Process Systems Engineering, 5. Process Dynamics, Control, Monitoring, and Identification

ULLMANN'S ENCYCLOPEDIA OF INDUSTRIAL CHEMISTRY

Krist V. Gernaey, Technical University of Denmark, Department of Chemical and Biochemical Engineering, Lyngby, Denmark

Jarka Glassey, Newcastle University, Faculty of Science, Agriculture and Engineering, Newcastle upon Tyne, United Kingdom

Sigurd Skogestad, Norwegian University of Science and Technology, Department of Chemical Engineering, Trondheim, Norway

Stefan Krämer, INEOS, Köln, Germany

Andreas Weiß, INEOS, Köln, Germany

Sebastian Engell, Technical University of Dortmund, Department of Chemical Engineering, Dortmund, Germany

Efstratios N. Pistikopoulos, Imperial College London, Department of Chemical Engineering, London, United Kingdom

David B. Cameron, IBM Global Business Services, Kolbotn, Norway

10.1002/14356007.o22_o09

1. Introduction

The focus of this keyword is on the exciting field of process dynamics, process control, process monitoring and process identification. This is a very broad field which is applied all across the process systems engineering (PSE) community. This keyword is structured such that it has focus on a number of key areas within this field.

In Chapter 2 special attention is paid to process monitoring applications and development in pharmaceutical production and food production. There have been major changes in those application areas, where the introduction of on-line measurement systems has received quite some attention in recent years. Process instrumentation is briefly covered in general terms, followed by an overview of some of the most frequently used monitoring tools. Short case studies illustrate the application of those tools.

Chapter 3 is introduced by means of standard definitions and terms. Key publications on plant-wide control are briefly summarized, followed by a comparison and critical discussion of two systematic procedures for design of plant-wide control systems. Most of the plant-wide control ideas can be transferred to batch production systems.

Chapter 4 focuses on batch production systems, for example, in the pharmaceutical and the polymer production industry. Following a basic definition of a batch production system, common methods for batch production management are introduced. Quality control of batch production systems is crucial in order to obtain an efficient production system, and therefore methods and tools for inferring information about batch production processes are briefly described as well. Finally, optimal operation of single batch processes and batch to batch control are introduced.

Chapter 5, on multiparametric programming and its application within model predictive control (MPC), starts with providing an overview of the most important developments in this area. The theory behind multiparametric programming is introduced, and its importance for the practical application of MPC and "MPC on a chip" technology is highlighted through a few illustrative examples. This section ends with a short discussion of future developments in the area.

The Chapter 6 focuses on on-line application, since dynamic simulators are increasingly used within the operation of chemical and petroleum production processes, to name a few

examples. Starting from the 1980s, the contribution provides a brief historical perspective of dynamic process modeling, followed by the description of the main software requirements for a typical architecture that allows on-line simulation. Technical and organizational challenges in using on-line simulation are highlighted, and applications of the technology are described.

2. Process Monitoring

2.1. Introduction

Monitoring process performance is a critical requirement in any manufacturing process as producing quality products within specification reproducibly is a prerequisite of an economically viable process. Without effective monitoring and control strategy, as key requisite, a capable manufacturing process, could not be successful. Monitoring is essential for various aspects of the control strategy–the quality of raw materials is usually tested on intake, process equipment often has to be rigorously qualified (e.g., in the highly regulated pharmaceutical or food industries), environment is controlled by implementing manufacturing-area classification where relevant, waste is treated prior to release and the quality of the final product is tested before release. Initiatives, such as quality by design (QbD) and a supporting enabling technology of process analytical technology (PAT) championed by the US Food and Drug Administration (FDA) in the pharmaceutical industry, aim to shift the focus for manufacturing from end-product quality testing to building the quality in the process. Such a shift in emphasis would not be possible without reliable and effective monitoring. Indeed PAT has been defined as "a system for designing, analyzing, and controlling manufacturing through timely measurements (that is, during processing) of critical quality and performance attributes of raw and in-process materials and processes, with the goal of ensuring final product quality" [1]. Traditional process control strategies based upon information from laboratory assays and supervisory computer systems (SCADA) are routinely used to regulate process operation and correct for disturbances resulting from raw material variations through to production plant variations. If PAT can provide additional information on disturbances and deviations, giving greater plant insight, then the effects of disturbances can be reduced and quality control tightened. However, greater benefits are to be gained by the systematic use of PAT tools in process development to increase fundamental understanding and more robust definition of the design and control space of the process operation.

An analogy in the food industry in terms of the importance of effective monitoring procedures can be seen in the hazard analysis critical control point (HACCP) food safety standard, which is now widely incorporated into national food safety legislation of many countries. The seven basic principles of HACCP implementation consist of [2]:

1. Conduct hazard analysis, considering all ingredients, processing steps, handling procedures, and other activities involved in a foodstuff's production
2. Identify critical control points (CCPs)
3. Define critical limits for ensuring the control of each CCP
4. Establish monitoring procedures to determine if critical limits have been exceeded and define procedure(s) for maintaining control
5. Define corrective actions to be taken if control is lost (i.e., monitoring indicates that critical limits have been exceeded)
6. Establish effective documentation and record-keeping procedures for developed HACCP procedure
7. Establish verification procedures for routinely assessing the effectiveness of the HACCP procedure, once implemented

Clearly effective monitoring is critical to ensuring product quality regardless of the type of manufacturing industry. Essential components of effective monitoring include representative measurement and a robust representation of the obtained information, allowing appropriate action to be taken.

2.2. Critical Process Parameter Measurement

A complete review of specific process instrumentation for critical parameter measurement is

beyond the scope of this section and the emphasis will be placed on the characteristics of measurements to be used in a critical parameter control scheme. These characteristics raise important questions that must be answered prior to sensor specification and they lead to the establishment of specific protocols that need to be followed during sensor use. Such characteristics would be equally applicable to established as well as emerging PAT measurement methodologies. The key considerations for a sensor are:

Accuracy and Resolution. A useful sensor provides measurement at an appropriate accuracy for the control task. If, for example, a temperature is to be controlled in the range of ± 0.1 °C then the measurement must be significantly more accurate than that. If that was not the case, the actual process may be subject to larger deviations, although it may appear that the process is controlled within this range.

Precision is the probability of obtaining the same value with repeated measurements on the same system and it is particularly important in the longer term operations. For instance, sensor drift from calibration can cause deterioration in system performance because the desired values are not achieved. Drift is often inevitable, so it is important to know the rates of likely drift so that recalibration can be performed as necessary.

Sensitivity is defined as the ratio between the sensor output change ΔS and the given change in the measured variable Δm (sensitivity $S = \Delta S/\Delta m$). If the critical control parameter value changes, it is important that the sensor responds to such a change.

Reliability. Sensors provide information which is acted upon either by process operators in a "human in the loop" control scheme or directly by closed-loop control schemes. When operators use the information, there is some opportunity for human interpretation of the results. Failed sensors are more difficult to detect in a hardware-based closed-loop scheme. If the information is essential and a sensor fails, then implications on operation can be severe. Reliability is a function of the failure rate, of the failure type, ease of maintenance and repair, and physical robustness. Redundancy and planned maintenance programs to maintain the sensors are required to maintain reliability.

Response time is defined as the time required for a sensor output to change from its previous state to a final settled value within a tolerance band of the correct new value. The dynamic sensor characteristics are important as the sensor must respond significantly faster than the process. If a sensor has a long response time it may indicate an "average" value rather than the actual process value.

Practicality. The environment within a process may be particularly demanding—for instance, the sensors may be exposed to high temperatures or pressures. Whilst a sensor may in theory measure the variable of interest in ideal conditions, the range of the operational environment could render it incapable of functioning or may influence reliability.

Cost. Sophisticated instrumentation is now available for process monitoring with PAT, but the price can be high. However, the benefits gained can be significant if sensor information leads to raw material/resource savings or increases productivity. A cost benefit analysis should be performed to assess whether the instrumentation is appropriate.

A significant issue to be addressed in effective monitoring is the placement of a sensor (Fig. 1) as it influences the frequency of available measurements. Theory dictates that for a

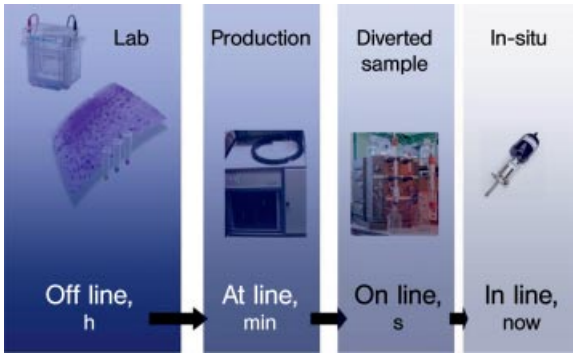

Figure 1. Sensor classification based on placement and speed of response

measurement to be of value it must be sampled above a certain minimum frequency. Often instruments are used on-line (say temperature or pH) or they can be multiplexed to save cost, but the frequency of information supply is limited because the instruments must serve several vessels (e.g., mass spectrometer measurements). However, it is off-line sample analysis where problems with low frequency measurement are most likely to arise.

Initiatives such as PAT lead to increased use of sophisticated sensor technology, such as near infrared spectroscopy (NIR), which requires more powerful data interpretation and monitoring tools.

2.3. Monitoring Tools

During the 1920s, the control charting methodology as the fundamental tool to understand and address variability, the foundation of so-called statistical process control, was developed [3]. Visualizing the variability is central to its reduction and statistical tools, such as cause and effect diagram, flow chart, Pareto chart, histogram, run chart, scatter diagram, and control charts, are often used. Histograms, flow charts, run charts, and scatter diagram compile the data to show the overall picture while Pareto diagrams are used to show problem areas. However, these methods do not indicate limits within which the process is to operate. The univariate SPC methodology uses charts with upper control limits known as "UCL", lower control limits known as "LCL" and means denoted as $\bar{X}$ or $\bar{R}$ for individual process variables. The basic principles of control charts, control limit settings, moving average charts, exponentially weighted moving average (EWMA) and cumulative sum (or CUSUM) control charts are described in [4] and illustrated by means of a case study of a mean particle size monitoring in a crystallization unit operation in the pharmaceutical industry [4].

Whilst univariate SPC can be very effective and has been used widely, it fails to account for the interactions between process variables and thus to recognize off-specification behavior. Also, univariate charts may indicate off-specification behavior in terms of one process variable, but to identify the cause of the fault conditions the interpretation of multiple charts is required. Finally, nonsteady-state behavior, process dynamics, time delays, etc. cause univariate charts to be inappropriate. Since most industries collect large amounts of data, multivariate statistical process control procedures are now considered to be an appealing approach to process monitoring and variability reduction.

2.3.1. Data Compression Methods for Multivariate Statistical Process Control (MSPC)

Multivariate SPC methods [5, 6] are based on fundamental concepts of principal component analysis (PCA) and partial least squares (PLS), also known as projection to latent structures. PCA [7] generates a new group of uncorrelated variables (principal components, PCs). The approach transforms matrix containing measurements from n process variables, $[\boldsymbol{X}]$, into a matrix of mutually uncorrelated PCs, t_k (where $k = 1$ to n) which are transforms of the original data into a new basis defined by a set of orthogonal *loading* vectors, $\boldsymbol{p}_k$. The individual values of the principal components are called *scores*. The transformation is defined by Equation (1):

$$[X] = \sum_{k=1}^{np<n} t_k p_k^T + E \tag{1}$$

The loadings are the eigenvectors of the data covariance matrix, $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$. The $\boldsymbol{t}_k$ and $\boldsymbol{p}_k$ pairs are ordered so that the first pair captures the largest amount of variation in the data and the last pair captures the least. This means that fewer PCs are required to describe the relationship than original process variables. The compression of data allows a visualization of the compressed data for the purpose of feature extraction and thus enables the analysis of interacting process variables that are the cause of process deviations.

PLS [8] is a tool suitable whenever plant variables can be partitioned into cause ($\boldsymbol{X}$) and effect ($\boldsymbol{Y}$) values. The algorithm operates by projecting the cause and effect data onto a number of latent variables and then modelling the relationships between these new variables (the so-called inner models) by single-input–single-output linear regression as described by

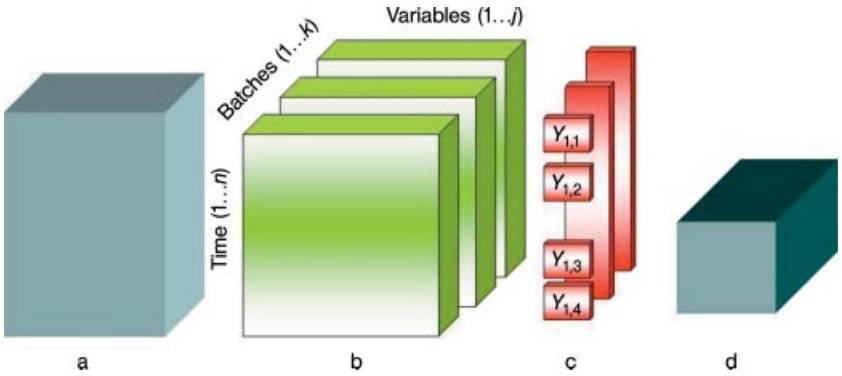

Figure 2. Typical data structure in a batch manufacturing process
a) Raw materials; b) Online data; c) Quality data; d) DSP data

Equations (2) and (3):

$$\boldsymbol{X} = \sum_{k=1}^{np<nx} \boldsymbol{t}_k \boldsymbol{p}_k^T + \boldsymbol{E} \quad \text{and} \quad \boldsymbol{Y} = \sum_{k=1}^{np<nx} \boldsymbol{u}_k \boldsymbol{q}_k^T + \boldsymbol{F}^* \tag{2}$$

where $\boldsymbol{E}$ and $\boldsymbol{F}^*$ are residual matrices, np is the number of inner components that are used in the model and nx is the number of causal variables.

$$\boldsymbol{u}_k = b_k \boldsymbol{t}_k + \varepsilon_k \tag{3}$$

where b_k is a regression coefficient, and ε_k refers to the prediction error.

2.3.2. Multiway MSPC

Batch processes typically exhibit nonlinear characteristics that may limit the effectiveness of conventional linear PCA and PLS procedures. Whilst nonlinear MSPC techniques have been developed and applied successfully [9], the transformation of batch data has proved to be a more effective option. The most common form of data transformation, termed multiway PCA and PLS, was initially proposed by [5]. Since then, for example, the technique was applied by [10] to monitor faults in automotive engine performance. The detection of faults by measuring particular chemicals from mixtures using electronic nose based on gas chromatography-mass spectrometry (GC-MS) was investigated by [11].

The concept of multiway PCA and PLS is a relatively straightforward extension where deviations from mean trajectories rather than steady-state are considered [5]. Figures 2 and 3 illustrate the principle for a typical set of

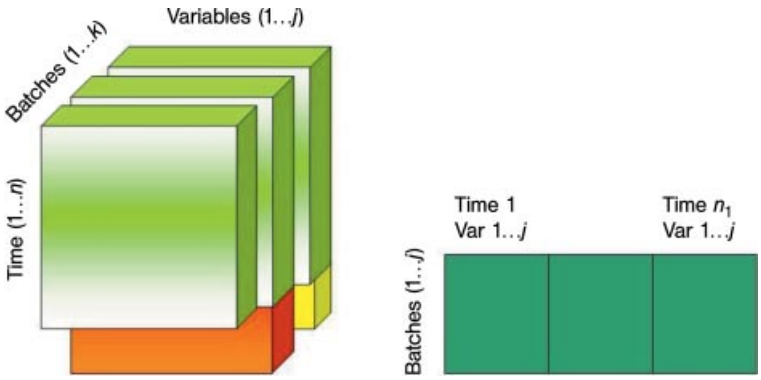

Figure 3. One possible multiway decomposition of on-line data
Batch 1: time = 1...n_1; Batch 2: time = 1...n_2; Batch 3: time = 1...n_3
$n_1 < n_3 < n_2$

operational process data where data of various size and frequency may be collected at various stages of processing.

Quality data on raw materials used in several batches, from which data is monitored over time from several sensors will need to be linked with quality data monitored during the batch at various frequencies for various quality attributes and merged with on-line data available from downstream processing unit operations.

Given that the duration of each batch is likely to differ, as indicated in Figure 3, the data from each batch is often considered only until the shortest run length. For each variable the mean trajectory over all the batches used in model building is calculated and removed from each process measurement. This effectively removes the major nonlinearity from the data and leaves a zero mean trajectory for each variable. The individual data matrices from each batch are unfolded into a single unfolded data matrix as depicted in Figure 3 and PCA can be applied to this unfolded data matrix.

2.4. Seed Quality Monitoring Case Study

A typical example of the data structure depicted in Figure 2 is taken from the bioprocessing industry. In order to monitor the quality of seed cultivations used for starting the manufacturing process in a range of valuable biological products, such as antibiotics, a number of process variables are measured. These include respiratory data, as well as information about the operating conditions, such as agitation, pH, temperature, etc. In this case study, 20 lots of data from the seed stage of pilot-scale antibiotic cultivations were available and only the airflow and respiratory data were used in analyses as other variables were tightly controlled. The data matrix for MPCA analysis was then constructed as indicated in Figure 3. Figure 4 depicts the plot of the resulting PC1 against PC2 and illustrates the degree of separation within this cluster. In Figure 4 (o) represent batches that ultimately resulted in low final stage productivity while (+) represent the high productivity batches. Tentative clusters of high- and low-productivity batches can be seen even at cursory inspection, for example, along the vertical line representing

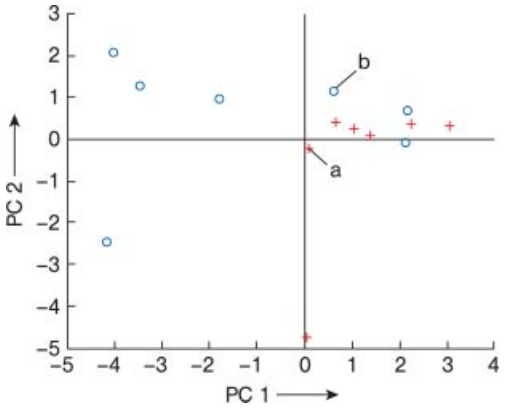

Figure 4. Plot of PC1 vs PC2 of a MPCA model for seed quality monitoring
a) High productivity; b) Low productivity
98% variance captured 5PCs

the PC2 axis. Although, based on such a simple separation, three of the low final-productivity seed batches would cluster within the "high" cluster, this may be an entirely plausible scenario. The seed could have had the same characteristics as those seeds ultimately resulting in high productivity, i.e., a "good" seed, but problems could have arisen during the final fermentations, which potentially could have led to reduced productivity.

These results demonstrate that it is possible to extract features from seed data that relate to the final productivity and thus to indicate the quality of a particular seed before inoculating the production vessel at the pilot plant scale.

2.5. Alternative Methods

Unfolding the data and reducing the length of each batch to that of the shortest one may significantly reduce the monitoring effectiveness of MPCA and a number of alternative methods have been developed over the years to address this issue. For example, see [12] for an application to on-line steady-state identification in polymer injection molding start-up process. There are also a number of alternative methods of data interpretation, such as parallel factor analysis (PARAFAC). The performance of several algorithms for fitting the PARAFAC model was compared by [13]. These include alternating least squares (ALS), direct trilinear

decomposition (DTLD), alternating trilinear decomposition (ATLD), self-weighted alternating trilinear decomposition (SWATLD), pseudo-alternating least squares (PALS), alternating coupled vectors resolution (ACOVER), alternating slice-wise diagonalization (ASD) and alternating coupled matrices resolution (ACOMAR).

A further category of methods include nonlinear data representation techniques ranging from the nonlinear forms of the multivariate data analysis methods above to the various forms of artificial neural networks (ANN) that have been proven effective in monitoring a variety of processes ranging from fermentations [14], object tracking [15], wastewater treatment [16] to monitoring the thermal performance of heat exchangers [17].

One particular type of ANN, referred to as radial basis function (RBF) network, has been proven to provide an efficient monitoring tool. RBF neural networks consist of three layers of nodes interconnected in a feed-forward manner, as shown in Figure 5 for two outputs and a limited interconnection illustrated to retain a reasonable level of picture clarity.

The first layer distributes the input data into the hidden layer of the network. The hidden nodes perform a nonlinear transformation of the input data [18]. Usually, the Gaussian function is used, as described by the Equation (4):

$$a_h = \exp\left[\frac{-\|x-c_h\|^2}{\beta_h^2}\right] \tag{4}$$

where a_h is the activation of the h^{th} processing unit in the hidden layer in response to the input vector '$\boldsymbol{x} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$'; c_h and β_h represent the position of the center and the cluster widths in the input space of the unit h, respectively.

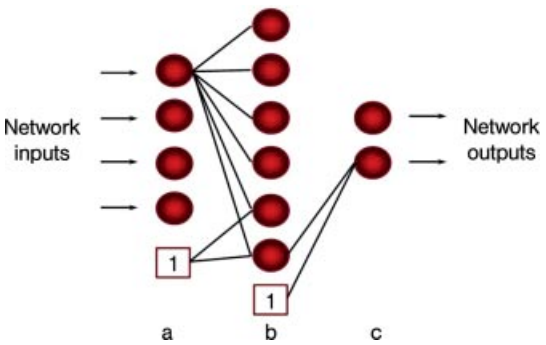

Figure 5. Radial basis function (RBF) neural network architecture
a) Input layer; b) Hidden layer; c) Output layer

The hidden layer outputs are weighted and summed in the output nodes. The response of the j^{th} output node, y_j, is given by Equation (5).

$$y_j = \sum_{h=1}^{H+1} W_{j,h} a_h + \theta \tag{5}$$

where $W_{j,h}$ is the weight between the hidden node h and the output node j. The bias node is represented by θ and has the value of 1 [19].

The major advantage of neural networks is that they are able to "learn" from the information that is presented. This means, however, that a suitable training data set is crucial for a good performance. The importance of the size and quality of the training data set in ANN modeling has been reported extensively in literature [20]. Other important issues in the development of RBF models are the selection of the network inputs and the most suitable architecture, i.e., the number of RBF units and the number of nearest neighbors to be used. Whilst the selection of inputs is usually accomplished by using process knowledge [21], prediction errors and cross validation are most frequently used to select the network topology [19].

Once the topology is defined, the network can be trained, i.e., the unit centers, unit widths, and weights are calculated, for example, by using Moody and Darken's three step approach [22]:

1. The unit centers, c, are determined by the k-means clustering algorithm, which divides the training data into subsets. Each subset is related to a cluster center, according to the similarities of the data. These similarities are determined by the distance between two data points. The algorithm minimizes an objective function E, which is usually the total squared Euclidean distance between the K training points in each cluster and the H cluster centers, according to Equation (6):

$$E = \sum_{h=1}^{H} \sum_{k=1}^{K} \boldsymbol{M}_{hk} \|c_h - x_k\|^2 \tag{6}$$

In Equation (6), $\boldsymbol{M}_{hk}$ is a $H \times K$ matrix called the membership function or cluster

partition. Each column contains a single 1 that identifies the processing unit to which a given training point belongs, and zeros are assigned elsewhere [21].

Once this is achieved, each cluster is associated with one RBF unit and the cluster centers become the unit centers c. Each center is then compared with the input vector and the corresponding unit is activated according to the distance between the network input vector and the center.

2. After determining the unit centers, a P-nearest neighbors heuristic (Eq. 7) can be used to find the unit widths σ_h. The unit width should be determined so that it is greater than the distance to the nearest unit center. This allows the hidden unit to activate at least another hidden unit. Consequently, any point within the bounds of that unit will be able to significantly activate more than one unit, improving the fit of the desired outputs.

$$\sigma_h = \left(\frac{1}{p} \sum_{m=1}^{p} \|c - z_m\|^2 \right)^{1/2} \tag{7}$$

where 'z_m' represents the P-nearest neighbors of c.

3. The weights of the output layer are then calculated using a least squares-based method. The objective is to find the weights that minimize the squared norm of the residuals. The output layer nodes simply sum the outputs from the hidden layer.

After determining the parameters of the network, the local reliability can be measured by calculating the confidence limits for the model estimation at a given test point. This is the result of the weighted average of the local confidence intervals calculated for each RBF unit [18, 19].

2.6. RBF-Based Monitoring Case Study

RBF neural network modeling has been used to monitor a range of different processes. In this example, it is used to detect deviations in large-scale production of penicillin. A number of factors influence the behavior of a large-scale fermentation and a dynamic and nonlinear character of the bioprocess mean that simple monitoring of individual process variables does not allow the detection of a developing process deviation, unless there is a severe, obvious fault. However, detecting deviations early in the process is essential in order to ensure that the process is returned to normal behavior and to prevent economic consequences due to lower productivity or even a complete loss of the whole batch.

Data from a range of nominal and faulty large-scale penicillin production batches was available. These included measurements of feed rates, total feeds added, carbon dioxide evolution rate (CER) and oxygen uptake rate (OUR)—respiratory data reflecting the progress of the fermentation. The hypothesis in this case was that if the batch behaves nominally, i.e., no deviations occur, then a model developed using nominal batches only to predict the respiration data should be accurate and estimate CER and OUR with very low margin of error. When RBF models of the process were developed to predict CER and OUR, respectively, this was indeed observed (Fig. 6).

In this RBF model, the feed rates, total feed, and batch age were used as input variables, and 17 RBF units to predict CER with very low error, remaining within the 95% and 99% confidence limits. However, when this model was challenged with the feed data from batches encountering various faults, the errors in CER prediction violated both 95% and 99% confidence limits at some point during the batch as

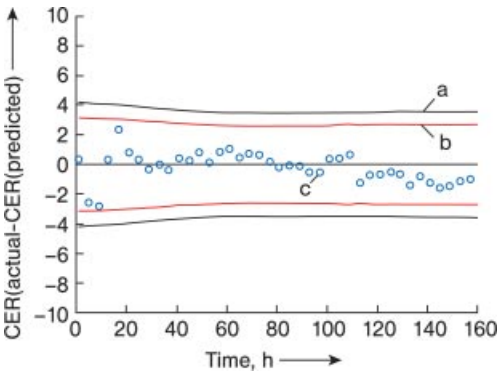

Figure 6. Error plot of RBF prediction of CER for a nominal penicillin production batch
a) 99% confidence limits; b) 95% confidence limits; c) CER error

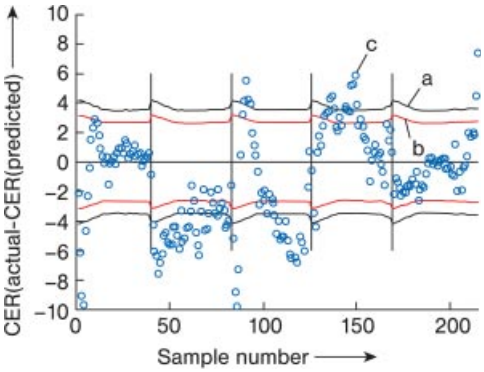

Figure 7. Error plot of RBF prediction of CER for five faulty penicillin production batches separated by vertical lines
a) 99% confidence limits; b) 95% confidence limits; c) CER error

shown in Figure 7 for five faulty batches separated by vertical lines.

The violations of the confidence limits could theoretically be caused by the RBF model extrapolating outside the range of the input data used for training (a frequent shortfall of ANN methodology) or by a biological variability causing real deviations of the process from the nominal behavior. The benefit of using RBF models is that check on maximum activity and probability density [19] provides a measure of extrapolation. In this case it clearly confirmed that the reason for the violation of the confidence limits is biological process variability.

However, establishing the reason for such deviations is not a straightforward matter. One of the limitations of ANN methodology is the fact that the interpretation of causal relationships is more difficult than with some of the more established linear methods, such as PCA and PLS. In some areas of bioprocessing, e.g., the manufacture of biologics for human consumption, the simple indication of process deviation, regardless of the underlying reason, is all that is required, as the strict regulatory requirements mean that the batch will have to be terminated and cannot be remedied. In such circumstances in particular the ANN-based monitoring can prove very effective.

There are a large number of other types of ANN models developed specifically for estimation of process variables or fault detection and clustering/classification. The various forms of neural networks used in diverse applications preclude detailed description of this methodology here, but extensive literature is available both on the principles and their various applications.

3. Plantwide Control

3.1. Introduction

A chemical plant may have thousands of measurements and control loops. By the term plantwide control it is not meant the tuning and behavior of each of these loops, but rather the control philosophy of the overall plant with emphasis on the structural decisions:

- Selection of controlled variables (CVs, "outputs")
- Selection of manipulated variables (MVs, "inputs")
- Selection of (extra) measurements
- Selection of control configuration (structure of overall controller that interconnects the controlled, manipulated, and measured variables)
- Selection of controller type (proportional–integral–derivative (PID), decoupler, model predictive control (MPC), linear–quadratic–Gaussian (LQG), ratio, etc.)

In practice, the control system is usually divided into several layers, separated by time scale (see Fig. 8).

Plantwide control thus involves all the decisions necessary to make a block diagram (used by control engineers) or a process and instrumentation diagram (used by process engineers) for the entire plant, but it does not involve the actual design of each controller.

In any mathematical sense, the plantwide control problem is a formidable and almost hopeless combinatorial problem involving a large number of discrete decision variables, and this is probably why the progress in the area has been relatively slow. In addition, the problem has been poorly defined in terms of its objective. Usually, in control, the objective is that the controlled variables (CVs, outputs) should remain close to their setpoints. However, what should be controlled? Which CVs? The answer

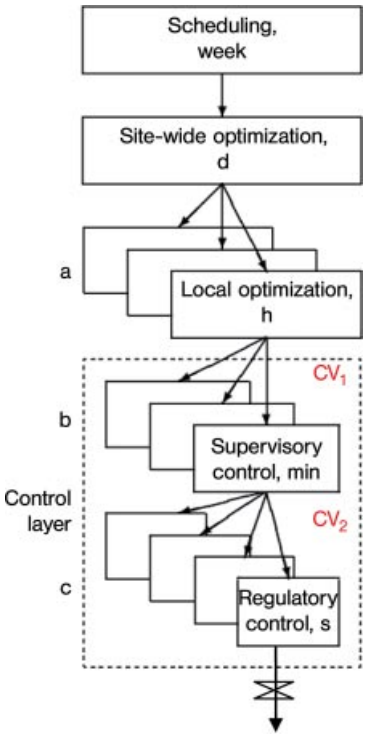

Figure 8. Typical control hierarchy in a chemical plant a) Real-time optimization (RTO); b) Model predictive control (MPC); c) Proportional–integral–derivative (PID) control

lies in considering the overall plant objective, which normally is to minimize the economic cost (= maximize profit) while satisfying operational constraints imposed by the equipment, market demands, product quality, safety, environment, and so on. The truly optimal "plant-wide controller" would be a single centralized controller which at each time step collects all the information and computes the optimal changes in the manipulated variables (MVs). Although such a single centralized solution is foreseeable on some simple processes, it seems to be safe to assume that it will never be applied to any normal-sized chemical plant. There are many reasons for this, but one important is that in most cases one can obtain acceptable control performance with simple structures where each controller block only involves a few variables, and such control systems can be designed and tuned with much less effort, especially when it comes to the modeling and tuning effort. After all, most real plants operate well with simple control structures. So how are systems controlled in practice? The main simplification is to decompose the overall control problem into many simpler control problems. This decomposition involves two main principles:

1. *Decentralized (local) control.* This "horizontal decomposition" of the control layer is mainly based on separation in space, for example, by using local control of individual process units
2. *Hierarchical control.* This "vertical decomposition" is mainly based on time scale separation, and in a process one typically has the following layers (see Fig. 8)
   - Scheduling (weeks)
   - Site-wide optimization (day)
   - Local optimization (hour)
   - Supervisory (predictive, advanced) control (minutes)
   - Regulatory control (seconds)

The upper three layers in Figure 8 deal explicitly with economic optimization and are not considered in this chapter. The focus is on the two lower control layers where the main objective is to track the setpoints specified by the layer above. A very important structural decision, probably more important than the controller design itself, is the choice of controlled variables (CVs) that interconnect the layers. More precisely, the decisions made by each layer (boxes in Fig. 8) are sent as setpoints for the controlled variables (CVs) to the layer below. Thus, indirectly optimization is considered because CVs should be selected that are favorable from an economic point of view.

Typically, PID controllers are used in the *regulatory control layer,* where "stabilization" of the plant is the main issue. In the *supervisory control layer*, one has traditionally used manual control or single-loop PID control, complemented by "advanced" elements such as static decouplers, feedforward elements, selectors, split-range controllers, and various logic elements. However, over the last 25 years, model predictive control (MPC) has gradually taken over as a unifying tool to replace most of these elements. In the (local) *optimization layer*, the decisions are usually executed manually, although real-time optimization (RTO) is used

for a few applications, especially in the refining industry.

The following decisions must be made when designing a plantwide control strategy:

1. Decision 1: Select "economic" (primary) controlled variables (CV_1) for the supervisory control layer
2. Decision 2: Select "stabilizing" (secondary) controlled variables (CV_2) for the regulatory control layer
3. Decision 3: Locate the *throughput manipulator* (TPM), that is, where to set the production rate
4. Decision 4: Select pairings for the stabilizing layer, that is, pair inputs (valves) and controlled variables (CV_2).

Decisions 1 and 2 are illustrated in Figure 9, where the matrices $\boldsymbol{H}$ and $\boldsymbol{H}_2$ represent a selection, or in some cases a combination, of the available measurements y.

This chapter deals with continuous operation of chemical processes, although many of the arguments hold also for batch processes.

3.2. Previous Work

Over the years, going back to the early work of Buckley [23] from DuPont, several approaches have been proposed for dealing with plantwide control issues. Nevertheless, taking into account the practical importance of the problem, the literature is relatively scarce. Larsson and Skogestad [24] provide a good review and divide into two main approaches. First, there are the *process-oriented* (engineering or simulation-based) approaches of [25–30]. One problem here is the lack of a really systematic procedure and that there is little consideration of economics. Second, there is the optimization or *mathematically oriented* (academic) approaches of [31–35]. The problem here is that the resulting optimization problems are intractable for a plantwide application. Therefore, a hybrid between the two approaches is more promising [24, 36–40].

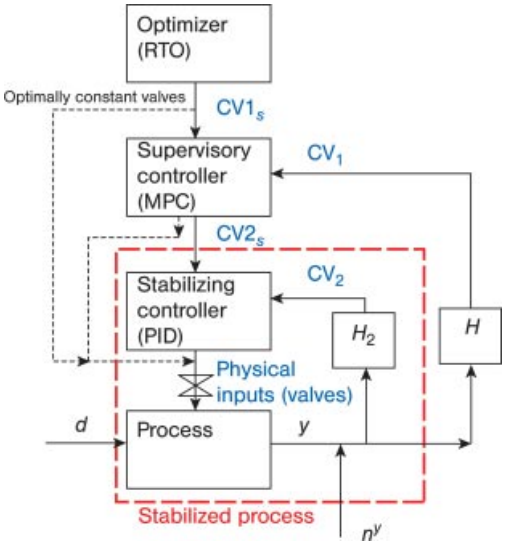

Figure 9. Block diagram of control hierarchy illustrating the selection of controlled variables ($\boldsymbol{H}$ and $\boldsymbol{H}_2$) for optimal operation (CV_1) and stabilization (CV_2)

The first really systematic plantwide control procedure was that of Luyben et al. [28, 29] which has been applied in a number of simulation studies. Luyben's procedure consists of the following nine steps

- L1: Establish control objectives
- L2: Determine control degrees of freedom
- L3: Establish energy management system
- L4: Set the production rate (decision 3)
- L5: Control product quality and handle safety, environmental, and operational constraints
- L6: Fix a flow in every recycle loop and control inventories
- L7: Check component balances
- L8: Control individual unit operations
- L9: Optimize economics and improve dynamic controllability

"Establish control objectives" in step L1 does not lead directly to the choice of controlled variables (decisions 1 and 2). Thus, in Luyben's procedure, decisions 1, 2, and 4 are not explicit, but are included implicitly in most of the steps. Even though the procedure is systematic, it is still heuristic and ad hoc in the sense that it is not clear how the authors arrived at the steps or their order. A major weakness is that the procedure does not include economics, except as an afterthought in step L9.

In this chapter, the seven-step plantwide control procedure of Skogestad [24, 39] is discussed. It was inspired by the Luyben procedure, but it is clearly divided into a top-down part, mainly concerned with steady-state

economics, and a bottom-up part, mainly concerned with stabilization and pairing of loops. SKOGESTAD's procedure consists of the following steps:

1. Top-down part (focus on steady-state optimal operation)
   - S1: Define operational objectives (economic cost function J and constraints)
   - S2: Determine the optimal steady-state operation conditions
   - S3: Select "economic" (primary) controlled variables, CV_1 (decision 1)
   - S4: Select the location of the throughput manipulator (TPM) (decision 3)
2. Bottom-up part (focus on the control layer structure)
   - S5: Select the structure of the regulatory (stabilizing) control layer (decisions 2 and 4)
   - S6: Select the structure of the supervisory control layer
   - S7: Select structure of (or need for) the optimization layer (RTO)

The top-down part (steps 1–4) is mainly concerned with economics, and steady-state considerations are often sufficient. Dynamic considerations are more important for steps 4–6, although steady-state considerations are important also here. This means that it is important in plantwide control to involve engineers with a good steady-state understanding of the plant. A detailed analysis in step S2 and step S3 requires that one has a steady-state model available and that one performs optimizations for the given plant design ("rating mode") for various disturbances.

3.3. Degrees of Freedom for Operation

The issue of degrees of freedom for operation, or control degrees of freedom, is often confusing and not as simple as one would expect. One issue is that the degrees of freedom change depending on where one is in the control hierarchy. This is illustrated in Figures 8 and 9, where the degrees of freedom in the optimization and supervisory control layers are not the physical degrees of freedom (valves), but rather the setpoints for the controlled variables in the layer below. The control degrees of freedom are often referred to as manipulated variables (MVs) or inputs. The physical degrees of freedom (dynamic process inputs) are called "valves", because this is usually what they are in process control.

Steady-State DOFs (u). A simple approach is to first identify all the physical (*dynamic*) degrees of freedom (valves). However, because the economics usually depend mainly on the steady-state, variables that have no or negligible effect on the economics (steady-state) should be subtracted, such as inputs with only a dynamic effect or controlled variables (e.g., liquid levels) with no steady-state effect.

$$\#\text{steady-state degrees of freedom } (u) = \#\text{valves} - \#\text{variables with no steady-state effect}$$

For example, even though a heat exchanger may have a valve on the cooling water and in addition have bypass valves on both the hot and cold side, it usually has only one degree of freedom at steady-state, namely the amount of heat transferred, so two of these three valves only have a dynamic effect from a control point of view.

In addition, we need to exclude valves that are used to control variables with no steady-state effect (usually, liquid levels). This is illustrated in the following example.

Example: DOFs for Distillation: A simple distillation column has six dynamic degrees of freedom (valves): feed F, bottom product B, distillate product D, cooling, reflux L, and heat input. However, two degrees of freedom (e.g., B and D) must be used to control the condenser and reboiler levels (M_B and M_D) which have no steady-state effect. This leaves four degrees of freedom at steady-state. For the common case with a given feed flow and a given column pressure, only two steady-state degrees of freedom remain. Thus, for the economic analysis in step S3, 2 controlled variables (CV_1) need to be selected associated with these. Typically, these will be the top and bottom composition, but not always.

3.4. Skogestad's Plantwide Control Procedure

Going through the Skogestad procedure in more detail, an existing plant is considered and it is assumed that a steady-state model of the process is available.

The top-down part is mainly concerned with the plant economics, which are usually determined primarily by the steady-state behavior. Therefore, although one is concerned about control, steady-state models are usually sufficient for the top-down part.

Step S1: Define Operational Objectives (Cost J and Constraints). A systematic approach to plantwide control requires that first the operational objectives are quantified in terms of a scalar cost function J [\$/s] that should be minimized (or equivalently, a scalar profit function, $P = -J$, that should be maximized). This is usually not very difficult, and typically it is:

J = cost feed+cost utilities (energy)−value products [\$/s]

Fixed costs and capital costs are not included, because they are not affected by plant operation on the time scale considered (ca. 1 h). The goal of operation (and of control) is to minimize the cost J, subject to satisfying the operational constraints ($g \leq 0$), including safety and environmental constraints. Typical operational constraints are minimum and maximum values on flows, pressures, temperatures, and compositions. For example, all flows, pressures, and compositions must be nonnegative.

Step S2: Determine the Steady-State Optimal Operation. Before the control system is designing the optimal way of operating the process should be considered. For example, a valve (e.g., a bypass) should always be closed. This valve should then not be used for (stabilizing) control unless one is willing to accept the loss implied by "backing off" from the optimal operating conditions.

To determine the steady-state optimal operation, a steady-state model should be obtained. Then the degrees of freedom and expected disturbances need to be identified, and optimizations for the expected disturbances should be performed:

1. Identify steady-state degrees of freedom (u): To optimize the process, the steady-state degrees of freedom (u) have to be identified as has already been discussed. Actually, it is the number of u's which is important, because it does not really matter which variables are included in u, as long as they make up an independent set
2. Identify important disturbances (d) and their expected range: Next, the expected range of disturbances (d) for the expected future operation have to be identified. The most important disturbances are usually related to the feed rate (throughput) and feed composition, and in other external variables such as temperature and pressure of the surroundings. Furthermore, changes in specifications and constraints (such as purity specifications or capacity constraints) and changes in parameters (such as equilibrium constants, rate constants and efficiencies) should be included as disturbances. Finally, the expected changes in prices of products, feeds, and energy need to be included as "disturbances".
3. Optimize the operation for the expected disturbances: Here, the disturbances (d) are specified and the degrees of freedom ($u_{\text{opt}}(d)$) are varied in order to minimize the cost (J), while satisfying the constraints. The main objective is to find the *constraints regions* (sets of active constraints) and the optimal nominal setpoints in each region.

Mathematically, the steady-state optimization problem can be formulated as

$$\begin{aligned} &\min_u J(u,x,d) \\ &\text{subject to:} \\ &\text{Model equations:} \quad f(u,x,d) = 0 \\ &\text{Operational constraints:} \quad g(u,x,d) \leq 0 \end{aligned}$$

Here u are the steady-state degrees of freedom, d are the disturbances, x are the internal states, $f = 0$ represents the mathematical model equations and possible equality constraints (like a given feed flow), and $g \leq 0$ represents the operational constraints (like a maximum or nonnegative flow, or a product composition constraint). The process model, $f = 0$, is often represented indirectly in terms of a commercial software package (process simulator), such as

Aspen or Hysis/Unisim. This usually results in a large, nonlinear equation set which often has poor numerical properties for optimization.

Together with obtaining the model, the optimization step S2 is often the most time consuming step in the entire plantwide control procedure. In many cases, the model may not be available or one does not have time to perform the optimization. In such cases a good engineer can often perform a simplified version of step S1–S3 by using process insight to identify the expected active constraints and possible "self-optimizing" controlled variables (CV_1) for the remaining unconstrained degrees of freedom.

A major objective of the optimization is to find the expected regions of active constraints. An important point is that one cannot expect to find a single control structure that is optimal because the set of active constraints will change depending on disturbances and economic conditions (prices). Thus, one should prepare the control system for the future, by using off-line analysis and optimization to identify regions of active constraints. The optimal active constraints will vary depending on disturbances (feed composition, outdoor temperature, product specifications) and market conditions (prices).

Generally there are two main *modes of operation* depending on market conditions:

- Mode I: Given throughput (buyers market). This is usually the "nominal" mode for which the control system is originally set up. Usually, it corresponds to a "maximize efficiency" situation where there is some "trade-off" between utility (energy) consumption and recovery of valuable product, corresponding to an unconstrained optimum.
- Mode II: Maximum throughput (sellers market). When the product prices are sufficiently high compared to the prices of raw materials (feeds) and utilities (energy), it is optimal to increase the throughput as much as possible. However, as one increases the feed rate, one will usually encounter constraints in various units, until eventually reaching the bottleneck where a further increase is infeasible.

Step S3: Select "Economic" (Primary) Controlled Variables, CV_1 (Decision 1). This is related to the implementation of the optimal operation points found in step S2 in a robust and simple manner. To make use of all the economic degrees of freedom (inputs u), as many economic controlled variables (CV_1) as there are inputs (u) need to be identified. In short, the issue is: What should be controlled?

1. Identify candidate measurements (y) and their expected static measurement error (n^y). In general, in the set y all inputs (valves) should be included to allow, for example, for the possibility of keeping an input constant.
2. Select primary (economic) controlled variables, $CV_1 = Hy$ (decision 1), among the candidate measurements (see Fig. 9), usually by selecting individual measurements. One needs to find one CV_1 for each steady-state degree of freedom (u)

For economic optimal operation, the rules for CV_1 selection are

1. Control active constraints
2. For the remaining unconstrained degrees of freedom: Control "self-optimizing" variables with the objective of minimizing the economic loss with respect to disturbances

The two rules are discussed in detail below. In general, step S3 must be repeated for each constraint region. To reduce the need for switching between regions, one may consider using the same CV_1's in several regions, but this is non-optimal and may even lead to infeasibility.

Control Active Constraints. In general, the obvious controlled variables to keep constant are the active constraints. The active constraints come out of the analysis in step S2 or may in some cases be identified based on physical insight. The active constraints are obvious "self-optimizing" variables and could be *input constraints* (in the set u) or *output constraints*.

Input constraints are trivial to implement; the input is set at its optimal minimum or maximum, so no control system is needed. For example, if a very old car is operated then optimal operation (defined as minimum driving time, $J = T$) may be achieved with the gas pedal at its maximum position.

For *output constraints*, a controller is needed, and a simple single-loop feedback controller is

often sufficient. For example, if there exists a better car then the maximum speed limit (say 80 km/h) is likely an active constraint and should be selected as the controlled variable (CV_1). To control this, one may use a "cruise controller" (automatic control) which adjusts the engine power to keep the car close to a given setpoint. In this case, the speed limit is a *hard constraint* and one needs to *back off* from the speed limit (say to a setpoint of 75 km/h) to guarantee feasibility if there is a steady-state measurement error (n^y) or a dynamic control error. In general, the backoff should be minimized because any backoff results in a loss (i.e., a larger $J = T$) which can never be recovered.

The backoff is the "safety margin" from the active constraint and is defined as the difference between the constraint value and the chosen setpoint:

$$\text{Backoff} = |\,\text{Constraint} - \text{Setpoint}\,|$$

In the car driving example: backoff = 5 km/h.

The active constraints should be selected as CVs because the optimum is not "flat" with respect to these variables. Thus, there is often a significant economic penalty if one "backs off" from an active constraint, so tight control of the active constraints is usually desired. If a constrained optimization method is used for the optimization, then the loss can be quantified by using the Lagrange multiplier λ associated with the constraint:

$$\text{Loss} = \lambda \times \text{backoff}$$

For input (valve) constraints, usually no backoff is needed, unless the input for stabilization is used in the lower regulatory (stabilizing) layer because one needs some range to use it for control. For output constraints two cases exist:

- *Soft output constraints* (only average value matters): Backoff = measurement error (bias n^y)
- *Hard output constraints* (must be satisfied at all times): Backoff = measurement error (bias n^y) + control error (dynamic)

To reduce the backoff, accurate measurements of the constraint outputs are necessary, and for hard output constraints one also needs tight control with a small dynamic control error. The *squeeze and shift rule* for hard output constraints indicates: By squeezing the output variation, the setpoint can be shifted closer to its limit (i.e., reduce the backoff). For soft output constraints, only the steady-state control error matters, which will be zero if the controller has integral action.

Control "Self-optimizing" Variable Which When Held Constant Keeps the Operation Close to the Optimum in spite of Disturbances. It is usually simple to identify and control the active constraints. The more difficult question is: What should the remaining unconstrained degrees of freedom be used for? Does it even make a difference what is controlled? The answer is "yes"!

As an example, optimal operation of a marathon runner is considered where the objective is to adjust the power (u) and to minimize the time ($J = T$). This is an unconstrained problem; a marathon runner cannot simply run at maximum speed ($u = u_{\max}$) as for a sprinter. A simple policy is constant speed (c_1 = speed), but it is not optimal if there are disturbances (d) caused by wind or hilly terrain. A better choice is to run with constant pulse (c_2 = pulse), which is easy to measure with a pulse clock. With a constant heart rate (c_2=constant), the speed (c_1) will increase when running downhill as one would expect for optimal operation, so pulse (c_2) is clearly a better self-optimizing variable than speed (c_1). Self-optimizing means that when the selected variables are kept constant at their setpoints, then the operation remains close to its economic optimum in spite of the presence of disturbances [40]. One problem with the feedback is that it also introduces a measurement error (noise) n^y which may also contribute to the loss (see Fig. 9).

In the following $CV_1 = c$. There are two main possibilities for selecting self-optimizing $c = \boldsymbol{H}y$:

1. Single measurements as CV_1's ($\boldsymbol{H}$ is a selection matrix with a single 1 in each row/column and the rest of the elements 0) are selected
2. Measurement combinations as CV_1's are used. Here, methods exist to find optimal linear combinations $c = \boldsymbol{H}y$, where $\boldsymbol{H}$ is a "full" combination matrix

In summary, the problem at hand is to choose the matrix $\boldsymbol{H}$ such that keeping the controlled variables $c = \boldsymbol{H}y$ constant (at a given setpoint c_s) gives close-to-optimal operation in spite of the presence of disturbances d (which shift the optimum), and measurement errors n^y (which give an offset from the optimum).

Quantitative Approaches. Are there any systematic methods for finding the matrix $\boldsymbol{H}$, that is, to identify self-optimizing c's associated with the unconstrained degrees of freedom? Yes, and there are two main approaches:

1. "Brute force" approach: Given a set of controlled variables $c = \boldsymbol{H}y$, one computes the cost $J(c,d)$ when c is kept constant ($c = c_s + \boldsymbol{H}n^y$) for various disturbances (d) and measurement errors (n^y). In practice, this is done by running a large number of steady-state simulations to try to cover the expected future operation. Typically, expected extreme values in the parameter space (for d and n^y) are used to compute the cost for alternative choice of the controlled variables (matrix $\boldsymbol{H}$). The advantage with this method is that it is simple to understand and apply and it works also for nonlinear plants and even for changes in the active constraint. Only one nominal optimization is required to find the setpoints. The main disadvantage with the method is that the analysis for each $\boldsymbol{H}$ is generally time consuming and one cannot guarantee that all important cases are covered. In addition, there exist an infinite number of choices for $\boldsymbol{H}$ so one can never guarantee that the best c's are found.
2. "Local" approaches: Based on a quadratic approximation of the cost. This is discussed in more detail in [41].

The main local approaches are:

- Maximum gain rule: The maximum gain rule says that one should control "sensitive" variables, with a large gain from the inputs (u) to $c = \boldsymbol{H}y$. This rule is good for prescreening and also yields good insight.
- Nullspace method: This method yields optimal measurement combinations for the case with no noise, $n^y = 0$. By simulations one must first obtain the optimal measurement sensitivity, $F = dy^{opt}/dd$. Then, assuming that the number of (independent) measurements y is the sum of the number of inputs (u) and disturbances (d), the optimal is to select $\boldsymbol{H}$ such that $\boldsymbol{H}F = 0$. Note that $\boldsymbol{H}$ is a nonsquare matrix, so $\boldsymbol{H}F = 0$ does not require that $\boldsymbol{H} = 0$ (which is a trivial uninteresting solution), but rather that $\boldsymbol{H}$ is in the nullspace of F^T.
- Exact local method (loss method): This extends the nullspace method to the case with noise and to any number of measurements. For details see [41].

For some practical applications of the nullspace method see [42].

Regions and Switching. New self-optimizing variables must be identified (off-line) for each region, and switching of controlled variables is required as one encounters a new region (on-line). In practice, it is easy to identify when to switch when one encounters a constraint. It seems less obvious when to switch out of a constraint, but actually one simply has to monitor the value of the unconstrained CVs from the neighboring regions and switch out of the constraint region when the unconstrained CV reaches its setpoint.

As an example, a recycle process is considered where it is optimal to keep the inert fraction in the purge at 5% using the purge flow as a degree of freedom (unconstrained optimum). However, during operation there may be a disturbance (e.g., increase in feed rate) so that the recycle compressor reaches its maximum load (e.g., because of constraint on maximum speed). The recycle compressor was used to control pressure, and since it is still optimal to control pressure, the purge flow has to take over this task. This means that one has to give up controlling the inert fraction, which will drop below 5%. In summary, one has gone from an unconstrained operating region (I) where the inert fraction is controlled to a constrained region (II) where the compressor is at maximum load. In region II, one keeps the recycle flow at its maximum. How does one know when to switch back from region II to region I? This is done by monitoring the inert fraction, and when it reaches 5% one switches back to controlling it (region I).

In general, one would like to simplify the control structure and reduce the need for switching. This may require using a suboptimal CV_1 in some regions of active constraints. In this case the setpoint for CV_1 may not be its nominally optimal value (which is the normal choice), but rather a "robust setpoint" which reduces the loss when operating outside the nominal constraint region.

Step S4. Select the Location of Throughput Manipulator (TPM) (Decision 3). The main purpose of a process plant is to transform feedstocks into more valuable products and this involves moving mass through the plant. The amount of mass moved through the plant, as expressed by the feed rate or product rate, is determined by specifying one degree of freedom, which is called the throughput manipulator (TPM). The TPM or "gas pedal" is usually a flow but not always, and it is usually set by the operator (manual control). Some plants, e.g., with parallel units, may have more than one TPM. The TPM is usually at a fixed location, but to get better control (with less backoff) one may consider moving the TPM depending on the constraint region.

Definition [44]: A TPM is a degree of freedom that affects the network flow and is not directly or indirectly determined by the control of the individual units, including their inventory control.

The TPM has traditionally been placed at the feed to the plant. One important reason is that most of the control structure decisions are done at the design stage (before the plant is built) where the feed rate is considered fixed, and there is little thought about the future operation of the plant where it is likely that one wants to maximize the feed (throughput). However, the location of the TPM is an important decision that links the top-down and bottom-up part of the procedure.

Where Should the TPM ("Gas Pedal") be Located for the Process?

In principle, the TPM may be located anywhere in the plant, although the operators often prefer to have it at the feed, so this will be the default choice. From a purely steady-state point of view, the location of the TPM does not matter, but it is important dynamically. First, it may affect the control performance (backoff from active constraints), and second, as soon as the TPM has been placed, the radiation rule (Fig. 10) determines the structure of the regulatory layer.

There are two main concerns when placing the throughput manipulator (TPM):

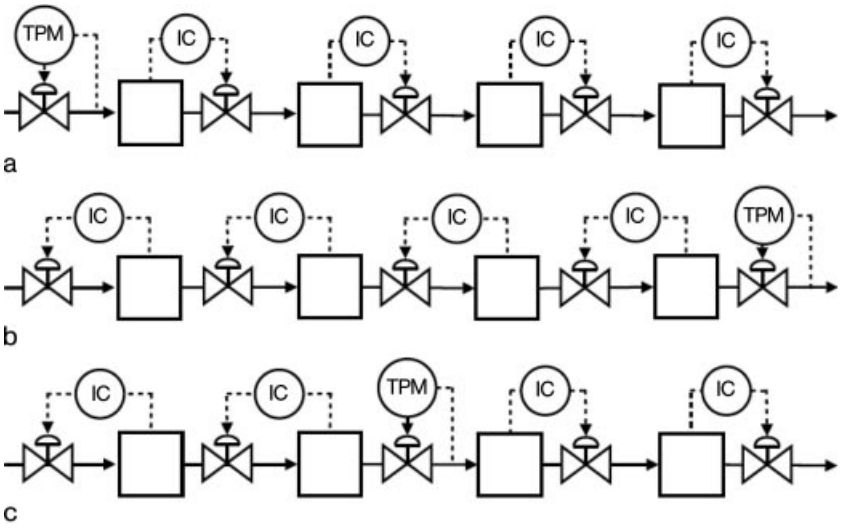

Figure 10. Radiation rule: Local consistency requires a radiating inventory control around a fixed flow (TPM) [43, 44]
a) TPM at inlet (feed): Inventory control in direction of flow; b) TPM at outlet (on demand): Inventory control in direction opposite to flow; c) General case with TPM inside the plant: Radiating inventory control

- Economics: The location has an important effect on economics because of the possible backoff if active constraints are not tightly controlled, in particular, for the maximum throughput case where tight control of the bottleneck is desired. More generally, the TPM should then be located close to the bottleneck to reduce the backoff from the active constraint that has the largest effect on the production rate.
- Structure of regulatory control system: Because of the radiation rule [43], the location of the throughput manipulator has a profound influence on the structure of the regulatory control structure of the entire plant (see Fig. 10).

An underlying assumption for the radiation rule, is that we want "local consistency" of the inventory control system [44]. This means that the inventory in each unit is controlled locally, that is, by its own in- or outflows. In theory, one may not require local consistency and allow for "long" inventory loops, but this is not common for obvious operational reasons, including risk of emptying or overfilling tanks, startup and tuning, and increased complexity.

Most plants have one "gas pedal" (TPM), but there may be more than one TPM for plants with parallel units, splits, and multiple alternative feeds or products. The feeds usually need to be set in a fixed ratio, so adding a feed usually does not give an additional TPM. For example, for the reaction A+B→ C, we need to have the molar ratio F_A/F_B close to 1 to have good operation with small loss of reactants, so there is only one TPM even if there are two feeds, F_A and F_B.

If only a part of the process is considered, then this part may have no TPM. Instead, there will be a given flow, typically a feed or product, that acts as a disturbance on this part process, and the control system must be set up to handle this disturbance. One may also view this as having the TPM at a fixed location. For example, for a utility plant the product rate may be given and in an effluent treatment plant the feed rate may be given. On the other hand, a closed recycle system, like the amine recycle in a CO_2 gas-treatment plant, introduces an extra TPM.

Moving the TPM During Operation. Preferably, the TPM should be in a fixed location. First, it makes it simpler for the operators, who usually are the ones who set the TPM, and, second, it avoids switching of the inventory structure, which should be "radiating" around the TPM (Fig. 10). However, since the TPM in principle may be located anywhere, it is tempting to use its location as a degree of freedom and move it to improve control performance and reduce backoff. The following rule is proposed:

To get tight control of the new active constraint and achieve simple switching, locate the TPM "close" to the next active constraint (such that the TPM can be used to achieve tight control of the constraint when it becomes active).

The rule is based on economic considerations with the aim of simplifying the required switching when the next capacity constraint becomes active. However, moving the TPM may require switching regulatory loops, which is usually not desirable.

Step S5. Select the Structure of Regulatory (Stabilizing) Control Layer. The main purpose of the regulatory layer is to "stabilize" the plant, preferably using a simple control structure with single-loop PID controllers. "Stabilize" means that the process does not "drift" too far away from acceptable operation when there are disturbances. The regulatory layer is the fastest control layer, and is therefore also used to control variables that require tight control, like economically important active constraints (recall the "squeeze and shift" rule, see step S3). In addition, the regulatory layer should follow the setpoints given by the *supervisory layer* (see below).

The main decision is step S5 are to (i) select controlled variables (CV_2) (decision 2) and (ii) to select inputs (valves) and "pairings" for controlling CV_2 (decision 4). Interestingly, decision (i) on selecting CV_2 can often be based mostly on steady-state arguments, whereas dynamic issues are the primary concern when selecting inputs (valves) and pairings.

No degrees of freedom have to be "used up" in the regulatory control layer because the setpoints CV_2's are left as manipulated variables (MVs) for the supervisory layer (see Fig. 9). However, one does "use up" some of the time

window as given by the closed-loop response time (bandwidth) of the stabilizing layer.

Step S5(a) Select "Stabilizing" Controlled Variables CV_2 (Decision 2). These are typically "drifting" variables such as inventories (level and pressure), reactor temperature, and temperature profile in distillation column. In addition, active constraints (CV_1) that require tight control (small backoff) may be assigned to the regulatory layer. On the other hand, it is usually not necessary with tight control of unconstrained CV_1's because the optimum is usually relatively flat.

To select systematically the stabilizing $CV_2 = \boldsymbol{H}_2 y$, one should consider the behavior of the "stabilized" or "partially controlled" plant with the variables CV_2 being controlled (see Fig. 9), taking into account the two main objectives of the regulatory layer:

- Local disturbance rejection (indirect control of primary variables CV_1): With the variables CV_2 controlled, the effect of the disturbances on the primary variables CV_1 should be small. This is to get "fast" control of the variables CV_1, which may be important to reduce the control error (and thus the backoff) for some variables, like active output constraints
- Stabilization (minimize state drift): More generally, the objective is to minimize the effect of the disturbances on the (weighted) states x. This is to keep the process in the "linear region" close to the nominal steady-state and avoid that the process drifts into a region of operation where it is difficult to recover. The advantage of considering some measure of all the states x is that the regulatory control system is then not tied to a particular control objective (CV_1) which may change with time, depending on disturbances and prices

When considering disturbance rejection and stabilization, it is the behavior at the closed-loop time constant of the above supervisory layer, which is of main interest. Since the supervisory layer is usually relatively slow, it is again (as with the selection of CV_1) usually sufficient to consider the steady-state behavior when selecting CV_2 (however, when selecting the corresponding valves/pairings in step 5b, dynamics are the key issue).

Step S5(b) Select Inputs (Valve) for Controlling CV_2 (Decision 4). Next, one needs to find the inputs (valves) that can be used to control CV_2. Normally, single-loop (decentralized) controllers are used in the regulatory layer, so the objective is to identify pairings. The main rule is to "pair close" so that the dynamic controllability is good with a small effective delay and so that the interactions between the loops are small. In addition, the following should be taken into account:

- "Local consistency" for the inventory control [44]. This implies that the inventory control system is radiating around the given flow
- Tight control of important active constraints (to avoid backoff)
- Variables (inputs) that may optimally saturate (steady-state), should be avoided as MVs in the regulatory layer, because this would require either reassignment of regulatory loop (complication penalty), or backoff for the MV variable (economic penalty)
- Reassignments (logic) in the regulatory layer should be avoided. Preferably, the regulatory layer should be independent of the economic control objectives (regions of steady-state active constraints), which may change depending on disturbances, prices, and market conditions. Thus, it is desirable that the choices for CV_1 (decision 1) and CV_2 (decision 2) are independent of each other.

In order to make the task more manageable, the choice of the regulatory layer structure, may be divided into step S5.1: Structure of inventory control layer (closely related to step S4) and step S5.2: Structure of remaining regulatory control system, but we here consider them combined.

Step S6. Select Structure of Supervisory Control Layer. The supervisory or "advanced control" layer has three main tasks:

Task 1. Control the Primary (Economic) Controlled Variables (CV_1) using as MVs the setpoints to the regulatory layer plus any remaining ("unused") valves (see Fig. 9).

- The supervisory layer may use "dynamic" degrees of freedom, including level setpoints, to improve the dynamic response (at steady-state these extra variables may be "reset" to their ideal resting values)
- The supervisory layer may also make use of measured disturbances (feedforward control)
- Estimators: If the primary controlled variables (CV_1) are not measured, typically compositions or other quality variables, then "soft sensors" based on other available measurements may be used for their estimation. The "soft sensors" are usually static, although dynamic state estimators (Kalman filter, moving horizon estimation) may be used to improve the performance. However, these are not common in process control, because the supervisory layer is usually rather slow

Task 2. Supervise the Performance of the Regulatory Layer. The supervisory layer should take action to avoid saturation of MVs used for regulatory control, which otherwise would result in loss of control of some "drifting" variable (CV_2).

Task 3. Switch Controlled Variables and control strategies when disturbances or price changes cause the process to enter a new region of active constraints.

Implementation. There are two main alternatives in terms of the controller used in the supervisory layer:

- "Advanced single loop control" = PID control with possible "fixes" such as feedforward (ratio), decouplers, logic, selectors and split range control (in many cases some of these tasks are moved down to the regulatory layer). With single-loop control an important decision is to select pairings. Note that the issue of finding the right pairings is more difficult for the supervisory layer because the interactions are usually much stronger at slower time scales, so measures such as the relative gain array (RGA) may be helpful.
- Multivariable control (usually MPC). Although switching and logic can be reduced when using MPC, it cannot generally be completely avoided. In general, it may be necessary to change the performance objective of the MPC controllers as we switch regions.

Step S7. Structure of (and Need for) Optimization layer (RTO) (Related to Decision 1). The task of the RTO layer is to update the setpoints for CV_1, and to detect changes in the active constraint regions that require switching the set of controlled variables (CV_1).

In most cases, with a "self-optimizing" choice for the primary controlled variables, the benefits of the RTO layer are too low to justify the costs of creating and sustaining the detailed steady-state model which is usually required for RTO. In addition, the numerical issues related to optimization are very hard, and even off-line optimization is difficult.

3.5. Comparison of the Procedures of Luyben and Skogestad

The most striking difference between the two procedures is that whereas the Skogestad procedure starts with economics (part I), the Luyben procedure does not explicitly include economics, except at the very last stage.

Step L1. Establish Control Objectives. By "control objectives", Luyben means the primary CVs but the Luyben procedure is unclear about how these should be selected. It is stated that "this is probably the most important aspect of the problem because different control objectives lead to different control structures", but the only guideline is that "these objectives include reactor and separation yields, product-quality specifications, product grades and demand determination, environmental restrictions, and the range of safe operating conditions."

In the Skogestad procedure, the first step is to define the cost function and the process constraints (step S1) and optimize the operation (step S2). The selection of CVs follows from this (step S3). The first thing is to control the active constraints. This will generally include product-quality specifications on valuable products (cheap products should often be overpurified to avoid losses of more valuable components), minimum product rates (demands), environmental and safety constraints, pressure and temperature constraints, and so on. For output constraints one may have to introduce a safety

factor ("backoff") which will imply an economic loss. To reduce the backoff for hard output constraints one wants tight control, which may imply that some of these variables are controlled in the regulatory layer.

Step L2 (and Step S2a). Determine Control Degrees of Freedom. This is an important step in both procedures, but in the SKOGESTAD procedure it comes before the selection of CVs, which is reasonable because we need to identify one CV for each degree of freedom. In addition, in SKOGESTAD'S procedure one distinguishes clearly between the steady-state degrees of freedom (step S2a) and the physical degrees of freedom (valves, step S5b).

LUYBEN states that most of the control degrees of freedom (valves) are used to achieve basic regulatory control of the process: "(i) set production rate, (ii) maintain gas and liquid inventories, (iii) control product qualities, and (iv) avoid safety and environmental constraints". He adds that "any valves that remain after these vital tasks can be utilized to enhance steady-state economic objectives or controllability".

This is in agreement with the SKOGESTAD procedure. Many of these variables are related to optimal active constraints. Control of gas inventories (pressures) is usually required to stabilize the plant (avoid drift), but note that one does not really "consume" any degrees of freedom here because the pressure setpoint can be used as a degree of freedom for effecting the economic (steady-state) operation. With liquid inventories (levels) the situation is a bit different because many liquid levels do not have a steady-state effect.

Step L3. Establish Energy Management System. It seems a bit unclear why this issue is so high up on the list in the LUYBEN procedure and what is so special about control of the energy system. Of course, an unstable exothermic reactor needs to be stabilized and selecting an appropriate sensitive variable (typically, a temperature) and pairing it with an input (MV) will be one of the first issues when designing the regulatory control system (step S5). However, since stabilizing control does not "use up" any degrees of freedom at steady-state, this may not be in conflict with the objectives of optimal economic operation, which is the third step (or actually Step S2b) in SKOGESTAD'S procedure.

Step L4 (= Step S4). Set the Production Rate. Note that in this work the terms "production rate" and "throughput" mean the same. As discussed in detail above, the location of the throughput manipulator (TPM) is very important, both for economic reasons (steady-state) and for dynamic reasons. For economic reasons, it should be close to the bottleneck in order to reduce the backoff when it is optimal to maximize production (sellers market) [39]. Dynamically, it determines the structure of the inventory control system, which is "radiating" around the TPM [43].

Traditionally, the main process feed has been selected as the "gas pedal" (TPM). However, LUYBEN et al. [29] recommend to locate it close to the reactor: "Establish the variables that dominate the productivity of the reactor and determine the most appropriate TPM". Again, the reasoning for focusing on the reactor is a bit unclear, and it is worth mentioning that the location of the TPM is also an important decision in plants with no reactor, like a gas processing plant [45]. Nevertheless, the reactor is obviously an important unit and will often be the bottleneck for the process. In addition, there is usually recycle around the reactor, and by locating the TPM in this recycle loop one can avoid the "snowball effect" and satisfy LUYBEN'S rule L6 of "fixing a flow in every recycle loop".

Step L5. Control Product Quality and Handle Safety, Environmental and Operational Constraints. LUYBEN says that we should "select the best variables to control each of the product-quality, safety and environmental variables", but he does not state what "best" is. He adds that "we want tight control of these important quantities for economic and operational reasons" which makes sense if these variables are optimally active constraints. Having performed an economic optimization (step S2b), it is then also easy to determine what these "best" variables are: They are active constraints when we operate the plant such that cost is minimized.

LUYBEN adds that "it should be noted that establishing the product-quality loops first,

before the material balance control structure, is a fundamental difference between our plantwide control procedure and BUCKLEY'S [23] procedure".

In this respect, the SKOGESTAD procedure is something in between the LUYBEN and BUCKLEY procedures. Similar to LUYBEN, it starts by identifying which variables should be controlled, which typically includes some active product-quality constraints. However, similar to BUCKLEY, in the SKOGESTAD procedure the design of the actual control system, including choice of pairings, starts with the "stabilizing" loops, including the material balance (inventory) control although it is recommended that control of active constraints that required tight control for economic reasons should be assigned to the regulatory layer.

Step L6. "Fix a Flow in Every Recycle Loop and Control Inventories". The recycle split adds a degree of freedom to the process, so it is possible to fix a flow in every recycle loop. This may be a good strategy from a regulatory and dynamic point of view, but not generally from an economic point of view. For example, if the throughput in the plant is increased then it will generally be economically optimal to increase all flows, including the recycles.

LUYBEN argues that fixing a flow in the recycle loop avoids the "snowball effect" where the recycle flow grows out of bound [46]. Note that the "snowball effect" is caused by having a unit, typically a reactor, which is too small compared to the desired throughput. This means that we are operating at high throughputs where the unit indirectly is the bottleneck of the process.

The systematic approach would be to perform an economic optimization with the throughput as a degree of freedom (step S2b), and from this the optimal control policy will follow (step S3), which will give the optimal way of handling the "snowballing". In some cases, maximum production may correspond to maximal recycle, which means that "snowballing" is optimal and the recycle flow should simply be fixed at its maximum (e.g., maximum gas recycle [29, 47]). In other cases, maximum recycle is not optimal because other constraints are reached, and one needs to use a flow in the recycle to control an optimally active constraint (e.g., use the column feed flow to control product composition in the simple recycle system studied by [46] and later by [48]). In yet other cases, there may be an "optimal maximum throughput" and one needs to identify a self-optimizing variable associated with the feed being an unconstrained degree of freedom.

Nevertheless, LUYBEN is obviously right that with the TPM located outside the recycle loop, and with all the flows inside the recycle loop on inventory (level or pressure) control, one may get "snowballing" inside the recycle loop if we feed more into the loop than its units can handle. "Snowballing" is clearly a "drifting mode" and it is a task of the regulatory control system to avoid drift (step S5). Snowballing is caused by the positive feedback in the recycle loop, and one way to break this loop is to follow LUYBEN and fix a flow in the recycle loop (including selecting it as a TPM). This forces the excess feed to exit the recycle loop. Another option, which is likely to be better economically, is to use one of the flows in the recycle loop to control some other variable, like a sensitive temperature or composition.

In summary, the importance of the "snowball" effect has probably been overemphasized in some of the literature on plantwide control. If it is actually a "problem", then it cannot be economically optimal, so it will automatically be avoided by following the procedure of SKOGESTAD. Nevertheless, one should be aware of the "snowballing" that may occur if all the flows inside the recycle loop are on level or pressure control.

L7. Check Component Balances. "Identify how chemical components enter, leave, and are generated or consumed in the process (downs drill). This is a very important issue, even for processes without reactions, and is included in step S5 (regulator control strategy) in the SKOGESTAD procedure.

L8. Control Individual Unit Operations. This step seems a bit arbitrary, as application of the previous steps will "automatically" lead to control of the individual units. Of course, it can be useful to compare the resulting control structure with common rules of thumb for individual units and consider changes if

it seems unreasonable. The SKOGESTAD procedure contains many steps where choices are made, so some iteration may be needed.

L9. Optimize Economics and Improve Dynamic Controllability. LUYBEN writes that "after satisfying all of the basic regulatory requirements, we usually have additional degrees of freedom involving control valves that have not been used and setpoints in some controllers that can be adjusted. These can be utilized either to optimize steady-state economic process performance or to improve dynamic response." This statement is true, but it is better to consider the economics much earlier. First of all, an economic analysis is generally needed in order to identify the optimal active constraints (in step L1), so one may as well identify good self-optimizing CVs for the remaining unconstrained degrees of freedom. Second, if one knows the self-optimizing variables, then one can take this into account when designing the regulatory control system.

3.6. Conclusion

Control structure design deals with the structural decisions of the control system, including what to control and how to pair the variables to form control loops. Although these are very important issues, these decisions are in most cases made in an ad hoc fashion, based on experience and engineering insight, without considering the details of each problem. Therefore, a systematic procedure for control structure design for complete chemical plants (plantwide control) is presented. It starts with carefully defining the operational and economic objectives, and the degrees of freedom available to fulfil them. Then the operation is optimized for expected future disturbances to identify constraint regions. In each region, one should control the active constraints and identify "self-optimizing" variables for the remaining unconstrained degrees of freedom. Following the decision on where to locate the throughput manipulator (TPM), one needs to perform a bottom-up analysis to determine secondary-controlled variables and a structure of the control system (pairing).

4. Process Control of Batch Processes

4.1. Introduction

A large number of products in the chemical industries are made in multiproduct and multipurpose plants containing batch reactors and other batch processes. Biological and biochemical processes, for example, which play an increasing role in the production of fine chemicals and pharmaceuticals, are almost exclusively produced in batch or semibatch (fed-batch). Reasons for using batch processes are that in batch reactors higher conversion can be reached compared to continuous stirred-tank reactors and that the throughput can be varied without varying the residence time. The flexibility of batch processes is in general much higher than for continuous production, for example, the production of different grades or different products in the same equipment is possible. Scale-up from the laboratory is often easier as it does not require changes of the types and the sequence of operations. Batch processes are robust to inaccurate and insufficient knowledge. A typical example is emulsion copolymerization which is very complex to model but has been operated successfully in the industry since the 1940s [49–52]. Furthermore, in contrast to continuous processes, solids can be handled more easily in batch processes.

The flexibility of batch plants may, however, lead to more unproductive periods of time and possible product cross-contamination. Control is often more challenging in comparison to continuous plants, as there is no fixed operating point. The need for planning and scheduling of production sequences, cleaning steps, and changeovers is typical for batch processes.

Batch processes often involve transformations which are substantially more complex than those realized in continuous processes, and comprehensive models of the processes often are not available. Therefore, the operation of such processes is to a large extent based on experience, and additional processing time or additional processing steps are used if the analysis of the product reveals quality problems. Such additional steps and varying batch times increase the complexity of planning and scheduling. Today, due to tougher competition,

advances in modeling and increasing computing power, model-based methods are increasingly employed to operate batch processes efficiently.

A definition of a batch process is given by [53]: A batch process is a process that leads to the production of finite quantities of material by subjecting quantities of input materials to an ordered set of processing activities over a finite period of time using one or more pieces of equipment. Semibatch processes where one or more substances are fed or withdrawn continuously during (part of) a batch run are a special and important class of batch processes. A batch plant is a chemical plant that contains one or more operations performed in batch.

Batch processes are defined by recipes. A recipe is the necessary set of information that uniquely defines the production steps that are needed to produce a specific product. It contains the amounts of raw materials and the processing instructions to make the product [53].

For batch process management, NAMUR and ISA have developed recommendations (NE33 [54] and NE59 [55]) and standards (S88 [53] and S95 [56]). Theoretical foundations and application examples of the optimization of single-batch runs are well covered in the open literature.

4.2. Batch Process Management

4.2.1. Recipe-Driven Operation Based on ANSI/ISA-88 (IEC 61512-1)

The US standard ANSI/ISA-88.01 "Batch Control" [53] which has been extended and accepted as the international standard IEC 61512 "Batch Control" [57, 58] proposes a standard batch control architecture and recommended practice for the implementation of batch control systems. It is based on the earlier NAMUR recommendation NE33 [54].

The standard defines a terminology that facilitates the understanding between the developers of *batch-control solutions* and the end users using the concept of *recipe-driven production* that describes how batch plants can be operated in a flexible, yet efficient manner. Figure 11 gives an overview of the models defined in IEC 61512-1 [57]. The layers in the figure are hierarchical and each entity may contain many instances of the entities of the lower layers and of the same layer. For example, an equipment module can be made up of many equipment and control modules.

The standard IEC 61512-1 defines three models, the process model, the physical model, and the control model.

The process model assists the engineer to answer the questions "What should be produced?" and "How should it be produced?" IEC 61512-1 defines that the process should be divided into *process stages*. Subdivisions of a process stage are one or more *process operations* with major processing activities. *Process actions* contain minor processing activities that are combined to realize a process operation. The process and the subprocesses are defined independent of the configuration of the actual equipment where the process is realized. Figure 11C shows on the right-hand side an example of a process with three different types of raw materials (A, B, C) and one product (D).

The physical model assists the engineer to answer the question "Where should it be done?" A batch process is run in a plant (called *process cell* in IEC 61512-1). In order to be able to map the requirements of the process model to the equipment it is important to divide the process cell into *units* based upon the piping and instrumentation diagram (P&ID) → Chemical Plant Design and Construction, Section 3.4.4. The subdivision is depicted in Figure 11B. The units may contain *equipment modules*. Equipment modules contain equipment and at least one actuator and possibly a control loop with one or more sensors. If an equipment module is complex, it might be helpful to divide it further into different *control modules*.

The standard considers three further layers (*enterprise*, *site*, and *area*) above the process cell that are not shown in Figure 11 because they are not directly relevant for batch-process control. The reader is referred to the recommendation NE59 [55] or the standard ANSI/ISA-95 [56]. These layers are addressed by enterprise resource planning (ERP) and manufacturing execution systems (MES). These systems offer functions such as recipe management, production planning, and overall equipment effectiveness monitoring.

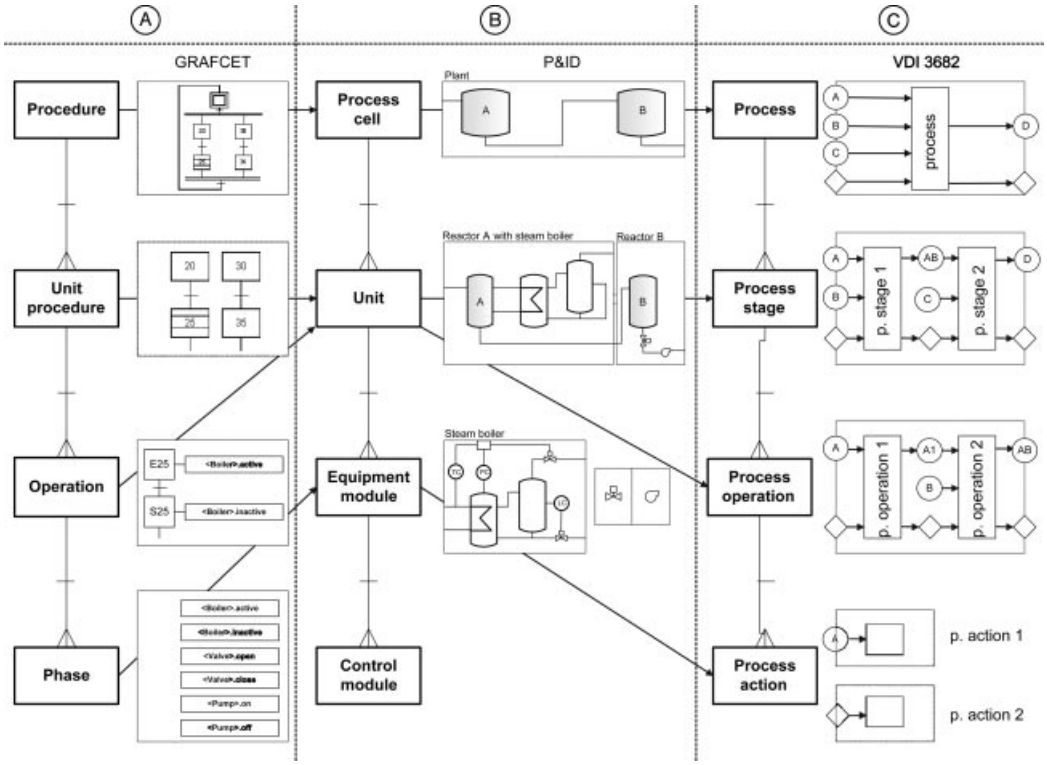

Figure 11. Models for the description of batch processes according to IEC 61512-1 [57]
A) Control model; B) Physical model; C) Process model

The control model assists the engineer to answer the question "What exactly should be done?" (Fig. 11C). The control model is built from *procedures*. The procedure is the highest level in the hierarchy and defines the strategy to perform the desired transformations, for example, "production of polystyrene". Procedures can be built using the standard EN 60848 GRAFCET or sequential function charts (SFC: IEC 61131-3). *Unit procedures* consist of an ordered set of operations to take place within a unit. An example for an operation would be "polymerization". *Operations* define major processing sequences, preferably between points where the process can be suspended such as "heat" or "charge". A *phase* specifies the exact commands that are sent to the controlled equipment and the conditions for these commands in terms of sensor readings or time elapsed. A sequence of phases realizes an operation.

This multiple hierarchical structure facilitates the understanding of the processes and in particular the use of the same equipment and the same control routines in different recipes as well as the use of different pieces of equipment to perform the same part of a recipe.

4.2.2. Recipes

IEC 61512-1 [57] defines a recipe as "the necessary set of information that uniquely defines the production requirements for a specific product" and recommends a hierarchy of recipes as shown in Table 1 to reduce the complexity to manageable parts and to maintain coherence of the different recipes. The hierarchical layers correspond to the scale of the application. Table 1 also explains the content that is expected on the different levels of the hierarchy.

Using all four types of recipe is the most complex situation in a distributed enterprise and not always necessary. Master and control recipes are used in most automated batch plants.

Table 1. Hierarchy of recipes

Recipe-type	Level
General recipe—	enterprise level:
	• basis for one or more site recipes
	• contains the processing information for a specific product
	• basis for enterprise-wide planning
Site recipe—	site level:
	• basis for one or multiple master recipes that are specific for process cells
	• contains site-specific data like raw material qualities, formulae which cover input–output material relationships
	• may be defined independently or be a refinement of a general recipe
Master recipe—	plant level:
	• defines the procedure depicted in Figure 11 for a specific processing cell
	• may be defined independently or be a refinement of a site recipe which is adjusted to a specific process cell
	• does not contain all information needed to produce a specific batch
Control recipe—	batch level:
	• contains all information to produce a batch of material in a processing cell.
	• is executed by the batch controller
	• may be changed during production by the operator to work around equipment or quality problems
	• is a refinement of a master recipe which is extended by data as, e.g., the set of equipment, the amount of material, and a unique batch ID

4.2.3. Control Hierarchy

The extended control hierarchy depicted in Figure 12 shows a hierarchical view of the control systems that are involved in the operation of a batch process and their interactions.

Safety control, often implemented in a fail-safe programmable logic controller PLC or emergency shutdown system (ESD), has the highest priority of control to guarantee the safety of humans, machines and the environment. Basic logic control implements elementary sequences and operational interlocks. Basic regulatory control establishes the desired processing conditions. It obtains the set points from advanced process control, from the recipes or from the operators. Sequential control establishes the logic of the control model, for example start-up and shut-down procedures. The production procedure is defined in the control recipe which is executed in the batch controller which can be part of the distributed control system (DCS), the MES, or even the ERP system.

4.2.4. Sequential and Logic Control

Control of batch processes is dominated by sequential control. Due to the nature of the process it is necessary to go through a number of steps from start to end. The output and the next state of a sequential control system depend on the current input as well as on its internal state. Practical implementations use SFC of GRAFCET as a graphical programming language. SFC is based on three elements: “steps”, “actions”, and “transitions” (see Fig. 13). GRAFCET or SFC can be used to specify the recipes as well as to specify the detailed control sequences.

4.2.5. Regulatory Control

Regulatory control in batch plants has to cope with continuously or abruptly varying behavior of the process. Often the sequential controllers change set-points, measurement and output ranges, or the parameters of

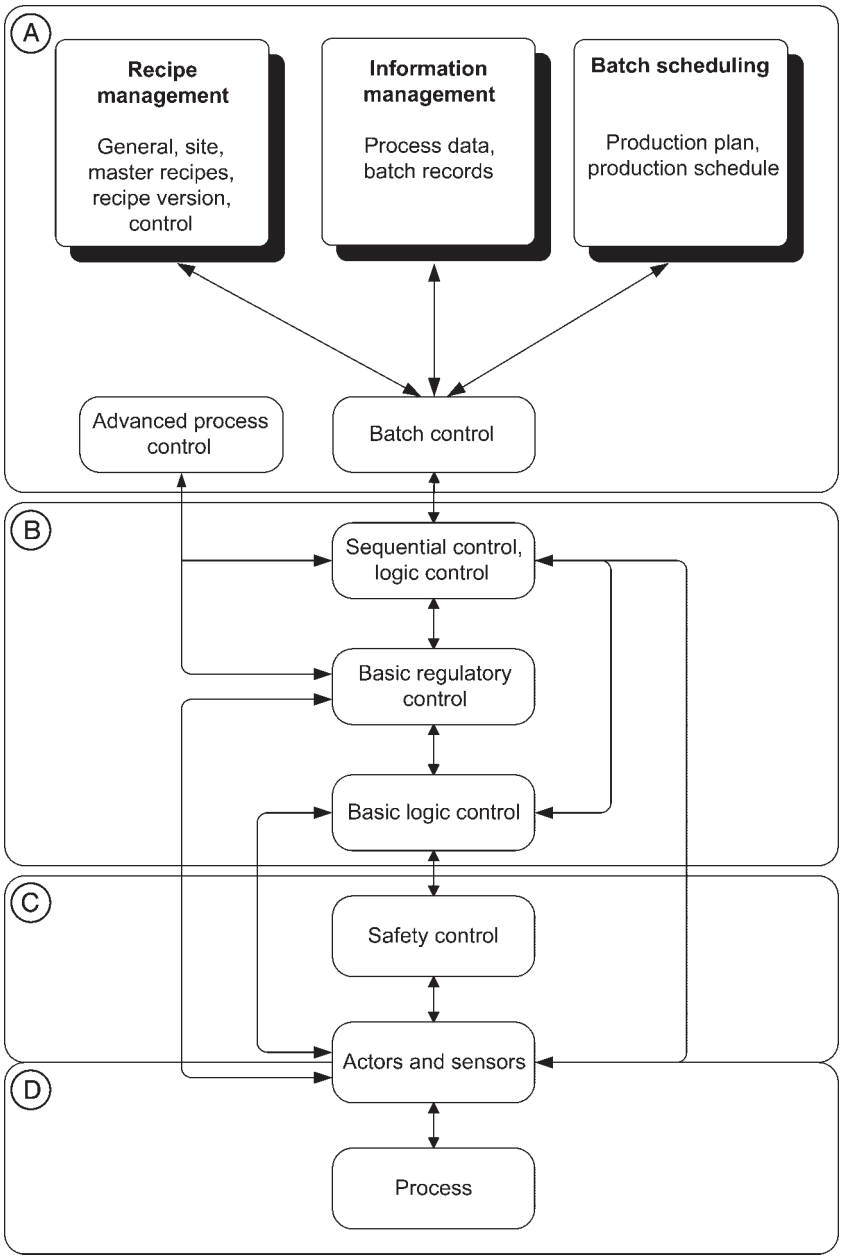

Figure 12. Extended control hierarchy: A control-oriented view on control systems in a chemical plant (expanded from [59]) A) Application station; B) DCS or PLC; C) Fail safe PLC; D) Process

proportional–integral–derivative (PID) controllers. Typically more disturbances are encountered in a batch plant compared to a continuous plant due to the variability and the discontinuity of the process.

From a controller design point of view, a time-varying behavior is added because batch and semibatch processes have no single operating point but follow a trajectory with constantly changing conditions of the reaction

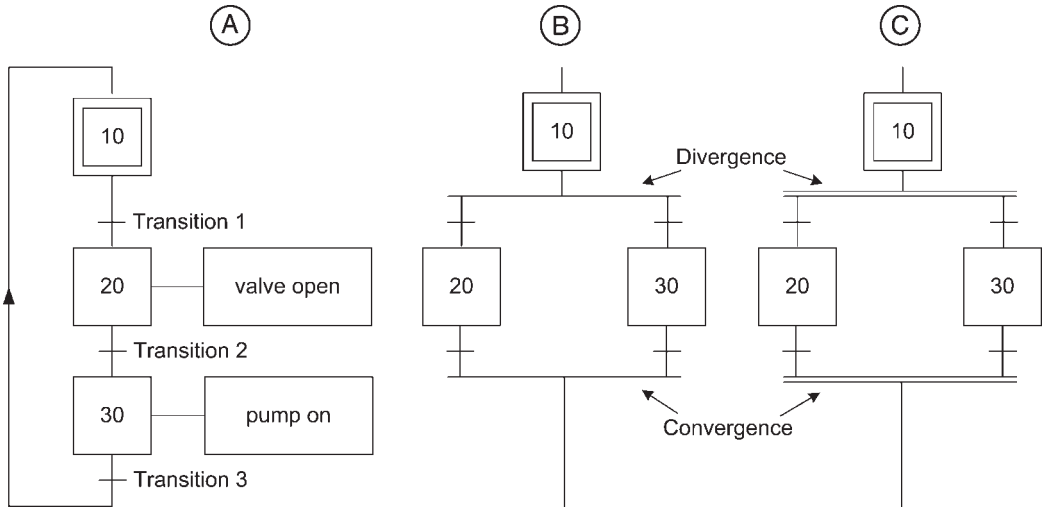

Figure 13. A simple SFC loop and transition types
A) Single-loop structure; B) Sequence selection; C) Simultaneous sequence

or the separation along which the parameters of linear controllers often need to be adjusted.

4.2.6. Planning and Scheduling in Multipurpose and Multiproduct Plants

Planning and scheduling is very important for multipurpose and multiproduct plants. In these plants several different kinds of recipes can be executed on different pieces of equipment to produce different products for varying market demands, sequentially and in parallel. This leads to a large number of decisions that have to been taken. Planning usually involves:

- Forecasting: Orders for raw materials with large lead-times are issued at the corporate level based on the forecast of the demand for products.
- Assignment of the orders of products: Orders are normally received at the corporate level and are then assigned to individual plants for (partial) production and shipment.
- Batch sizing and campaign planning: The number and sizes of the individual batches are determined and it is decided whether the production is performed in a campaign mode (many similar batches are produced in a sequence before a changeover to another product is made).

4.3. Quality Control and Batch-Process Monitoring

4.3.1. Measurement and Control of Quality Parameters

Batch processes are advantageous if robustness to unknown or unmeasured influences is required. This is often the case in chemical processes that are not well understood or in bioprocesses during which the microorganisms do not behave reproducibly. Measuring standard and nonstandard properties during and at the end of a batch is of crucial importance for batch control.

Typical measurements in chemical and biochemical processes that are available during the batch run are temperatures, pressures, volume or liquid level, and volumetric or mass flow rates. Quality measurements such as concentrations are often not available and the equipment required for such measurements is normally less robust and much more expensive than for standard measurements. This is why even for complex reaction systems quality measurements are often performed only at the end of the batch or even only every few batches by taking samples that are sent to laboratories for analysis. If the product does not meet the specifications, additional processing steps are required or the problem is solved

by blending it with products from different batches.

Typical examples of quality variables are:

- Concentrations and conversion in chemical and biochemical processes, for example, residual monomer content in polymers or stereochemical composition for pharmaceuticals.
- pH-value, conductivity, and other electrochemical measurements in applications that use or reduce acidic compounds.
- Polymer chain length distributions in polymerization processes.
- Particle size distributions in emulsion and dispersion-based processes such as emulsion polymerization or crystallization.

The measurement of quality variables often requires long measurement times and expensive equipment. Nowadays sensing technology is available that cover most quality parameters of interest. Examples are:

- Optical measurement techniques such as refractometers
- Paramagnetism for oxygen detection
- Gravimetry or coriolis meters for density measurement
- (Gas) chromatography (GC) or high-performance liquid chromatography (HPLC) for compositions
- Ultrasound
- Spectroscopic measurement devices (e.g., NIR-, IR-, NMR-, Raman-, UV-spectrometers)

Many of the above devices do not measure the quality variables directly and therefore need to be calibrated extensively, often using standards and statistical techniques. In many cases only samples are taken and analyzed in a laboratory. Due to the resulting time delay, such measurements can typically not be used for the control of the current batch but in batch-to-batch (or run-to-run) control.

The location of sensors in batch processes can be classified into *in situ*, *bypass*, and *sample*. In situ and bypass measurements are usually nondestructive whereas sampling always removes some product from the process. Examples for in situ sensors are a Raman probe or a pH probe in a reactor, for bypass a coriolis densitometer, and for sampling a gas chromatograph using a removed sample.

Table 2. Typical time delays of polymer quality measurements

Value measured	Device	Time delay
Conversion	gravimetry	10 min
Molecular mass distribution	gel permeation chromatography	1 h
Particle size distribution (PSD)	light scattering	10 min
Concentrations	gas chromatography	15 min
	or HPLC	10 min

The measurements can further be classified into online (in quality control any result that arrives within 60 s can be classified as on-line), automatic (also classified as at-line) where the result arrives automatically after a certain specific period of time without intervention, and off-line or laboratory, where samples are removed and analyzed in a separate location. For process control the sampling period and the time delay are important aspects.

Time delays as shown in Table 2 may pose a serious problem in the control of batch processes with short batch times. Alternatives with virtually no time delay are integral measurements, such as ultrasound and density measurements. The optimal device is one that provides the parameters of interest on-line and in situ while being inexpensive to buy and to run. Practically a compromise has to be found between the cost for installation and maintenance of the equipment, the effort for calibration, and the benefit of the measurement for the control strategy.

An alternative to the use of in situ measurements is the use of inferential measurements that employ process models. They can be realized as static maps (due to the time-varying nature of batch processes this can be problematic), as dynamic state estimation techniques, or by statistical methods.

4.3.2. Inferential Measurements

Measurements that are only available with large time delays but also quality indicators that can only be measured in a laboratory may be replaced by inferential measurements. "Inferential measurement" is the general term for quality parameters that are predicted using a

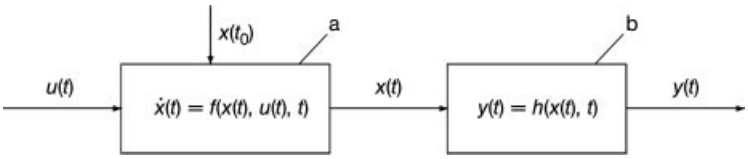

Figure 14. Block diagram of a state space system
a) Process; b) Sensors

linear or nonlinear model that is calibrated using measured process and laboratory data [60, 61].

The models are often black-box models → Biotechnology, 5. Monitoring and Modeling of Bioprocesses, Section 5.3. If a linear approach is sufficient, statistical techniques such as principle component analysis (PCA) and projection on latent structures (PLS) are applied. Nonlinear black-box models can have many forms; examples are neural networks or fuzzy logic.

Rigorous first-principles models can also be used but are often very complex and expensive to derive. Especially in batch processes the underlying phenomena might not even be known at all. Sufficiently accurate inferential measurements are usually more quickly obtained using data-based black-box model fitting. If process knowledge in the form of a physical model is available but some complex parts of the process are not well understood, black-box models can also be combined with physical models to create so-called grey-box models, for example, in emulsion polymerization [62].

4.3.3. State Estimation

In control theory, process parameters such as temperatures, pressures, volumes, and concentrations that change their values over time are called *process states*. The vector of the state variables is $\boldsymbol{x}(t)$, and $\boldsymbol{u}(t)$ is the vector of the manipulated variables of the process.

Figure 14 shows a process model as a block diagram. Some of the state variables can be measured easily (but the measurement devices are subject to disturbances) and some, for example, concentrations, compositions, molecular mass and particle size distributions, cannot–with justifiable effort–be measured on-line or can not at all be measured.

State estimation is a method to filter existing measurements and to estimate unknown and unmeasurable states of a system. If a good dynamic model (physical- or data-based) of the process exists, the easily available measurements can–under certain conditions–be used to estimate the unmeasured states. Figures 15 and 16 show two possible concepts of state estimation in the form of block diagrams. The superscript $\wedge$ indicates the estimated states.

Open-Loop Observer. The differential or algebraic states of a system that can be directly measured are used as inputs to a dynamic model which is simulated in parallel to the real process (Fig. 15).

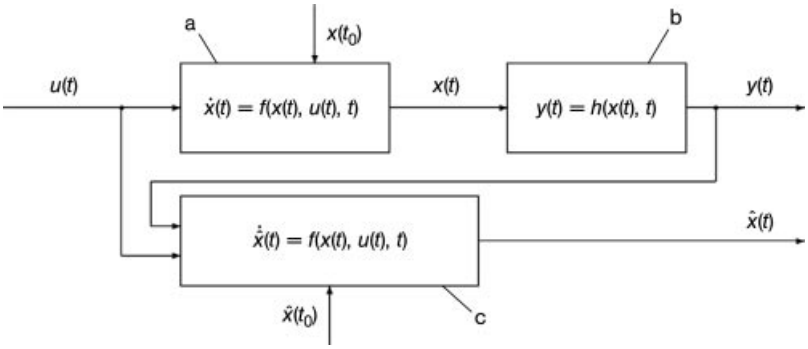

Figure 15. Block diagram of the open-loop observer
a) Process; b) Sensors; c) (Reduced) process model

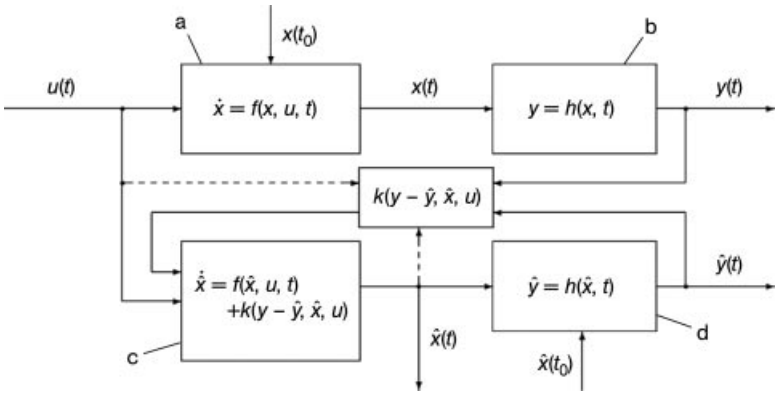

Figure 16. Block diagram of the closed-loop observer (dashed lines for nonlinear systems)
a) Process; b) Sensors; c) Process model; d) Sensor model

A typical example of an open-loop observer is the use of the estimated heat of reaction in a polymerization reactor to estimate the monomer concentrations [63]. The heat of reaction of the exothermic process is calculated using reaction calorimetry (see Section 4.3.4). This information is then used in a dynamic model of the process which is simulated in parallel to the process in order to predict the monomer concentrations.

Such an observer will follow the real states well if the initial conditions are known and the model is correct. Measurement noise may be amplified by the described concept and errors in the initial conditions and plant-model-mismatch lead to wrong estimates [64, 65].

Closed-Loop Observer. A (closed-loop) state estimator also simulates the process under consideration but the errors between the measured states $y(t)$ and the predicted measured states $y[\hat{c}](t)$ are used to correct the estimation as shown in Figure 16. The major results on linear state estimation were developed by [66–68]. Methods to check the observability of linear systems can be [66, 69, 70].

If the system is linear and observable and plant-model-mismatch is small, the correction term $k(y-\hat{y}, \hat{x}, u) = \boldsymbol{K}\cdot(y-\hat{y})$ which is linear in the error $y-\hat{y}$ will result in convergence of the estimated states to the true values. This observer is called the *Luenberger Observer*. The speed of convergence can be adjusted by the choice of the gain matrix $\boldsymbol{K}$. Large error gains give fast convergence but also amplify the measurement noise. The compromise between fast convergence and noise amplification is explicitly handled by the *Kalman filter* (→ Biotechnology, 5. Monitoring and Modeling of Bioprocesses, Section 5.3) which has the same structure as the observer in Figure 16. Here $\boldsymbol{K}$ is computed from the covariance matrices of normally distributed random disturbances that are assumed to act on the measurements (measurement noise) and on the evolution of the states (state noise).

In the nonlinear case the choice of $k(y-\hat{y}, \hat{x}, u)$ is by no means trivial. Frequently used solutions are based on linearizations around a fixed operating point yielding a linear observer, or on linearizations around the estimated state, for example, the extended Kalman filter (EKF, [71, 72]). As estimators based upon linearizations are difficult to tune and may even fail for strongly nonlinear plants, direct fully nonlinear estimation schemes have been proposed, for example, the moving horizon estimator (MHE, [73, 74]). Another option is the use of the unscented Kalman filter [75] or particle filters [76].

A practical example for the application of an EKF to a batch process is the estimation of the heat of reaction and of the heat-transfer coefficient simultaneously by heat-balance calorimetry.

4.3.4. Calorimetry

In monitoring and control of batch reactors, reaction calorimetry is widely used as a tool to

determine the heat that is produced by the reaction at a certain point in time → Thermal Analysis and Calorimetry, Chapter 2. The heat of reaction can be used in combination with the mass balances to determine the instantaneous and the cumulative conversion of the different species. For the safe operation of exothermic reactions, knowledge about the heat-transfer coefficient which governs the rate of energy transferred between the reactor content and the jacket is also crucial.

Classical reaction calorimetry is based on an energy balance of a perfectly mixed tank reactor. In real reactors isothermal conditions are desired, as most reactions can be run best at a certain temperature, but not always present, especially during the start-up phase. The calorimetric methods applied to production reactors can be subdivided into *heat-flow calorimetry* in which the jacket temperature dynamics are not considered and *heat-balance calorimetry* which incorporates the jacket temperature dynamics.

In heat-flow calorimetry only the differential equation for the reactor temperature is used to estimate the heat of reaction. In order to apply heat-flow calorimetry, the evolutions of the heat-transfer coefficient and of the heat-transfer area need to be known precisely.

In heat-balance calorimetry the heat balances of both the reactor and the jacket are considered. If the reactor temperature and the temperatures of the cooling fluid at the inlet and at the outlet of the reactor as well as the coolant flow rate are measured, the heat-transfer coefficient k and the heat of reaction can be estimated simultaneously [77–81].

In small reactors, often only heat-flow calorimetry can be used as the jacket inflow and outflow temperatures are almost identical. Similarly, the estimation of both the heat of reaction and the heat-transfer coefficient by heat-balance calorimetry fails if the temperature difference between the jacket inlet and the jacket outlet temperature is small.

In such cases an additional excitation of the temperature control system must be used. This can be done by the addition of a small sinusoidal oscillation to the reactor temperature set point and the exploitation of the gain and phase shifts between the jacket and the reactor temperatures, termed *oscillation calorimetry* [82]. An investigation of more general excitation signals can be found in [83].

When applying reaction calorimetry, practical aspects that need to be considered are:

- Heat losses: Every process that operates at a temperature higher than ambient temperature experiences heat losses. It is often possible to calibrate the reactor before the batch run and to assume that the heat losses are constant.
- Measurement accuracy and noise: For reaction calorimetry, the temperature differences are more important than the absolute values of the temperatures. If temperature sensors are calibrated well, these differences can be measured quite accurately. If the heat-transfer coefficient and the heat of reaction are estimated simultaneously, the measurement noise will be amplified because in essence the estimation is based on derivatives of the temperatures. Low pass filtering of the measurements of the results can be applied to obtain a smoother but slightly delayed estimate.
- Jacket flow rate used for control: In many applications the jacket flow rate rather than the jacket inlet temperature is used as a manipulated variable. If the range of the flow rates used is large, two aspects have to be considered. On the one hand, for small flow rates the jacket may not behave like a stirred tank and different approaches for the estimation are necessary [84]. On the other hand, the large span adds a nonlinearity that may require an on-line adjustment of the tuning of the estimator and of the controller [33].

4.3.5. Detection of Abnormal Situations and Statistical Process Control

The production of off-specification product may lead to a total loss of a batch. Advanced measurement and state-estimation methods help to detect abnormal batches early. There are, however, many batch processes where neither the necessary measurements nor the required physical models are available to employ traditional model-based control approaches. This motivated researchers to use statistical methods to monitor and control batch processes [85–90].

An established method in batch operation is (multiway) principle component analysis ((M) PCA) (→ Chemometrics, Section 9.1) and–if

one or many quality variables are measured during a batch run–the related projection on latent structures, also known as (multiway) partial least squares ((M)PLS) (→ Chemometrics, Section 11.2). These statistical techniques were first used in chemometrics and later found their way into the monitoring of continuous processes and have since been adapted to batch operation.

The use of PCA and statistical process control consists of four steps:

1. Historical data of normal or good process operation is collected and projected onto a required number of principal components using PCA
2. New process data is projected onto the subspace defined by the PCA loadings
3. The size of the projections is then compared to a predefined statistical upper and lower bound for the normal operation (normally 95% confidence bounds of HOTELLING'S T^2-distribution)
4. If the bounds are violated, contribution plots, which identify the variables with the largest influence on the observed deviation by displaying their contribution, are used to identify the possible cause of the problem and to enable the operators to take corrective actions

In batch processes the data consists of several time series of different variables for different batches. Two principal ways are possible for dealing with this problem. The difference is how the 3D data structure is unfolded and which method is used for the data analysis. Figure 17 depicts the two methods.

MPCA unfolds the 3D matrix such that the batches are sliced at each point in time. Therefore, the single batches cannot be identified directly after unfolding (Fig. 17A). One row represents the set of variables changing with time for one batch.

This way, the trajectory can be eliminated by subtracting each column mean from the

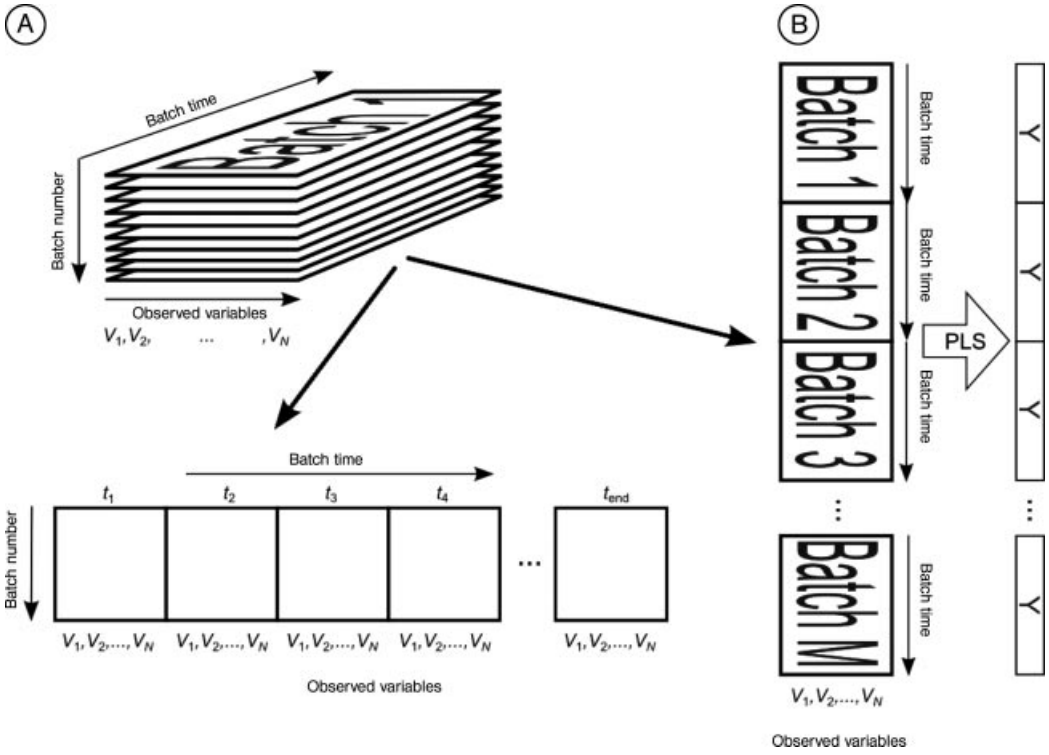

Figure 17. Unfolding of the 3D batch-data matrix
A) Multiway principle component analysis (MPCA); B) The batch fingerprint method

respective column. The matrix then contains the deviations from a mean trajectory. After variance scaling, PCA can be applied to the resulting matrix. Abnormal situation management and statistical process control in the manner described above is then possible. [85, 86]. For on-line monitoring, the empty cells of the matrix row of live data that is projected are filled with the last good values of the current batch.

The *batch fingerprint* method unfolds the matrix such that each column contains the values of one variable dependent on the batch run and the point in time in the run. Each row corresponds to a point in time in a specific batch (Fig. 17B). PCA cannot be applied to this matrix because of the autocorrelation of the data but PLS can be used to project the data onto quality variables resulting in a batch fingerprint [91]. During a batch run, the confidence band around the fingerprint should not be left. If this condition is violated, contribution plots can help show which variable is the likely cause and needs to be adjusted.

Both methods have problems with batches that are of different lengths. There are different methods to deal with this problem, an overview and a combination of both unfolding methods that also handles batches of different lengths can be found in the literature [90, 91].

Abnormal situation detection and batch-process control are powerful techniques for batch operation especially if a significant amount of historical data is available.

4.4. Optimal Operation of Single-Batch Processes

4.4.1. Trajectory Optimization

In batch processes nonlinear dependencies need to be considered more carefully than in continuous processes. Therefore, for the optimization of batch processes usually nonlinear physical models or simplifications of such models are employed. In contrast to the static optimization of steady-state operating points in continuous processes, optimization of batch processes always involves solving a dynamic constrained optimization problem. For batch processes this optimization is called *trajectory optimization*. As batch processes are transient, the whole trajectory must be optimized. Results of the optimization are:

- A trajectory of the manipulated variables $u(t)$ for t_0 to t_{end}
- The corresponding state trajectories $x(t)$ for t_0 to t_{end}

The definitions of the cost function and of the constraints must represent the goal and the limitations of the process adequately. The objective function should be economic in nature—the minimization of cost, the maximization of profit, the minimization of batch time, or a combination thereof can be considered. Constraints include operational limitations and product quality specifications as well as safety related limits. Constraints can be dealt with as:

- Hard constraints that have to be fulfilled, otherwise the optimization problem is not solved (the solution is not feasible).
- Soft constraints the violation of which is penalized in the objective function. Soft constraints do not have to be satisfied but should be close to being fulfilled for the optimal solution, which is assured by sufficiently large penalty terms.

For the solution of the resulting rather difficult [92] optimization problems, sequential, simultaneous or multiple-shooting techniques can be used. In sequential optimization (or single shooting [93, 94]), the inputs to the process are parameterized by piecewise constant or piecewise linear functions, where the intervals can also be parameters of the optimization. The process model is solved by a simulator, and the degrees of freedom are optimized by an optimizer (often a sequential quadratic programming (SQP) method [95]) in the outer loop. This method is robust if the process is stable, however, the satisfaction of path constraints (constraints along the trajectories of state variables) may be difficult. If the optimization does not lead to a feasible solution, the simulation of the last result is valid and can be used to reformulate the problem.

In the simultaneous or full discretization approach, the differential equations are solved together with the optimization problem by means of a parameterization of the trajectories

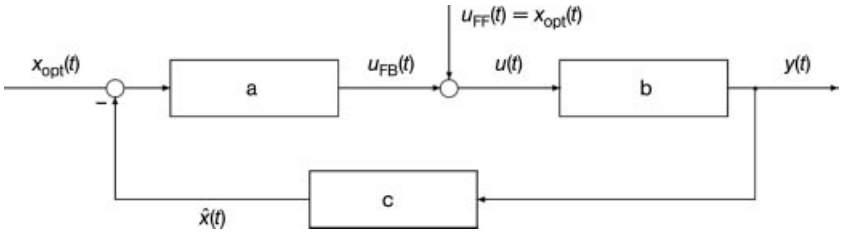

Figure 18. Control along a given trajectory with feed-forward elements
a) Controller; b) Process; c) Estimator/filter

of the inputs and of the states [96, 97]. The resulting large nonlinear optimization problem is solved by SQP or interior point methods [98, 99]. The model equations are only satisfied when the optimization has converged. The speed of convergence depends crucially on the precision of the Hessian of the optimization problem. If no feasible solution is obtained, the computed trajectories do not satisfy the model equations.

Multiple shooting divides the optimization horizon in several intervals and solves individual optimization problems on each interval by the sequential approach. Continuity of the solutions is imposed as additional constraints. By exploitation of the structure of the optimization problems, efficient and robust procedures result [100, 101]. Infeasibility or unsatisfactory solutions can often be attributed to over- or underconstraining the problem.

4.4.2. Implementation of the Optimized Trajectories

The optimized trajectories can be implemented in different fashions:

1. Feed-forward (open loop)The optimal trajectories of the manipulated variables are implemented and not modified if the states deviate from their nominal trajectories.
2. Decentralized control
    1. A set of measurable states are controlled such that they follow the optimized trajectories using SISO control with appropriately chosen manipulated variables.
    2. If not all manipulated variables are used, the remaining ones are implemented as computed by the optimization.
3. Trajectory following linear or nonlinear controllers
    - The optimal trajectories of the manipulated variables become feed-forward elements $u_{FF}(t)$
    - The trajectories of a subset of the state variables are controlled by modifications of the inputs:

$$u(t) = u_{FF}(t) + u_{FB}(x - \bar{x}_{ref})$$

Figure 18 illustrates the following of an optimal trajectory. x_{opt} and u_{opt} are results of the optimization. If the controller is a standard linear controller, its settings can be adjusted along the trajectory. The controller does not provide the total required change in $u(t)$ but only the fraction required to handle the unknown influences on the process.

In general, it is not clear which approach is preferable, because this depends on the effect of the deviations of the variables on the cost function and on the constraints. Constrained variables must always be controlled to maintain feasibility.

4.4.3. On-line Optimization

If the process parameters change significantly during a batch, on-line reoptimization might be required [102]. In on-line reoptimization the next input variables are computed by optimization based upon measured process information. The measurements are used in data reconciliation and state estimation to provide the current states and estimates of the disturbances. The optimizer uses this information to calculate the optimal inputs for the remainder of the batch run. This is called shrinking horizon control as

the horizon becomes shorter the further the batch run progresses. If a model is available that represents the process well and that can be employed for optimization this approach guarantees an optimal operation of the batch process. However, the computation times may be too large to use on-line reoptimization as the only controller of the process. In that case, it needs to be combined with trajectory tracking control. The optimization problem is solved with a lower sampling frequency and reacts to longterm disturbances that require the computation of a new trajectory. In between these reoptimizations, the last calculated optimal trajectory is implemented using a tracking controller that reacts to short term disturbances.

4.4.4. Optimal Control Along Constraints

A frequently occurring case of the optimization of batch processes is the minimization of the batch time in order to maximize the throughput of the plant. In order to achieve a time-optimal operation the process often has to be run at the constraints of certain parameters, for example of the heat generation which must not exceed the heat removal capacity. Quite often, a time-optimal operation can be achieved by tracking constraints.

Considering the exothermic reaction A to B as an example, the time-optimal operation policy for infinite heat removal capacity is to add all of reactant A at the beginning of the batch, then to heat the reactor up to reaction temperature as fast as possible and then perform the reaction up to the specified minimum conversion. In practice the resulting heat of reaction can usually not be removed by the cooling system. As the heat removal is the limiting factor, in the optimal case A is fed such that the current heat of reaction is equal to or slightly less than the maximum heat removal. For optimal productivity the process is thus driven along this path constraint. A detailed discussion of this approach is provided by [103–105].

When batch processes are driven along the heat removal constraint, safety margins become a very important issue because a cooling failure can lead to a thermal runaway. The maximum feed rate should be constrained such that in the case of a cooling failure the reaction of the unconverted raw materials in the reactor heats the reactor contents only up to a temperature below the thermal runaway or below triggering relief systems.

4.4.5. Golden Batch Approach

The golden batch approach is an established practical method. A golden batch is a batch with very good results in terms of the specific objectives, such as product quality, batch time, and energy consumption. This batch is then used as a template for further batches.

A data historian or a database provides the trajectories of the relevant process parameters of the golden batch. These trajectories define the optimal trajectories for following batches. Their values and the deviations of the current values from the golden trajectories are displayed during other batch runs. The operators will try to drive the current batch close to the golden batch. Using trajectory tracking by feedback control, this process can be automated. If the relevant process parameters are measured, displayed and controlled, near optimal batch operation is guaranteed.

In practice, not all relevant process parameters are measured during the batch. This is why the golden batch method should be combined with estimation techniques, for example, a PLS that estimates important quality parameters during the batch run, because much historical data is usually available and relationships between measured and quality variables are often hidden in the data.

Furthermore, it has to be possible to actually follow the golden batch trajectories of the different variables. In large batch or semibatch tanks the number of measured variables is often large, for example, several temperatures and pressures, flow rates as well as the stirrer torque, while the number of manipulated variables may be small, possibly only the coolant temperature, the stirrer speed, and the feed flow rates.

In such situations the methods of statistical process control described in the Section 4.3.5 can be applied. As the models are trained using batches that have been classified as good batches, the methods inherently contain the golden batch method. Classical trajectory tracking controllers are typically not applied in this

Table 3. Summary of the methods for practical optimal control of single batches

Method	Effort	Result quality
Rigorous dynamic (re)optimization	high	very good if model is precise
Control along limiting constraints	medium	very good if optimum is at constraints and disturbances are not too large
Golden batch approach	low (old recipe) high (new recipe)	good if golden batch can be tracked
Feedforward/feedback approach	low (old recipe) medium (new recipe)	good as long as disturbances are not too large

case. The operator takes corrective action assisted by the contribution plots.

4.5. Batch-to-Batch Control

4.5.1. General

The motivation of batch-to-batch control is the lack of measurements of product quality indicators during the batch runs. In most industrial batch processes quality variables are measured only at the end of the batch. Batch-to-batch control is a discrete-time control strategy which incorporates a feedback loop using the measurements at the end of the batch to change the settings for the next batch [106]. By analyzing the last run, the batch-to-batch controller manipulates the recipe of the next run to achieve a better operation. Batch-to-batch control is sensible for processes in which the same product is produced regularly, that are difficult to handle, and that have a tendency to drift from the optimal operation. If, for example, fouling is a problem, the controller will react by increasing the stirrer speed or the cooling. If a fouling problem is known and a cleaning schedule for the reactor exists, feed-forward control elements should be employed so that it does not take several batch runs for the batch-to-batch controller to realize that the fouling has disappeared.

4.5.2. Iterative Batch-to-Batch Optimization

Model-based optimization and batch-to-batch control can be combined into an efficient and robust scheme for processes where the cost function and the constraints can be measured at the end of the batch. The key idea is to use a gradient-based optimization to compute optimal operating parameters based upon a process model, and to compensate for the inevitable mismatch between the model and the behavior of the real plant by correction terms. These correction terms are empirical gradients of the cost function and of the constraints that are obtained from the measurement information about past batches. This scheme was first proposed in [107] for the unconstrained case and later extended to the constrained case and applied to batch chromatography in [108]. In [109] it was demonstrated for a case study where a batch reactor was controlled using a simplified model of the chemical reaction that this partly data-based and partly model-based iterative scheme performed better than a data-based adaptation of the parameters of the (structurally incorrect) model. The drawback of the method is that the computation of the gradients of the cost function and of the constraints with respect to the operating parameters requires several batch runs until the true optimum is reached and a batch run at a suboptimal operating point may be required to obtain sufficient information on the gradients. Several schemes for how to choose the set-points during the course of the optimization are discussed in [108]. On the other hand, convergence is much faster than for a purely data-driven batch-to-batch optimization and the resulting operating point is feasible and optimal which is not be the case if the optimization relies only on the model.

4.6. Summary

Batch processes are often used if process robustness to insufficient knowledge is required. A batch process can be adapted on-line, the batch time can be increased or decreased or recipes can be modified slightly. The

insufficient knowledge on the other hand implies that good models capable of predicting the process behavior, as they are often required for advanced control and optimization approaches, are either not available or very expensive to develop.

Control of batch processes is dominated by logic control. Commonly agreed standards cover the structuring of batch plants, recipe-based production, and logic control. These standards provide the necessary means to realize and to use software for recipe-driven batch operation. However, the combination of logic control and continuous behavior during a batch results in hybrid system dynamics. The combined optimal design of logic and continuous controllers including optimal trajectory planning and its tracking is a challenging problem [110].

Batch processes are always dynamic and exhibit nonlinear dynamic behavior. This implies that classical linear control theory is often only applicable with significant enhancements such as feed-forward of the desired trajectories of the manipulated variables and gain scheduling control. Batch run optimization requires optimal trajectory planning which gives rise to challenging dynamic process control and process optimization problems.

While the standards on batch control are now generally agreed and have resulted in significant standardization of batch logic control systems that are sold by many control system vendors, the areas of optimal logic control design, state estimation, and abnormal situation detection as well as optimal batch control are open research areas. Some of these aspects such as optimal trajectory design are well understood and ready for implementation in industrially validated software products, others such as an optimization that includes the switching (logic) control as well as the continuous dynamics are not yet mature.

A central aspect remains the development of robust dynamic models with good prediction capabilities to employ the advanced methods described above. New methods that achieve good results without very precise models, e.g., as proposed in [108] have only recently been developed and are an interesting alternative to classical model-based methods in batch-process optimization.

5. Model Predictive Control: Multiparametric Programming

5.1. Introduction

Multiparametric programming has emerged in the last decade as an important optimization-based tool for systematically analyzing the effect of uncertainty and variability in mathematical programming problems. Its importance has been widely recognized and many significant advances have been established both on the theory and application of multiparametric programming in engineering problems such as control and optimization. The adoption of multiparametric programming in model-based control and specifically model predictive control (MPC) has created a new field of research in control theory and applications, known as multiparametric model-based predictive control or explicit control.

Multiparametric programming is a technique that, in an optimization framework with an objective function to minimize, a set of constraints to satisfy and a number of bounded parameters affecting the solution, obtains computationally inexpensively the exact mapping of the optimal solution profile in the space of the parameters. As it is illustrated in Figure 19, the optimal solution mapping (or explicit solution) consists of:

- The objective function and the optimization variables as functions of the parameters
- The space of parameters (known as *critical regions*) where these functions are valid

The optimization can then be replaced by its optimal solution mapping and the optimal solution for a given value of the parameters can be computed efficiently by performing simple function evaluations, without the need to solve the optimization. The advantage to replace optimization by simple and efficient computations has given multiparametric programming wide spread recognition and has triggered significant advances in its theory and applications.

Multiparametric–linear programming (mp–LP) algorithms, based on the Simplex algorithm were first investigated by [111–113], when the parameters are present both in the coefficients of the objective function and the right-hand side of

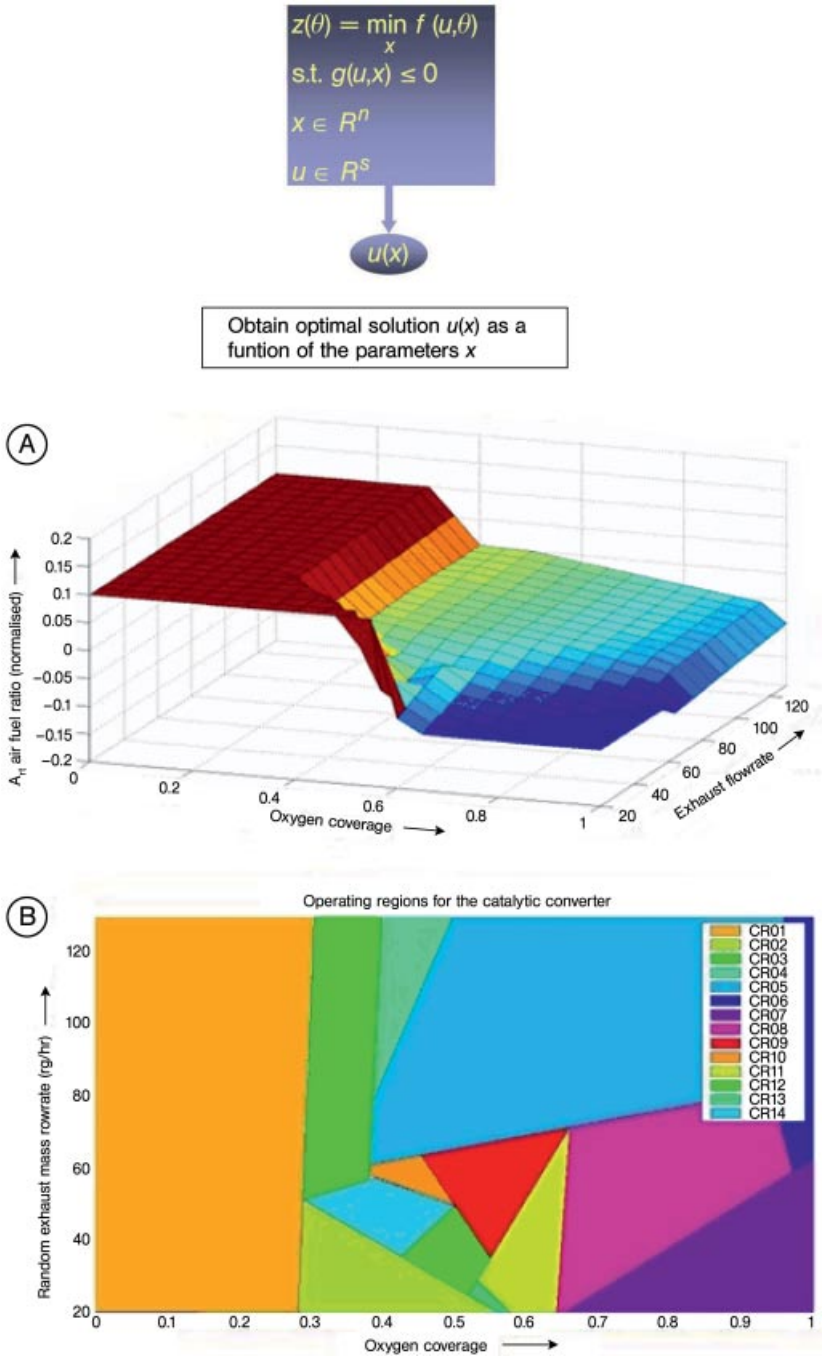

Figure 19. Multiparametric programming
A) Optimal look-up function; B) Critical regions

the constraints while the same problem was treated in [114] by applying sensitivity analysis. An mp–LP framework for flexibility analysis in process design problems under uncertainty is presented in [115]. The general framework of multiparametric MPC (mp–MPC), the theory, and related applications were presented for the first time in [116]. Multiparametric–quadratic programming (mp–QP) algorithms by explicitly were investigated by [117, 118] solving the KKT optimality conditions. Algorithms for multiparametric–mixed-integer linear programming (mp–MILP) problems with scalar parameters were developed in [119, 120], while the nonscalar mp–MILP problems were investigated in [121, 122]. An algorithm for multiparametric–mixed-integer quadratic programming (mp–MIQP) problems was introduced for the first time by [117]. Methods for multiparametric–mixed-integer nonlinear programming (mp–MINLP) problems with scalar parameters were developed by [120, 123], while the more general nonscalar case was first treated by [124]. Finally, the first algorithms for multiparametric global optimization and multiparametric dynamic optimization where introduced in [125, 126], respectively. The advances in multiparametric programming theory and its applications in advanced model-based control are the subject of a two volume textbook that has appeared recently in the literature [127, 128].

Multiparametric programming has found many applications especially in the area of process engineering such as process design, optimization, and control. However, the most significant one has been established in the area of model-based control and specifically MPC. Traditionally, MPC obtains the control actions on a process by solving repetitively an on-line optimization problem based on the prediction of the future system behavior. Despite MPC's advantage to handle process constraints and multivariable processes, its applications has been rather limited due to the demanding computational requirements of on-line optimization. mp–MPC, on the other hand, is an advanced control method that uses multiparametric programming methods to solve the on-line optimization problem of MPC and obtain the exact mapping of the optimal control variables as functions of the state variables. The main advantage of this approach is that it replaces the on-line optimization of MPC with simple function evaluations that require a smaller on-line computational effort in comparison with on-line optimization. This advantage has made it possible for MPC to be implemented on simple computational hardware such as microchips, paving the way for many advanced control applications in chemical, energy, automotive, aeronautics, and biomedical systems. The concept of replacing the on-line optimization via the exact mapping of its optimal solutions, has become known as "on-line optimization via off-line optimization" while the ability of mp–MPC to be implemented on the simplest possible hardware has become known as the "MPC-on-a-Chip" technology. These concepts as well as the framework for the design and implementation of explicit MPC are illustrated in Figure 20.

The explicit solution of the discrete-time constrained linear quadratic control with multiparametric programming was first studied

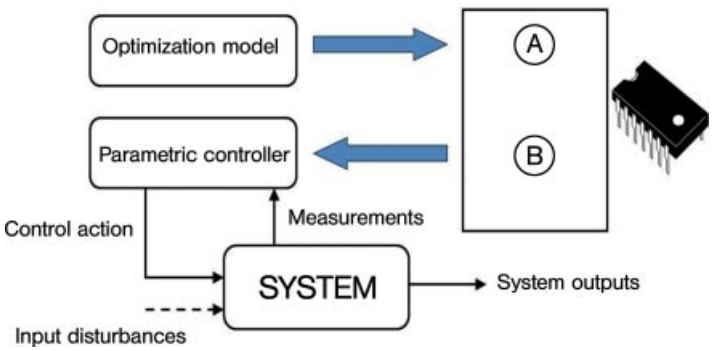

Figure 20. mp–MPC and the MPC-on-a-Chip technology

in [118] and paved the way for the development of mp–MPC. The solution for the continuous-time linear quadratic control problem was provided by [129]. In the area of robust mp–MPC [130] provided two methods for deriving the explicit solution for linear quadratic control of system with disturbances while in [131] a min–max robust mp–MPC algorithm was developed for systems with parametric model uncertainties. Two algorithms for multiparametric dynamic programming were presented in [132, 133], while multiparametric nonlinear MPC (NMPC) was first investigated in [134–136].

The developments in multiparametric programming and mp–MPC theory were soon followed by equally important developments in applications. Industrial applications of mp–MPC include the control of an air separation process [137] and active train valve control for the Lotus experimental engine while applications in biomedical systems include control of insulin delivery for type 1 diabetes. These are the first two milestone applications of mp–MPC in industrial and automotive applications and have served as platforms for proving the concept of MPC-on-a-Chip. Ongoing applications on multiparametric programming and MPC such as the control of proton exchange membrane (PEM) fuel cells, hybrid systems for hydrogen generation, hydrogen storage in metal–hydride bed, navigation and control of multiple unmanned air vehicles (MUAV) highlight the importance of multiparametric programming and MPC in engineering applications. These applications highlight the potential of mp–MPC for implementation in various processes and systems and its possible value for commercialization. Hence, two patents [138, 139] have emerged recently on the MPC-on-a-Chip technology.

The objectives of this article are to overview recent advances in multiparametric programming and mp–MPC theory and applications, and provide future directions to the opportunities and challenges for research in multiparametric programming and MPC.

5.2. Multiparametric Programming Theory

Despite the major advances in the theory of multiparametric programming, still many unresolved issues exist for many classes of multiparametric programming problems. It is evident from the relevant literature that a major research effort has been made so far for the study and development of mp–LP and mp–QP algorithms, with the rest of the problems receiving less attention. This is due to the many applications of multiparametric programming, such as linear control applications, where the main focus is on linear problems, and to the complexity and difficulty to solve explicitly mp–MINL and/or dynamic optimization problems. However, the ongoing developments in:

- Process optimization and control
- Nonlinear, continuous-time and/or hybrid systems
- Hierarchical decision making and control

have created new challenges for multiparametric programming. Recent advances in:

- mp–NLP
- Bilevel/multilevel, and hierarchical programming
- Constrained dynamic programming
- Global optimization of mp–MILP

are overviewed here and the perspective research opportunities in the unexplored areas of mp–MPC are exposed.

5.2.1. Multiparametric Nonlinear Programming

Developments in the methods of mp–NLP have not followed the rapid progress in the developments of mp–LP and mp–MILP methods. However, there are some significant advances in mp–NLP algorithms. Previous work in mp–NLP was focused on the development of outer mp–LP approximations within a prescribed approximation error of the underlying mp–NLP problem [140]. A number of novel results have also been established recently. In [141] a geometric, vertex-based algorithm is introduced for obtaining a piecewise affine approximation of the explicit solution of an mp–NLP problem. Ongoing research in mp–NLP methods is currently focusing on the development of quadratic approximations for mp–NLP problems.

5.2.2. Bilevel/Multilevel, Hierarchical Programming

Bilevel programming falls within the class of hierarchical optimization problems, where one optimization problem (outer- or upper-level problem) is constrained by another optimization problem (inner- or lower-level). Bilevel programming problems have found applications in game theory and hierarchical decision making, while there is a potential for applications in hierarchical control. In recent developments [142], multiparametric programming has been used as a tool for solving bilevel programming problems. Depending on the type of the outer and inner problems, multiparametric programming has been applied for the following bilevel programming problems:

- A QP formulation for the outer and inner problems
- A LP formulation for both the outer and inner problems
- A outer QP problem and an inner LP problem
- An outer LP problem and inner QP problem
- For mixed-integer bilevel problems

In [142] a global optimization framework was established for bilevel programming that (i) recasts and solves the inner optimization problem as multiparametric programming, where the parameters are the optimization variables of the outer optimization and (ii) transforms the bilevel problem into a single level convex optimization problem. More recently, [143] applied a similar framework for mixed-integer bilevel programming where the inner problem is first reformulated to its vertex polyhedral convex hull representation and then to a multiparametric programming problem using convex underestimators problem. The bilevel problem is then transformed to simple convex optimization problem.

5.2.3. Constrained Dynamic Programming

Dynamic programming (DP) has been a popular method for the optimization of multistage decision processes, with many applications found in decision making, operation research, and optimal control. Its main advantage is the ability to break a multistage problem into solving a sequence of smaller size stage-wise optimization problems and obtain the optimal decisions as policies (functions) of the state of the underlying system. Although DP is a well-established methodology, there are still issues with the solution of multistage optimization problems especially in the presence of constraints and parameter variations. This case had not been fully treated previously in the relevant research. Multiparametric programming has been used to solve the constrained DP problem [132] of linear quadratic multistage problems. Each of the stagewise optimization problem is solved as a multiparametric quadratic program where only the optimization variables, parameters, and constraints at the current stage are considered. However, with this approach the convexity of the original problem is lost since the objective function is piecewise quadratic [132, 133]. Global optimization methods have to be employed then to solve the stagewise optimization, which usually lead to overlapping critical regions. A method is shown in [133] as a convex multiparametric quadratic problem where the decisions of each stage are derived as explicit functions of the states of the stage, where no critical region overlapping occur.

5.2.4. Global Optimization of Multiparametric Mixed-Integer Linear Programming

The general mp–MILP problem (Fig. 21) focuses on MILP problem with parameters in the coefficients of the objective function and the right-hand side of the linear inequalities. Previous methods have been focused on the simple mp–MILP problem where no parameters appear in the coefficients of the objective function while recently in [144] the general mp–MILP problem was treated. The general algorithm for solving mp–NILP problems is based on a procedure that iterates between the solutions of a master optimization problem and a slave optimization problem (Fig. 21). The master problem is formed as a MINLP where the minimization is over all variables including the parameters. The slave problem is formed as a multiparametric nonlinear program, by substituting the integer

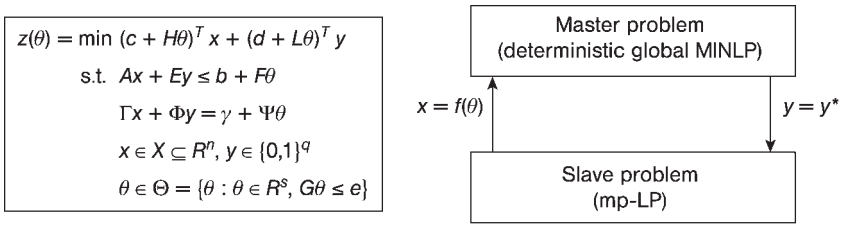

Figure 21. The general mp-MILP problem and the master-slave formulation

solution of the master problem in the mp–MILP problem, where the objective function contains bilinear terms of the parameters and the continuous optimization variables. The master problem can be solved to global optimality since a solution of the integer variables is required. The challenge here is to avoid the need for global optimization for the slave problem thus reducing the computational effort of the master–slave iterating process. In [144] it is shown that global optimization of the slave mp–NLP problem can be avoided and the slave problem can be solved as a simple mp–LP problem.

5.3. Explicit/Multiparametric MPC Theory

Past and current research has mainly focused on the theoretical and algorithmic developments in the areas of linear explicit MPC and robust explicit MPC while some important results in the theory of hybrid, continuous-time and nonlinear explicit MPC have also been reported in the literature. The recent developments in mp–/explicit MPC theory are focused mainly in the following areas:

- Explicit MPC and model order reduction
- Explicit nonlinear MPC (NMPC)
- Robust explicit MPC

These developments are overviewed here and the future research directions in the explicit MPC theory will be discussed next.

5.3.1. Explicit Control and Model Order Reduction

The purpose of model order reduction methods is to provide approximating reduced-order models (with a reduced number of state variables) for large-scale processes. In the case of mp–MPC, the reasons for applying model order reduction methods are:

- Insufficient available memory for solving the mp–MPC problem off–line
- The desire to reduce the time in which the explicit solution of mp–MPC is obtained
- The need to reduce the size of the explicit solution (smaller number of critical regions and smaller number of parameters) in order to speed-up on-line calculations

In these cases, a reduced-order model of the real large-scale process can be directly used for the design of reduced-order mp–MPC [145]. However, since the reduced-order models are only approximations of the real process, the optimality and feasibility of the reduced mp–MPC is not guaranteed [141]. A systematic method that combines balanced truncation model reduction and mp–MPC design was developed by [141], which obtains the minimum order of the reduced-order model for which the resulting reduced-order mp-MPC controller ensures the optimality and feasibility for the large-scale system. This is the first reported work in model order reduction and mp–MPC to deal with the issue of the optimality and feasibility of reduced-order multiparametric controllers. It is also the first work to introduce the concept of combined model reduction and mp–MPC techniques in order to resolve these issues.

5.3.2. Robust Explicit MPC

There is an undisputed need for robust explicit MPC methods for the design of explicit controllers for dynamic systems with bounded disturbances and model uncertainties. Explicit

MPC controllers designed with the use of nominal dynamic models cannot guarantee feasibility, in terms of constraint satisfaction, and system stability when disturbances and/or model uncertainties are present. The challenge here is to develop algorithms for the design of robust explicit MPC controllers, which guarantee constraint feasibility and robust stability for any values of the uncertainty. The recent research of robust explicit MPC has been focused on the design of robust explicit MPC controllers for linear dynamic systems with additive disturbances in the linear state-space models and/or parametric model uncertainties in the system matrices. The case of robust explicit MPC of linear systems with additive disturbances was first examined in [130]. The design of robust explicit MPC for linear systems with model parametric uncertainties and linear objective functions was also investigated in [131]. Robust mp–QP methods were investigated by [146], based on the previous work on robust optimization [147, 148], for solving the robust explicit MPC of linear systems with parametric model uncertainties with a quadratic objective function (robust linear quadratic control). Recently, a novel framework for robust explicit MPC of uncertain systems was developed [133, 149, 150] by using combined constrained dynamic programming and robust optimization methods. The proposed approach is based on the following three–step algorithm:

- The underlying optimization MPC problem is recast as a multistage optimization problem
- The multi–stage optimization problem is reduced to smaller single-stage optimization problems by applying constrained dynamic programming, where only the controls, states, and constraints at the current stage are considered
- The single-stage problem is solved with robust multiparametric optimization methods to derive the control variables as an explicit function of the states

5.4. MPC-on-a-Chip–Applications

The significant advances in multiparametric programming and mp–MPC were followed by a number of important applications. Many of these applications involve the design and implementation of multiparametric controllers for real, complex processes where the available control hardware and software is limited for advanced control applications. The main areas of application of multiparametric programming and mp–MPC include (i) process engineering, (ii) heat networks, (iii) automotive, (iv) aeronautics, (v) biology systems, (vi) scheduling, (vii) waste management, (viii) power electronics, (ix) gas–liquid separation, and (x) oil and gas processes.

Three milestone applications that showed the potential of the MPC-on-a-Chip technology are the following ones:

- Process systems: small air separation plants for the production of nitrogen [137]
- Automotive systems: active valve train control [151]
- Biomedical systems: insulin delivery for type 1 diabetes [152]

The first application falls within the area of medium-scale processes with medium dynamics while the last two applications fall within the area of small-scale, portable processes with fast dynamics. In all these three applications the main issues for the control design and implementation was the available control hardware, which was mainly based on the use of microchips, while the last two applications also faced issues with fast dynamics and small sampling times. mp–MPC was successfully applied in all three cases while important performance improvements were also reported [137, 151, 152]. Figures 22 and 23 demonstrate the implementation of explicit MPC for the small air separation plant and the insulin delivery applications. These applications were used as proofs of concept for demonstrating the simplicity and effectiveness of explicit MPC.

There is currently a number of ongoing applications of mp–MPC which are presented below with the details of their related projects:

- Hybrid pressure swing adsorption/membrane hydrogen separation
- Hydrogen storage based on metal-hydride beds
- Fuel cells
- Unmanned air vehicles (UAV) and biomedical systems

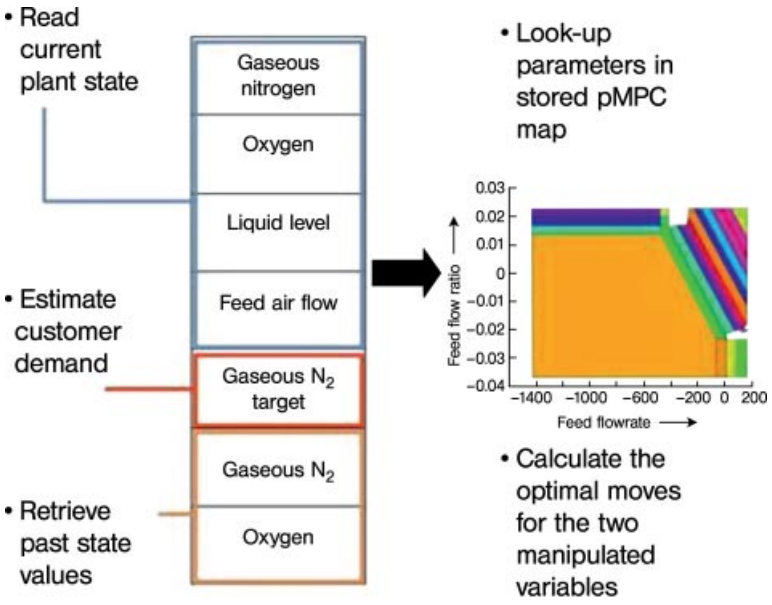

Figure 22. Explicit MPC for small air separation plants [137]

The objectives of these applications are to (i) apply and evaluate the results of the recent developments in the theory of multiparametric programming and mp–MPC, (ii) investigate new methods for mp–MPC to address application-specific issues (e.g., new state and disturbances estimation techniques are necessary for the control of UAV) and (iii) to evaluate the future potential of mp–MPC for a wide range of systems.

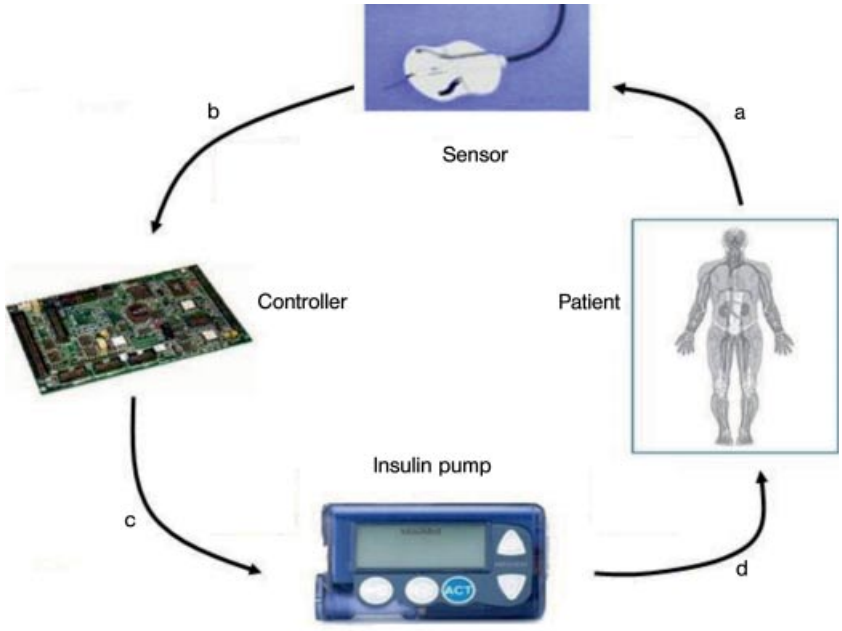

Figure 23. Explicit MPC for insulin delivery for type 1 diabetes [152]

5.5. A Framework for Multiparametric Programming and Explicit MPC

The recent advances in multiparametric programming include the introduction of a unified framework for multiparametric programming and explicit MPC controllers. Figure 24 shows an illustration of the main idea of this framework and the steps required for the design of explicit MPC. A high-fidelity dynamic model is used to provide a detailed description of the process. A reduced-order model is then obtained from the high-fidelity model by applying model reduction or identification methods and is used to form the MPC problem. The MPC problem is solved by applying multiparametric programming to obtain the explicit controller. In the last step of the algorithm the explicit controller is applied on the high-fidelity model and is tested to identify possible deviations from the desired behavior or possible infeasibilities (since the explicit controller is derived from an approximating model and not the actual high-fidelity model). If necessary, the procedure is repeated until the desired behavior is achieved for the explicit controller. Each of these tasks can be performed off-line and on-line computations are not required either for the design or the validation of the controller. When the validation is completed, the controller is implemented on the real system. The high-fidelity dynamic model is important in this framework since it is used to represent the real process. The explicit controller validation is performed on this model to ensure accuracy and feasibility. Hence, the advantage of the proposed framework is that it allows for the design of "tailor–made" explicit controllers, which can be tested off-line based on high-fidelity model.

5.6. Concluding Remarks and Future Outlook

The main advantage of multiparametric programming and mp–MPC is their ability to replace the on-line optimization in an MPC framework, with computationally inexpensive function evaluations, that can be applied on simple computational hardware. This paves the way for many advanced control applications not only in the area of large- and medium-scale processes, where advanced control and MPC has been traditionally applied, but also for small-scale systems such as portable devices and equipment, where advanced control methods had not yet found applications due to the insufficient computational power required for their implementation.

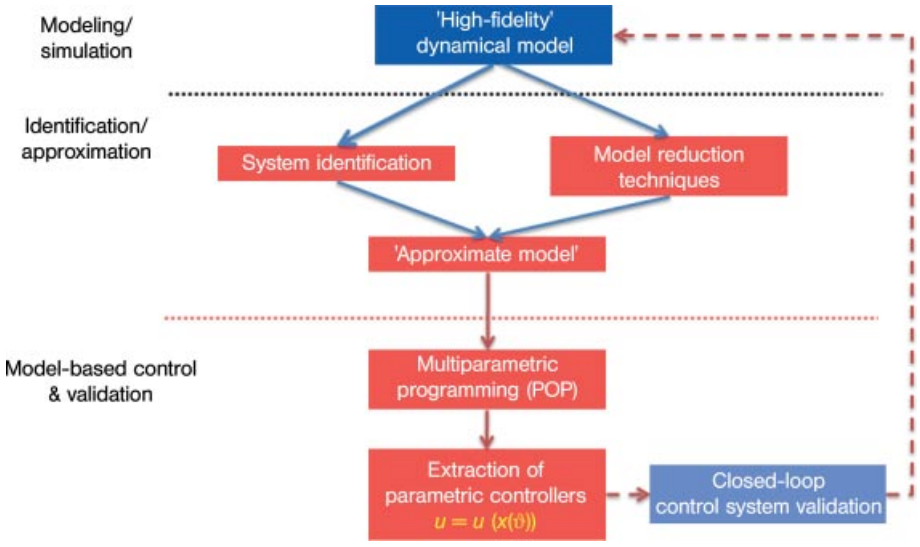

Figure 24. A framework for multiparametric programming and explicit MPC

Future research opportunities in multiparametric programming include: (i) multiparametric dynamic optimization (mp-DO) of continuous-time multistage dynamic systems (also involving 0-1 variables, i.e., mp-MIDO problems), (ii) global optimization/multiparametric programming of nonlinear programs–current research has mostly focused on linear and mixed-integer linear programs, and (iii) revisiting the fundamentals of optimization theory (or investigate new ones) since many issues in multiparametric programming are common issues of standard optimization as well. Future research opportunities in explicit/mp–MPC include: (i) robust explicit/mp–MPC of hybrid and continuous-time systems which is an area with limited attention in the relevant literature, (ii) explicit/mp–NMPC (most of the current work address the linear system case only) and (iii) model reduction, identification and explicit/multi-parametric control. Finally, future opportunities for the application of explicit/mp–MPC include medium-scale processes such as small air separation plants, PSA units and fuel cell systems and small-scale systems such as portable devices and equipment, for which the available control hardware is mainly based on microprocessor and/or microchip technology and its available computational power, cannot support on-line optimization computations. mp–MPC and the MPC-on-a-Chip is particularly suitable for this type of systems, since the simple function evaluations involved in the implementation of mp–MPC, allow for its implementation on the simplest control hardware such as microprocessors and microchips.

6. On-Line Applications of Dynamic Process Simulators

6.1. Introduction and Historical Background

6.1.1. Modeling Dynamic Simulation

A dynamic simulator is a mathematical description of the time-varying physical behavior of a production facility. In many cases, this production facility is a chemical production facility, since on-line application of dynamic simulation is quite far developed in the chemical industry. However, in principle it can also be a food or, a pharmaceutical production facility. Dynamic models do not simply represent the phase, flow, and reaction behavior of the material in a process. They need also to model the behavior of the processing equipment. Considering, for example, a large processing vessel made of steel or a valve the following questions may arise: How long does it take to warm up this vessel from a cold start? How quickly will the valve close? Will it close so quickly that a pressure surge will rupture a hose and result in a leak? Do one need to install a device to slow down the valve? In addition, the control and safety systems in the facility must be part of the model. The interaction between the process and its control, sequences, and safety logic is often the most important (and interesting) part of the dynamic behavior of the facility.

A dynamic process model that takes account of process, equipment, and control is therefore a detailed and complex artefact. Fortunately, the unit operations concept can be applied to dynamic models so that a model of a large facility can be built up as a network of models of individual processing operations, equipment, and control algorithms.

6.1.2. Historical Perspective: From Design and Training to Full Lifecycle Operations

Today's tools and methods for dynamic simulation build on academic work from the late 1970s and early 1980s [153, 154] that resulted in commercial tools that became available in the late 1980s and early 1990s. Tools for operator training used a modular approach. Some examples were OTISS, G-PURS, Trainer, ProSim, and CADAS, whereas equation oriented tools (e.g., SpeedUp [154], gProms and Massbal [155]) were primarily used for engineering. A second generation of simulation tools, that took advantage of the personal computer, became available from the mid 1990s. These second-generation products have been developed and remain the dominant commercial tools today. A nonexhaustive list of products includes K-Spice (Kongsberg Group), INDISS (RSI-Simcon), Aspen Dynamic Modeler (AspenTech), gProms (Process Systems

Enterprise), Unisim (Honeywell), and Hysys (AspenTech).

The general process simulators listed above are complemented by specialized dynamic simulators that model some specific part of the production chain. Thus, specialized models are used for refinery reactors, olefin crackers, and other–often proprietary–reactions. In the upstream petroleum industry, multiphase pipeline simulators, such as OLGA [156], PIPEPHASE [157] and LedaFlow [158] simulate the dynamic behavior of oil, gas, and water mixtures in long pipelines, oil well bores and complicated subsea fluid connection networks.

Generic dynamic simulation tools and languages, such as Matlab/Simulink and Modelica, have to date had limited impact for dynamic process simulation. This is primarily due to their limited ability to access thermodynamic and physical property data. This situation may change with the adoption of the CAPE-OPEN interface standards for thermodynamics and physical properties.

Initial applications used detailed, slow, equation-based calculations (using, for example, SpeedUp or gProms) to solve design problems, and simplified modular simulations for operator training. As the price of computers decreased and computing power increased the fidelity of simulators used for training was improved to a degree where they were accurate enough to be used for design calculations. At the same time, the rise of client-server computing allowed interactive, graphical configuration of the simulation models. The first generation of simulators used a batch work process: first configure a model, then do the simulation calculations, and finally visualize the results. The best modern tools provide a single user interface for model configuration, simulation, and display of results.

By the mid-1990s, as reliable commercial tools became available for process simulation, work began on seeing how these models could be applied to actual process operations. Increased use of process historical databases and open-architecture control systems meant that dynamic process data–trends and events–could be collected, stored, and analyzed. This raised questions such as: How good are the simulators we have built? Do they actually predict the behavior that is observed in the facility? Can our model be tuned to match the observed data? At the same time, the adoption of the object linking and embedding (OLE) for process control (OPC) standard meant that a simulator-based application could connect to a control system, simply and cheaply, using a standard interface, rather than the expensive, proprietary protocols that dominated in the 1980s and 1990s. This provided a way of using a simulator, in real-time, with real process data to perform calculations that would help a process operator. Initial applications began with relatively limited models, such as a single multiphase pipeline and its reception facilities or a batch polymerization reactor. As confidence grew and computing speed increased, the scope of the models and systems has grown.

6.2. Architecture for On-Line Simulation

A typical architecture for an on-line simulator is shown in Figure 25. Raw process measurements and data for synchronization are read from the

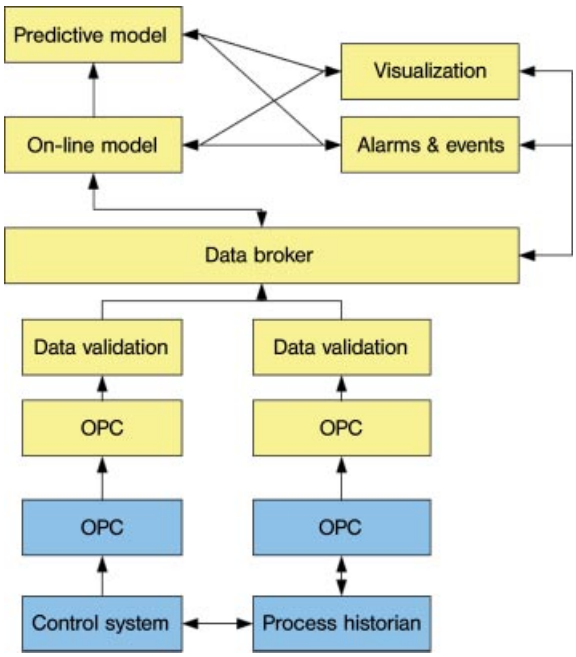

Figure 25. Architecture for on-line simulation applications

facility's control system or process historian. These measurements are then validated–using simple checks and rules–before being used as input to the dynamic simulator. The dynamic data uses selected measurements as fixed boundary values and the synchronization data are used to ensure that the equipment in the simulation has the same settings and status as the equipment in the actual plant. Other measurements can then be used to tune the model and to detect abnormal conditions, such as leakage or equipment fouling. The choice of boundary conditions, synchronization variables, and tuning measurements depends on the application. Each application described in Sections 6.4, 6.5, 6.6 gives examples of these different types of variables.

As noted above OPC is important as an enabler for on-line simulation. The OPC standards for data acquisition (DA) and historical data acquisition (HDA) mean that systems can be developed that can read data from essentially all modern control systems and process historians using a common data protocol. A new version, OPC unified architecture (UA), has just been released and is likely to rapidly replace the older standards [159].

The on-line simulator reads data from the raw data source using OPC. The communication link then applies data validation and replacement to ensure that missing or erroneous data is not used in calculations. This is discussed in more detail in Section 6.3.2 below. The processed data is made available to the on-line simulator via a *data broker* component.

The *on-line model* runs in synchronization with the process. It provides data for display to a visualization interface and can also raise alarms and events. The model regularly saves snapshots of its state. These snapshots can be used to start another copy of the dynamic simulator–a predictive or look-ahead simulator. The predictive simulator is used to evaluate the future effect of operator actions. Its purpose is to run in as short a time as possible so that an operator can be warned about undesirable effects of a planned action. The *predictive model* can display its results in the visualization system and can also generate alarms and events. Predictive simulations can be run automatically–at a fixed time interval or when an operation is done–or on demand.

6.3. Challenges in the Use of Dynamic Process Models for On-Line and Real-Time Applications

On-line simulation is technically challenging. It involves trying to relate an imperfect process simulation to a real industrial process with inaccurate measurements, communication glitches, corporate bureaucracy, limited budgets, and wear and tear on equipment. The challenges can be sorted into the following categories:

- Data security and corporate information policy
- Data communications and quality
- Synchronization with operations
- Model quality–stability and accuracy
- Thermodynamics

6.3.1. Data Security and Corporate Information Policy

On-line simulation applications interact with safety-critical systems and handle confidential information related to production and efficiency. For these reasons, an on-line simulator is subject to rigorous data security requirements [160]. Modern production facilities use a three-tier network architecture to enforce security (Fig. 26).

An on-line simulator provides information for operators and uses raw data from the process control system. The on-line simulator is usually not allowed to write data to the control system.

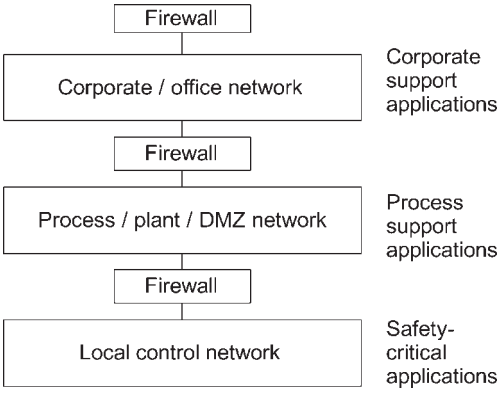

Figure 26. The three-tier security architecture

However, it is common to place the on-line system in the process network, or demilitarized zone. This enables the system to communicate quickly and efficiently with the control system. However, this approach makes it more difficult to use data outside the production facility and to provide remote support.

The alternative approach is to place the simulation application in the corporate network, with a connection to the control system via a process historian. This approach makes corporate access to data and remote support easier. However, the process historian can introduce delays and unwanted filtering into the data used by the on-line simulator. Furthermore, process historians are often configured not to store all the types of data needed by an on-line simulator.

6.3.2. Data Communications and Quality

A successful on-line simulator requires good process data and must implement mechanisms so that it is robust–keeps operating as well as it can–when data communication is lost or there are problems with measurement sensors, transmitters, or signal converters. Mechanisms that can be used are:

- Simple validation and replacement of single signals. The signal is checked against its maximum value, minimum value, maximum allowable rate of change, and status reported from the data source. Bad values can be replaced by the last good value or a specified override value. Lack of change in a measurement should also be detected as a problem. This can indicate a communication failure, a too-wide range for reporting on change or a poor configuration of data compression in a process historian.
- Logical checks on relationships between process variables. Process measurements can be checked against each other to ensure that they are consistent. For example, during normal operation, the pressure downstream a pump must be higher than the suction pressure.
- Data reconciliation calculations, where mass, momentum and energy balances are used to correct a set of process measurements so that they are consistent with the balances.

6.3.3. Synchronization

An on-line simulator can only be synchronized with equipment units that report their status to a control system or process historian. The problems this causes can be illustrated by the very simple example shown in Figure 27.

The control system can supply values for the flow measurement, the flow controller set point and status, the flow controller output, and the isolation valve position (open or closed). However, the manual bypass valve and drain valve are not usually connected to a control system. They are operated by a field operator. A mismatch will occur if one of these valves is open in the plant while it is closed in the model. If the drain valve is opened the simulated flow downstream the isolation valve will be greater than the actual value. If the bypass is opened, the flow controller output will be different in the simulation and the process.

An on-line simulator must handle field operator actions elegantly. In an ideal world, one could hope that field operators, as part of the standard procedure, opened the corresponding manual valve in the on-line model as they opened the real valve in the process. This is unlikely in practice. However, as unmanned facilities become more common, this problem may become less of an issue. All valves will be remotely manipulated.

However, the best an on-line system can do in the presence of unimplemented field operator actions is to report discrepancies between observed and measured conditions. In the case shown above, a discrepancy between measured and observed flows or pressures downstream of the isolation valve could provide a valuable

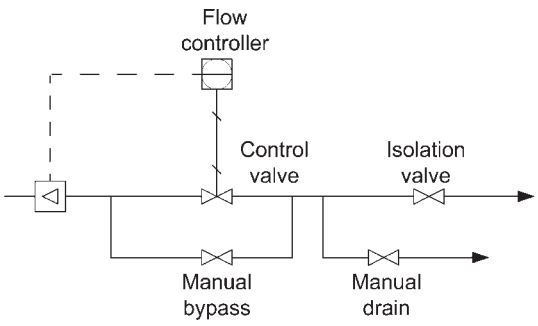

Figure 27. A simple example of model synchronization

indication that a drain valve that should not be open is open.

6.3.4. Model Quality

"All models are wrong, but some models are useful" is a quote from the famous statistician George Box. Dynamic process simulators have been proved to be useful. The fundamental physics and thermodynamics embedded in the models also mean that they are often correct enough for process design and operations. However, the model is always wrong, and will tend to drift away from the process unless it is tuned. Fortunately, because the model incorporates fundamental physics, we know the parameters that account for uncertainty in the model. These are the empirical design constants for the equipment: the friction factor for a pipe, the heat-transfer coefficient for a heat exchanger, and the separation efficiency of a column.

A model can be tuned automatically, and recursively, by using chosen measurements to slowly adapt the chosen parameter so that the residual between this measurement and the model's estimate is driven towards zero. Any parameter estimation algorithm can be used, but it is important that the parameter is changed slowly, so that the short-term dynamic predictions from the model are not disturbed by vigorous changes in parameters. A simple PID controller is often a suitable tuning algorithm. For example, a controller can be used to drive the difference in observed and modeled pressure drop over a pipe section to zero by manipulating the friction factor on the pipeline. Similarly, a temperature difference across a heat exchanger can be used to tune the heat-transfer coefficient in the exchanger. The estimated parameters are useful indicators of the performance of the equipment. An increase in friction factor or decrease in heat-transfer coefficient can indicate fouling or blockage of equipment.

6.3.5. Thermodynamics

A final challenge to on-line modeling lies in the thermodynamic calculations of the model. There are two challenges that need to be addressed:

- The composition of the material to the process may not be known or may not match the current composition
- Thermodynamic models are empirical and are by definition inaccurate and limited

The first problem arises because composition analyses are slow, expensive, delayed, and inaccurate. Consider a model of an oil and gas production facility. The composition of the feed to model is the composition of the oil, gas, and water at the bottom of the oil well. This can only be determined by well tests or samples taken while drilling. These samples are taken infrequently. This means that an on-line model will tend to lose accuracy over time if feed composition changes. In addition, the amounts of oil, gas, and water and the density of each phase can be measured by a multiphase flow meter installed on the well. These provide useful information–but these meters often have low accuracy and poor reliability.

Fortunately, the available measurements around an oil well (pressures, temperatures, flows of each phase, and valve positions) can be used to detect and compensate for uncertainty in the fluid composition. For example, if a well begins to produce water, while the model assumes that there are only hydrocarbons in the fluid, discrepancies will appear in the pressure drop and temperature change over the well bore and over the so-called choke valve at the top of the well. These discrepancies can be used to adjust the feed composition to avoid model drift.

The second problem–inherent inaccuracy of thermodynamic models–can only be handled by careful engineering work. An on-line model provides a systematic tool by which engineers are able to validate the accuracy of the chosen thermodynamic methods and try out alternative methods.

6.4. Pipeline Management and Leak-Detection

The first example is a system for monitoring the behavior of long gas or liquid pipelines. These pipelines are used to transport crude oil, natural gas, processed petroleum products, and water over long distances. Complex networks of pipelines are common and substantial energy

is used to compress the gas or pump the liquid through these pipelines. A long pipeline also has substantial capacity. This means that, for example, a natural gas utility can use a gas transmission line as a storage buffer. This requires timely, accurate estimates of the amount of material in the pipeline.

These pipeline networks are monitored and controlled using a supervisory control and data acquisition system (SCADA). Until recently, with the advent of broadband communications, these SCADA systems contained few measurements and had a long sampling period. The measurements available are usually pressure and temperature at each end of a pipeline segment, valve positions, compressor or pump status and flow rates at inlet, outlet, and custody-transfer points.

A pipeline management and leak detection system [161] uses a dynamic simulation of a pipeline or pipe network. This model is run in synchronization with available SCADA measurements, valve positions, and equipment status. This information can then be used to:

- Calculate the inventory–the amount of material–in the pipeline.
- Predict the speed and estimate the arrival time of scrapers that are sent through the pipeline to clean the pipe and inspect the integrity of the pipeline.
- Track batches of fluid in petroleum product transportation lines. The operator is warned when a new batch for fluid is expected. They can then prepare for the arrival of the new material, minimizing the amount of off-spec material.
- Monitor operations for dangerous conditions, such as vacuum formation, liquid hammer and high velocity (which can lead to erosion of the pipe wall).
- Detect leakage and determine an approximate location of the leak.

Model-based systems for monitoring single-phase pipelines became commercially available in the late 1990s. They are based on a one-dimensional, distributed-parameter model of the pipeline. Leakage is identified by detecting a statistically-significant discrepancy between the observed and modeled value of a chosen pressure in a pipe segment. The calculated pressure profile can then be examined, with a sensitivity analysis, to determine the size and location of the leak. This approach is statistical and requires well-maintained and accurate pressure and flow sensors to be effective. This is often difficult in practice, and for this reason, model-based leakage detection has fallen into disfavor to be replaced by methods based on empirical modeling or acoustic analysis of the pipeline. However, for the other capabilities, a simulation-based system remains the best approach.

The best source for further information on this area is the web site to the pipeline simulation interest group [162].

6.5. Management of Multiphase and Subsea Oil Production

A related application arose out of challenges posed by deep-water off-sea oil production during the 1990s. Prior to this time, off-shore oil production had occurred on platforms that were placed over production wells and were often built standing on the sea bed. The oil, gas, and water produced were separated on the platform. Water was dumped to sea or reinjected into the oilfield. Gas was flared or piped away and oil was shipped in a pipeline or in a tanker.

As oil exploration moved into deeper water, this type of production became too expensive. New designs were needed, where the oil and gas wells were placed on the sea bed, often many kilometers from a floating production platform or a processing facility on land. The production facility received oil, gas, and water from many wells. The long pipelines between the wells and the platform had to convey a multiphase mixture or oil, gas, and water (Fig. 28). The behavior of such mixtures is complex, posing challenges for safe and effective operation. For example, at low production rates, large amounts of liquid(L) slugs can accumulate in the pipeline. This accumulation can continue until enough back pressure has developed to force the liquid through the line. This sudden rush of liquid can overwhelm the capacity of the processing facility. Another problem can occur when pipelines are run in cold conditions. At certain temperatures and pressures, water can react with natural gas to form hydrates, which are hard, ice-like

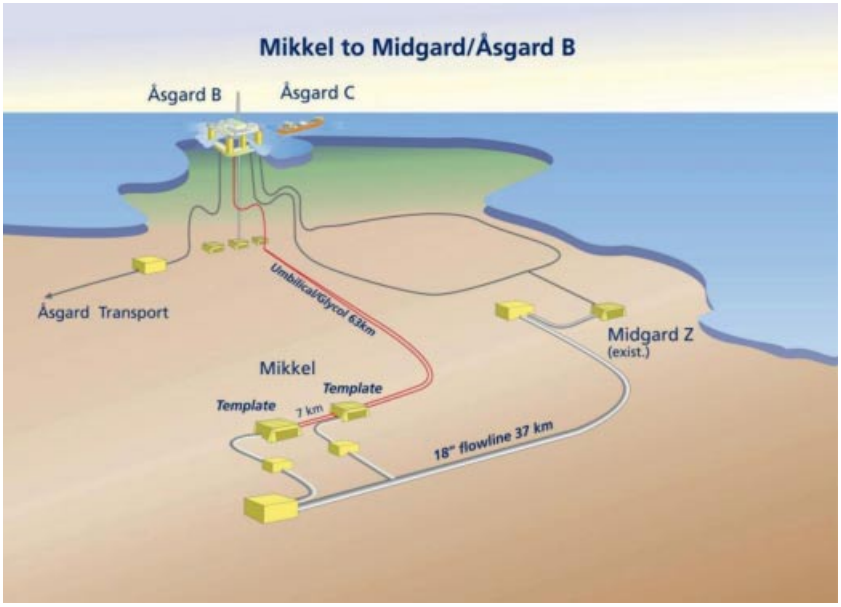

Figure 28. A typical subsea development incorporating a long (37 km) multiphase pipeline (used with permission by Anne Lise Tveit/Statoil ASA)

solids. These will block the pipeline, stopping production and requiring complicated, sometimes risky actions to be taken to unblock the pipe. The pipe can also be blocked by hydrocarbon wax and heavy asphaltenes in the oil, high production rates can lead to erosion of the pipe wall, or high concentrations of salty water can lead to corrosion of the pipeline.

In addition, measurements are usually only available at the ends of the pipeline. Pressure and temperature measurements are usually available. Multiphase flow meters of various types are used to measure flow, gas fraction, and water cut (fraction of liquids as water). These meters are under constant development, but are inaccurate and difficult to maintain.

A specific discipline has developed within petroleum engineering, called *flow assurance,* which specializes in designing and operating multiphase pipelines. Oil companies have invested in experimental work and mathematical modeling to develop specialized dynamic models of multiphase flow [156, 158]. These models allow designers to size pipelines, design insulation, size processing facilities, and specify operating procedures so that slugging and blockage problems can be avoided. However, pipelines must be operated properly if slugging and blockage is to be avoided. Inhibitors, such as methanol or glycols can be injected to avoid hydrate formation. Wells can be opened or closed in a way that does not lead to slugging, low fluid-temperatures or high water-concentrations. These operations are much easier if the operator knows what is happening inside the pipeline, rather than just what is happening at either end. The operator needs to know the temperature, pressure, flow, gas fraction, water cut, and inhibitor concentration along the pipeline. This information is only available by running the multiphase model as an on-line simulator.

Furthermore, proper operational decisions can only be made if the multiphase flow is simulated together with the production wells and the part of the production facilities that handles and separates the material coming out of the multiphase pipelines. This is because control actions in these parts of the process are critical for ensuring proper behavior in the pipeline.

Pioneering work is described in [163] for the Troll Oseberg Gas Injection pipeline. This work described the fundamental elements of an

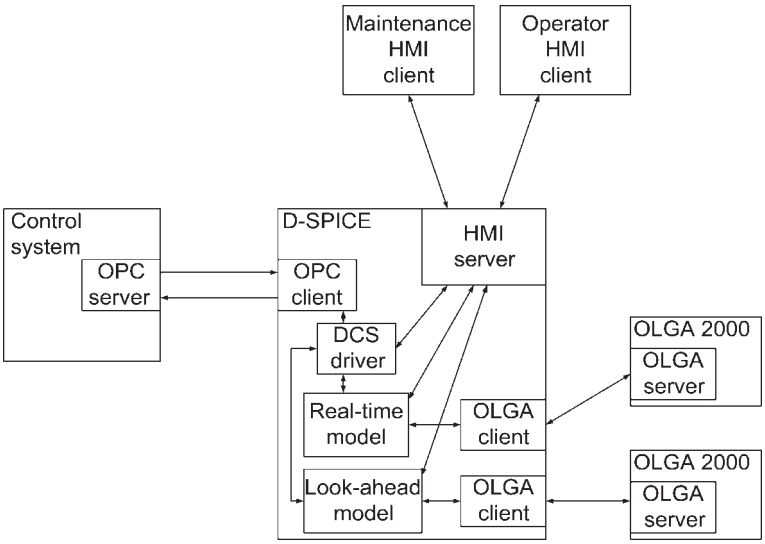

Figure 29. System architecture for a typical multiphase process monitoring system [164]

on-line multiphase system: synchronization with process data, a predictive model, a hierarchical decomposition of the model to allow parallel computation, algorithms to detect leaks and hydrate blockage, and algorithms to tune the model.

A commercial application was developed for the Troll gas multiphase pipeline to land in 1996, but the first large-scale commercial application of a multiphase pipeline management system was for a gas field in Egypt in 2001. This application is described by [164]. The system architecture for this system is shown in Figure 29.

The system used the OLGA multiphase simulator to model the wells and subsea pipelines and the D-SPICE dynamic simulator to provide system integration and model the on-shore receiving facilities.

The system provided a dedicated operator interface that allowed the operator to see both measured data and predicted data. The example screen in Figure 30 shows the screen used to manage conditions around the slug catcher–the vessel that handles large accumulations of hydrocarbon liquids in the pipeline.

This application successfully tracked operation after start up and has been expanded to include all additional oil fields that have been connected to the on-line facility later. Since this time, a pipeline management system like this has become standard equipment for subsea oil and gas developments.

6.6. The On-Line Facility Simulator

Finally, experience gained in running pipeline models on-line provided a basis for running an entire production facility model on-line. This work is described in [165]. In this project, a simulator that had already been delivered to an oil company as a training tool was run in synchronization with the real facility's control system. The experience obtained from this implementation is summarized in Section 6.3.

One of the main findings of this project is that an on-line simulator is best exploited if it generates results for use by off-line simulators. These off-line simulators are used for engineering and training. Valuable engineering time can be saved if an on-line simulator can be used to track actual process behavior. The on-line simulator can then provide configuration files for the off-line simulators so that they represent actual process conditions. This requires a way of archiving and securely distributing simulator configurations and relevant process data. Tools from business computing (service-oriented architecture, SOA) and

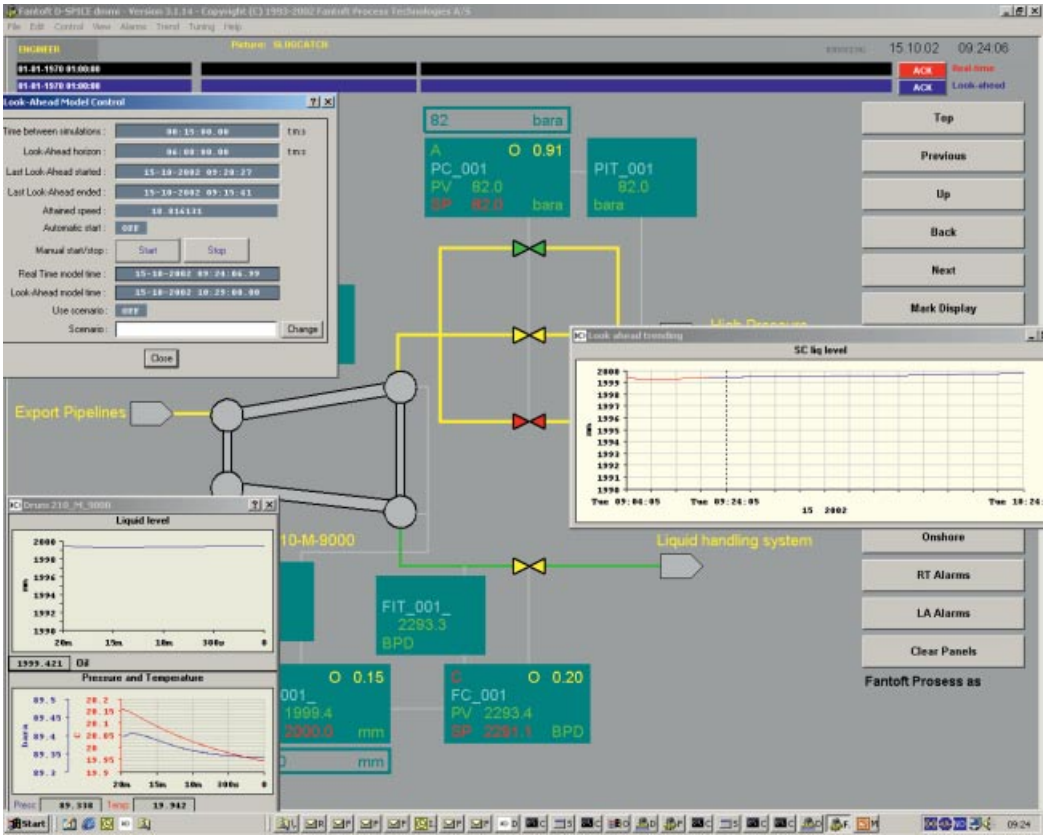

Figure 30. A screen dump of the user interface showing the look-ahead trend and look-ahead simulator control panel [164]

document management (XML databases) can be used for this purpose.

6.7. Conclusion and Future Directions

On-line, dynamic process simulation is a mature technology for specific applications such as pipeline leakage detection and flow assurance. Use of large models for operations support and process troubleshooting is less mature, but has been proven in realistic applications.

Challenges that remain are related to improving the accuracy of the model and using the model for optimization. Operational decisions are always of the form "What should I do?" Unfortunately, process simulators only answer this question indirectly. They actually answer the question: "If you do this, what will happen?" Direct answers to the "What should I do?" question require optimization calculations. These types of calculations currently require massive computer resources or simplified process models. This area therefore remains a fruitful area for research and software development. Much useful information about applications can be obtained from vendor web sites. A useful review of the field is also given by [166].

References

1 US Food and Drug Administration: Guidance for Industry. PAT—A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, FDA, Rockville 2004, http://www.fda.gov (accessed 7 December 2011).

2 K. Ropkins, A.J. Beck: "Evaluation of worldwide approaches to the use of HACCP to control food safety", *Trends Food Sci. & Technol.* **11** (2000) 10–21.

3 W.A. Shewhart in E. Deming (ed.): *Statistical Method from the Viewpoint of Quality Control*, General Publishing Company, Toronto 1986.

4 P. Mohan, J. Glassey, G.A. Montague: *Pharmaceutical operations management*, McGraw-Hill, New York 2006.

5 P. Nomikos, J.F. MacGregor: "Monitoring of batch processes using multi-way principal component analysis", *AICHE J.* **40** (1994) 1361–1375.

6 C.C.F. Cunha et al.: "An Assessment of Seed Quality and Its Influence on Productivity Estimation in an Industrial Antibiotic Fermentation", *Biotech. Bioeng.* **78** (2002) no. 6, 658–669.

7 C. Chatfield, A.J. Collins: *Introduction to Multivariate Analysis*, Chapman & Hall, London 1980.

8 P. Geladi, B.R. Kowalski: "Partial least-squares regression: a tutorial", *Anal. Chim. Acta* **185** (1986) 1–17.

9 D. Dong, T.J. McAvoy: "Multi-stage batch process monitoring" *Proceedings of the American Control Conference*, Seattle, Washington 1996, pp. 1857–1881.

10 K. Cho et al.: "Novel Classifier Fusion Approaches for Fault Diagnosis in Automotive Systems", *IEEE Trans. Inst. Meas.* **58** (2009) 602–601.

11 C. Buriana et al.: "MS-electronic nose performance improvement using the retention time dimension and two-way and three-way data processing methods", *Sensors and Actuators B* **143** (2010) 759–768.

12 Y. Yao, C. Zhao, F. Gao: "Batch-to-Batch Steady State Identification Based on Variable Correlation and Mahalanobis Distance", *Ind. Eng. Chem. Res.* **48** (2009) 11060–11070.

13 N.M. Faber, R. Bro, P. Hopke: "Recent developments in CANDECOMP/PARAFAC algorithms: a critical review", *Chemometrics and Intelligent Laboratory Systems* **65** (2003) 119–137.

14 M. Feng, J. Glassey: "Physiological State Specific Models in Estimation of Recombinant *Escherichia coli* Fermentation Performance", *Biotech. Bioeng.* **69** (2000) 494–593.

15 R.V. Babu, S. Suresh, A. Makur: "Online adaptive radial basis function networks for robust object tracking", *Computer Vision And Image Understanding* **114** (2010) no. 3, 297–310.

16 L. Luccarini et al.: "Formal verification of wastewater treatment processes using events detected from continuous signals by means of artificial neural networks. Case study: SBR plant", *Environmental Modelling & Softwar*e **25** (2010) no. 5, 648–660.

17 C.K. Tan et al.: "Artificial neural network modelling of the thermal performance of a compact heat exchanger", *Appl. Therm. Eng.* **29** (2009) no. 17, 18, 3609–3617.

18 J.A. Leonard, M.A. Kramer, L.H. Ungar: "Using radial basis functions to approximate a function and its error bounds", *IEEE Transactions on Neural Networks* **3** (1992) no. 4, 624–627.

19 J.A. Leonard, M.A. Kramer, L.H. Ungar: "A neural network architecture that computes its own reliability", *Comp. Chem. Eng.* **16** (1992) no. 9, 819–835.

20 L. Al-Haddad, C.W. Morris, L. Boddy: "Training radial basis function neural networks: effects of training set size and imbalanced training sets", *J. Microbiological Methods* **43** (2000) 33–44.

21 M.R. Warnes et al.: "Application of Radial Basis Function and Feedforward Artificial Neural Networks to the *Escherichia coli* Fermentation Process", *Neurocomputing* **20** (1998) 67–82.

22 J. Moody, C.J. Darken: "Fast learning in networks of locally-tuned processing units" *Neural Computation* **1** (1989) 281–294.

23 P.S. Buckley: *Techniques of Process Control*, J. Wiley & Sons, New York 1964.

24 T. Larsson, S. Skogestad Plantwide control: "A review and a new design procedure", *Modeling, Identification and Control* **21** (2000) 209–240.

25 F.G. Shinskey: *Distillation control: For productivity and energy conservation*, 2nd ed., McGraw-Hill, New York 1984, p. 364.

26 J.M. Douglas: *Conceptual Design of Chemical Processes*, McGraw-Hill, New York 1988.

27 J.J. Downs: "Distillation Control in a Plantwide Control Environment", in W.L. Luyben (ed.): *Practical Distillation Control*, Van Nostrand Reinhold, New York 1992, pp. 413–439.

28 M.L. Luyben, B.D. Tyreus, W.L. Luyben: "Plant-wide Control Design Procedure", *AIChE J.* **43** (1997) 3161–3174.

29 W.L. Luyben, B.D. Tyreus, M.L. Luyben: "*Plantwide Process Control*", McGraw-Hill, New York 1998.

30 N.V.S.N.M. Konda, G.P. Rangaiah, P.R. Krishnaswamy: "Plantwide control of industrial processes: An integrated framework of simulation and heuristics", *Ind. Eng. Chem. Res.* **44** (2005) no. 22, 8300–8313 (2005).

31 L.T. Narraway, J.D. Perkins: "Selection of control structure based on economics", *Comp. and Chem. Eng.* **18** (1993) S511–S515.

32 J.E Hansen et al.: "Control structure selection for energy integrated distillation column", *J. Proc. Control* **8** (1998) 185–195.

33 I.K. Kookos, J.D. Perkins: "An Algorithmic method for the selection of multivariable process control structures", *J. Proc. Control* **12** (2002) 85–99.

34 R. Chen, T.J. McAvoy: "Plantwide control system design: Methodology and application to a vinyl acetate process", *Ind. Eng. Chem. Res.* **42** (2003) no. 20, 4753–4771.

35 S. Engell: "Feedback control for optimal process operation", *J. Proc. Control* **17** (2007) 203–219.

36 E.M. Vasbinder, K.A. Hoo: "Decision-based approach to plantwide control structure synthesis", *Ind. Eng. Chem. Res.* **42** (2003) 4586–4598.

37 J.D. Ward, D.A. Mellichamp, M.F. Doherty: "Insight from Economically Optimal Steady-State Operating Policies for Dynamic Plantwide Control", *Ind. Eng. Chem. Res.* **45** (2006) 1343.

38 A. Zheng, R.V. Mahajanam, J.M. Douglas: "Hierarchical procedure for plantwide control system synthesis", *AIChE J.* **45** (1999) no. 6, 1255–1265.

39 S. Skogestad: "Control structure design for complete chemical plants", *Comp. Chem. Eng.* **28** (2004) no. 1, 2, 219–234.

40 S. Skogestad: "Plantwide control: the search for the self-optimizing control structure", *J. Proc. Control* **10** (2000) 487–507.

41 V. Alstad, S. Skogestad, E.S. Hori: "Optimal measurement combinations as controlled variables", *J. Proc. Control* **19** (2009) 138–148.

42 J.J. Downs, S. Skogestad "An industrial and academic perspective on plantwide control" *Annual Reviews in Control* **35** (2011) 99–110.

43 R.M. Price, C. Georgakis: "Plantwide regulatory control design procedure using a tiered framework", *Ind. Eng. Chem. Res.* **32** (1993) 2693–2705.

44 E.M.B. Aske, S. Skogestad: "Consistent inventory control", *Ind. Eng. Chem. Res.* **48** (2009) no. 44, 10892–10902.

45 E.M.B. Aske, S. Strand, S. Skogestad: "Coordinator MPC for maximizing plant throughput", *Comp. Chem. Eng.* **32** (2008) 195–204.

46 W.L. Luyben: "Snowball Effects in Reactor Separator Processes with Recycle", *Ind. Eng. Chem. Res.* **33** (1994) 299–305.

47 A. Araujo, S. Skogestad: "Control structure design for the ammonia synthesis process", *Comp. Chem. Eng.* **32** (2008) no. 12, 2920–2932.

48 T. Larsson, M. Govatsmark, S. Skogestad, C.C. Yu: "Control structure selection for Reactor, Separator and Recycle Processes", *Ind. Eng. Chem. Res.* **42** (2003) , 1225–1234;

49 J. Forcada, J.M. Asua: "Modeling of unseeded emulsion copolymerisation of styrene and methyl methacrylate", *J. Polym. Sci.: Part A: Polym. Chem.* **28** (1990) 987–1009.

50 J. Forcada, J.M. Asua: "Emulsion copolymerization of styrene and methyl methacrylate. II. Molecular weights", *J. Polym. Sci.: Part A: Polym. Chem.* **29** (1991) 1231–1242.

51 R.G. Gilbert: *Emulsion Polymerization: A mechanistic approach*, Academic Press, New York 1995.

52 J.M. Asua: "A new model for radical desorption in emulsion polymerization", *Macromolecules* **36** (2003) no. 16, 6245.

53 ISA, ANSI/ISA-88.01-1995: Batch control part 1: Models and terminology, Technical report, ISA, The Instrumentation, Systems, and Automation Society, 1995.

54 NAMUR, NE33: Requirements to be met by systems for recipe-based operations, Technical report, NAMUR, 1992.

55 NAMUR, NE59: Functions of the operation management level in batch oriented production. Technical report, NAMUR, 1996.

56 ISA, ANSI/ISA-95.00.01-2000: Enterprise-control system integration part 1: Models and terminology. Technical report, ISA, The Instrumentation, Systems, and Automation Society, 2000.

57 IEC, Batch control part 1: Models and terminology–IEC 61512-1, Technical report, International Electrotechnical Commission, Geneva, Switzerland, 1997.

58 IEC, Batch control. Part 2: Data structures and guidelines for languages–IEC 61512-2, Technical report, International Electrotechnical Commission, Geneva, Switzerland, 2001.

59 R.H. Perry, D.W. Green, J.O. Maloney: *Perry's Chemical Engineers' Handbook*, 7th ed., McGraw-Hill, New York 1999.

60 G.A. Montague, M.T. Tham, P.A. Lant: "Estimating the immeasurable without mechanistic models", *Trends Biotechnol.* **8** (1990) 82.

61 G.A. Montague, M.T. Tham, A.J. Morris, P.A. Lant: "Soft-sensors for process estimation and inferential control", *J. Proc. Control* **1** (1991) 3–14.

62 A. D'Anjou et al.: "Model reduction in emulsion polymerization using hybrid first principles/artificial neural networks models", *Macromol. Theory Simul.* **12** (2003) 4256.

63 L.M. Gugliotta et al.: "Estimation of conversion and copolymer composition in semicontinuous emulsion polymerization using calorimetric data", *Polymer* **36** (1995) no. 10, 2019–2023.

64 S. Krämer et al.: "Online monitoring of semi-continuous emulsion copolymerization: Comparing constrained extended kalman filtering to feed-forward calorimetry" *DYCOPS-6*, IFAC, Jejudo Island, Korea 2001, pp. 263–268.

65 S. Krämer: *Heat Balance Calorimetry and Multirate State Estimation Applied to Semi-Batch Emulsion Copolymerisation to Achieve Optimal Control*, Shaker Verlag, Achen 2005.

66 R.E. Kalman: "On the general theory of control systems", *First International Congress on Automatic Control*, Moscow 1960, pp. 481–492.

67 R.E. Kalman, R. Bucy: "New results in linear filtering and prediction", *Trans. ASME, Ser. D* **83** (1961) 98–108.

68 D.G. Luenberger: "Observing the state of a linear system", *IEEE Trans. Mil. Electron.* **8** (1964) 74–80.

69 E.G. Gilbert: "Controllability and observability in multivariable control systems", *SIAM Control Ser. A* 1963, 128–151.

70 M.L.J. Hautus: "Controllability and observability of linear autonomous systems", *Proc. Kon. Akad. Wetensci. Ser. A* 1969, no. 72, 443–448.

71 A.H. Jazwinski: *Stochastic Processes and Filtering Theory*, Academic Press, New York 1970.

72 A. Gelb: *Applied Optimal Estimation*, The M.I.T. Press, Massachusetts Institute of Technology, Cambridge 1974.

73 K.R. Muske, J.B. Rawlings, J.H. Lee: "Receding horizon recursive state estimation", *Proceedings of the American Control Conference*, San Francisco 1993, pp. 900–904.

74 D.G. Robertson, J.H. Lee, J.B. Rawlings: "A moving horizon-based approach for least-squares estimation", *AIChE J.* **42** (1996) no. 8, 2209–2224.

75 S. Julier, J. Uhlmann: "Unscented filtering and nonlinear estimation", *IEEE Proceedings* 2004, 401–422.

76 J.B. Rawlings, B.R. Bakshi: "Particle filtering and moving horizon estimation", *Comp. Chem. Eng.* 2006, 1529–1541.

77 D. Bonvin, P. de Valliere, D.W.T. Rippin: "Application of estimation techniques to batch reactors–I.: Modelling thermal effects", *Comp. Chem. Eng.* **13** (1989) no. 1/2, 1–9.

78 H. Schuler, C.U. Schmidt: "Calorimetric-state estimators for chemical reactor diagnosis and control: Review of methods and applications", *Chem. Eng. Sci.* **47** (1992) 899–915.

79 J. Valappil, C. Georgakis: "A systematic approach for the use of Extended Kalman Filters in batch processes", in S. Yurkovich (ed.): *Proc. Am. Control Conf.* 1999, 1143–1147.

80 B.J. Guo et al.: "Nonlinear adaptive control for multivariable chemical processes", *Chem. Eng. Sci.* **56** (2001) 6781–6791.

81 S. Krämer: "Determining the best reaction calorimetry technique: Theoretical development", *Comp. Chem. Eng.* **29** (2005) 349–365.

82 A. Tietze, I. Lüdtke, K.-H. Reichert: "Temperature oscillation calorimetry in stirred tank reactors", *Chem. Eng. Sci.* **51** (1996) no. 11, 3131–3137.

83 W. Mauntz et al.: "Neue Auswertungsalgorithmen und optimierte Anregung für die Temperaturoszillationskalorimetrie", *Chem. Ing. Tech.* **80** (2008) no. 1–2, 215.

84 S. Krämer, R. Gesthuisen: "Simultaneous estimation of the heat of reaction and the heat transfer coefficient by calorimetry: Estimation problems due to model simplification and high jacket flow rates: Theoretical development", *Chem. Eng. Sci.* **60** (2005) 4233–4248.

85 P. Nomikos, J.F. MacGregor: "Monitoring batch processes using multiway principal component analyses", *AIChE J.* **40** (1994) no. 8, 1361–1375.

86 K.A. Kosanovich: "Improved process understanding using multiway principal component analysis", *Ind. Eng. Chem. Res. Am. Chem. Soc.* **35** (1996) 138–146.

87 M.J. Piovoso, K.A. Hoo: "Multivariate statistics for process control", *IEEE Control Systems Magazine* **22** (2002) no. 5, 8–9.

88 C. Ündey, A. Çinar: "Statical monitoring of multistage, multiphase batch processes", *IEEE Control Systems Magazine* **22** (2002) no. 5, 40–52.

89 E. Martin, J. Morris, S. Lane: "Monitoring process manufacturing performance", *IEEE Control Systems Magazine* **22** (2002) no. 5, 26–39.

90 T. Kourti: "Multivariate dynamic data modelling for analysis and statistical process control of batch processes, start-ups and grade transitions", *J. Chemom.* **17** (2003) 93–109.

91 J.-M. Lee, C.Y. Yoo, I.-B. Lee: "On-line batch monitoring using different unfolding method and independent component analysis", *J. Chem. Eng. Jpn.* **36** (2003) no. 11, 1384–1396.

92 R.F. Hartl, S.P. Sethi, R.G. Vickson: "A survey of the maximum principles for optimal control problems with state constraints", *SIAM Rev.* **37** (1995) no. 2, 181–218.

93 V.S. Vassiliadis, R.W.H. Sargent, C.C. Pantelides: "Solution of a class of multistage dynamic optimization problems. 1. Problems without path constraints", *Ind. Eng. Chem. Res.* **33** (1994) no. 9, 2111–2122.

94 V.S. Vassiliadis, R.W.H. Sargent, C.C. Pantelides: "Solution of a class of multistage dynamic optimization problems. 2. Problems with path constraints" *Ind. Eng. Chem. Res.* **33** (1994) no. 9, 2123–2133.

95 P.E. Gill, W. Murray, M.H. Wright: *Practical Optimization*, Academic Press, London 1981.

96 J.E. Cuthrell, L.T. Biegler: "On the optimization of differential-algebraic process systems", *AIChE J.* **33** (1987) no. 8, 1257–1270.

97 J.E. Cuthrell, L.T. Biegler: "Simultaneous optimization and solution methods for batch reactor control profiles", *Comp. Chem. Eng.* **13** (1989) no. 1/2, 49–62.

98 A. Forsgren, P.E. Gill, M.H. Wright: "Interior methods for nonlinear optimization", *SIAM Rev.* **44** (2002) no. 4, 525–597.

99 A. Wächter, L.T. Biegler: "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming", *Mathematical Programming* **106** (2006) 25–57.

100 M. Diehl et al.: "Real-time optimization and nonlinear model predictive control of processes governed by differential-algebraic equations", *IFAC Symposium: Advanced Control of Chemical Processes*, Pisa 2000.

101 D.B. Leineweber et al.: "An efficient multiple shooting based reduced sqp strategy for large-scale dynamic process optimization. Part 1: Theoretical aspects", *Comp. Chem. Eng.* **27** (2003) no. 2, 157–166.

102 A. Cruse et al.: "Batch process modeling and optimization", in E. Korovessi A.A. Linninger (eds.): *Batch Processes*, Marcel Dekker, New York 2005, pp. 305–380.

103 B. Srinivasan, D. Bonvin: "Interplay between identification and optimization in run-to-run optimization schemes", *Am. Control Conf.*, AACC, Anchorage 2002, pp. 2174–2179.

104 B. Srinivasan, S. Palanki, D. Bonvin: "Dynamic optimization of batch processes. I. Characterization of the nominal solution", *Comp. Chem. Eng*, **27** (2003) 1–26.

105 B. Srinivasan et al.: "Dynamic optimization of batch processes. II. Role of measurements in handling uncertainty", *Comp. Chem. Eng.* **27** (2003) 27–44.

106 E.D. Castillo, A.M. Hurwitz: "Run-to-run process control: Literature review and extensions", *J. Quality Technol.* **29** (1997) 184–196.

107 P. Tatjewski: "Iterative optimizing set-point control-the basic principle redesigned", *15th Triennial IFAC World Congress*, Barcelona 2002.

108 W. Gao, S. Engell: "Iterative set-point optimization of batch chromatography", *Comp. Chem. Eng.* **29** (2005) 1401–1409.

109 B. Chachuat, B. Srinivasan, D. Bonvin: "Adaptation strategies for realtime optimization", *Comp. Chem. Eng.* **33** (2009) 1557–1567.

110 S. Engell et al.: "Continuous-discrete interactions in chemical processing plants", *IEEE Proceedings* **88** (2000) 1050–1068.

111 S. Gass, T. Saaty: "Parametric objective function (part 1)", *J. Oper. Res. Soc. Am.* **2** (1954) no. 3, 316–319.

112 S. Gass, T. Saaty: "Parametric objective function (part 2)", *J. Oper. Res. Soc. Am.* **3** (1955) no. 4, 395–401.

113 T. Gal, J. Nedoma: "Multiparametric linear programming", *Management Science* **18** (1972) no. 7, 406–422.

114 A.V. Fiacco: *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*, Academic Press, New York 1983.

115 V. Bansal, J.D. Perkins, E.N. Pistikopoulos: "Flexibility analysis and design of linear systems by parametric programming", *AIChE J.* **46** (2000) no. 2, 335, 2000.

116 E.N. Pistikopoulos: *Parametric and stochastic programming algorithms for process synthesis, design and optimization under uncertainty*, Aspen World, Boston, MA, 1997.

117 V. Dua, N.A. Bozinis, E.N. Pistikopoulos: "A multiparametric programming approach for mixed-integer quadratic engineering problems", *Comp. Chem. Eng.* **26** (2002) 715–733.

118 E.N. Pistikopoulos et al.: "On-line optimization via off-line optimization tools", *Comp. Chem. Eng.* **26** (2002) 175–185.

119 Y. Ohtake, N. Nishida: "A Branch-and-Bound algorithm for 0-1 parametric Mixed-Integer programming", *Operations Research Letters* **4** (1985) no. 1, 41–45.

120 A. Pertsinidis: On the parametric optimization of mathematical programs with binary variables and its application in the chemical engineering process synthesis, PhD thesis, Department of Chemical Engineering, Carnegie-Mellon University, Pittsburg 1992.

121 J. Acevedo, E.N. Pistikopoulos: "A multiparametric programming approach for linear process engineering problems under uncertainty", *Ind. Eng. Chem. Res.* **36** (1997) no. 3, 717–728.

122 V. Dua, E.N. Pistikopoulos: "An algorithm for the solution of multiparametric mixed integer linear programming problems", *Annals of Operations Research* **99** (2000) 123–139.

123 J. Acevedo, E.N. Pistikopoulos: "A parametric minlp algorithm for process synthseis problems under uncertainty", *Ind. Eng. Chem. Res.* **35** (1996) no. 1, 147.

124 V. Dua, E.N. Pistikopoulos: "Algorithms for the solution of multiparametric mixed-integer nonlinear optimization problems" *Ind. Eng. Chem. Res.* **38** (1999) no. 10, 3976–3987.

125 V. Dua, K.P. Papalexandri, E.N. Pistikopoulos: "Global optimization issues in multiparametric continuous and mixed-integer optimization problems", *J. Global Optimization* **30** (2004) , 59–89.

126 V. Sakizlis, J. Perkins, E.N. Pistikopoulos: "An algorithm for multiparametric dynamic optimization", *ICOTA'01*, Hong Kong 2001.

127 E.N. Pistikopoulos, M. Georgiadis, V. Dua: *Multiparametric Programming: Theory, Algorithms and Applications*, vol. **1**, Wiley-VCH Verlag, Weinheim 2007.

128 E.N. Pistikopoulos, M. Georgiadis, V. Dua: *Multiparametric Model-Based Control: Theory and Applications*, vol. **2**, Wiley-VCH Verlag, Weinheim 2007.

129 V. Sakizlis, J.D. Perkins, E.N. Pistikopoulos: "Explicit solutions to optimal control problems for constrained continuous-time linear systems", *IEE Proceedings: Control Theory and Applications* **152** (2005) no. 4, 443–452a.

130 V. Sakizlis Et al.: "Design of robust model-based controllers via parametric programming", *Automatica* **40** (2004) 189–201.

131 A. Bemporad, F. Borrelli, M. Morari: "Min-max control of constrained uncertain discrete–time linear systems", *IEEE Trans. Aut. Con.* **48** (2003) 1600–1606.

132 M. de la Pena et al.: "A dynamic programming approach for determining the explicit solution of linear mpc controllers", *43rd IEEE Conference on Decision and Control* **3** (2004) 2479–2484.

133 N. Faisca et al.: "A multi-parametric programming approach for constrained dynamic programming problems", *Optimization Letters* **2** (2008) 267–280.

134 A. Johansen: "On multiparametric nonlinear programming and explicit nonlinear model predictive control", *41st IEEE*

Conference on Decision and Control, Las Vegas, Nevada, USA, 2002.

135 A. Bemporad: "Multiparametric nonlinear integer programming and explicit quantized optimal control", *42nd IEEE Conference*, Maui, Hawaii, Dec. 2003.

136 V. Sakizlis et al.: "Towards the design of parametric model predictive controllers for non-linear constrained systems", in R. Findeisen, F. Allgöwer, L. Biegler (eds.): *Assessment and Future Directions of Nonlinear Model Predictive Control*, vol. **358**, 2007.

137 J. Mandler et al.: "Parametric model predictive control of air separation", *International Symposium on Advanced Control of Chemical Processes*, ADCHEM, Gramado, Brazil, 2006.

138 E.N. Pistikopoulos, N.A. Bozinis, V. Dua, J.D. Perking, V. Sakizlis: EP 1399784, 2004.

139 E.N. Pistikopoulos, N.A. Bozinis, V. Dua, J.D. Perking, V. Sakizlis: US 7433743, 2008.

140 V. Dua, E.N. Pistikopoulos: "An outer-approximation algorithm for the solution of multiparametric minlp problems", *Comp. Chem. Eng.* **22** (1998) 955–958.

141 E.N. Pistikopoulos et al.: "Nonlinear multiparametric model-based control", *International workshop on Assessment and Future Directions of Nonlinear Model Predictive Control*, 2008.

142 N.P. Faisca et al.: "Parametric global optimisation for bilevel programming", *J. Global Optimization* **38** (2007) no. 4, 609–623.

143 L.F. Dominguez, E.N. Pistikopoulos: "Global optimization of mixed-integer bi-level problems via multi-parametric programming", *10th International Symposium on Process Systems Engineering*, 2009.

144 N. Faisca, V.D. Kosmidis, B. Rustem, E.N. Pistikopoulos: "Global optimization of multi-parameric milp problems", *Journal of Global Optimization* **45** (2009) 131–151.

145 T. Johansen: "Reduced explicit constrained linear quadratic regulators", *IEEE Trans. Aut. Con.* **48** (2003) no. 5, 823–828.

146 K.I. Kouramas, V. Sakizlis, E.N. Pistikopoulos: "Design of robust model predictive controllers via parametric programming", *Encyclopedia of Optimization* (2009) 677–687.

147 A. Ben-Tal, A. Nemirovski: "Robust solutions of linear programming problems contaminated with uncertain data", *Math. Prog.* **88** (2000) 411–424.

148 X. Lin, S. Janak, C. Floudas: "A new robust optimization approach for scheduling under uncertainty", *Comp. Chem. Eng.* **28** (2004) 1069–1085.

149 E.N. Pistikopoulos, K.I. Kouramas, N.P. Faisca: "Robust multiparametric model-based control", *19th European Symposium on Computer Aided Process Engineering*, Cracow, Poland 2009.

150 E.N. Pistikopoulos, K.I. Kouramas, C. Panos: "Explicit robust model predictive control", *International Symposium on Advanced Control of Chemical Processes (ADCHEM)*, Istanbul 2009.

151 V. Kosmidis et al.: "Output feedback parametric controllers for an active valve train actuation system", *45th IEEE Conference on Decision and Control*, Dec. 2006, pp. 4520–4525.

152 P. Dua: "*Model Based and Parametric Control for Drug Delivery Systems*", PhD thesis, Department of Chemical Engineering, Imperial College London, London 2005.

153 R. Gani, J. Perregaard, H. Johansen: "Simulation strategies for design and analysis of complex chemical processes", *Trans I. Chem. E.* **68** (1990) Part A, 407–417.

154 C. Pantelides: "SpeedUp–Recent advances in process simulation", *Comp. Chem. Eng.* **12** (1988) no. 7, 745–755.

155 C. Shewchuk: "MASSBAL MKII: New process simulation system", *Pulp Pap. Can.* **88** (1987) no. 5, T161–T167.

156 K.H. Bendiksen et al.: "The Dynamic Two-Fluid Model Olga: Theory and Application". *SPE Production Engineering*, May 1991, Houston pp. 171–180.

157 J. Tingas, R. Frimpong, J. Liou: "Integrated reservoir and surface network simulation in reservoir management of southern North Sea gas reservoirs", *1998 SPE European Petroleum Conference*, The Hague, Netherlands, 20–22 October 1998.

158 H. Laux et al.: "Multidimensional Simulations of Multiphase Flow for Improved Design and Management of Production and Processing Operation", *Offshore Technology Conference*, Houston, Texas, 5–8 May 2008.

159 M. Hollender: *Collaborative Process Automation Systems*, Research Triangle Park, NC, ISA 2010.

160 ANSI/ISA-99.02.01-2009: *Security for Industrial Automation and Control Systems: Establishing an Industrial Automation and Control Systems Security Program*, Research Triangle Park, NC, ISA, 2009.

161 D.B. Cameron, R.J. Ødegaard, E. Glende: "On-line Modeling in the petroleum industry: Successful applications and future perspectives", in R. Gani, S. Bay Jørgensen (eds.): *11th European Symposium on Computer Aided Process Engineering*, Elsevier, Amsterdam 2001, pp. 111–116.

162 Pipeline simulation interest group, http://www.psig.org (accessed 3 January 2012).

163 A. Ek, et al.: "Monitoring Systems for Multiphase Gas-Condensate pipelines", *22nd Annual OTC*, Houston, Texas, 7–10 May, 1990.

164 M. Hyllseth, D. Cameron, K. Havre: "Operator training and operator support using multiphase pipeline models and dynamic process simulaton: sub-sea production and on-shore processing", in A. Kraslawski, I. Turunen (eds.): *13th European Symposium on Computer Aided Process Engineering*, Elsevier, Amsterdam 2003, pp. 425–430.

165 D.B. Cameron, C. Larsson, I.L. Sperle, H. Nordhus" "VALMUE: Linking process operations with training and engineering through on-line simulation and simulation data management" in M. Ierapetriou, M. Bassett, S. Pistikopoulos (eds.): *Proceedings of the Fifth International Conference on Foundations of Computer Aided Process Operations (FOCAPO)*, Cambridge, MA 2008.

166 J.A. Romagnoli, P.A. Rolandi: "Model-centric technologies for support of manufacturing operations", in W. Marquardt, C. Pantelides (eds.): *16th European Symposium on Computer Aided Process Engineering and 9th International Symposium on Process Systems Engineering*, Elsevier, Amsterdam 2006, pp. 63–70.