 and LD3 plot, clusters 1 and 4 are separated as well.

The hidden patterns in different high lactate clusters can be visualized by the heatmap of every feature. For instance, Fig.7 shows the heatmap of packed cell volume (PCV) in residual space, it can be seen that cluster 2 has different pattern among others.

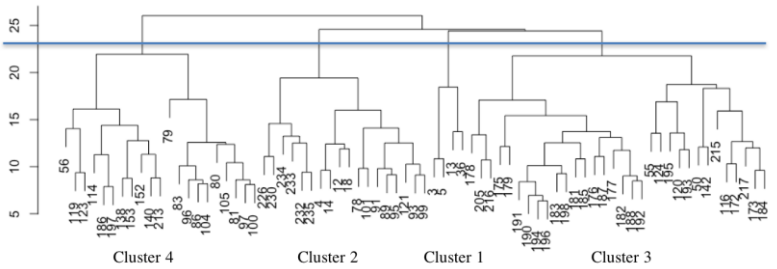


Figure 4. Hierarchical clustering tree

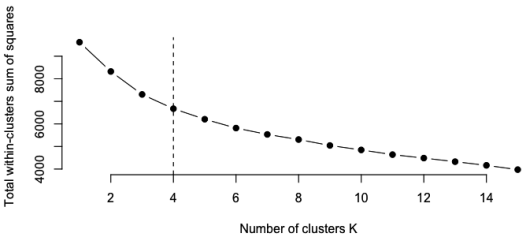


Figure 5. Elbow method to select the number of clusters

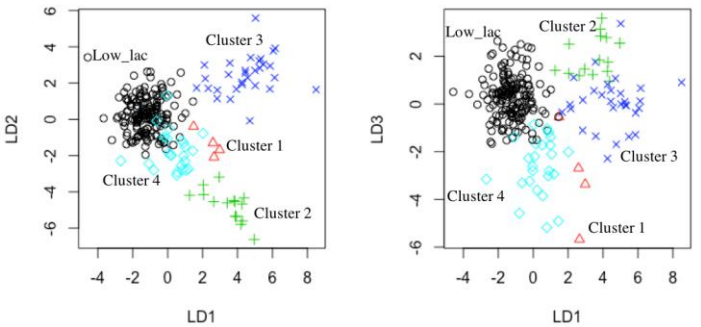


Figure 6. Overplot low lactate with high lactate clusters by LDA

3.3. Step-wise data analysis

In previous section, we discovered that there are four final high lactate clusters. Each cluster is potentially related to a specific reason. Though Fig.7 is helpful to better understand the process, these intra-batch features are mainly related to the cell physiology and chemical environment, which cannot be adjusted by the operators.

A more useful way is to identify the potential causes that can be adjusted in an early stage and make prediction before the production run starts. Therefore, in this section, we focus on the potentially adjustable features in inoculum N-1 train runs. These features are selected based on process knowledge, mainly from initial states, offline measurements, and media preparation data.

23 features are selected and LDA is used to reduce dimension, the corresponding results are in Fig.8. It can be observed that most of the batches in cluster 2 are not overlapped with others, meaning that the anomaly source for cluster 2 batches may happen in inoculum run. To differentiate the abnormal batches with the normal ones, we then use the first three LDs to train a support vector machine (SVM) classifier with RBF ($\gamma = 2$) kernel. In this model, batches in cluster 2 are labelled as abnormal and all others are as normal. The decision boundaries calculated by SVM are also plotted in Fig.8. This model has a good performance with type 1 error as 2.3 % and type 2 error as 20.6 %.

Further, we use box plot to represent features that cause cluster 2 to be separated, two features are selected as examples. It can be observed in Fig.9, low Na2CO3 per media volume and low initial osmolality are the most likely features that differentiate the high lactate batches in cluster 2.

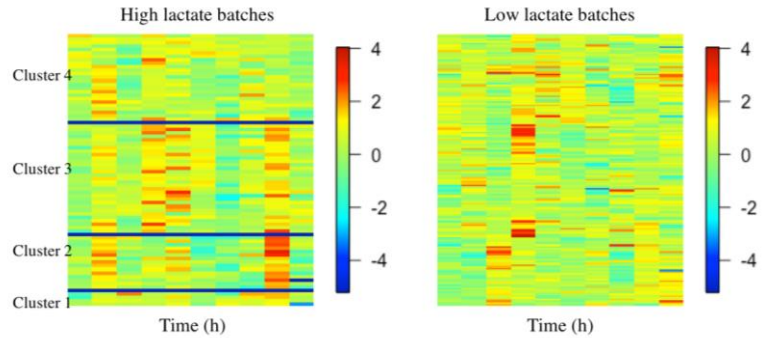


Figure 7. Hidden patterns of PCV in different clusters

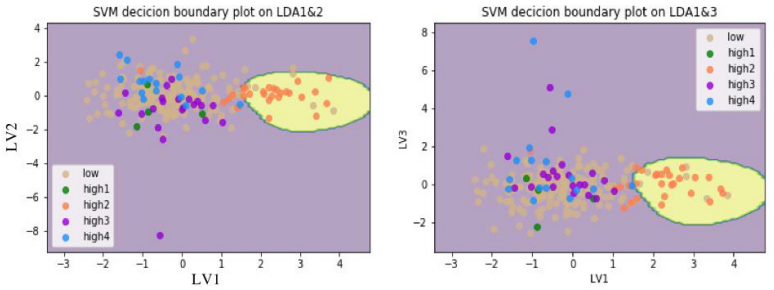


Figure 8. SVM boundary plot after LDA dimension reduction

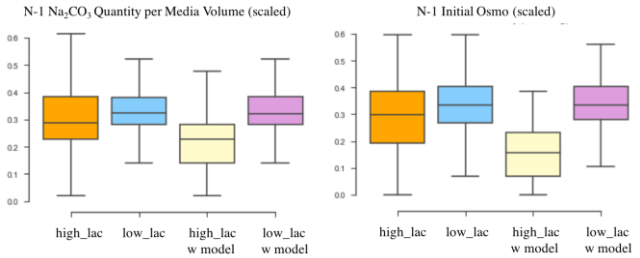


Figure 9. Box plot of the potential high lactate causes

4. Conclusions and future work

In this paper, we analyze the manufacturing variability in a multi-stage cell culture process with historical data. To find the hidden patterns within the data, we first adopt PCA based monitoring to find the latent space in which data is most separated, then apply clustering method. High lactate batches can be grouped into four different clusters. To make early prediction before the production run starts and to uncover potential causes of high lactate, we then adopt a nonlinear SVM classifier to analyze data in inoculum stage. In this way, the causes for cluster 2 batches are identified.

This modelling method provides new insights on process characteristics to understand the sources of process variability and help researchers form a data-driven experimental plan to confirm the root cause. Shortly, we will continue working on finding the high lactate causes related to other clusters. Other data sets such as the thaw data and raw material information will also be explored.

References

Charaniya, Salim, Huong Le, Huzefa Rangwala, Keri Mills, Kevin Johnson, George Karypis, and Wei-Shou Hu. "Mining manufacturing data for discovery of high productivity process characteristics." Journal of biotechnology 147, no. 3 (2010): 186-197.

Kodinariya, Trupti M., and Prashant R. Makwana. "Review on determining number of Cluster in K-Means Clustering." International Journal 1, no. 6 (2013): 90-95.

Mulukutla, Bhanu Chandra, Andrew Yongky, Simon Grimm, Prodromos Daoutidis, and Wei-Shou Hu. "Multiplicity of steady states in glycolysis and shift of metabolic state in cultured mammalian cells." PloS one 10, no. 3 (2015): e0121561.

Nomikos, Paul, and John F. MacGregor. "Multi-way partial least squares in monitoring batch processes." Chemometrics and intelligent laboratory systems 30, no. 1 (1995): 97-108.

Nomikos, Paul. "Detection and diagnosis of abnormal batch operations based on multi-way principal component analysis World Batch Forum, Toronto, May 1996." ISA transactions 35, no. 3 (1996): 259-266.

Qin, S. "Statistical process monitoring: basics and beyond." Journal of chemometrics 17, no. 8 - 9 (2003): 480-502.

Qin, S. Joe. "Survey on data-driven industrial process monitoring and diagnosis." Annual Reviews in Control 36, no. 2 (2012): 220-234.

Zagari, Francesca, Martin Jordan, Matthieu Stettler, Hervé Broly, and Florian M. Wurm. "Lactate metabolism shift in CHO cell culture: the role of mitochondrial oxidative activity." New biotechnology 30, no. 2 (2013): 238-245.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50412-2

A System Identification Enhanced Phenotype Phase Plane Analysis

Matthew Hilliard[a], Andrew Damiani[a], Q. Peter He[a] and Jin Wang[a*]

[a] *Dept. of Chemical Engineering, Auburn University, Auburn, AL, 36849, US*
wang@auburn.edu

Abstract

Genotype-phenotype relationship is fundamental to biology, and predicting different phenotypes based on the sequenced genome is one of the main goals for genome-scale metabolic model development. Phenotype phase plane (PhPP) analysis is a powerful tool to provide a global perspective on the genotype-phenotype relationship, and to help characterize different metabolic phenotypes. However, the traditional PhPP analysis is based on shadow price analysis, which provides limited information to characterize different phenotypes. In this work, we propose a system identification (SID) enhanced PhPP, which not only obtains information that could be obtained through shadow price analysis, but also provides additional information that helps characterize different phenotypes. The effectiveness of the SID-PhPP approach is demonstrated using an illustrative cell model and a core metabolic network model of *E. coli*.

Keywords: Phenotype phase plane analysis, System identification, Genome-scale metabolic model, Shadow price.

1. Introduction

Systems biology is a rapidly evolving discipline that seeks to determine how complex biological systems function by integrating experimentally derived information through mathematical modeling. Genome-scale metabolic network models represent the link between the genotype and phenotype of the organism. In essence, a genome-scale metabolic model (GEM) is a comprehensive functional database of the organism's cellular metabolism (King et al., 2015; Sánchez and Nielsen, 2015), which consists of a set of metabolites, metabolic reactions (*i.e.*, stoichiometric matrix), and constraints. These models provide a holistic view of the organism's metabolism, and constraint-based metabolic flux analyses, such as flux balance analysis (FBA), have been used extensively to study genome-wide cellular metabolic networks (Kauffman et al., 2003). One major application of GEM is to predict different growth phenotypes (*e.g.*, how fast cells grow, what products are excreted) in various genetic and environmental conditions. Developed by the Palsson lab, phenotype phase plane (PhPP) analysis is a powerful tool that uses FBA with the GEM to provide a global perspective on the genotype-phenotype relationship, and to help characterize different metabolic phenotypes (Edwards et al., 2002). In PhPP analysis, the classification of different phenotypes is based on the shadow price of various metabolites, where a phenotype is defined by all the conditions where a metabolite's shadow price stays constant. However, shadow prices only describe how each metabolite affects the objective function of FBA, without providing any information on how different reactions interact with each other within the same phenotype. Therefore, when PhPP is applied to analyze GEMs, it is usually combined with other *in silico* simulations and gene deletions to

identify and explain the fundamental differences among different phenotypes (Duarte et al., 2004; Edwards et al., 2001).

To address this challenge, the system identification (SID) based framework that we previously developed for genome-scale metabolic model analysis (Damiani et al., 2015) is extended to enhance PhPP, which is termed SID-enhanced PhPP (SID-PhPP) analysis. In the SID framework, we first perturb the metabolic network through designed input sequences, *i.e.*, designed *in silico* experiments; then we apply multivariate statistical analysis tools such as principal component analysis (PCA) to analyze the *in silico* results in order to extract information on how such perturbations propagate through the network; finally, we visualize the extracted knowledge against the network map to provide meaningful and intuitive interpretations of the *in silico* results. The SID framework has been successfully applied to genome-scale model evaluation, as well as guiding the development of an improved genome-scale model of *Scheffersomyces stipites* (Damiani et al., 2015; Damiani et al., 2017). In this work, we first briefly review the traditional PhPP analysis and introduce the SID- PhPP method, using the same illustrative example provided in the original PhPP work (Edwards et al., 2002); then we use the core metabolic network model of *E. coli* to demonstrate the effectiveness of the proposed SID-PhPP approach, and show how a "hidden" phenotype that share the same set of shadow prices with another phenotype could be identified by the SID-PhPP approach.

2. Traditional PhPP Analysis

Metabolic phenotypes can be viewed as different metabolic network utilization profiles. Within each phenotype, all points (culture conditions) share the same set of activated pathways, and the same set of excreted products. In other words, the activated reactions and/or excreted products are different among different phenotypes. The traditional PhPP analysis provides a global view of each phenotype in genome-scale models by breaking down the metabolic behavior in distinct phases through shadow price analysis. Mathematically, shadow price of metabolite i is defined as $\gamma_i = -\frac{dZ}{db_i^v}$ where Z is the objective function of FBA and b_i^v denotes the additional availability for the metabolite (Edwards et al., 2002). In the traditional PhPP analysis, a phenotype is defined by all the culture conditions, defined by two independent variables (typically the carbon and oxygen source), where metabolites' shadow prices stay constant.

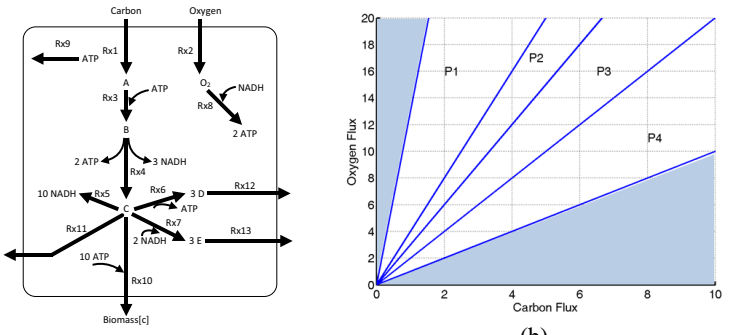


Figure 1. (a) The network map of an illustrative cell model; (b) Phenotype phase plane of the cell model.

Table 1. Shadow prices of different metabolites in different phenotypes

	Carbon	O_2	ATP	B	NADH	C	D	E	Biomass
P1	-1.30	0.10	0.00	-1.30	-0.10	-1.00	-0.33	-0.40	0.00
P2	-0.21	-0.17	-0.09	-0.30	-0.01	-0.09	0.00	-0.04	0.00
P3	-0.05	-0.23	-0.09	-0.14	0.05	-0.09	0.00	0.00	0.00
P4	0.50	-0.50	0.00	0.50	0.50	-1.00	-0.33	0.00	0.00

Here we use a simplified cell model to illustrate how PhPP analysis works. Figure 1 (a) shows the network map of the illustrative example, which was first introduced in (Edwards et al., 2002). Figure 1 (b) plots "phenotype phase plane" of the model cell, which exhibits 4 different phenotypes (*i.e.*, P1~P4). The boundaries between different phenotypes are determined by the shadow prices, with a few of them listed in Table 1. It is worth noting that, according to PhPP, a zero shadow price indicates the metabolite would be secreted because it has no effect on the objective function (Edwards et al., 2002).

3. The SID Enhanced PhPP Analysis

The SID-PhPP is proposed in this work as an alternative approach to characterize different phenotypes. Figure 2 provides an overview of the SID-PhPP. The same illustrative example from Section 2 is used to demonstrate how SID-PhPP works. First, the designed one-dimensional *in silico* experiments are conducted to determine the boundaries of the different phenotype phases. As shown in Figure 3 (a), one set of designed experiments is to hold carbon uptake flux constant while varying oxygen update. Next, PCA was applied to analyze the resulted flux matrixes, and the resulted scores corresponding to the first principal component (PC) are plotted in Figure 3 (b). Each linear segment in the score plot indicates one phenotype, and Figure 3 shows that the phenotype boundaries determined through SID-PhPP exactly match those determined through shadow price analysis. It is worth noting that SID-PhPP does not require running simulations that cover the whole plane. Instead, only few sets of one-dimensional *in silico* experiments are needed to determine the phenotype boundaries.

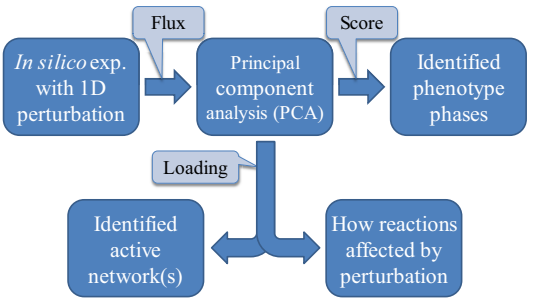
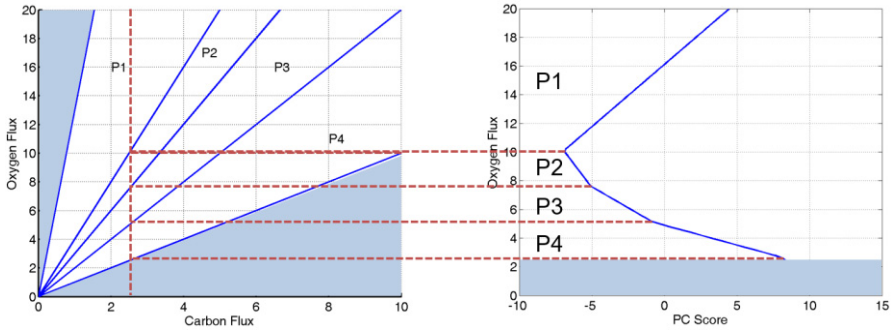


Figure 2. Overview of the SID enhanced PhPP

In the next section, we use the *E. coli* core model as an example to illustrate how SID-based framework can help characterize different phenotypes, as well as to discover "hidden" phenotypes that share the same set of shadow prices with another phenotype.



(a) (b)

Figure 3. (a) Designed in silico experiments at fixed carbon uptake and varying oxygen uptake; (b) Phenotype boundaries determined through PC scores.

4. Application to *E. coli* core model

4.1. E. coli core model

The central carbon metabolic model of *E. coli* proposed by Orth et al. (2010) contains 95 reactions and 72 metabolites, with major reaction pathways including glycolysis, both branches of the pentose phosphate pathway (PPP), TCA cycle, and electron transport chain. The two substrates chosen for this study are glucose and oxygen. Five by-products could be excreted are CO_2, lactate, formate, ethanol, and acetate. The biomass reaction is used as the objective function.

4.2. The traditional PhPP analysis

First, the traditional PhPP was applied to predict phenotypes based on the *E. coli* core model, which results in 4 different phenotypes (P1~P4) as shown in Figure 4 (a), with different by-products excreted (Table 2). The shadow prices of a few key metabolites are listed in Table 3. It is worth noting that although formate has shadow price of 0 for all 4 phenotypes, it is only excreted in P3 and P4, not in P1 or P2. For ethanol, although its shadow price is not zero for P4, FBA simulations show that ethanol is excreted in P4; similarly for acetate, although its shadow price is not zero for any phenotypes, it is excreted in P3 and P4 confirmed by FBA simulations. This finding indicates that the statement about zero shadow price by Edwards et al. (2002) is inaccurate.

Table 2 By-products excreted in different phenotypes

Phase	By-products
P1	CO2
P2	CO2
P3	Acetate, Formate, CO2
P4	Acetate, Formate, Ethanol

Table 3 Shadow prices of some key metabolites in different phenotypes

Metabolite	P1	P2	P3	P4
Glucose	-0.1373	-0.0459	-0.0325	-0.0304
Oxygen	0.0229	-0.0230	-0.0325	-0.0359
CO2	0	0	0	-0.0014
Formate	0	0	0	0
Acetate	-0.0458	-0.0026	-0.0027	-0.0028
Ethanol	-0.0686	-0.0128	-0.0054	-0.0028

4.3. The SID-enhanced PhPP analysis

SID-PhPP is performed to analyze the same model, with results shown in Figure 4 (b). It can be seen that SID-PhPP identified 5 different phenotypes. Three of them (P1, P2 and P4) are the same as those identified by the traditional PhPP, while P3 identified by the traditional PhPP is split into two phenotypes (P3' and P3") by SID-PhPP. It is confirmed that the shadow prices of different metabolites in P3' and P3'' are exactly the same as those of P3 listed in Table 3. According to the traditional PhPP, P3' and P3" should be the same phenotype. However, SID-PhPP analysis shows that, for *in silico* experiments conducted in P3, the resulted scores form two linear segments, indicating two different phenotypes. To determine whether the two linear segments represent two different phenotypes, we visualized the SID-PhPP analysis results against the network map for P3' and P3'' in Figure 5 (a) and (b), respectively. Figure 5 clearly shows that P3' and P3'' are two distinct phenotypes, with different activated reaction pathways (oxidative PPP active in P3', while inactive in P3") and different by-product excretion patterns (acetate only in P3', while acetate and formate in P3"), although they share the same set of shadow prices.

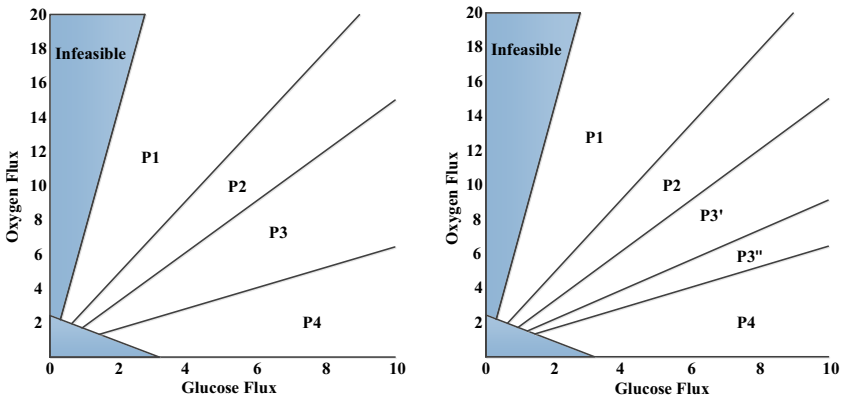


Figure 4. (a) The traditional PhPP reveals four different phenotypes; (b) The SID-PhPP reveals five different phenotypes.

5. Conclusions

Genome-scale metabolic network models represent the link between the genotype and phenotype of the organism, and phenotype phase plane analysis is a prevalent method to characterize different phenotypes predicted by genome-scale model. However, certain limitations exist for the original PhPP analysis approach. For example, as we have shown using the *E. coli* core model, the metabolites with zero shadow prices are not always excreted, while the metabolites with non-zero shadow prices could be excreted. This is because the shadow prices only quantify the effect of the overall flux of a metabolite on the FBA objective function, without taking the interactions among different activated pathways into account. However, it is the interactions among different activated pathways that determine the function and behaviour of a given phenotype. To address this limitation, we propose SID enhanced PhPP in this work. In the SID-PhPP, the designed *in silico* experiments are conducted to determine the boundaries between different phenotypes with significantly reduced computation. More importantly, SID-PhPP provides additional information that help fully characterize

different phenotypes. The effectiveness of the SID-PhPP approach is demonstrated using an illustrative example and a core metabolic network model of *E. coli*. Our results show that SID-PhPP is a powerful tool for revealing the metabolic capability of a metabolic network model, and enables better and more intuitive systems level understanding which provides the foundation for improved strain development.

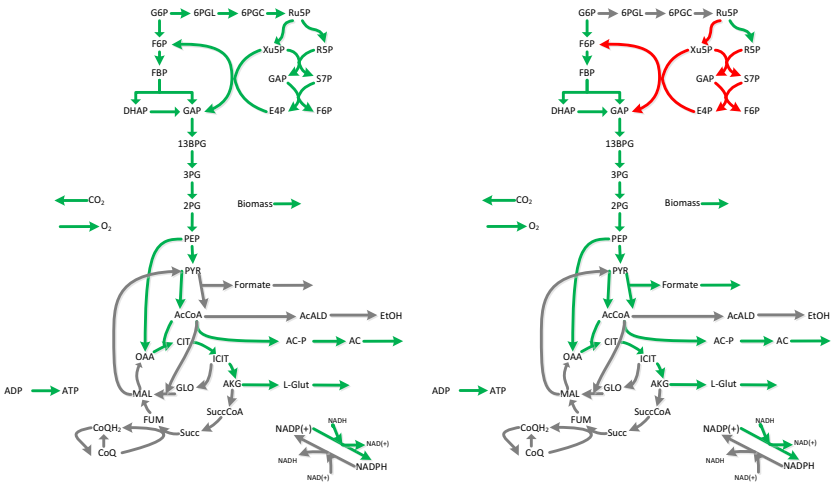


Figure 5. Different activated pathways and by-product excretion patterns in P3' (a) and P3'' (b) with an increasing oxygen uptake, where green arrows indicating increasing flux, red arrows indicating decreasing flux.

References

A Damiani, QP He, J. Wang, 2017. A System Identification Based Framework for Genome-Scale Metabolic Model Validation and Refinement. In: . *Proc. 10th IFAC Int. Symp. Adv. Control. Chem. Process.*, pp. 13013–13018.

AL Damiani, QP He, TW Jeffries, J. Wang, 2015. Comprehensive evaluation of two genome-scale metabolic network models for Scheffersomyces stipitis. *Biotechnol. Bioeng.* **112**:1250–1262.

NC Duarte, BØ Palsson, P Fu. 2004. Integrated analysis of metabolic phenotypes in Saccharomyces cerevisiae. *BMC Genomic-* **5**.

JS Edwards, RU Ibarra, BO Palsson. 2001. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.* **19**:125–130.

JS Edwards, R Ramakrishna, BO Palsson. 2002. Characterizing the metabolic phenotype: a phenotype phase plane analysis. *Biotechnol. Bioeng.* **77**:27–36.

KJ Kauffman, P Prakash, JS Edwards. 2003. Advances in flux balance analysis. *Curr. Opin. Biotechnol.* **14**:491–496.

ZA King, CJ Lloyd, AM Feist, BO Palsson. 2015. Next-generation genome-scale models for metabolic engineering. *Curr. Opin. Biotechnol.* **35**:23–29.

JD Orth, RM Fleming, BO Palsson. 2010. Reconstruction and use of microbial metabolic networks: the core Escherichia coli metabolic model as an educational guide. *EcoSal plus*.

BJ Sánchez, J Nielsen. 2015. Genome scale models of yeast: towards standardized evaluation and consistent omic integration. *Integr. Biol.* **7**:846–858.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50413-4

Uncovering New Opportunities from Frequency Regulation Markets with Dynamic Optimization and Pyomo.DAE

Alexander W. Dowling[a*] and Bethany L. Nicholson[b]

[a]*Department of Chemical and Biomolecular Engineering, University of Notre Dame, Notre Dame, Indiana, 46556, USA*
[b]*Center for Computing Research, Sandia National Laboratories, Albuquerque, New Mexico, 87185, USA*
[*]*alex@dowlinglab.org*

Abstract

Real-time energy pricing has caused a paradigm shift for process operations with flexibility becoming a critical driver of economics. As such, incorporating real-time pricing into planning and scheduling optimization formulations has received much attention over the past two decades (Zhang and Grossman, 2016). These formulations, however, focus on 1-hour or longer time discretizations and neglect process dynamics. Recent analysis of historical price data from the California electricity market (CAISO) reveals that a majority of economic opportunities come from fast market layers, i.e., real-time energy market and ancillary services (Dowling et al., 2017).

We present a dynamic optimization framework to quantify the revenue opportunities of chemical manufacturing systems providing frequency regulation (FR). Recent analysis of first order systems finds that slow process dynamics naturally dampen high frequency harmonics in FR signals (Dowling and Zavala, 2017). As a consequence, traditional chemical processes with long time constants may be able to provide fast flexibility without disrupting product quality, performance of downstream unit operations, etc. This study quantifies the ability of a distillation system to provide sufficient dynamic flexibility to adjust energy demands every 4 seconds in response to market signals. Using a detailed differential algebraic equation (DAE) model (Hahn and Edgar, 2002) and historic data from the Texas electricity market (ECROT), we estimate revenue opportunities for different column designs. We implement our model using the algebraic modeling language Pyomo (Hart et al., 2011) and its dynamic optimization extension Pyomo.DAE (Nicholson et al., 2017). These software packages enable rapid development of complex optimization models using high-level modelling constructs and provide flexible tools for initializing and discretizing DAE models.

Keywords: electricity markets, demand response, smart manufacturing, nonlinear programming, distillation

1. Introduction: Frequency Regulation Markets

Modern electricity infrastructures use complex hierarchical wholesale markets to coordinate generation, loads, and transmission networks. This results in temporal and spatial variations in prices set at multiple timescales (e.g., 1-hour prices in day-ahead markets, 5- to 15-minute prices in real-time markets) that directly impact large electricity producers/consumers including traditional utility companies, aggregators, and some

industrial sites. Dowling et al. (2017) and Zhang and Grossmann (2016) provide thorough background in electricity markets and emphasize emerging opportunities for industrial participants. Many recent studies focus on extending planning and scheduling formulations to account for time-sensitive energy prices.

Ancillary services, i.e., spinning/non-spinning reserves and frequency regulation capacity, are critical to provide contingency at fast timescales (minutes to seconds) and ensure grid reliability. Resources providing reserve capacity are compensated for making a commitment to either increase generation or shed load if dispatched. Often this dispatch occurs once a week or less and can provide new revenue opportunities for flexible industrial sites. Zhang et al. (2015) explore optimal scheduling strategies of reserve capacity under uncertainty for air separation units. Analysis of California market data shows that frequency regulation offers significantly greater revenue opportunities, but also requires more intimate integration with the grid (Dowling et al., 2017). Resources providing FR capacity must adjust their energy production/demands every 4 seconds in accordance with grid dispatch. As such, this market opportunity has received significantly less attention as it requires careful consideration of system dynamics. Dowling and Zavala (2017) formulated FR capacity allocation as a dynamic optimization problem,

$$
\begin{aligned}
&\min \int_0^{t_f} \phi\big(z(t), u(t), d(t)\big) dt \\
&\text{s.t.,} \quad \dot{z} = f(z(t), u(t), d(t)) \\
&\qquad g(z(t), y(t), d(t)) \leq 0 \\
&\qquad y(t) = h(z(t)) \\
&\qquad d(t) = r_+ \beta_+(t) - r_- \beta_-(t) \\
&\qquad 0 \leq r_+, r_-
\end{aligned}
\tag{1}
$$

where $z(t)$ represents the states, $y(t)$ represents the observations, $u(t)$ represents the controls, and $d(t)$ represents the disturbances. The (non)linear functions $f(\cdot)$, $g(\cdot)$, and $h(\cdot)$ capture the system dynamics, operating restrictions, and transformation from states to observations, respectively. In this formulation, the dispatch signals coming from the grid for regulation up and down are encoded in β_+(t) and β_-(t), respectively, and are treated as disturbances. The economic objective is often $\phi = r_+\pi_+ + r_-\pi_-$ where r_+ and r_- are the committed regulation up and down capacities during an hour and π_+ and π_- are the respective capacity prices. Thus, problem (1) seeks to maximize the revenue from frequency regulation capacity (i.e., maximizes the size of the disturbance) while determining a control strategy that maintains operational feasibility. This formulation is a generalization of Fares et al. (2014), which focuses exclusively on grid-scale redox flow batteries. Other emerging FR market participants include aluminium smelters (Zhang and Hug, 2014) and commercial HVAC systems (Lin et al., 2015).

2. Mathematical Models and Software Implementation

Many industrial sites have on-site utility plants that can directly participate in electricity markets. In this context, providing FR capacity requires either modulating electricity

demands of equipment such as pumps, arc furnaces, etc. (Zhang and Hug, 2014) or modulating steam production rates. Using the general formulation (1), Dowling and Zavala (2017) characterized revenue opportunities for a methanol-water distillation system providing FR by adjusting the energy input (reboiler steam flow rate). A key limitation of this analysis is that it relies on the simple two-input two-output transfer function model from Wood and Berry (1973) for a laboratory-scale system. Instead, the present work combines optimization formulation (1) with more detailed differential algebraic equation models. Thus, we seek to extend the results to industrial scale systems and explore how design decisions (e.g., holdup capacities) impact system dynamics and thus revenue opportunities. We start by considering the dynamic model from Hahn and Edgar (2002) adapted to a methanol-water separation with 8 trays:

Total condenser (tray 1): $\frac{dx_1}{dt} = \frac{1}{A_{cond}} V(y_2 - x_1)$ (2)

Rectifying trays (i = 2, …, 4): $\frac{dx_i}{dt} = \frac{1}{A_{tray}} [L_r(x_{i-1} - x_i) - V(y_i - y_{i+1})]$ (3)

Feed tray (i = 5): $\frac{dx_i}{dt} = \frac{1}{A_{tray}} [F\, x_f + L_r\, x_{i-1} - L_s\, x_i + V(y_i - y_{i+1})]$ (4)

Stripping trays (i = 6, …, 9): $\frac{dx_i}{dt} = \frac{1}{A_{tray}} [L_s(x_{i-1} - x_i) - V(y_i - y_{i+1})]$ (5)

Partial reboiler (tray 10): $\frac{dx_i}{dt} = \frac{1}{A_{reb}} [L_s\, x_{i-1} - (F - D)x_i - V\, y_i]$ (6)

Condenser mass balance: $V = L_r + D$ (7)

Saturated liquid feed quality: $L_s = L_r + F$ (8)

Reflux ratio: $R = \frac{L_r}{D}$ (9)

Constant relative volatility (i = 2, …, 10): $\alpha = \frac{y_i(1-x_i)}{(1-y_i)\, x_i}$ (10)

Reboiler energy balance: $Q = \Delta H_{vap} V$ (11)

where index *i* indicates the tray, $x(t), y(t)$ are mole fractions of methanol, F and x_f are the constant feed flow rate and composition, $L_r(t)$ and $L_s(t)$ are the liquid flow rates in the rectifying and stripping sections, respectively, $V(t)$ is the vapor flow rate, $D(t)$ is the distillate product flow rate, $R(t)$ is the reflux ratio, $Q(t)$ is the reboiler duty, and ΔH_{vap} is the constant heat of vaporization (40 kJ/mol). This model assumes constant total molar holdups ($A_{cond} = 0.5,\ A_{tray} = 0.25,\ A_{reb} = 1.0$ mol), constant molar overflow, and constant relative volatility ($\alpha = 4.02$). Eqns (2) – (8) arise from mass balances, Eqn (9) is a definition, and Eqn (10) captures vapor-liquid equilibrium. Eqns (2) – (10) are from Hahn and Edgar (2002), whereas Eqn (11) was added to calculate the reboiler energy input. Similar to Dowling and Zavala (2017), we use the reboiler duty to track the regulation signal as follows:

$$Q(t) = \beta_+(t)\, r_+ - \beta_-(t)\, r_- + \bar{Q} \quad (12)$$

where $Q(t)$ is the time-varying reboiler duty, $\beta_+(t)$ and $\beta_-(t)$ are the regulation tracking signals (input data), r_+ and r_- are the regulation capacity commitments (decision-variables) and $\bar{Q}$ is the reboiler duty offset (decision-variable). For most industrial sites, the distillation system would not be bidding by itself into the market, but instead the electricity generation of the onsite utility system would be modulated to track the FR signal. This would cause variations in the steam production flowrate, which would be absorbed by the flexibility of several steam consumers. The present study focuses on the distillation system and does not consider dynamics of the utility plant. Thus Eqn (12) is a simplification and corresponds with an electrically heated reboiler. Using operational

bounds on the distillate $0.92 \leq y_1(t) \leq 1.0$, bottoms $0.0 \leq x_{10}(t) \leq 0.02$, and reflux ratio $0.1 \leq R(t) \leq \rho$ and Eqns (2) – (12), we express the FR market participation problem in the form of (1). Unit conversions are omitted for brevity.

For each hour of data, we solve a nonlinear program (NLP) to determine the regulation capacity commitments, $\boldsymbol{r_+, r_-}$, and optimal reflux ratio control profile, $\boldsymbol{R(t)}$, that maximizes FR market revenue while satisfying path constraints (i.e., product quality bounds). Input data were obtained from the ERCOT website[1] and include the regulation capacity prices, $\boldsymbol{\pi_+, \pi_-}$, which are set in 1-hour intervals and the regulation dispatch signals, $\boldsymbol{\beta_+(t), \beta_-(t)}$, which are available in 4-second intervals. These models are implemented in Pyomo and Pyomo.DAE. Pyomo is an algebraic modeling language built on top of the Python programming language and Pyomo.DAE is an extension to Pyomo for representing and solving dynamic optimization problems. Moreover, Pyomo.DAE allows us to directly implement the derivatives and differential equations in Eqns (2) – (12), exploit DAE simulation tools for initialization, and automate the transformation of the continuous-time model into a discrete approximation. For this particular problem, we choose a backwards Euler discretization (although several other discretization schemes, including collocation over finite elements, are also natively supported in Pyomo.DAE) with a 4-second time step to match the resolution of $\boldsymbol{\beta_+(t), \beta_-(t)}$. A 1-hour horizon resulted in a NLP with approximately 32,000 variables (with bounds) and equality constraints. IPOPT (Wächter and Biegler, 2006**)** was used to solve these problems, requiring up to 5 minutes per instance.

3. Results and Discussion

The previously described NLP was solved for each hour of March 1st, 2016 using historical market data $\pi_+, \pi_-, \beta_+(t), \beta_-(t)$ as input parameters. Figure 1 shows the time-evolution of the bottoms composition, distillate composition, and reflux ratio. For the bottoms composition and the reflux ratio, the upper bounds are active during the first 1-hour horizon.

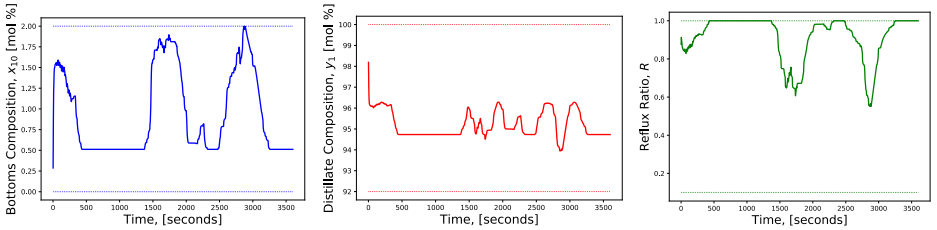


Figure 1. Composition of the bottoms (left), composition of the distillate (center), and reflux ratio (right) as a function of time with $\rho = 1$. Dashed lines indicate operational bounds.

A key advantage of model Eqn (2) – (12) and formulation (1) is the ability to explicitly connect design decisions with revenue opportunities from FR markets. Interestingly, we find that increasing the reboiler holdup A_{reb} from 1.0 mol to 10.0 mol only increases revenues by 3 %. This trend makes sense as larger holdups give rise to slower system dynamics which further increases the ability to reject fast market signals. We also find that revenues are extremely sensitivity to the maximum reflux ratio ρ as shown in Figure 2.

[1] http://www.ercot.com/mktrules/pilots/frrs

Furthermore, realizing the revenues for $\rho = 10$ requires dramatic changes in the reflux ratio, which may not be acceptable to operators. This is because of the constant holdup and constant molar overflow assumptions in the system model. The vapor flow rate V and thus reboiler duty Q are tightly coupled with the reflux ratio R, and thus ρ acts as an implicit bound on Q and thus regulation capacity. Furthermore, the chosen composition operating bounds are such that the upper bound on R limits the FR capacity and not the system dynamics. Again, this explains why revenues weakly depends on A_{reb}. These insights are not available from the transfer function model previously analysed.

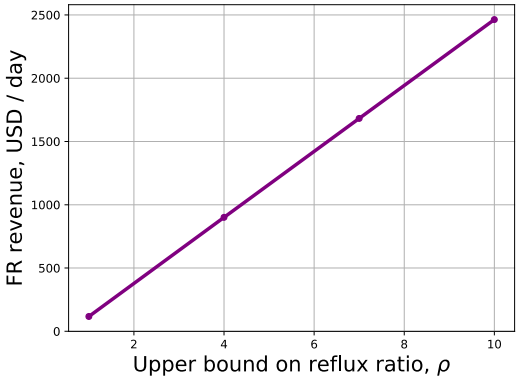


Figure 2. Sensitivity analysis of revenue with respect to the upper bound for reflux ratio (ρ). Revenue is scaled for nominal reboiler duty of 1 MW.

4. Conclusions and Future Directions

Opportunities in FR markets are emerging for chemical industrial systems. Their inherently slow dynamics filter out high-frequency disturbances from market signals. This paper expands upon a dynamic optimization framework that explicitly calculates revenue opportunities from historic market data and dynamic process models. We believe that the presented modelling and optimization tools are an essential first step to understanding and exploiting revenue opportunities ranging from \$100 to \$2,500 / day for a 1 MW (nominal) system. In the presented distillation case study, sensitivity analysis is used to understand how design decisions impact system dynamics and market revenues. In particular, we find that column diameter (which limits reflux ratio) is much more important than reboiler holdup capacity. Pyomo and Pyomo.DAE facilitate rapid model development, debugging, initialization, and automation of data-driven analysis, which has historically been difficult for dynamic optimization with algebraic modelling languages such as GAMS, AMPL, AIMMS, etc.

As future work, we plan to consider more detailed dynamic models with variable material holdups, such as those from López-Negrete et al (2013). We hypothesize that the current case study underestimates the flexibility and thus revenue opportunities of distillation systems because holdup capacities are fixed. We anticipate that it is possible to further increase the range of feasible energy inputs by strategically controlling holdups levels. An industrial FR capacity provider would likely exploit flexibility from several sources. We plan to investigate other equipment (e.g., pumps) as well as plant-wide coordination schemes. Moreover, the present analysis assumes perfect information and may overestimate realistic market revenues. As future work, we plan to characterize the impact

of market uncertainty on chemical manufacturing systems, and anticipate PySP, a separate Pyomo extension for stochastic programming, will be especially helpful (Watson et al, 2012). Finally, we plan to expand the case study to balance new market revenue against operating costs from more dynamic operation such as equipment wear-and-tear.

References

A. W. Dowling, R. Kumar, V. M. Zavala (2017). A multi-scale optimization framework for electricity market participation. Applied Energy, 190, pp. 147-164.

A. W. Dowling, V. M. Zavala (2017). Economic opportunities for industrial systems from frequency regulation markets. Computers & Chemical Engineering, in press. 10.1016/j.compchemeng.2017.09.018

R. L. Fares, J. P. Meyers, M. E. Webber (2014). A dynamic model-based estimate of the value of a vanadium redox flow battery for frequency regulation in Texas. Applied Energy, 113, pp. 189 – 198.

J. Hahn, T. F. Edgar (2002). An improved method for nonlinear model reduction using balancing of empirical gramians. Computers & Chemical Engineering 26, pp. 1379-1397.

W. E. Hart, J. P. Watson, D. L. Woodruff (2011). Pyomo: modeling and solving mathematical programs in python. Mathematical Programming Computation, vol. 3, no. 3, pp. 219–260.

Y. Lin, P. Barooah, S. Meyn, T. Middelkoop (2015). Experimental Evaluation of Frequency Regulation From Commercial Building HVAC Systems. IEEE Trans. Smart Grid, 6 (2) pp. 776 – 783.

R. López-Negrete, F. J. D'Amato, L. T. Biegler, A. Kumar (2013). Fast nonlinear model predictive control: Formulation and industrial process applications. Computers & Chemical Engineering, vol. 51, no. 0, pp. 55-64.

B. L. Nicholson, J. D. Siirola, J. P. Watson, V. M. Zavala, L. T. Biegler (2017). Pyomo.DAE: A modeling and automatic discretization framework for optimization with differential and algebraic equations. Mathematical Programming Computation, 10.1007/s12532-017-0127-0

A. Wächter, L.T. Biegler (2006). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Mathematical Programming, 106, pp 25-57.

J.-P. Watson, D. L. Woodruff, W. E. Hart (2012). PySP: modeling and solving stochastic programs in Python. Math. Prog. Comp., 4 (2) pp. 109-149.

R. K. Wood, M. W. Berry (1973). Terminal composition control of a binary distillation column. Chemical Engineering Science 28 (9), pp. 1707-1717.

Q. Zhang, I. E. Grossman (2016). Enterprise-wide optimization for industrial demand side management: Fundamentals, advances, and perspectives. Chemical Engineering Research and Design, 116, pp. 114-131.

Q. Zhang, I.E. Grossman, C. F. Heuberger, A. Sundaramoorthy, J. M. Pinto (2015). Air separation with cryogenic energy storage: Optimal scheduling considering electric energy and reserve markets. AIChE Journal 61 (5), pp. 1547-1558.

X. Zhang, G. Hug (2014). Optimal regulation provision by aluminum smelters. IEEE PES General Meeting – Conference Exposition, pp. 1 – 5.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50414-6

Design Support of Smart Energy Systems based on Locally Available Resources: A Case Study in Isolated Islands in Japan

Yasunori Kikuchi[a,b*], Kotaro Ouchida[b], Yuichiro Kanematsu[a], Tatsuya Okubo[a,b]

[a] *Presidential Endowed Chair for ''Platinum Society'', The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*
[b] *Department of Chemical System Engineering, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*
kikuchi@platinum.u-tokyo.ac.jp

Abstract

Unused renewable resources can be collected in some regions as, for example, agricultural or forestry residues. To corroborate such locally available resources with energy systems mainly composed of fossil resources, technologies for their sourcing, transferring, storing, and operating should be systematically selected and integrated into a system. In this paper, we tackled with the design support of smart energy systems, where locally available resources are utilized and collaborated with fossil-based conventional systems. To demonstrate the design of energy system, a case study in isolated islands in Japan was conducted, where the power grids are not connected with other islands and can be regarded as microgrids. The solutions for the design of regional energy systems are generated through the consideration of local availability of renewable resources. As the technologies utilizing local resources, photovoltaic, biomass-derived combined heating and power, and biomass-derived boilers were taken into account, and existing developed models were utilized in this paper for the analysis of designed smart energy systems. The design results demonstrated that the utilization of locally available biomass can become strong options for the isolated islands whereas strategic reconstruction of the industrial cooperating networks should be incorporated. Bridging the gaps of the objective functions and constraints among existing industries should be addressed by computer-aided visualization and simulation as a design support.

Keywords: Micro grid, Bagasse, Timber, combined heating and power.

1. Introduction

The design of energy systems has become an issue in all over the world. A single optimal system cannot be suggested, because the availability of infrastructure and resources should be considered locally. Unused renewable resources can be collected in some regions as, for example, agricultural or forestry residues. For example, the utilization of forest biomass as energy source has become practical in some regions in European Union (EU), where forest biomass supplies over 50 % of renewable energies (Bentsen and Felby 2012). To collaborate such locally available resources with energy systems mainly composed of fossil resources, technologies for their sourcing, transferring, storing, and operating (Kato et al., 2016) should be systematically selected and integrated into a system. Simulation-based design support is necessitated for specifying possible solutions

considering multiple requirements on energy such as low carbon resources, high stability, and high economic performance.

In this study, we tackled with the design support of smart energy systems, where locally available resources are utilized and collaborated with fossil-based conventional systems. To demonstrate the design of energy system, a case study in isolated islands in Japan is conducted, where the power grids are not connected with other islands and can be regarded as microgrids. The solutions for the design of regional energy systems are generated through the consideration of local availability of renewable resources. As the technologies utilizing local resources, photovoltaic, biomass-derived combined heating and power (CHP), and biomass-derived boilers are taken into account, and existing developed models on cane sugar mills (Ouchida et al., 2017) and woody biomass (Kanematsu et al., 2017a) are utilized in this paper for the analysis of designed smart energy systems.

2. Materials and method

2.1. Concepts for smart energy systems in isolated islands

The meaning of "smart" energy systems has been discussed with diversities on their sourcing, operability, resilience, security, and responses for demands (Kikuchi, 2017a). The common premises can be regarded as the utilization of renewable resources which are usually treated as less stability, controllability, and usability than fossil resources. Energy systems should be able to equip somehow redundancy with reasonable socioeconomic costs to accept the fluctuation of renewable resources. Especially in isolated islands where industries have difficulties in procurements of fossil fuels, the utilization of locally available resources such as the residues of agriculture and forestry has been desired by local governments and industries.

In Tanegashima, an isolated island in Japan, a cane sugar mill is operated to produce raw sugar, where a bagasse-based CHP plant is utilized and a plenty of unused energy is generated (Kikuchi et al., 2016). In other words, the bagasse can provide energy not only for sugar milling but also other energy demand. For example, the processing potato factories for Shochu or starch needs around 100 to 130 °C which is not high temperature heat and can be supplied by the excess low-pressure steam in the sugar mill (Kikuchi et al., 2016). The developing thermal energy storage and transport system can also be employed to decrease the fossil fuel utilized in such factories (Fujii et al., 2016). At the same time, woody biomass can also supply the heat with such temperature. Because appropriate forestation and harvesting has been needed for sustainability in ecosystem (Oosawa et al., 2016), the timber demand for fuel is one of the option. The cane sugar mill is only operated for five months each year, from December to April, the harvesting season for sugarcane. The expanded use of bagasse-based CHP that burns excess bagasse and/or forest biomass in off-season of sugar production can be considered. The installation of a decentralized biomass CHP is also possible as the option for energy systems.

2.2. Case examples of design problem

Adequate resource allocation should be addressed for integrated design of multiple energy systems due to the limitation of resource within one region. Design support of smart energy system should be able to make a consensus on such allocation based on some evaluation indicators such as environmental impacts (Kanematsu et al., 2017c) and

social values (Kikuchi et al., 2017b). In this paper, case examples on forest resources were shown for demonstrating the design problems on locally available resources. Forest resource availability should be estimated based on the long-term forest reformation against the maldistribution of age-class (Oosawa et al., 2016). Two CHP plants were taken into account in this case study: in-house CHP at cane sugar mill (CHP1) and decentralized CHP located at city center (CHP2). CHP1 has the specific capacity for meeting the heat and power demands at cane sugar mill, e.g., 1,750 kW_e in Tanegashima (Kikuchi et al., 2016). On the other hand, the CHP2 at city center can be adjusted for the timber availability by adopting small packaged CHP units, i.e., 40 kW_e power + 70 kW_{th} heat per unit, and their parallelization (Kanematsu et al., 2017a). The power is sold for the micro grid, while the heat is only utilized for the demand at cane sugar mill (Kikuchi et al., 2016) in CHP1 and connected buildings around CHP2, which was estimated by the data conversion of investigated actual heat demands in Tanegashima as shown in Figure 1. The greenhouse gas emission from the life cycle (LC-GHG) of energy facilities was evaluated in this paper.

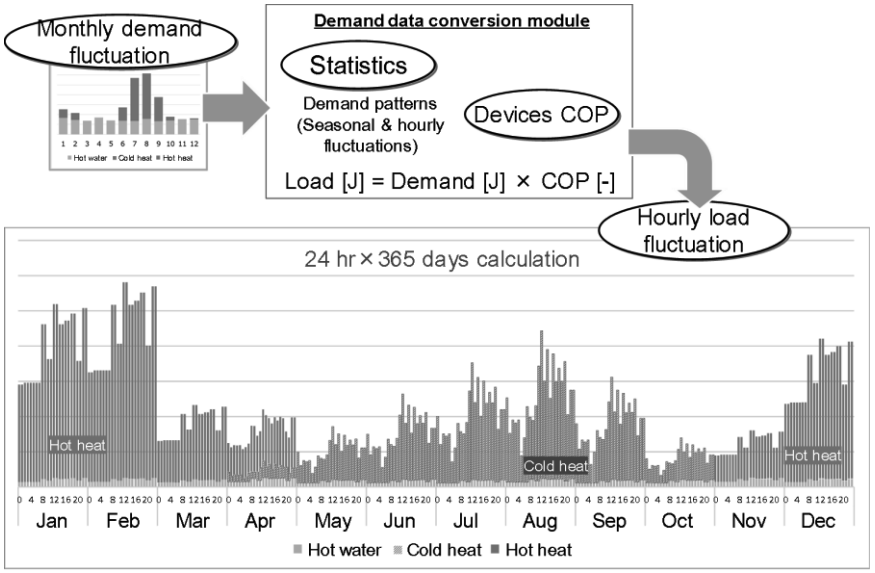


Figure 1 Data conversion of monthly demand of heat in Tanegashima, where the monthly data was investigated from actual buildings. The coefficient of performance (COP) of installed devices was also obtained from the investigation.

3. Results and discussion

3.1. LC-GHG emission reduction by timber utilization

Figure 2 shows the LC-GHG results for the forest resource allocation in Tanegashima. The LC-GHG can be reduced by the installing decentralized small-packaged CHP units in CHP2, which can replace both of heat and power around CHP2. The fuel needs for heating were larger in CHP2 than CHP1. While such local energy demand should be addressed in the design of CHP, the power can replace the fossil fuel burned at the centralized power plants by CHP1 and CHP2. As the LC-GHG of micro grid has relatively higher than that of power grids in the other areas of Japan, the locally available

resources have an advantage in isolated islands rather than those in the other areas. Thus, the increase of small-packaged CHP units resulted in the decrease of LC-GHG because their power conversion efficiency is higher than that of CHP1 which utilize back pressure steam turbine. To pay back the LC-GHG induced by the forestry to keep the reforestation at Tanegashima, ten, nine, and eight units of CHPs are required in CHP2 in three resource availability cases, i.e., Figure 2 (a), (b), and (c), respectively.

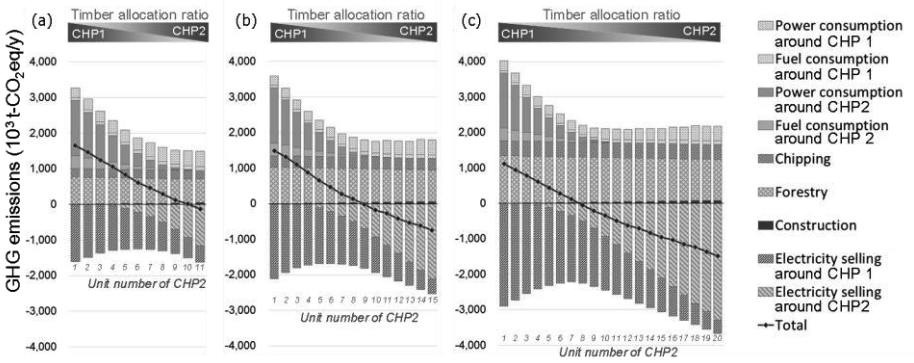


Figure 2 LC-GHG based on the forest resource allocation in Tanegashima to CHP1 and CHP2, where three patterns of resource availability were considered: (a) 15,700 m^3-timber/y, 21,100 m^3-timber/y, and 27,600 m^3-timber/y. Although GHG emitted from fossil resources was considered in LC-GHG, that from the feedstock of biomass was not included on this graph.

3.2. Design support needs for decision makers

The resource allocation should be addressed on the basis of the consensus among related decision makers. Figure 3 schematically shows the related decision makers for design problems on locally available forest resources (Kanematsu et al., 2017b). For the forestry workers, the sustainable forest management is an emerging issue to keep the forest in Japan (Oosawa et al., 2016), while the stable wood demand should be guaranteed to start the dynamic forestry plans to mitigate the maldistribution of age-class of trees. For the energy suppliers, the stable fuel supply plan is strongly needed before constructing energy plants. On that meaning, CHP1 is preferable to start the biomass-derived energy supply because CHP1 is existing plant for cane sugar mill. Based on Figure 2, however, the CHP2 has larger emission reduction than CHP1 partly because it can adjust the capacity of heating and power generation and has higher power conversion efficiency.

3.3. Scenario planning for energy systems in a region

Design support of smart energy system at local systems should be able to visualize the benefits and risks on technology implementation. Cooperative plans are strongly needed for obtaining the locally available resources, otherwise the misunderstandings among players in energy systems could be occurred and resulted in the failure of design. Computer-aided approach can visualize the scenarios of technology implementation. The number of installed small-packaged CHP units in CHP2 shown in Figure 2 can become useful in scenario planning with forestry revitalization.

The phase of forestry management for supplying timber should be monitored and taken into account in the introduction of CHP devices. At that time, the larger unit number of

CHP2 can reduce more GHG emission as shown in Figure 2. To enhance the profitability of energy company of CHP2, the fuel wood from the outside of Tanegashima can be transferred/imported and utilized in CHP2. Because the transferred/imported fuels cannot circulate the domestic economy and support domestic forestry management, their amount should be gradually decreased along with the increase of domestic timber supply. The scenario should be planned with such consideration on the local efforts on the resource circulation.

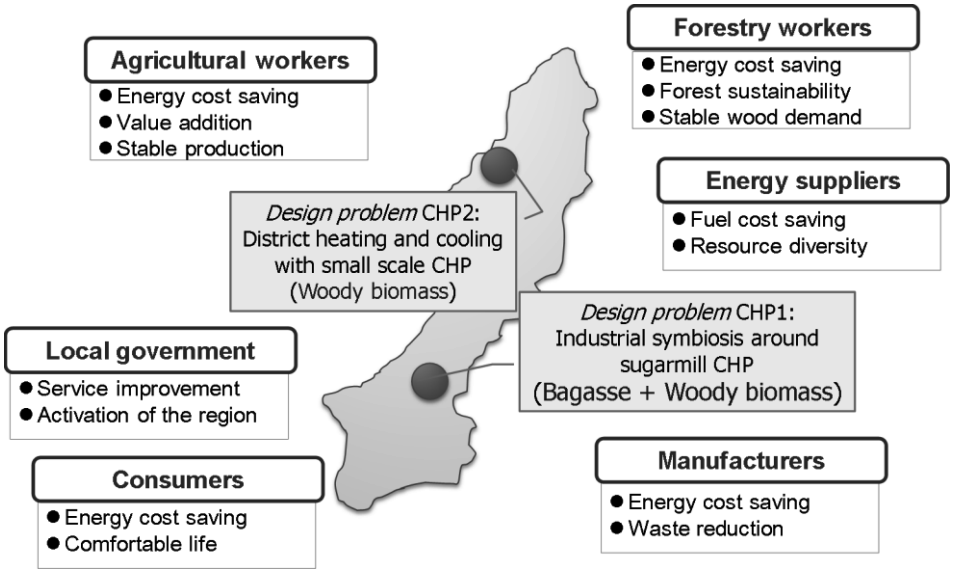


Figure 3 Related decision makers for design problems on locally available biomass

4. Conclusions

The local industries and ecosystems should be considered for designing "smart" energy systems. The utilization of locally available resources should involve the potential players in the region such as manufacturers, farmers, forestry workers, and local governments. The resource availability is significantly related with their cooperation. At that time, computer-aided approach can be employed as the visualization of the benefits and risks of technology implementation. For the deep decarbonization, such technology-oriented analysis is not enough and sociotechnical analysis should be conducted (Geels et al., 2017). The visualization of the benefits and risks is definitely required steps for system design to reveal the behaviours and performances of designed systems. Additionally, actual design support for smart energy systems need the roles of communication with related stakeholders through computer-aided design tools.

Acknowledgements

The authors also acknowledge the support given by Shinko Sugar Mill Company, Nishinoomote-shi, Nakatane-cho, Minamitane-cho, and the Tanegashima Forestry Cooperative. Part of this study was financially supported by the Ministry of Agriculture, Forestry and Fisheries of Japan (Rural Biomass Research Project, BUM-Cm6500), the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number 16H06126

(Young Scientists A), 15H01750 (Scientific Research A), and 26285080 (Scientific Research B). Activities of the Presidential Endowed Chair for “Platinum Society” in the University of Tokyo are supported by the KAITEKI Institute Incorporated, Mitsui Fudosan Corporation, Shin-Etsu Chemical Co., ORIX Corporation, Sekisui House, Ltd., and East Japan Railway Company.

References

N. S. Bentsen and C. Felby, 2012. Biomass for energy in the European Union - a review of bioenergy resource assessments. Biotechnology for biofuels, 5, 1, p.25.

S. Fujii, Y. Kanematsu, Y. Kikuchi, T. Nakagaki, 2016, Material and heat flow analysis in thermal energy storage and transport system utilizing unused heat from bagasse boiler, Mechanical Engineering Journal, 3(5), 16-00334

F.W. Geels, B.K. Sovacool, T. Schwanen, S. Sorrell, 2017, Sociotechnical transitions for deep decarbonization. Science, 357(6357), 1242-1244.

Y. Kanematsu, K. Oosawa, T. Okubo, Y. Kikuchi, 2017a, Designing the scale of a woody biomass CHP considering local forestry reformation: A case study of Tanegashima, Japan, Applied Energy, 198, 160-172.

Y. Kanematsu, T. Okubo, Y. Kikuchi, 2017b, Activity and data models of planning processes for industrial symbiosis in rural areas, KagakuKogakuRonbunshu, 43(5) 347-357

Y. Kanematsu, K. Oosawa, T. Okubo, Y. Kikuchi, 2017c, A design of rural energy system by industrial symbiosis considering availability of regional resources, Computer-Aided Chemical Engineering, 40, 1987-1992

Y. Kato, M. Koyama, Y. Fukushima, T. Nakagaki (Edit.), 2016, "Energy Technology Roadmaps of Japan; Future Energy Systems Based on Feasible Technologies Beyond 2030", Springer: Tokyo

Y. Kikuchi, 2017a, Simulation-based approaches for design of smart energy system: A review applying bibliometric analysis, Journal of Chemical Engineering of Japan, 50(6) 385-396

Y. Kikuchi, M. Nakai, K. Oosawa, Y. Kanematsu, K. Ouchida, T. Okubo, 2017b, A computer-aided socio-technical analysis on national and regional energy systems considering local availability of renewable resources, Computer-Aided Chemical Engineering, 40, 2485-2490

Y. Kikuchi, Y. Kanematsu, M. Ugo, Y. Hamada, T. Okubo, 2016, Industrial Symbiosis Centered on a Regional Cogeneration Power Plant Utilizing Available Local Resources: A Case Study of Tanegashima, Journal of Industrial Ecology, 20(2) 276-288.

Y. Kikuchi, S. Kimura, Y. Okamoto, M. Koyama, 2014, A scenario analysis of future energy systems based on an energy flow model represented as functionals of technology options, Applied Energy, 132 586-601.

K. Oosawa, Y. Kanematsu, Y. Kikuchi, 2016, Forestry and Wood Industry, in Y. Kato, M. Koyama, Y. Fukushima, T. Nakagaki (Edit.), 2016, "Energy Technology Roadmaps of Japan; Future Energy Systems Based on Feasible Technologies Beyond 2030", Springer: Tokyo

K. Ouchida, Y. Fukushima, S. Ohara, A. Sugimoto, M. Hirao, and Y. Kikuchi, 2017, Integrated design of agricultural and industrial processes: A case study of combined sugar and ethanol production. AIChE Journal, 63(2), 560-581.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50415-8

A Mathematical Programming Approach to Optimal Design of Smart Distributed Energy Systems

Evgenia Mechleri* and Harvey Arellano-Garcia

Department of Chemical and Process Engineering, University of Surrey, Guildford, United Kingdom
e.mechleri@surrey.ac.uk

Abstract

The UK is committed to reducing its greenhouse gas emissions by at least 80% by 2050, relative to 1990 levels. For this to happen, we need to transform the UK economy while ensuring secure, low-carbon energy supplies to 2050. The future electricity distribution system, known as smart grid, will integrate advanced digital meters, distribution automation, communication systems and distributed energy resources. There has been a lot of discussion about the importance of the Internet of Things (**IoT**) in future smart grids and smart cities stating that IoT offers many applications and can be used to integrate efficiency renewable energy sources in the smart grid by making the electricity grid more robust and scalable.

This study will focus on the development of an integrated IoT-Distributed energy systems (**DES**) model for the efficient energy management of a microgrid under the integration of the intermittent renewable energy resources. In this work, we expand the definition of flexible options to include demand and supply together with design and operation strategies using internet of things (IoT). Our framework brings weather data and sensor information into a virtual energy plant optimisation model that connects supplier and consumer to optimise potential flexibility gaps arising from supply and demand mismatch. The problem is posed as a hybrid mixed-integer linear programming (MILP) optimisation model combining flexibility analysis and optimal synthesis for integrating energy supply and demand, where environmental information is added to each stage. Finally, we combine traditional mathematical programming approaches such as flexibility analysis and optimal network synthesis and within a single optimisation framework combining IoT and urban DES.

Keywords: Distributed energy systems (DES), Internet of things (IoT), mixed-integer linear programming (MILP).

1. Introduction

In the last century, there has been an overwhelming rural-urban migration, with the movement of people from the countryside to the city, this has resulted in urban growth and urbanisation. Consequently, and along with the modernisation of everyday life, the demand for energy has, since then, been increasing greatly (Giordano et al., 2011). In parallel, the way energy has been produced over the years, and still is today, with a great use of fossil fuels, has contributed largely to high emission rates of carbon into the atmosphere (Muslikhin et al., 2016).

The power industry needs to rearrange the energy supply chain distribution to not depend solely on the processes currently used in the energy sector, which is likely to face a collapse shortly. In a whole, the energy sector requires the optimisation of all four stages of the supply chain: generation, transmission, distribution and consumption (Wang et al., 2014).

The power grid consists of a vast interconnected framework infrastructure with the main purpose of distributing electricity from power plants to consumers. Recently, even though information and control frontiers have been through significant changes, the advance of power systems was not able to keep up with technology development and innovation. In this context along with the need for a fully upgraded energy utilisation, distribution, transmission and generation, an idea of the smart grid has emerged as the power grid system new key factor (Federal Energy Regulatory Commission 2012). Of the most important concepts for the success of this new way of energy distribution, is the idea of a "prosumer", a user that when connected to the power grid is able to produce or consume energy within the concept of the smart grid. This would be made possible by smart meters installed at the user's' place, which consists of a device that characterises the Smart grid is and is connected to the power utility, determining a two-way communication between them. In order to achieve this communication, the idea of integrating IoT with the smart grid has emerged. The IoT is a new communication and network paradigm in which a variety of items or objects become an integral part of the Internet. In other words, the objects become "smart objects" which are equipped with microprocessors and transceivers, which makes them able to communicate with each other and provide intelligent services to users autonomously (Bui et al., 2012). The IoT has been realised as one of the most promising networking paradigms that bridge the gap between the cyber and physical world.

Several studies have already been done encompassing IoT applied in homes, buildings, cars, health and cities. A model structure of optimisation of energy based in IoT and realised in a "green technology" has been proposed by Huang et al. (Huang et al., 2014). Kim et al., (Kim et al., 2015) proposed an IoT-based home energy management system considering a home area with your appliances and photovoltaic boards as distributed energy sources. Although there has been a lot of discussion around the idea of IoT and how it can be deployed in smart-grids, only a small number of research efforts have practically addressed the integration of smart-grids with IoT. The main aim of this work is to develop an optimisation model for a virtual energy plant that connects supplier and consumer using IoT and urban Distributed Energy Systems. In order to provide an optimal design for a distributed energy system and fill the gaps that may arise from supply and demand discrepancy, the focus of the research will be on the integration between the power grid with environmental information, energy storage technologies and renewable energy sources using IoT.

The remainder of this paper is formulated as follows: In section 2 the power grid with its power flow equations, the energy sources and the loads are mathematically modelled, and the objective function of the optimisation problem is deduced, in section 3 a case study of a microgrid is presented and section 4 summarises conclusions and future research directions.

2. Methodology

In order to optimise a distributed urban energy system, the demand response problem will be formulated with an approach with convex optimization. First, the power flow equations will be defined, then the system with all the utility and devices of a distributed energy system.

2.1. Power flow equations

In order to characterize the power flow on a Power grid, a diagram for the branches, buses (nodes) and connection between them and the ground is used, as presented in Figure 1. The y_{jk} denotes for a branch linking bus on position j to bus k, the admittance value y_k for the element at bus k and I_k for the current injection on bus bar k.

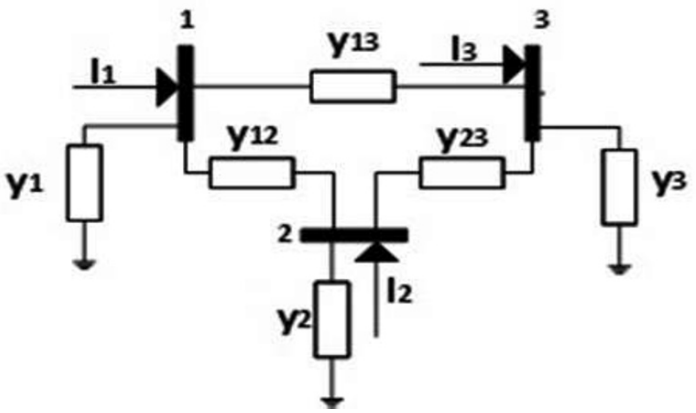


Figure 1. Electrical system transmission diagram

2.2. Mathematical system model

In this subsection the mathematical system model for building up the energy management system is presented. It is assumed a distributed energy system (DES) with heterogeneous types of energy resources where there is an Energy Management System (EMS) responsible for interact with all the active devices to the DES. The EMS acts at an operational level, gathering all the data collected by the devices on the buses, smart meters and power controls\monitor on the energy source and with all the data given acts as a controller of the power energy sources. Consider an urban energy grid with an arrangement of distributed generators units denoted by $G = \{g_1, g_2, g_3, \dots, g_g\}$, and distributed storage units denoted by $B = \{b_1, b_2, b_3, \dots b_b\}$, and buses indexed as N by $i = \{1, \dots, n\}$ and controllable loads indicated by $L = \{l_1, l_2, l_3, \dots, l_l\}$. The system is assumed to operate in discrete time, where the time is represented by $t \in \{0, 1, 2, \dots\}$ and time interval Δt.

We consider two kinds of Distributed Generator (DG) units in the grid: non- dispatchable renewable distributed generator units denoted by $g \in G_r$ and dispatchable conventional distributed generator units denoted by $g \in G_c$. For each DG, its complex power output is denoted by $s_r(t) = p_r(t) + iq_r(t)$ and $s_g(t) = p_g(t) + iq_g(t)$ for renewable energy source and conventional generator unit respectively. The active and reactive power are

$p_g(t)\ and\ q_g(t)$. Similarly we have modelled the traditional distributed generation fossil fuel motor unit, the renewable distributed generator taking into account photovoltaic units and wind turbines and also the battery system model and export to the grid which are not presented due to space limitations.

The demand requests from the user are assumed to be unknown and stochastic. The demand is given to the EMS by the user on each time t. It is considered a demand side management to the grid just about the production of energy. The load demand has a fixed value on time t. The customers loads need to be supplied interely for the grid. Where the complex power is denoted by $s_l(t) = p_l(t) + iq_l(t)$.

The Objective function of the system is defined as $C(t)$. The total cost of the power used on the system is the sum of the cost of the different decentralised units, the price of energy acquisition from the grid and the energy lost on the system.

$$C(t) = \sum_{g\in G} C_g(p_g(t)) + \sum_{b\in B} C_b(p_b(t)) + C_o(t, p_o(t) + \sum_{rg\in R} C_{rg}\left(p_{rg}(t)\right) + \varphi \sum_{(i,j)\in E} r_{ij} l_{ij}(t) \quad (1)$$

where φ is a coefficient to penalize the power lost, C_g, C_b, C_o and C_{rg} are the generation cost of traditional DG unit, of the battery, the price of the energy acquisition from the grid and the cost of the renewable generator, $p_b(t)$ is the active power of the battery, $p_o(t)$ can be negative or positive. $p_o(t)$ positive means that the urban grid is buying power from the external grid, $p_o(t)$ negative means that the urban grid is selling power to the main grid, $p_{rg}(t)$ is the power output of the renewable energy unit.
The objective of the model and energy management system (EMS) is to minimise the DES operational cost, while satisfying the power loads.

3. Results and discussion

Figure 2, presents displays the layout of a real distributed energy system in Tohoku Area, Japan. This DES consists of four power generation sources and one battery bank working as a storage energy unit. The energy sources are divided in two groups of diesel combustion conventional motor generator connected to bus 1 and bus 9 (output power of 6000KW and 1000KW respectively); two renewable generation unit being one of the solar energy (maximum output power 1000KW) and other wind energy (maximum output power possible 3000KW) connected to the bus 13 and 14 respectively. The battery bank has a maximum capacity of storage 500KW, and it is connected to the bus 7.
This is consideredan online DES, and it is connected to the external grid (national grid). The load on each bus corresponds the maximum values demanded by the users connected to that bus. The day starts at 12 am in the simulation. Because the disponibility of data, the time interval Δt is comprehended in intervals of one hour. The day is denoted by T= {1,2, 3…,24}, where each t denotes the hour T.

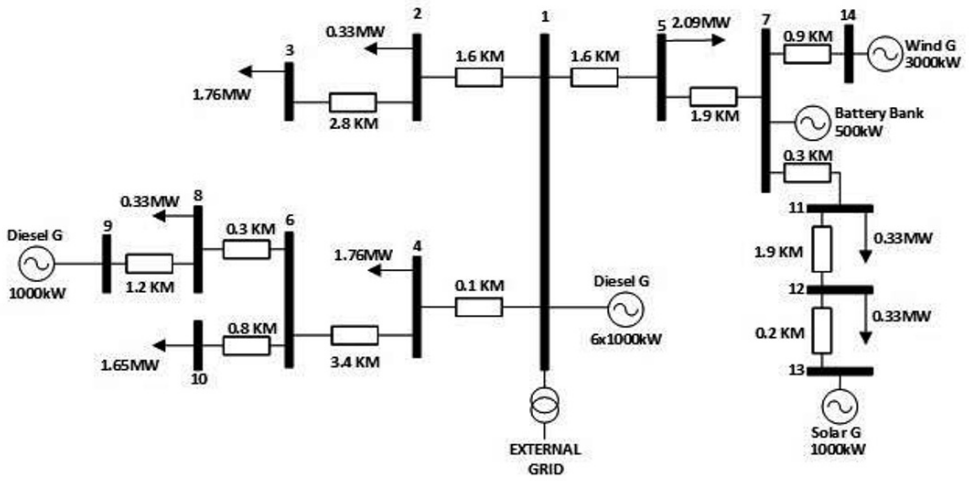


Figure 2. Distributed energy system-case study

The mathematical model described in chapter 3 with the considerations cited above was submitted and run on a commercial Software. A computer Intel core i5, 2GB RAM was responsible for running. The time spent on the activity to run for an interval T= {1,2,3,…,24} is 12 minutes and 35 seconds. Figure 3, shows the results provided by the mathematical model for an interval of 24 hours.

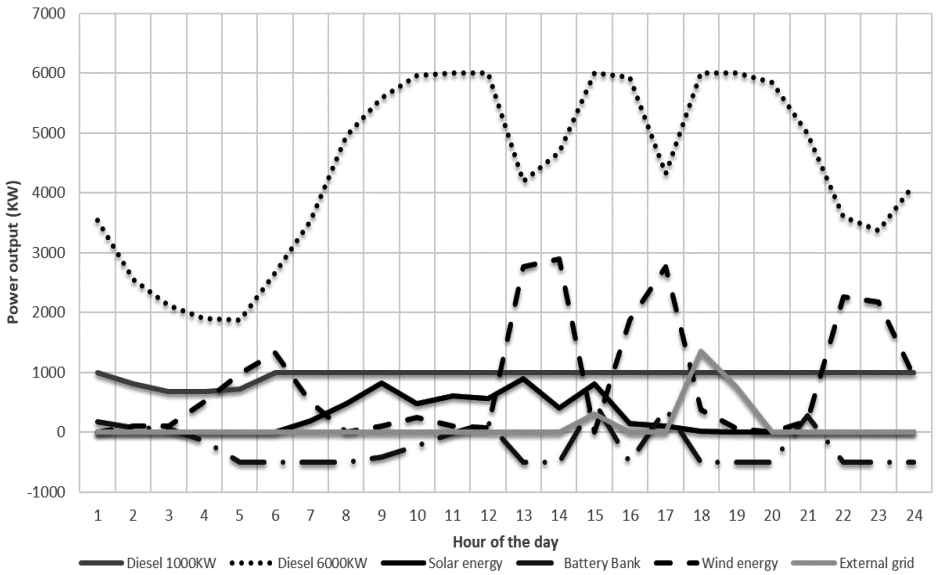


Figure 3. Power output of DES in one-day period

The wind energy and solar energy source are used most of the time for power production. In relation to the battery bank, the values show the energy level of the battery bank with positive and negative values. When the values are positive the battery is discharging and injecting power to the grid, and when negative the battery is charging. It is important to mention that the model´s objective function finds a better solution looking for all the

interval of 24 hours, so it is better for the system to keep the energy battery storage and then discharge on interval time above when the renewable energy is not available (i.e. hour of the day 14 and 20). As expected, the electricity is purchased from the external grid only when the load demand is superior of the DES internal production and when this happens all the DES production is utilised. In relation to the DG diesel source, the results show that the optimal solution is always scheduling the energy's sources to afford the load on the bus near the sources. (i.e. Hour of the day between 2 and 5).

4. Conclusions

A distributed energy system was studied addressing the demand response issue. A mathematical model was formulated to describe the power flow relationship between the power grid utilities and the power production of renewable energy sources and storage on a virtual power network. The final mathematical model was simulated with a real case scenario of a distributed energy system. This study has followed an approach to model a distributed energy system with an IoT structure, considering real-time data and the The IoT characteristic of collecting real-time data and capacity of command transmission will be an important support on a future smart power grid.

References

Bui, N., Castellani, A., Casari, P., & Zorzi, M. (2012). The internet of energy: A web- enabled smart grid system. Network, IEEE, 26(4), IEEE Network, July-August 2012, Vol.26(4).

Federal Energy Regulatory Commission Staff Team, "Assessment of demand response and advanced metering," Fed. Energy Regul. Commission, Tech. Rep., 2012

Giordano V, Gangale F, Fulli G, Jiménez MS (2011) Smart Grid projects in Europe. JRC Reference Reports by the Joint Research Centre of the European Commission.

Jun Huang, Yu Meng, Xuehong Gong, Yanbing Liu, & Qiang Duan. (2014). A Novel Deployment Scheme for Green Internet of Things. Internet of Things Journal, IEEE, 1(2), 196-205.

Kim, J., Byun, J., Jeong, D., Choi, M., Kang, B., & Park, S. (2015). An IoT-Based Home Energy Management System over Dynamic Home Area Networks. International Journal of Distributed Sensor Networks, 11(10), 1-15.

Muslikhin E, Marpanaji M, IzzuddinMahali M. Internet of things for survey of renewable energy potential (SREP) as the basis for hybrid power plant development. MATEC Web of Conferences. 2016;54, DOI: 10.1051/matecconf/20165406003.

Yu Wang RM, Shiwen Mao RM, Nelms RM. Distributed online algorithm for optimal real-time energy distribution in the smart grid. Internet of Things Journal, IEEE. 2014;1(1):70-80.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50416-X

Dynamic Real-Time Optimization of Air-Conditioning Systems in Residential Houses with a Battery Energy Storage under Different Electricity Pricing Structures

Moataz N. Sheha, Kody M. Powell*

University of Utah, Department of Chemical Engineering, 50 S Central Campus Drive, Salt Lake City, UT 84112, United States of America
kody.powell@utah.edu

Abstract

This paper investigates potential cost savings in operating residential houses air-conditioning systems through dynamic real-time optimization (D-RTO). Standard design data collected from BEopt (Building Energy Optimization) software were used in Matlab/Simulink to simulate cooling energy consumption of a house model in Salt Lake City, Utah. Two different electricity pricing structures were employed; time-of-use (TOU) and real-time pricing (RTP). The D-RTO was formulated as a linear programming problem with the air conditioning temperature setpoint, the cooling energy, the battery state of charge, the charging energy and discharging energy of the battery being decision variables of the model.

The D-RTO determines day-ahead values for the dynamic variables of the system based on the weather forecast and energy price signals. The D-RTO uses a model predictive control (MPC) like approach in which the problem is solved for 24 hours in advance, but the solutions are implemented on a receding horizon, where the prediction interval moves forward by one hour each time step. Results show significant energy cost reductions for the optimized cases with a battery energy storage versus the cases without a battery under each pricing structure. Also, both pricing structures are ranked based on their capabilities for cost and energy savings and peak shifting.

Keywords: Dynamic real-time optimization; Smart Grid; Matlab/Simulink; Battery energy storage; Air conditioning systems.

1. Introduction

Air conditioning systems in residential houses consume big amounts of electricity and is usually accompanied with peak demand issues. Because of this, the residential sector has enormous potential to be a key player in the future of grid management [1]. Thus, optimizing the use of air conditioning systems is a topic that earned global interest.

Different modeling techniques have been implemented through the years to study the effect of optimizing the electricity consumption through air conditioning systems [1]–[5]. Passive thermal energy storage might not be enough to achieve the purpose of cost reduction and peak shifting under variable energy prices, thus, incorporating an energy storage system (e.g. a battery) is something that needs to be studied.

This paper presents a simple model for optimizing the cooling energy cost in residential houses through dynamic real-time optimization (D-RTO) incorporating a battery energy storage system under two different electricity pricing structures; time-of-use (TOU), and real-time pricing (RTP). The work demonstrates that dynamic optimization with a battery storage system can decrease the energy cost and shift the peak energy consumption periods towards off-peak price hours better than optimization without an energy storage system.

The rest of this paper is organized as follows. Section 2 present an overview of the system. Section 3 discusses the formulation of the model. Section 4 is devoted to the results of the study. Section 5 presents the conclusion.

2. Overview of the System

The system modelled in this study consists of four main elements. The first element is the heat transfer between the house and the surroundings to simulate the indoor temperature of the house. TMY3 (Typical Metrological Year, version 3) data set for Salt Lake City, Utah, were used to represent the weather in the surroundings of the house.

The second element is the optimizer used to minimize the cost of electricity for the house under two different electricity pricing structures; Time of Use (TOU), and Real Time Pricing (RTP).

The third element is the battery model and the fourth element is the air conditioning system represented by a PI controller that adjusts the required cooling energy based on the temperature setpoint obtained from the optimizer every hour. The formulation of the system is discussed in details in section III.

3. Model Formulation

3.1. Heat Transfer between the House and the Surroundings

Modeling the heat transfer through the boundaries of a house requires knowledge of the properties of the elements of the house involved in this process (e.g. thermal resistance and heat capacity). These elements include; the exterior walls of the house, windows, door, ceiling, floor slab, and all the interiors of the house (e.g. furniture, kitchen appliances, etc…).

In order to simulate realistic data, standard properties for the walls, windows, door, ceiling, and floor slab of the house were obtained from BEopt (Building Energy Optimization) software to represent a standard design for the house. The TMY3 weather file for Salt Lake City, Utah, were obtained from BEopt as well.

The following 1st order ODE was used to simulate the heat transfer through the boundaries of the house:

$$A\frac{dT}{dt} = B(T_{amb} - T_{house}) - Q + C \quad (1)$$

Where $\frac{dT}{dt}$ is the rate of change of indoor temperature with time in K/s, T_{amb} is the ambient temperature in K, and T_{house} is the indoor temperature of the house in K.

"A" is a constant which accounts for all the thermal masses within a house including the thermal mass of air, exterior walls, ceiling, interiors, and floor slab. All the heat capacities of the aforementioned elements were assumed to be constant over the operating temperature range. "A" has units of J/^{o}C.

"B" can be defined as: $B = \sum_{i=1}^{n} U_i A_i$, where U_i is the overall heat transfer coefficient of "i" and A_i is the heat transfer area relevant to "i". "B" accounts for the heat transfer coefficients and heat transfer areas of the exterior walls, windows, door, ceiling, and a fitting term to account for the neglected items (e.g. heat transfer through the floor slab driven by the ground temperature which was assumed to have a minor effect). "B" has units of W/^{o}C.

"Q" represents the cooling energy required per unit time and has units of Watts and "C" is a fitting parameter to account for any other neglected sources of heat (e.g. heat transfer through floor slab and heat radiation from the sun). C has units of Watts.

The proposed heat transfer model has been validated against the results of BEopt to make sure that the values for indoor temperature and the cooling energy obtained from Matlab/Simulink model and BEopt at every time step are close enough to each other.

3.2. Cost Optimization

The cost optimization problem was formulated as a linear programming problem with 120 decision variables; the temperature setpoint of the air conditioning system within each house for 24 hours, the cooling energy required for each of them for 24 hours, the battery state of charge for 24 hours, the charging energy for the battery for 24 hours, and the discharging energy of the battery for 24 hours. The temperature setpoint and the battery state of charge does not appear in the objective function but appears in the constraints.

The problem solves for the temperature setpoint and the cooling energy for 24 hours in advance, but the solutions are implemented on a receding horizon, where the prediction interval moves forward by one hour each time step. This is because there is no guarantee that the indoor temperature will reach the setpoint every hour, so the prediction needs to be updated continuously to account for that. The objective function is given by the following form:

$$min \sum_{i=1}^{24} EP_i \times (CEC_i + P_{C_i} - P_{DC_i}) \tag{2}$$

Where EP_i is the electricity price at "i" in \$/kWh, CEC_i is the cooling electricity consumption at hour "i" in kWh, P_{C_i} is the charging energy of the battery at hour "i", and P_{DC_i} is the discharging energy of the battery at hour "i".

The optimization problem is subject to two equality constraints at every hour during the 24 hour period which makes them a system of 48 equality constraints. 24 of these equality constraints are a numerically integrated form of (1) using Backward Euler's method for every hour and has the following form:

$$(A + \Delta tB)T_{i+1} - AT_i + \Delta tQ_{i+1} = \Delta tBT_{amb} + \Delta tC \tag{3}$$

Where T_{i+1} is the temperature setpoint at time "i+1", Q_{i+1} is the cooling energy at time "i+1", T_i is the temperature setpoint at time "i" (i.e. the initial condition for time "i+1), Δt is the time step, and A, B, C are the same parameters defined in (1). The cooling electricity consumption is related to the cooling energy consumption by the following relation:

$$CEC = \frac{Q}{SEER} \tag{4}$$

Where SEER is the Seasonal Energy Efficiency Ratio defined as the cooling output divided by the electric energy input. Each air conditioning system employed in this study has its own SEER rating.

The other 24 equality constraints are a numerically integrated form of the battery function. The battery function is given as follows:

$$\frac{d(SOC)}{dt} = \eta_C P_C - \frac{P_{DC}}{\eta_{DC}} \tag{5}$$

Where $\frac{d(SOC)}{dt}$ is the rate of change of the state of charge of the battery with time in kWh per unit time. η_C and η_{DC} are the charging and discharging efficiencies of the battery, respectively. Consequently, the 24 constraints for the battery each is a numerically integrated form of (5) using Backward Euler's method at every hour and has the following form:

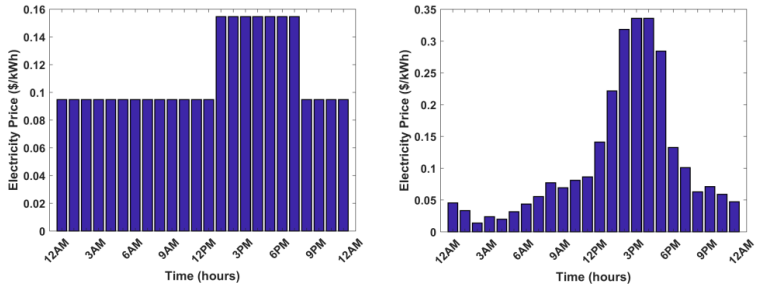
$$SOC_{i+1} - SOC_i + \frac{\Delta t}{\eta_{DC}} P_{DC} - \Delta t \eta_C P_C = 0 \tag{6}$$

Each of the 120 decision variables is subject to a lower and an upper bound, where the temperature setpoint for each house is subject to $T_{min} < T_i < T_{max}$, the cooling energy is subject to $0 < Q_i < Q_{max}$, the state of charge is subject to $0 < SOC_i < SOC_{max}$, the battery charging energy is subject to $0 < P_{C_i} < P_{C_{max}}$ and the battery discharging energy is subject to $0 < P_{DC_i} < P_{DC_{max}}$. For the house simulated in this study, a maximum state of charge of 1.5 kWh was used [6].The operating range of the temperature setpoint

expands during unoccupied periods. It was assumed that half the house was unoccupied from 8 AM till 6 PM.

Two different electricity pricing structures were used and both were built on the basis of an average daily price of 0.1122 $/kWh. The pricing structures used are defined below:

- Time of Use (TOU): Where electricity prices usually have 2 or 3 different values per day depending on the time of the day. In this study, 2 different prices were employed every day. The average electricity price for residential houses in Utah is 0.1111 $/kWh [7]. Peak hours in Utah during summer are from 1 to 8 PM, the price of electricity during the off-peak hours reduces by 0.016334 $/kWh and increases during peak hours by 0.04356 $/kWh [8]. Using the average value and the peak and off-peak modifications, the TOU price was taken as 0.0947 $/kWh for the off-peak hours and 0.1546 $/kWh for the peak hours. This would result in an average price per day of 0.1122 $/kWh. The TOU pricing structure that was used is presented in Figure 1a.
- Real-Time Pricing (RTP): where electricity price varies every hour during the day based on the electricity consumption. In this study, we employed 24 different electricity prices one for each hour and used the same values through the entire period of summer. The used RTP pricing structure is depicted in Figure 1b.



(a) TOU Pricing Structure (b) RTP Pricing Structure

Figure 1: Employed TOU and RTP Pricing Structures

3.3. Air Conditioning System

This paper studies only the cooling systems of the houses, nothing related to the heating system was included because the house is assumed to be gas heated. The cooling system was formulated as feedback PI controller. The controller takes the house indoor temperature and the setpoint as inputs and then adjusts the cooling energy of the house (output) based on how far is the indoor temperature from the temperature setpoint.

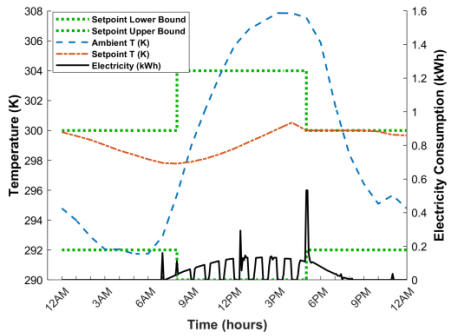
The system was solved dynamically in a real-time optimization (D-RTO) form using Matlab/Simulink. D-RTO uses a model predictive control (MPC) like approach in which the problem is solved for 24 hours in advance, but the solutions are implemented on a receding horizon, where the prediction interval moves forward by one hour each time step.

4. Results

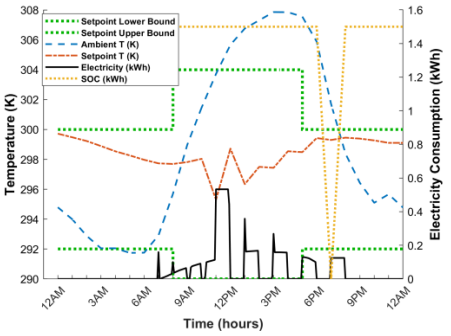
Figure 2a through 2d illustrate the results for the two pricing structures; TOU (2a and 2b) and RTP (2c and 2d), respectively, by presenting hourly cooling energy consumption for the house in a hot summer day in August vs. the temperature setpoint and the ambient temperature. The state of charge for the cases with a battery storage system is plotted as well. The electricity consumption is plotted over intervals of five minutes for a 24 hour period. In Figure 2a, it can be noticed that the peak electricity consumption is at 5 PM, which is during the peak hours of TOU (Figure 1a). This means that the optimizer resulted

in a peak consumption during a peak hour. So, what happened here is that the optimizer could not shift the peak energy consumption outside the peak hours. Instead, the optimizer tried to maintain low energy values during the peak hours, but eventually, the temperature setpoint became very high with the ambient temperature being high as well, therefore, the setpoint dropped around one degree resulting in a peak energy consumption during the peak hours of TOU. This is because the thermal masses of the house were not enough to store sufficient amount of energy to optimize the consumption. In Figure 2b, where a battery is being used as an energy storage system, the system is optimized and the peak consumption is shifted towards an off-peak hour for TOU (11 AM). Notice the plot of the state of charge of the battery where the battery discharges at the last peak hour preventing the air conditioning system from consuming more energy from the grid.

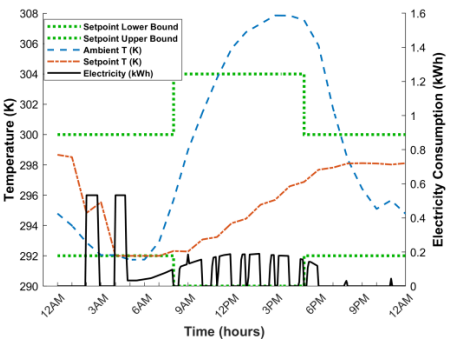
In Figures 2c and 2d, it can be illustrated that RTP was able to optimize the system in Figure 2c without a battery and had no peak consumptions during peak hours (Figure 1b). But still Figure 2d shows that using a battery storage system resulted in decreasing the peak consumption at 2 AM and 4 AM.



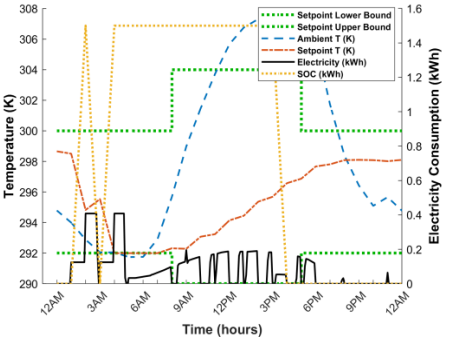
(a) Setpoint and cooling energy for TOU without a battery



(b) Setpoint and cooling energy for TOU with a battery



(a) Setpoint and cooling energy for RTP without a battery



(b) Setpoint and cooling energy for RTP with a battery

Figure 2: Setpoint and cooling energy for a 24 hour period for all cases

Table 1 shows the total cost, total consumption and percentage savings for all cases. An interesting observation from Table 1 is that however RTP without a battery was enough to optimize the system efficiently, it still resulted in higher energy consumption and higher price than TOU without a battery which was not optimized efficiently. The same observation can be drawn for the cases with a battery storage system. This means that TOU pricing structure with a battery storage system shall be the favoured operating strategy over the other studied cases.

Table 1: Total Energy Cost and Total Energy Consumption for All Cases during Summer Period (May through September)

Case	Energy Cost ($)	Energy Usage (kWh)	Energy Cost Savings (%)	Energy Usage Savings (%)
TOU without battery	106.1	996.3	-	-
TOU with battery	88.6	886	16.5	11.07
RTP without battery	214	1,752.2	-	-
RTP with battery	182.9	1,590.8	14.5	9.21

5. Conclusion

Dynamic real-time optimization of air conditioning systems is necessary for peak shifting purposes as well for reducing electrical energy consumption and electrical energy cost. Two different electricity pricing structures were studied. TOU pricing structure and RTP pricing structure both with and without a battery energy storage system. Results showed that TOU pricing structure with a battery storage system was able to shift the peak towards off peak hours and resulted in more energy and cost reduction than RTP pricing structure.

In future work, distributed generation (e.g. solar panels) will be incorporated to the system to study the effect of distributed generation on the optimization of air conditioning systems in residential houses with the presence of a battery energy storage system and for grid frequency regulation.

References

[1] W. J. Cole, K. M. Powell, E. T. Hale, and T. F. Edgar, "Reduced-order residential home modeling for model predictive control," *Energy Build.*, vol. 74, pp. 69–77, 2014.

[2] W. J. Cole, J. D. Rhodes, W. Gorman, K. X. Perez, M. E. Webber, and T. F. Edgar, "Community-scale residential air conditioning control for effective grid management," *Appl. Energy*, vol. 130, pp. 428–436, 2014.

[3] A. G. Thomas *et al.*, "Intelligent residential air-conditioning system with smart-grid functionality," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 2240–2251, 2012.

[4] K. M. Tsui, S. C. Chan, G. Observation, R. Algorithms, S. F. P. Gas, and E. Holding, "Demand response optimization for smart home scheduling under real-time pricing," *IEEE Trans. Smart Grid*, vol. 4, no. 8, pp. 1812–1821, 2002.

[5] T. Hubert and S. Grijalva, "Modeling for Residential Electricity Optimization in Dynamic Pricing Environments," vol. 3, no. 4, pp. 2224–2231, 2012.

[6] "Humless Home Mini | Solar Home Battery Backup," 2017. [Online]. Available: https://humless.com/home-series/mini/. [Accessed: 03-Feb-2018].

[7] "Residential Price Comparison," 2017. [Online]. Available: https://www.rockymountainpower.net/about/rar/rpc.html. [Accessed: 24-Sep-2017].

[8] "Time of Day," 2017. [Online]. Available: https://www.rockymountainpower.net/ya/po/otou.html. [Accessed: 24-Sep-2017].

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50417-1

Maximum groundwater level for urban development: Evaluation of different calculation methods in Western Australia

Tara Zirakbash[a*], Anastasia Boronina[b], Martin Anda[a], Parisa A. Bahri[a]

[a] *Murdoch University, 90 South St, Murdoch WA 6150, Australia*
[b] *GHD, 999 Hay St, Perth WA 6000, Australia*
T.Zirakbash@Murdoch.edu.au

Abstract

This paper closely evaluates three conventional methods of calculating the maximum groundwater level to be used for design purposes in Perth, Western Australia, in order to identify the shortcomings. The results of all of the methods are incorporated into ArcGIS in a study area in Perth for the interpretation of groundwater changes in calculated levels and to assess their reliability. The range of differences in results by application of the methods is presented at monitoring locations and other places over the study area.

A comparison of the methods shows that, despite similar principles in calculation, the results are very different, thus creating uncertainty for decision making and risks for urban structures, people and the environment. The results vary as a consequence of different density and types of data used in each method, the inconsistent period of data selection, impact of anthropogenic activities and climate change, as well as the physical properties of the aquifer, such as slope. In areas of high groundwater levels and in major engineering projects, it is critical to more accurately calculate the maximum groundwater levels.

Therefore, a combination of detailed data analysis in order to estimate maximum groundwater level accurately at monitoring locations at certain times is proposed, along with using a numerical simulation of physical processes to reduce and define uncertainty for planning. This can pave the way towards the development of a reliable method for the calculation of maximum groundwater levels.

Keywords: Maximum groundwater level, urban development, groundwater contours.

1. Introduction

Urban development impacts groundwater levels and, in turn, groundwater level variations may result in major risks to infrastructure and the environment, which can cause economic losses. For example, subsidence, inundation of road surfaces, drying surface water bodies or promoting insect breeding are all possible consequences of variation in groundwater levels.

Urbanisation in Perth, Western Australia (WA) causes a higher runoff and recharge rate and, subsequently, groundwater levels rise (Appleyard, 1995; Barron, et al., 2013). Despite the drying climate, in Perth, groundwater levels near the surface (3-10m) are not as affected by a decline in rainfall due to the impact of urbanization (McFarlane et al., 2012). In order to avoid the adverse impact of high groundwater level below the road and building structure, various processes are put in place such as constructing drainage, importing sand fill, and using water tolerant designs or materials.

Furthermore, there is a compromise between the need to reduce costs, whilst still ensuring the protection of infrastructure and groundwater users, both human and environmental (C. G. Davies et al., 2017). Inaccuracies in defining the groundwater level may result in unacceptable risks or over conservatism in the construction of urban structures that will potentially cause inequitable lot fill, costly design and materials (Boronina, 2015).

Groundwater planning and design in urban development depends on maximum groundwater levels. However, until now, no standardized method exists to determine maximum groundwater levels (Boronina et al, 2015; Fürst et al., 2015).

2. Methods

Three methods that have been used in WA to determine maximum groundwater levels are evaluated to assess the reliability of results. All of these methods are based on interpolation, which is a common method for groundwater level assessments.

2.1. Perth Groundwater Atlas (PGA)

Perth Groundwater Atlas is an online spatial dataset provided by the WA Government to help drillers, irrigators, and private householders to establish groundwater bores in the water table aquifer within the Perth metropolitan area. The "Groundwater Contours Historical Maximum" layer is 1-meter interval groundwater contours based on measured maximum groundwater level records from the early 1950's through to the 1970's, which were hand drawn interpolations from 1: 2,000 map sheets. Due to the fact that the contours are not related to any particular period of groundwater level records and because of changes in groundwater and natural surface levels that can occur over time, these contours are indicative only, providing merely a regional overview.

2.2. Average Annual Maximum Groundwater Levels (AAMGL)

The use of the AAMGL has been advocated by the WA Government for new drainage works, road construction and pavement design (J. Davies et al., 2004). There is no explicit rule in respect to calculating AAMGL in the WA guidelines, and anomalies occur when this method is used by different practitioners. Usually, in order to calculate AAMGL at a site, onsite groundwater monitoring is undertaken for 12 to 18 months to capture maximum groundwater level. This short-term data is then modified with long-term AAMGL at a regional groundwater monitoring bore close to the site. The difficulty in defining AAMGL is discussed by Davies et al (2004), which is related to undefined period of data over which the AAMGL shall be calculated, the method of correlation with long term data to a particular site and lack of consideration for existing drainage systems.

2.3. Design Groundwater Level (DGWL) methodology for infrastructure projects

The Gateway WA project is a road project undertaken by Main Roads Western Australia. A new methodology was applied for the project, which is based on hydrogeology concepts and incorporates various types of data, such as groundwater levels in open drains or surface water bodies (Boronina et al, 2015). In this method, a year in the history that corresponds to a reasonable maximum is selected, with consideration of changes in groundwater levels owing to the construction of drainage. At all locations, maximum groundwater level for the same year was estimated and groundwater contours were created by interpolation of groundwater elevation from point data, with a clear level of confidence in monitoring locations based on the quality of the data.

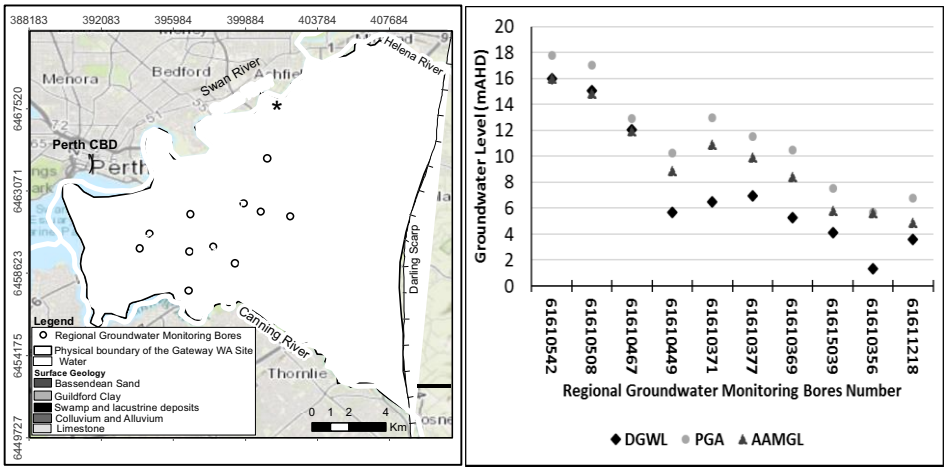


Figure 1 Surface geology and locations of regional groundwater monitoring bores (DWER)(left) and differences of methods at regional groundwater monitoring bore locations (right)

3. Study Area

The study site is related to the Gateway WA and it is located in the Perth basin, comprising a multi-layered unconfined aquifer, covering an area of 185 Km^2 (Figure 1, left). The site is bordered by Darling Scarp to the East, Helena River to the North-East, Swan River to the North and West and Canning River to the South.

Rainfall data in the Perth Airport station inside the study area shows a downward trend over the last few decades (from 900mm in 1945 to below 700mm in 2016).

The dominant 'Superficial formations' (late Tertiary to Quaternary sediments) in the study area is Bassendean Sand, comprising of quartz sand, and Guildford Clay, consisting of low permeability material and a perched water table (Davidson, 1995). The regional groundwater flow direction is from Darling Scarp at East towards the Swan River at West.

4. Data Collection and Analysis

Spatial data including "Groundwater Contours Historical Maximum" layer and all the groundwater monitoring bore data and information, is collected from the WA Government. The regional monitoring bores in the superficial aquifer with long-term records (>30 years), up to the present, are selected for further analysis (Figure 1).

At the construction time of the Gateway WA project, many bores were drilled in the superficial aquifer for monitoring. The water level data and groundwater contours created by the DGWL method for the Gateway project are provided by Main Roads WA. The water level measurements are predominantly during 2013 and 2014 and it is specified that in most locations annual maximum groundwater levels occurred in October.

For the AAMGL calculation of Gateway WA project bores, those with groundwater level measurement in Oct 2013 are selected and their maximum level is linearly modified with AAMGL of the regional monitoring bores with long-term records to estimate AAMGL at the locations of the selected bore at the construction site. Based on available data, these estimated water levels along with calculated AAMGL at the regional monitoring bores are used to generate groundwater contours.

Three groundwater contours resulted from the PGA, DGWL and AAMGL methods are integrated into ArcGIS. In order to standardize the contours, all of them are changed to raster data and subtracted from each other to understand their range of differences. They are also subtracted from the digital elevation model (DEM, a 2008 LIDAR survey) in order to determine the locations that show the groundwater level above the surface.

5. Results and Discussion

Although groundwater contours are all created based on interpolation between groundwater elevation at points inside the study area, Figure 2. (A) shows that, due to different levels of density of the points and different periods of selection used in each method, they bring about very different results and their accuracy is not clear.

The comparison of methods at monitoring locations is depicted in Figure 1, and bore number 61610542 is presented as an example in Figure 3. Maximum groundwater level calculated based on each method at this bore is shown on the groundwater level data (Figure 3). AAMGL value is AAMGL of measured data from 1983 to 2017 and maximum groundwater level values for the PGA and DGWL methods are extracted from their contours at the bore location. The same principle is applied for comparing maximum groundwater levels of three methods at monitoring locations, as shown in Figure 1 (right).

It can be concluded that, if the groundwater level trend is downward owing to the drier climate, drainage design or abstraction, and if long-term data is available (in this analysis, 33 years), there is rarely a risk of annual exceedance of groundwater level from the long-term AAMGL at monitoring points (Figure 3). Moreover, Figure 1 (right), reveals that, except for the bores which are used in the calculation of DGWL, at bores which are not used in the calculation, the DGWL method underestimates the maximum groundwater levels and PGA overestimates them in all bores. In accordance with the different period of selection, overestimation of maximum levels in the PGA method could originate from climate change and human intervention. Anthropogenic activities could have a big influence that should be taken into account. For example, there is about a one-meter decline in the groundwater level in the late 1960's at bore number 61610542, which is related to drain construction (Boronina et al, 2015). Underestimation of maximum level in DGWL at monitoring points farther away from the construction site could be originated from a lower density of data, along with different physical properties at those locations.

Surfaces made from groundwater elevations of each method subtracted from DEM of the area are presented in Figure 2 (B). According to the negative values which are representative of the area and the amplitude of the groundwater level above the ground, DGWL is the most efficient method (-3.7-0 m) and AAMGL resulted in the least reasonable values (-13.5-0 m). It is related to the density and types of data used in the creation of these surfaces and the fact that AAMGL is not corrected by topography.

A comparison of methods over the study area in Figure 2 (C) shows that the range of differences between methods is quite large (10 and 15m) which, for detailed designs, is unacceptable. The largest difference between the PGA and DGWL methods (5-8.2m) might be related to the change in topography, due to the steeper slope in this part.

The DGWL method for infrastructure projects has some advantages compared with PGA and AAMGL, which have brought about better results. For example, different types of data and a change in the groundwater level trend based on anthropogenic activities are considered in this method.

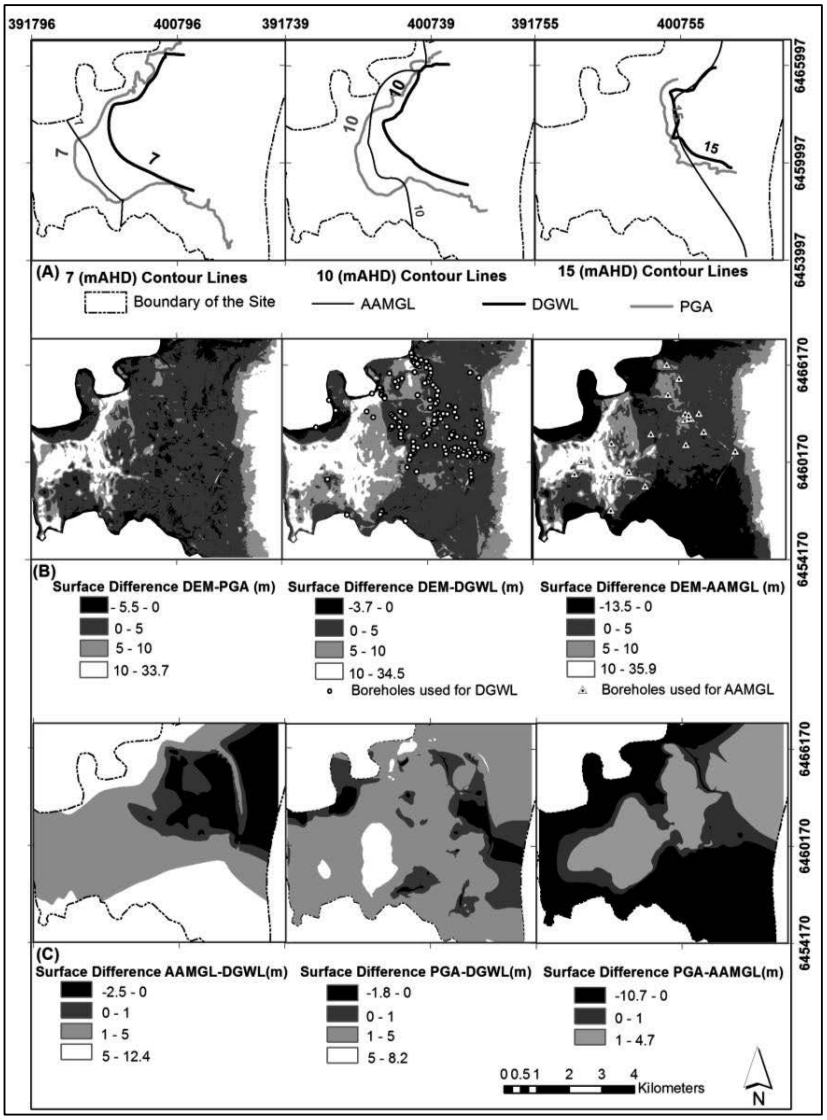


Figure 2 Groundwater contours created by different methods (A), Subtraction of surfaces generated from DEM (B), Surface difference resulted from subtraction of two methods (C).

However, interpolation doesn't account for physical processes and optimum bore density and focal locations for monitoring according to the site complexity is not clear.

Based on this comparison, a similar approach of the DGWL method that includes all types of data and an accurate estimation of maximum groundwater level at monitoring locations, in combination with numerical modelling to consider physical properties, is proposed. In numerical modelling, simulation at the steady state for a certain time derived from data analysis is suggested, with a focus on reducing the difference between the modelled level with calculated or measured levels at points, as well as estimating uncertainty for planning.

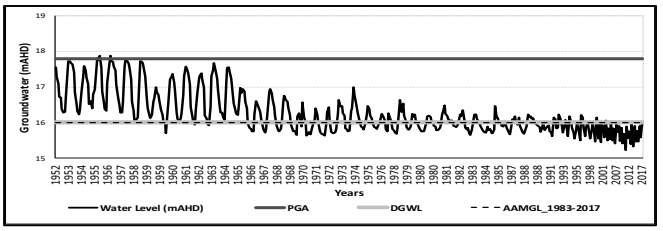


Figure 3 Groundwater level data and values calculated based on methods at bore 61610542

6. Conclusion

The comparison of three methods for the calculation of maximum groundwater level in WA shows that, in spite of a similar theory in the basis of interpolation, there is a significant spatial difference in their results. The DGWL that incorporates different types of data could present more reliable results spatially, and the AAMGL method, despite its high efficiency at monitoring points, could result in the least reliable surface. The greatest differences between results correspond with the different density of data, period of selection and the change in topography. For engineering projects and detailed designs, more accuracy in maximum groundwater level calculations is advised, as this could prevent damages and risks related to exceeding water levels from their allowable limits and a rise in costs as a result of the damages or conservatism in design. To achieve higher accuracy, a similar approach to the DGWL, along with a simulation of maximum level through numerical modelling to reduce uncertainty, is proposed.

7. Acknowledgement

The Authors gratefully acknowledge the Government of Western Australia, Department of Water and Environmental Regulation and Main Roads WA for providing data.

References

A. Boronina, R. Clayton, C. Gwynne, 2015, Recommended methodology for determination of design groundwater levels, Australian Geomechanics Journal, 50(3), 11-21.

C. G. Davies, R. Vogwill, C. Oldham, 2017, Urban subsurface drainage as an alternative water source in a drying climate, Australasian Journal of Water Resources, 1-12.

D. McFarlane, R. Stone, S. Martens, J. Thomas, R. Silberstein, R. Ali, G. Hodgson, 2012, Climate change impacts on water yields and demands in south-western Australia, Journal of Hydrology, 475, 488-498.

J. Davies, S. Martens, A. Rogers, 2004, Review of water table elevation on the Swan Coastal Plain and the AAMGL concept. Paper presented at the Institute of Public Works Engineering Australia (IPWEA) Western Australia Devision State Conference, Perth, Western Australia.

J. Fürst, A. Bichler, F. Konecny, 2015, Regional frequency analysis of extreme groundwater levels, Groundwater, 53(3), 414-423.

O. Barron, M. Donn, A. Barr, 2013, Urbanisation and shallow groundwater: predicting changes in catchment hydrological responses, Water resources management, 27(1), 95-115.

S. Appleyard, 1995, The impact of urban development on recharge and groundwater quality in a coastal aquifer near Perth, Western Australia, Hydrogeology journal, 3(2), 65-75.

W. A. Davidson, 1995, Hydrogeology and Groundwater Resources of the Perth Region, Western Australia. Perth: Western Australia Geological Survey.

Mario R. Eden, Marianthi Ierapetritou and Gavin P. Towler (Editors) Proceedings of the 13th International Symposium on Process Systems Engineering – PSE 2018
July 1-5, 2018, San Diego, California, USA
https://doi.org/10.1016/B978-0-444-64241-7.50418-3

Extracting Heuristics for Designing Sustainable Built Environments by Coupling Multiobjective Evolutionary Optimization and Machine Learning

Xinyu Liu[a], Bhavik R. Bakshi[a*]

[a]*William G. Lowrie Department of Chemical and Biomolecular Engineering, The Ohio State University, Columbus OH 43210, USA*
bakshi.2@osu.edu

Abstract

In sustainable process design, computational optimization methods are often used to find a single or a set of Pareto optimal or near-optimal solutions. However, knowledge about what makes the solutions optimal can be more valuable as it can be formulated as heuristics for guiding future designs of similar systems. In this paper, we propose a methodology that couples multiobjective optimization and machine learning to extract design heuristics. The methodology has been demonstrated by a case study on sustainable residential system design. Potential heuristics for sustainable buildings include: 1) keep the house as warm as comfortable in the summer and as cold as comfortable in the winter; 2) place vegetable garden, shading trees, and rain water harvesting system around the house. This further indicates the benefits of the synergistic design between technologies and ecosystems. The approach of combining optimization with learning is general enough to be applied to other systems, ranging from local manufacturing processes to urban systems.

Keywords: Heuristic extraction; multiobjective optimization; machine learning; sustainable built environment

1. Introduction

In sustainable process design (SPD), it is common to use multiobjective optimization (MOO) for understanding the trade-offs between various objectives and obtaining Pareto curves. However, what may be more valuable is to extract knowledge about what makes the solutions optimal, which can result in design heuristics. The motivation for finding such heuristics is because solving optimization problem can be expensive in terms of both computational power and human expertise. Therefore, instead of repetitively performing similar tasks, a more practical approach would be to discover design heuristics that can better explain optimality and guide the future design of similar systems without detailed optimization. To obtain such insights, additional analyses are required to go beyond just computing the optima (Deb and Srinivasan, 2008).

Conventionally, heuristics are obtained through human experience. For example, we can develop distillation sequences based on previous experience (Seider et al., 2017), look at points on Pareto curves and search for common characteristics (Urban and Bakshi, 2013), or represent different problem identities in closed-form mathematical expressions (Deb and Srinivasan, 2008). However, these approaches might become intractable when facing large dataset or complex problems. Heuristics extraction has also been a research topic in the field of machine learning (ML) and knowledge

discovery. There are two approaches for heuristics extraction in ML. The first is to apply a symbolic learning algorithm such as decision trees directly; while the second is to extract tree-structured representations of trained models (Al Iqbal, 2012).

On the other hand, most SPD problems ignore the role played by nature in supporting engineering systems. Similar to technological systems, ecosystems also have limits in terms of goods and services available. Ignoring these limits while designing systems may result in unintended harm, including ecological degradation. Methods for making decisions toward environmental sustainability need to encourage satisfaction of two necessary but not sufficient conditions: 1) goods and services demanded from ecosystem should not exceed the ecological carrying capacity; 2) the proposed solution should not shift the problem outside the system boundary. The TES-LCA methodology explicitly accounts for the capacity of ecosystems to provide the goods and services demanded in the life cycle across multiple spatial scales (Liu and Bakshi, 2017).

In this work, we extract heuristics from sustainable designs of the built environment. The rest of this paper is organized as follows. A methodology for heuristic extraction by coupling optimization outputs with ML is proposed in Section 2. This methodology is demonstrated by a case study on sustainable residential system design via TES-LCA methodology.

2. Methodology

Our systematic approach to extract design heuristics couples optimization outcomes and ML algorithms. As indicated in Figure 1, multiobjective optimization can be used to obtain not only the Pareto front, but also determine whether a solution is optimal or not, along with its associated decision variables and objectives values. This information about decision variables that correspond to optimal and non-optimal solutions can be the input to ML for the extraction of knowledge about what makes the solutions optimal. This can be formulated either as a classification or a regression problem. In the classification context, solutions are labelled either as optimal or non-optimal; while in the regression context, solutions are labelled by their objectives values. Decision variables are identified as heuristics if they contribute to a more thorough classification or explain more variability of objective values.

3. Case Study

3.1. System Description

The methodology has been applied to the design of a single-family residential system located in Columbus, Ohio. The house is assumed to be 185 m^2 with no basement

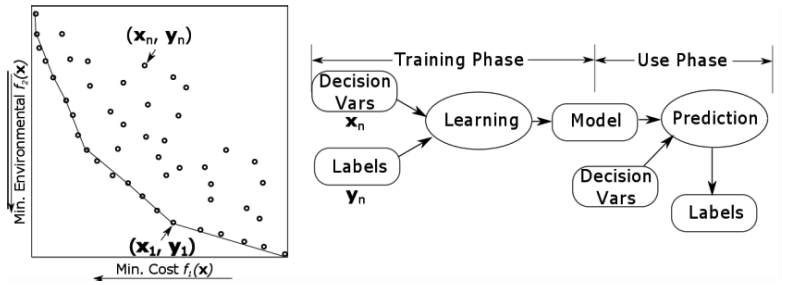


Figure 1: Coupling optimization outputs with ML for heuristics extraction

construction (Urban and Bakshi, 2013). Our scope only considers the operational stage of the residential house, without considering its construction and dismantlement stages. During the operational stage, the house mainly needs to purchase utilities from public supplies to sustain its demand. Therefore, it is reasonable to estimate the house's life cycle impacts with those of utility supplies, namely, water, natural gas, and electricity. We compare different design alternatives to the base-case residential system with no changes involved. Therefore, we mainly focus on optimizing the changes in the life cycle impacts associated with purchased goods and utilities. The functional unit of this study is its occupation time of 10 years and the amount of electricity, natural gas, and water that are required during this period. To prevent the impacts from shifting outside the boundary, we consider designs at Scope 1 (local), Scope 2 (local + utility generation) and Scope 3 (life cycle). The design variables are specified in Table 1. The bold values are associated with the base-case. Three objectives are minimized simultaneously, namely net carbon footprint, net water footprint, and relative life cycle cost. Nine optimization scenarios are considered by combining three variable types (Technological, Techno-Ecological, Techno-Eco-Behavioral) and the three scopes (1, 2, and 3).

Table 1: Design Variables

Cat	Variables	Carbon	Money	Water	Type	Values
Tech	Cooling efficiency	√	√	NR	Cont.	**3.0** - 3.9
	Furnace	√	√	NR	Cont.	**0.8** - 0.96
	Roof (reflectivity)	√	√	NR	Disc.	**0.7**, 0.45
	Wall (insulation)	√	√	NR	Disc.	**R13**, R21
	Ceil (insulation)	√	√	NR	Disc.	**R30**, R49
	Toilet	NR	√	√	Disc.	**Conv**, dual flush
	Sink	NR	√	√	Disc.	**Conv**, low flow
	Shower	NR	√	√	Disc.	**Conv**, low flow
	Clothes washer	√	√	√	Disc.	**Conv**, EnergyStar
	Dish washer	√	√	√	Disc.	**Conv**, EnergyStar
	PV	√	√	NR	Cont.	**0.0** - 0.3
Env	Corn	√	√	NR	Cont.	**0.0** - 1.0
	Shading tree	√	NR	NR	Disc.	**0**, 9
	Carbon tree	√	√	NR	Disc.	**0**, 1, 2,· · · , 20
	Lawn	√	√	NR	Cont.	**0.0** - 1.0
	Pond	√	√	√	Cont.	**0.0** - 1.0
Behav	Heat set point	√	NR	NR	Cont.	18.3 - **22.2** (◦C)
	Cooling set point	√	NR	NR	Cont.	**22.2**- 26.7 (◦C)

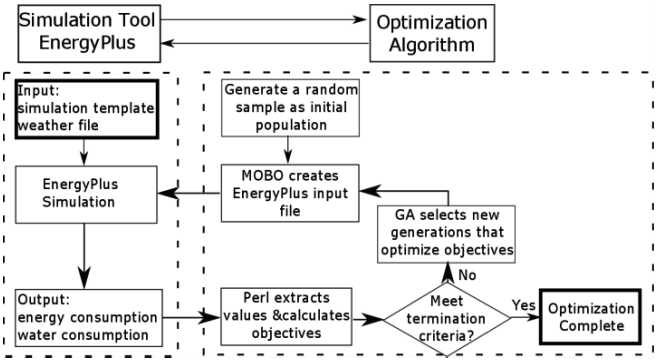


Figure 2: Simulation-based optimization flow chart

3.2. Simulation-based Optimization

This study utilizes a simulation-based optimization methodology that couples building simulation software Energyplus and an optimization algorithm non-dominated sorting genetic algorithm-II (NSGA-II) to seek optimal designs (Palonen et al., 2013). NSGA-II is chosen due to its suitability to derivative-free optimization and its ability to archive all points evaluated in the optimization. Figure 2 shows the flow chart of the simulation-based optimization process. Energyplus simulates the utility usage in the building and Perl language extracts such information to calculate optimization objectives. NSGA-II selects a new generation that optimizes the objectives, which is used to create new simulation files that are again passed to Energyplus simulator. The optimization process completes when the termination criteria are meet.

3.3. Heuristic Extraction

In this case study, the problem is formulated as a classification problem, using the decision tree algorithm, which has the ability to detect interactions and identify groups that have similar outcomes, along with the associated predictor variables (Neville, 1999). The predictor variables that lead to a more thorough classification can be determined and identified as design heuristics. Decision variable values of an instance are used as input features; while whether the evaluated instance is on the Pareto curve is used as output label. The decision tree classifiers are trained and pruned by applying hyperparameter grid search method. A 5-fold cross validation is performed with a train/test ratio of 4. All trained classifiers obtained ~0.9 area under receiver operating characteristic curve score, which is a commonly used performance indicator for binary classifiers. Feature importance is then computed as the normalized total entropy reduction brought by that feature. However, since the decision tree algorithm only provides the absolute value of feature importance, linear support vector machines have also been fitted to each of the nine scenarios to determine whether the decision variables contribute positively or negatively to optimality.

3.4. Results

The results for each scenario are shown in Figures 3, 4, and 5, respectively. Of all design variables in Table 1, only those whose importance rank is in the top ~50 % are shown in the bar plots. Black bars indicate that increasing variable value contributes negatively to optimality while grey bars indicate the opposite. The sums of variable importance grouped by their categories are also provided to facilitate the comparison

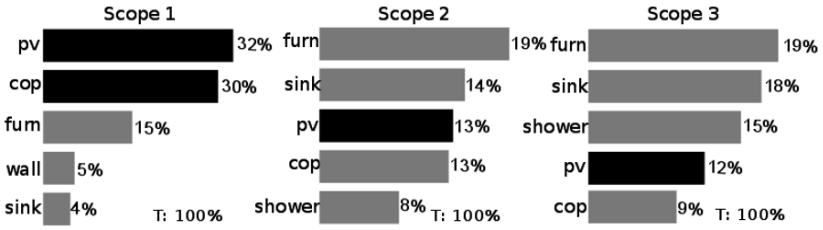


Figure 3: Designs with only Technological Variables

between different categories. From Figure 3, it can be inferred that of 11 technological decision variables, the top 5 most influential variables remain similar across scopes. Also, increasing the percentage of roof covered by photovoltaics would contribute negatively to optimality across all scopes. This is because the cost of photovoltaics is too high to offset the electricity saving. However, with increasing scope, the emission reduction from electricity generation renders the contribution from photovoltaics installment less negative.

From Figure 4, it can be inferred that the significance of technological and ecological variables is similar: technological variables contribute less than 60%, while ecological variables more than 40%. This also justifies the ecological variables in our design problem. Moreover, we can observe that the improvement of coefficient of performance (cop) is not beneficial at scope 1 because of its high cost; however, it is beneficial at large scopes. This shows the necessity to consider different scopes to prevent the impacts from shifting outside the boundary. In other words, even though some changes are not promising from a local perspective, it might be beneficial from a life cycle perspective and vice versa. Furthermore, existence of ecosystems on-site, e.g. vegetable garden, shading tree, and rain water harvest system, are beneficial across scales. The same trend is also shown in Figure 5.

Additionally, Figure 5 indicates that technologies may not be the only option, since their importance generally decreases with increasing scopes. Also, we can infer that "cheap" technologies, such as sink and shower faucet, are beneficial across variables types and scopes, and thus should be implemented; while "expensive" technologies, such as photovoltaics installation, are the opposite, and thus should be implemented with caution. Moreover, it can be observed that heating set point becomes more prominent with increasing scope. This is because the effect of utility saving becomes more important when upstream processes are considered. Furthermore, as indicated by all three scopes, we should keep the temperature cooler in winter and warmer in summer.

4. Conclusions

In this paper, we propose a methodology that couples multiobjective optimization and ML to extract design heuristics. The methodology has been demonstrated by a case study of designing a residential system consisting of a house and its yard. Potential heuristics for sustainable buildings include: 1) keep the house as warm as comfortable in the summer and as cold as comfortable in the winter; 2) place vegetable garden, shading trees, and rain water harvest system around the house; 3) implement cheap technologies first such as efficient showerheads and faucets before expensive ones such

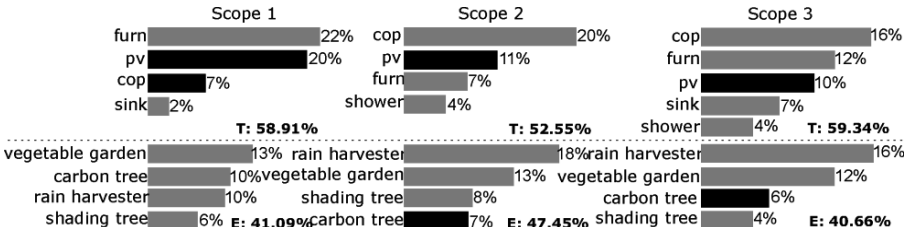


Figure 4: Designs with Techno-Ecological Variables

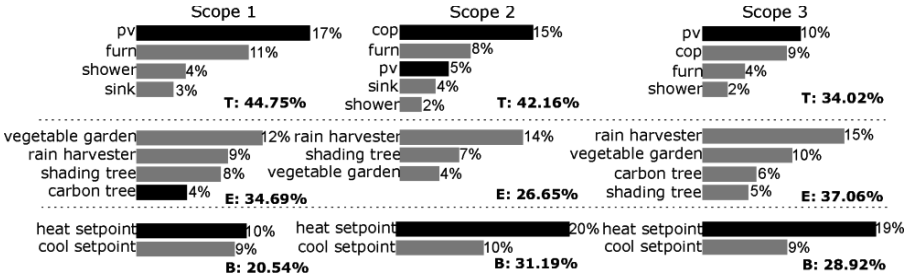


Figure 5: Designs with Techno-Eco-Behavioral Variables

as solar panels. This further indicates the benefits of the synergistic design between technologies and ecosystems and justifies the use of TES-LCA methodology. Coupling optimization outcomes with ML algorithm helps to extract useful design heuristics for guiding the future design of similar systems. This approach is general enough to be applied to other systems, ranging from local manufacturing processes to urban systems.

Acknowledgements

Partial funding for this work was provided by the National Science Foundation (CBET-1444745 & CBET 1336872).

References

M. R. Al Iqbal. 2012. Rule extraction from ensemble methods using aggregated decision trees. In: International Conference on Neural Information Processing. Springer pp 599-607.

K. Deb, A. Srinivasan. 2008. Innovization: Discovery of innovative design principles through multiobjective evolutionary optimization. Multiobjective Problem Solving from Nature pp 243-262.

X. Liu, B. R. Bakshi. 2017. Ecosystem services in life cycle assessment while encouraging techno-ecological synergies. J Ind Ecol (Accepted).

P. G. Neville. 1999. Decision trees for predictive modeling. SAS Institute Inc 4.

M. Palonen, M. Hamdy, A. Hasan. 2013. MOBO a new software for multiobjective building performance optimization. In: Proceedings of the 13th Internationcal Conference of the IBPSA pp 2567-2574.

W. D. Seider, D. R. Lewin, J. D. Seader, S. Widagdo, R. Gani, K. M. Ng. 2017. Product and Process Design Principles: Synthesis, Analysis and Design, 4th Edition. Wiley.

R. A. Urban, B. R. Bakshi. 2013. Techno-ecological synergy as a path toward sustainability of a North American residential system. Environmental Science & Technology 47(4): 1985-1993.

Author Index

M

This page intentionally left blank