SURROGATE-BASED OPTIMIZATION OF A FLEXIBLE INTEGRATED BIOREFINERY

Yuqing Luo, Marianthi Ierapetritou*
Department of Chemical and Biomolecular Engineering, University of Delaware
Newark, DE 19716

Abstract

Biomass conversion facilities are subject to many uncertainties, such as the feedstock supply, product demand, and chemical prices. The integrated biorefinery makes use of different technologies and feedstock types to improve both the profitability and flexibility of the process. The flexibility index has been developed as a means of quantifying the process's capacity to handle different sources of uncertainties. However, the evaluation of the flexibility index requires solving the bi-level optimization problem, which is difficult to be directly incorporated into the process design problem. In this work, a two-stage stochastic programming problem is built for the rational design of biorefinery that is environmentally friendly and economically viable. Different data-driven surrogate models are constructed for the biorefinery flexibility index. They are embedded in the integrated biorefinery design problem as the flexibility requirement constraints. The result shows that the rectified linear unit (ReLU) neural network model achieves better performance than the support vector regression (SVR) and Lasso regression due to the versatile fitting ability and the resulting MILP formulation.

Keywords

Flexibility index, Surrogate model, Neural network, Life cycle assessment, Integrated biorefinery.

Introduction

Utilizing biomass as the feedstock for chemical production has been established as an effective strategy to reduce greenhouse gas emissions. However, there are multiple levels of uncertainties associated with the biomass feedstock and its conversion technologies at the early stage of development. Ignoring these uncertainties may render the designs and operations sub-optimal or even infeasible (Bhosekar, et al., 2021). Hence, it is essential to consider uncertainties in biorefinery optimization to allow for additional flexibility (Sahinidis, 2004). Stochastic programming is a commonly used formulation to incorporate uncertainty information into optimization (Martín and Martínez, 2015). Probability distributions of uncertain parameters are first sampled into a set of scenarios. Then, the deterministic equivalent problem is solved by optimizing the expected values of the objective functions (Bhosekar, et al., 2021).

On the other hand, the flexibility index has been proposed to quantify the maximum parameter ranges that a process can tolerate for feasible operation. The classic definition of flexibility index is based on the maximum hyperrectangle inscribed in the feasible region. The vertex enumeration method has been applied to convex problems. However, this method becomes less effective in high-dimensional and more general non-convex problems (Swaney and Grossmann, 1985). The active set method is developed to introduce the binary variable that indicates whether the constraints are active and solve the problem more efficiently (Grossmann and Floudas, 1987). The concept of flexibility index has been introduced in the

* To whom all correspondence should be addressed. Email: mgi@udel.edu

process design problems, such as selecting the operating ranges (Ochoa, et al., 2021). The flexibility index of the supply chain has also been analyzed to demonstrate the trade-off between costs and flexibility (Wang, et al., 2016).

Surrogate embedded optimization has been applied in many areas of process system engineering. Bhosekar and Ierapetritou (2020) utilized the support vector machine as the surrogate for feasibility constraints and adopted the vertex formulation to add flexibility index requirement of the processes Artificial neural network (ANN) surrogate model has also been embedded in the optimization for case studies in compressors, fermentation, and chemical process operations (Schweidtmann and Mitsos, 2019). Adaptive sampling strategy has been implemented using both ANN and Kriging surrogate models to optimize computationally expensive processes (Metta, et al., 2021, Rogers and Ierapetritou, 2015). Kim and Boukouvala (2020) extended Gaussian Process and ANN to mixed-integer surrogate models by using one-hot encoding for optimization.

Rectified linear unit (ReLU) is one of the most widely used activation functions in the ANN models. This activation function has demonstrated the ability to capture the nonlinearity of the model while not adding too much model complexity (Katz, et al., 2020a).

This work incorporates the requirement for process flexibility into the economical and sustainable process design. Three different data-driven surrogate models, namely support vector regression (SVR) with radial basis function (RBF) features, polynomial features, and feed-forward neural network with ReLU activation function, are built for the flexibility index constraints of the biorefinery.

Two-stage Stochastic Programming for Biorefinery

The two-stage stochastic programming formulation with recourse actions is utilized for the biorefinery optimization problem. The first-stage design decisions, such as the technology choices and capacities, are made before uncertainty realization. Then, the actual feedstock supply, product demand, and prices are observed. The second-stage operational-level decisions include the actual production activity and stream flowrates, ensuring process feasibility and improve performance.

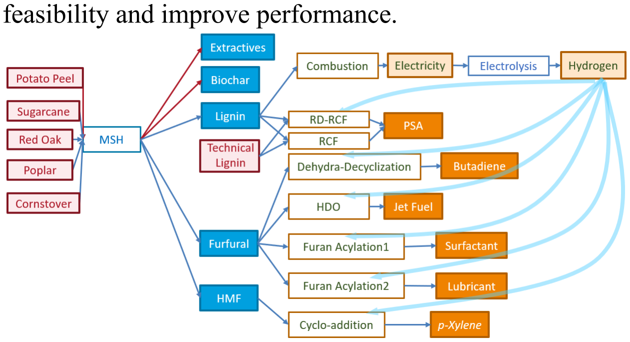

Figure 1. Biorefinery Superstructure

The objective of the superstructure optimization problem is to maximize the expected profit and minimize the expected global warming potential (GWP) while having certain flexibility. The expected profit includes the revenue of selling products/byproducts minus both the first-stage capital investments and second-stage feedstock and operating costs. The annualized capital investment and fixed operating costs (such as administration and maintenance) demonstrate the famous "economy-of-scale" and follow the power-law relationship (Eq. (1) and Eq. (2)) (Tribe and Alpine, 1986).

$$\boldsymbol{CAPEX_{annual}} = \sum_{m\in M} \boldsymbol{C_0(m)} \cdot \left(\frac{\boldsymbol{Q(m)}}{\boldsymbol{Q_0(m)}}\right)^{\boldsymbol{\alpha(m)}} \quad (1)$$

$$\boldsymbol{OPEX(\omega)} = \sum_{m\in M} \left[\boldsymbol{O_0(m)} \cdot \left(\frac{\boldsymbol{Q(m)}}{\boldsymbol{Q_0(m)}}\right)^{\boldsymbol{\beta(m)}} + \boldsymbol{R_0(m)} \cdot \left(\frac{\boldsymbol{x(m,\omega)}}{\boldsymbol{Q_0(m)}}\right)\right] \quad (2)$$

where $C_0(m)$ is the annualized capital cost basis at a plant scale of $Q_0(m)$, $Q(m)$ is the actual designed capacity, and $\alpha(m)$ is the capital cost scaling exponent for technology $m \in M$. $O_0(m)$ and $R_0(m)$ are the known fixed and variable operational costs, and $\beta(m)$ is the operating cost scaling exponent for technology $m \in M$. $x(m,\omega)$ is the actual production activity in the process unit m in scenario $\omega \in \Omega$. The total profit is calculated by Eq. (3):

$$\boldsymbol{Profit(\omega)} = \sum_{i\in I} \boldsymbol{\mu(i,\omega)} \cdot \boldsymbol{\rho(i,\omega)} - \boldsymbol{OPEX(\omega)} - \boldsymbol{CAPEX_{annual}} \quad (3)$$

where $\mu(i,\omega)$ is the purchase or selling price, $\rho(i,\omega)$ is the flow rate of the feedstock, intermediate or product $i \in I$ in scenario $\omega \in \Omega$. The profit objective is piece-wise linearized using the function interpolation by lookup of grid points (Misener and Floudas, 2010).

The global warming potential (GWP) is chosen as the environmental impact of interest since it is the most popular environmental indicator (Luo and Ierapetritou, 2020). Equation (4) includes the emissions associated with production activity, upstream raw material, and utility extraction, avoided burden from selling byproducts, and plant construction.

$$\boldsymbol{GWP(\omega)} = \sum_{m\in M} \boldsymbol{\pi_p(m)} \cdot \boldsymbol{x(m,\omega)} + \sum_{i\in I} \boldsymbol{\pi_u(i)} \cdot \boldsymbol{\rho(i,\omega)} + \sum_{s\in S} \boldsymbol{\pi_e(s)} \cdot \boldsymbol{\gamma(s,\omega)} + \sum_{m\in M} \boldsymbol{\pi_c(m)} \cdot \boldsymbol{Q(m)} \quad (4)$$

where $\pi_p(m)$, $\pi_u(i)$, $\pi_e(s)$, and $\pi_c(m)$ are the greenhouse emissions during the production stage, upstream raw material extraction, electricity generation from energy source $s \in S$, and plant construction. Here, $\gamma(s,\omega)$ is the electricity usage from energy source $s \in S$ in scenario ω.

Aspen Plus process flowsheets are first developed for each technology by setting the experimental yields in the reactor units and designing appropriate separation units. The conversion coefficients of each process unit are extracted to capture the inlet and outlet flowrate for feedstocks, intermediates, and products $i \in I$. Equation (5) is the mass balance for each process unit:

$$\boldsymbol{\rho(i,\omega) = \sum_{m\in M} v(i,m)\cdot x(m,\omega)} \tag{5}$$

The parameter $\boldsymbol{v(i,m)}$ is the conversion coefficient, representing the consumption or production of chemical $i \in I$ in process $m \in M$. A negative $\rho(i,\omega)$ indicates i is a feedstock or intermediate that has to be purchased from the suppliers, while a positive $\rho(i,\omega)$ means i is sold to the market. Uncertainties in pretreatment and other process yields could also be captured in the conversion coefficients. This formulation is similar to the matrix-based LCA calculation, which can easily expand and include more processes and technologies as they become available or of interest.

The binary variable, $\boldsymbol{y(m)}$, is used to indicate whether technology m is chosen or not. When this technology is chosen, its capacity is then limited by the possible minimum and maximum capacities, $Q^-(m)$ and $Q^+(m)$, as illustrated in Eq. (6).

$$\boldsymbol{y(m)\cdot Q^-(m) \le Q(m) \le y(m)\cdot Q^+(m)} \tag{6}$$

Additionally, some feedstocks (and thus technologies) may share the same facility, which is captured by equating their capacities (Eq. (7)).

$$\boldsymbol{Q(m) = Q(m')}\ \ \forall\, \boldsymbol{m, m'} \text{ that share the same facility} \tag{7}$$

There are also bounds on the operational decisions, such as the flow rate of feedstocks could not exceed their maximum supply, and the maximum production should be less than the market demands in each case. Equation (8) imposes the lower and upper bounds ($\rho^-(i,\omega)$ and $\rho^+(i,\omega)$) on the material flowrates.

$$\boldsymbol{\rho^-(i,\omega) \le \rho(i,\omega) \le \rho^+(i,\omega)} \tag{8}$$

Moreover, the actual production activity in each production unit should not exceed its designed capacity. Equation (9) thus connects the first- and second-stage decisions of stochastic programming.

$$\boldsymbol{x(m,\omega) \le Q(m)} \tag{9}$$

The aforementioned two-stage stochastic programming model could readily consider other operational level constraints, such as the service level of the supply chain.

Flexibility Evaluation for Integrated Biorefinery

The biorefinery operational flexibility is quantified by the max-min-max flexibility index formulation (Eq. (10)) (Swaney and Grossmann, 1985).

$$\begin{gathered}\boldsymbol{F = \max \delta}\\ \boldsymbol{s.t.\ \max_{\theta\in T(\delta)} \psi(d,\theta) \le 0}\\ \boldsymbol{\psi(d,\theta) = \min_{z}\ \max_{j} f_j(d,z,\theta)}\\ \boldsymbol{T(\delta) = \{\theta: \theta^N - \delta\cdot\Delta\theta^- \le \theta \le \theta^N + \delta\cdot\Delta\theta^+\}}\end{gathered} \tag{10}$$

where F is the flexibility index, $\psi(d,\theta)$ is the feasibility function that indicates the existence of constraint violations, and $T(\delta)$ is the hyperrectangular that allows feasible uncertain parameters to vary without correlations. The interactions between different uncertain parameters could be captured by ellipsoidal or diamond shape uncertainty sets (Pulsipher and Zavala, 2018).

Here, $f_j(d,z,\theta)$ is obtained by substituting equality constraints and eliminating state variables in inequality constraints. In this model, d is the design variable (e.g., choice of technologies $y(m)$ and capacities $Q(m)$), θ is the uncertain parameters that affect the process feasibility (e.g., supply or demands, $\rho^-(i,\omega)$ or $\rho^+(i,\omega)$, and conversion coefficients, $v(i,m)$), and z is the recourse/control actions (e.g., production activity $x(m,\omega)$ and flowrates $\rho(i,\omega)$).

The active-set formulation of the flexibility index is used to convert the problem into a single level MINLP (Eq. (11)) (Floudas and Grossmann, 1987).

$$\begin{gathered}\mathbf{F(d)} = \min_{\theta,z,\delta,s_j,\lambda_j,y_j} \boldsymbol{\delta}\\ \mathbf{s.t.}\ \ \boldsymbol{s_j + f_i(d,z,\theta) = 0,\ \forall j \in J}\\ \boldsymbol{\sum_{j\in J} \lambda_j = 1,}\\ \boldsymbol{\sum_{j\in J} \lambda_j \cdot \frac{\partial f_i}{\partial z} = 0,}\\ \boldsymbol{\lambda_j - y_j \le 0,}\\ \boldsymbol{s_j - U(1 - y_j) \le 0,}\\ \boldsymbol{\sum_{j\in J} y_j \le n_z + 1,}\\ \boldsymbol{T(\theta) = \{\theta: \theta^N - \delta\cdot\Delta\theta^- \le \theta \le \theta^N + \delta\cdot\Delta\theta^+\}}\\ \boldsymbol{\delta \ge 0;\ \ y_j = 0, 1;\ \ s_j,\ \lambda_j \ge 0,, \forall j \in J}\end{gathered} \tag{11}$$

where s_j is the slack variable for each constraint f_i, λ_j is the Lagrange multiplier, y_j is the binary variable as an indicator for whether the inequality constraint f_i is active, and n_z represents the number of control variables.

Since the conversion coefficients are uncertain, the resulting active-set problem still contains bilinear terms in the KKT condition for the mass balance. Hence, the global MINLP solver BARON is used to solve the flexibility index problem when the process design is fixed.

Surrogate Modeling of Process Flexibility

A total of 1600 Latin hypercube sampling is performed in MATLAB on the capacity of each production unit. Then, the MATLAB-GAMS interface was used to solve each sampled case and records the flexibility index. The Scikit-learn and TensorFlow machine learning packages in Python are used to build and assess different SVR, Lasso regression, and neural network data-driven surrogate models (Eq. (12)).

$$\hat{\boldsymbol{F}}(\boldsymbol{d}) \approx \mathbf{F}(\mathbf{d}) = \min_{\boldsymbol{\theta},\boldsymbol{z},\boldsymbol{\delta},\, s_j,\, \lambda_j,\, y_j} \boldsymbol{\delta} \tag{12}$$

The fitted model parameters are used in the GAMS optimization program to build the surrogate constraints for the flexibility index requirement.

Support Vector Regression (SVR) with RBF Kernels

The SVR model (Eq. (13)) is easy to train and flexible enough to introduce different types of nonlinear kernels (Nandi, et al., 2004).

$$\hat{\boldsymbol{F}}^{\boldsymbol{RBF}}(\boldsymbol{x}) = \sum_{s \in SV} \boldsymbol{a_s}\, \boldsymbol{e}^{-\gamma \|x - v_s\|^2} + \boldsymbol{b} \tag{13}$$

where v_s is the *s*th support vector obtained from model training, ‖·‖ represents the Euclidean distance, and γ is the hyperparameter associated with the length scale.

However, using all features obtained in the sampled designs tends to overfit the data. Thus, feature selection is performed by sequentially removing features from the set of all the candidates. As the number of features reduces from 7 to 4, the fitting error of the training data does not increase significantly while the error in the testing set drops, meaning that the three features removed correspond to the processes less relevant to the process flexibility. Although the SVR with RBF kernels are able to approximate different types of functions very well, the highly nonlinear and non-convex nature of this model becomes an obstacle in the design optimization problem.

Lasso Regression with 3rd-order Polynomial Features

Lasso regression is a popular regularization technique in model fitting to reduce the model complexity and help avoid overfitting. This method adds an *L1*-norm penalty term of model coefficients to the objective function (Terrell, 2022). After cross-validation of the highest power of the model, the 3rd-order polynomial model has the best performance.

However, a large number of non-convex terms are still introduced into the Lasso regression when using all 3rd-order polynomial features (e.g., bilinear and trilinear terms). These non-convex terms are difficult to relax when treated as optimization constraints. To reduce the non-convexity of the surrogate model, trilinear terms ($x_i x_j x_k$) are removed from the regression, and only the bilinear terms are retained in Eq. (14) to capture the interactions.

$$\hat{\boldsymbol{F}}^{\boldsymbol{3rd}}(\boldsymbol{x}) = \sum_i \boldsymbol{a_i\, x_i} + \sum_i \boldsymbol{b_i\, x_i^2} + \sum_i \sum_{j>i} \boldsymbol{c_{i,j}\, x_i x_j} + \sum_i \boldsymbol{d_i\, x_i^3} + \boldsymbol{h} \tag{14}$$

Feed-forward Neural Network (FFNN)

The artificial neural network model has been successfully implemented in many areas, including chemical process design and control (Katz, et al., 2020b). In this case study, the piece-wise linear activation function, ReLU, is used to capture the nonlinearity of the model. The first layer of the neural network, $z^{(1)}$, contains the input vector. Then, the outputs of the *k*th layer ($k \geq 2$) are calculated by Eq. (15):

$$\boldsymbol{z^{(k)}} = \boldsymbol{h_k}(\boldsymbol{W^{(k-1)} z^{(k-1)}} + \boldsymbol{b^{(k-1)}}) \tag{15}$$

Here, the second layer of the ANN model is the hidden layer with the ReLU activation function (Eq.(16)).

$$\boldsymbol{h_2}(\boldsymbol{x}) = \boldsymbol{ReLU}(\boldsymbol{x}) = \mathbf{max}\,(\mathbf{0}, \boldsymbol{x}) \tag{16}$$

The last layer (third layer) just applies the linear combination of $z^{(2)}$ to make the final prediction (Eq.(17)).

$$\boldsymbol{z^{(3)}} = \boldsymbol{W^{(2)} z^{(2)}} + \boldsymbol{b^{(2)}} \tag{17}$$

The only hidden layer with the ReLU activation function (second layer) could be converted into the mixed-integer linear constraints with the big-M formulation.

Surrogate-Embedded Optimization for Process Design

The surrogate-embedded optimization for process design is formulated in Eq. (18) using the different data-driven surrogate models of $\hat{F}(d)$:

$$\begin{gathered} \max_{\boldsymbol{y,Q,\rho,x}} \mathbb{E}_{\boldsymbol{\omega \in \Omega}}[\boldsymbol{Profit}(\boldsymbol{\omega})] \\ \boldsymbol{s.t.\ Constraints}\ (\mathbf{1}) - (\mathbf{9}) \\ \mathbb{E}_{\boldsymbol{\omega \in \Omega}}[\boldsymbol{GWP}(\boldsymbol{\omega})] \leq \boldsymbol{\varepsilon_2} \\ \hat{\boldsymbol{F}}(\boldsymbol{d}) \geq \boldsymbol{\Delta_1} \end{gathered} \tag{18}$$

The uncertain parameters used in the stochastic programming are biomass supply, chemical prices, and electricity supply from different sources since the empirical distributions could be estimated from historical data. On the other hand, the uncertainties in conversion coefficients and products' supply and demand are considered in the flexibility index subproblem to ensure feasibility without making assumptions on the actual parameter probability distribution. The resulting model is an MINLP problem with nonlinearity in the economic objective function and flexibility requirement constraints. After piece-wise linearization of the capital and fixed operating cost, the model is converted to a MILP problem if the surrogate model $\hat{F}(d)$ is also formulated by linear equations with integer variables. If the surrogate model is nonlinear, this design optimization problem becomes challenging to solve.

Case Study Results and Discussion

The epsilon constraint method is utilized to build the Pareto curve of the biorefinery's profit and GWP trade-off by varying the parameter ε_2 in Eq. (18). Δ_1 is the minimum flexibility index level required for the biorefinery, which is chosen as 0.3 for the case study.

Since the rest of the biorefinery design problem is MILP, the feed-forward neural network surrogate with ReLU activation function will retain the model linearity and be solved by the efficient CPLEX solver. When embedded into the GAMS optimization model, each neural introduces one extra integer variable in the ReLU function. On the other hand, the SVR and Lasso regression models are nonlinear and non-convex, requiring the use of a global MINLP solver, such as BARON. All optimization models are implemented in GAMS 33.1 on a computer with Intel Xeon E-2274G CPU @ 4.00GHz 32 GB RAM.

The biorefinery design problem that maximizes the process profit while maintaining a minimum flexibility index level is solved using the three surrogate constraints mentioned above, and the results are compared in Table 1. This is the first point on the multi-objective optimization Pareto front when ε_2 is the lowest.

Table 1. Computation results using different surrogate models for flexibility index

Surrogate Model	SVR with RBF	3rd-order Lasso	FFNN with ReLU
Training R^2	95%	89%	98%
Testing R^2	94%	88%	97%
Solver	BARON	BARON	CPLEX
Design Problem Solution time	3600 s	285 s	6 s
Gap	7%	0.5%	0%

The SVR model with RBF kernel is able to fit the flexibility index model with relatively low training and testing errors. Nonetheless, at least 66 support vectors have to be included even after tuning the model's hyperparameter. The fitted SVR model with RBF kernels is highly non-convex, making the upper bound updating progress very slow in the global optimization of the MINLP problem. During the course of optimization, the upper bound of profit does not improve even after 3600s, leaving an optimality gap of 7%. Hence, it is not practical to embed SVR surrogate models with RBF kernels in the design problem even though it provides accurate model prediction.

On the other hand, through cross-validation, a polynomial regression model with all 3rd features is capable of approximating the flexibility index. However, even with only 4 input variables, there are 16 trilinear non-convex terms, which also poses challenges to obtaining good relaxations. If only the bilinear terms are retained, and other trilinear interaction terms are ignored, the model could capture part of the trend in the flexibility index (testing R^2=88%) while still making the optimization problem solvable.

With only one hidden layer of 45 neurons, the ReLU feed-forward neural network model is able to fit the flexibility index very well. More importantly, this model does not introduce any nonlinearity or a large number of binary variables into the original MILP formulation. Therefore, this MILP model is solved efficiently in 6 s using the CPLEX solver.

Next, ten equidistant points from the Pareto curve of the GWP and profit are obtained using the ReLU feed-forward neural network surrogate. The new Pareto front with flexibility requirement is obtained within 41 s by gradually increase ε_2 and solve the MILP of Eq. (18) (orange points in Figure 2). These Pareto points are then benchmarked against the designs without flexibility consideration (Pareto solutions of Eq. (18) without the constraint on $\hat{\boldsymbol{F}}(\boldsymbol{d})$, blue points in Figure 2). As illustrated in Figure 2, a clear trade-off exists between the economic and environmental performances. This is because the candidate technologies that produce high value-added products are also energy-intensive and emit more greenhouse gases during the production stage. The comparison between both Pareto curves also demonstrates that the biorefinery capacities tend to be over-designed to improve the process flexibility. This capacity rise allows the biorefinery to treat more raw materials from different sources in case of extremely low reaction conversion and insufficient biomass supply. Inevitably, it causes higher emissions during the plant construction and more capital investment.

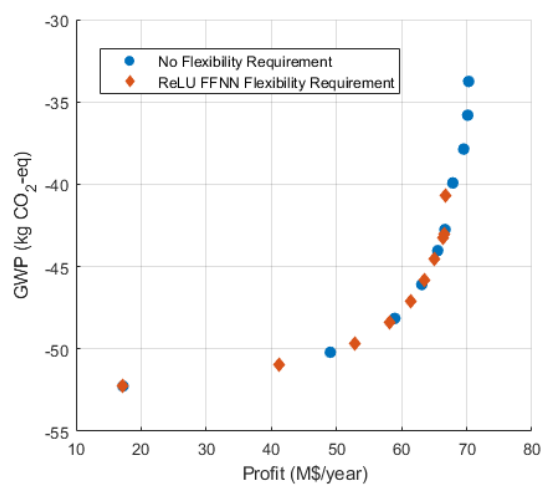

Figure 2. Pareto Curve of Biorefinery's GWP and Profit Using FFNN surrogate model

Meanwhile, the Pareto curve points with better GWP performance correspond to the biorefinery configurations with higher plant capacities and thus high flexibility index. They already have a high enough flexibility index to satisfy the flexibility requirement and need no capacity adjustment. Consequently, these points collapse into the old Pareto curve that does not consider flexibility during its design.

Conclusions

This work formulated the biorefinery design problem with flexibility index requirements under multiple uncertainty sources. Stochastic programming formulation is adopted for the overall integrated process design and aims for better average economic and environmental performance under different supply and price scenarios. Flexibility index constraints to reduce the operational risks by accounting for process-feasibility-related parameters (demand and conversion) that are difficult to obtain empirical distributions. This framework is able to handle different uncertain parameters without exaggerating the uncertainties in biorefinery's profitability and emissions by introducing assumptions on parameters' distribution.

Different data-driven surrogate models are developed for the process flexibility index and embedded into the two-stage stochastic programming biorefinery design problem as the flexibility requirement constraints. Although the SVR model has higher model prediction accuracy owning to the RBF kernel, it also brings the non-convexity into the overall process design optimization. The 3rd-order polynomial Lasso regression model with fewer nonconvex terms makes the surrogate-embedded optimization solvable but has lower prediction accuracy. In both cases, the nonlinearity and non-convexity of the surrogate model result in a computationally expensive MINLP optimization problem that requires the use of a global solver.

On the other hand, the feed-forward neural network with the ReLU activation function can fit the flexibility index well with only one hidden layer. This piece-wise linear activation function was reformulated with binary variables keeping the optimization problem as MILP. With this data-driven surrogate model choice, the biorefinery design problem with flexibility index requirement could be solved efficiently with relatively high accuracy.

Acknowledgments

The authors gratefully acknowledge financial support from the National Science Foundation (NSF GCR CMMI 1934887) and the U.S. Department of Energy (DE-EE000788-7.6).

References

Bhosekar, A., et al., (2021) Multiobjective Modular Biorefinery Configuration under Uncertainty, *Industrial & Engineering Chemistry Research*, 60, 12956-12969

Bhosekar, A.; Ierapetritou, M., (2020) Modular Design Optimization using Machine Learning-based Flexibility Analysis, *J. Process Control*, 90, 18-34

Floudas, C. A.; Grossmann, I. E., (1987) Synthesis of flexible heat exchanger networks with uncertain flowrates and temperatures, *Comput. Chem. Eng.*, 11, 319-336

Grossmann, I. E.; Floudas, C. A., (1987) Active constraint strategy for flexibility analysis in chemical processes, *Comput. Chem. Eng.*, 11, 675-693

Katz, J., et al., (2020a) Integrating deep learning models and multiparametric programming, *Comput. Chem. Eng.*, 136, 106801

Katz, J., et al., (2020b) The Integration of Explicit MPC and ReLU based Neural Networks, *IFAC-PapersOnLine*, 53, 11350-11355

Kim, S. H.; Boukouvala, F., (2020) Surrogate-based optimization for mixed-integer nonlinear problems, *Computers & Chemical Engineering*, 140, 106847

Luo, Y.; Ierapetritou, M., (2020) Comparison between different hybrid life cycle assessment methodologies: a review and case study of biomass-based p-xylene production, *Ind. Eng. Chem. Res.*, 59, 22313-22329

Martín, M.; Martínez, A., (2015) Addressing Uncertainty in Formulated Products and Process Design, *Ind. Eng. Chem. Res.*, 54, 5990-6001

Metta, N., et al., (2021) A novel adaptive sampling based methodology for feasible region identification of compute intensive models using artificial neural network, *AIChE J.*, 67, e17095

Misener, R.; Floudas, C. A., (2010) Piecewise-Linear Approximations of Multidimensional Functions, *J. Optim. Theory Appl.*, 145, 120-147

Nandi, S., et al., (2004) Hybrid process modeling and optimization strategies integrating neural networks/support vector regression and genetic algorithms: study of benzene isopropylation on Hbeta catalyst, *Chem. Eng. J.*, 97, 115-129

Ochoa, M. P., et al., (2021) Novel flexibility index formulations for the selection of the operating range within a design space, *Comput. Chem. Eng.*, 149, 107284

Pulsipher, J. L.; Zavala, V. M., (2018) A mixed-integer conic programming formulation for computing the flexibility index under multivariate gaussian uncertainty, *Computers & Chemical Engineering*, 119, 302-308

Rogers, A.; Ierapetritou, M., (2015) Feasibility and flexibility analysis of black-box processes Part 1: Surrogate-based feasibility analysis, *Chem. Eng. Sci.*, 137, 986-1004

Sahinidis, N. V., (2004) Optimization under uncertainty: state-of-the-art and opportunities, *Comput. Chem. Eng.*, 28, 971-983

Schweidtmann, A. M.; Mitsos, A., (2019) Deterministic Global Optimization with Artificial Neural Networks Embedded, *Journal of Optimization Theory and Applications*, 180, 925-948

Swaney, R. E.; Grossmann, I. E., (1985) An index for operational flexibility in chemical process design. Part I: Formulation and theory, *AIChE Journal*, 31, 621-630

Terrell, E., (2022) Estimation of Hansen solubility parameters with regularized regression for biomass conversion products: An application of adaptable group contribution, *Chem. Eng. Sci.*, 248, 117184

Tribe, M. A.; Alpine, R. L. W., (1986) Scale economies and the "0.6 rule", *Eng. Costs Prod. Econ.*, 10, 271-278

Wang, H., et al., (2016) Flexibility analysis of process supply chain networks, *Comput. Chem. Eng.*, 84, 409-421