I&EC research
Industrial & Engineering Chemistry Research

pubs.acs.org/IECR

Article

Multifeedstock and Multiproduct Process Design Using Neural Network Surrogate Flexibility Constraints

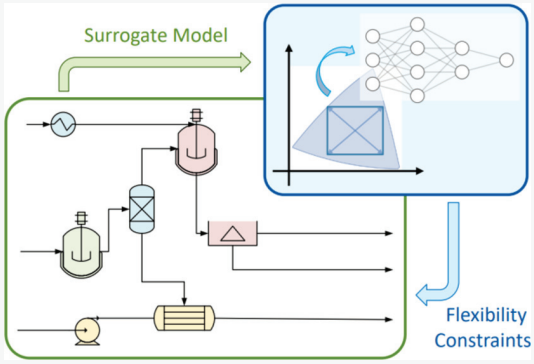
Yuqing Luo and Marianthi Ierapetritou*

Cite This: *Ind. Eng. Chem. Res.* 2023, 62, 2067−2079

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Biorefineries are designed to utilize a combination of various technologies to transform biomass derived raw materials into different value-added products. This strategy has been highlighted in the literature for reducing waste, increasing profitability, and improving the process resilience to uncertain biomass feedstocks. In this work, a two-stage stochastic programming (TSSP) model is developed to maximize profit and minimize emissions under different sources of uncertainties. Data-driven surrogate models are built for biorefinery's flexibility index (FI) to quantify and improve its operational flexibility. The neural network with rectified linear unit (ReLU) activation function is established as the appropriate surrogate model because it closely approximates the flexibility index while retaining the mixed-integer linear characteristics of the overall design formulation. Moreover, the stochastic programming demonstrates the magnitude of environmental impact uncertainty quantitatively in each scenario using empirical price/demand/supply uncertainty information, which cannot be addressed by the traditional Pedigree-based life cycle assessment (LCA) uncertainty analysis.



1. INTRODUCTION

Sustainable engineering solutions for environmental systems, flexible manufacturing, and circular economy are among the challenges and opportunities facing chemical engineers.[1] A promising solution is to use biomass and replace fossil fuels for chemical production to reduce greenhouse gas emissions due to carbon sequestration during plant growth. Several efficient and selective processes have been designed to produce *p*-xylene from cellulose,[2] furfural,[3] other value-added products[4] from hemicellulose, as well as phenols from lignin.[5] The biorefinery concept is developed to enable the efficient use of all biomass components to increase process economic viability. The biorefinery design requires rationally choosing different feedstocks, technologies, and products on the strategic and operational levels, which is typically solved by a superstructure optimization framework.[6,7] Bartling et al. designed the integrated poplar biorefinery that combined reductive catalytic fraction, enzymatic hydrolysis, and fermentation.[8] The superstructure formulation was also utilized for detailed technology and module selection in the biorefinery that produces advanced fuel from ethanol.[9,10] A greenhouse gas emission comparison of different yellow poplar forest residuals as the feedstock for integrated biorefinery operation was demonstrated based on the process simulation.[11] Dickson et al. applied the macroalgae biorefinery superstructure model to choose promising pathways, operating conditions, and wastewater treatment network.[12]

Life cycle assessment (LCA) has been widely applied to evaluate the environmental impacts of chemical production and support sustainable process development. Nevertheless, the parameters and data used for LCA are often susceptible to uncertainties, which severely affect decision-making.[13] One common LCA uncertainty evaluation method is performing Monte Carlo simulation to propagate the parameter distributions, often from the Pedigree method.[14] Heijungs raised concerns about the incompatibility of Pedigree-based distributions with large-scale Monte Carlo simulation.[15] Even though the Pedigree method provides a good data quality indicator, the reliability of probability distributions is compromised due to the subjective rating of data quality.[16] The importance of using actual measurement or providing an empirical basis for more accurate LCA uncertainty evaluation has been emphasized in the literature.[16,17] However, such empirical LCA uncertainty data are scarce and time-consuming to gather.[14] Recently, technology choices and thus the material flow rates under parameter uncertainties have been modeled in consequential LCA with optimization tools.[18] Nevertheless,

Special Issue: In Honor of Babatunde A. Ogunnaike

Received: August 19, 2022
Revised: December 14, 2022
Accepted: December 28, 2022
Published: January 17, 2023

this model uses the arithmetic mean of each linear programming scenario solution for LCA calculation, which is less effective than the two-stage stochastic programming with recourse actions in capturing the actual plant and supply chain behavior.[13]

In addition to LCA uncertainties, there is also no surprise that most of the data used for the early stage biorefinery design come with significant uncertainties. Although pilot plant can help reducing the process performance uncertainties during scale-up, it is important to screen promising technologies in the early stage to save time and resources. In this case, benchtop experimental yields and separation efficiencies are utilized to larger scale plants, which is subject to intrinsic variations, inaccurate measurement, and unexpected scale-up engineering problems.[19,20] The feedstock supply, composition, and chemical prices are volatile and could affect plant profitability.[21] Ignoring these uncertainties may render the designs and operations suboptimal or even infeasible.[22,23] Stochastic programming is commonly used to incorporate uncertain parameter distributions as scenarios.[24] The deterministic equivalent problem is then solved by optimizing the expected values of the objective functions within the scenario set.[25] For example, Bhosekar et al. applied the two-stage stochastic programming (TSSP) formulation for a biorefinery superstructure optimization to illustrate the effects of uncertainties on the trade-off between process profit and emissions.[25] Moreover, flexibility index analysis has been established as a measure of the maximum parameter ranges that a process can tolerate for feasible operation.[26] The classic definition of flexibility index is based on the maximum hyperrectangle inscribed in the feasible region, and the vertex enumeration method has been applied to convex problems.[27] However, this method becomes less effective in high-dimensional and general nonconvex problems.[27] Therefore, the active-set approach is developed to introduce binary variables indicating whether the constraints are active so that the problem could be solved more efficiently.[28] The concept of flexibility index has been applied to select the operating ranges of processes[20] and to analyze the cost-flexibility trade-off of the supply chain.[29] Zhang and Grossmann further demonstrated the process design under uncertainty with flexibility constraints using the idea of affinely adjustable robust optimization and applied the new problem formulation for a petrochemical complex case study.[26]

Surrogate-embedded optimization has been applied to many process design and optimization problems. For instance, Bhosekar and Ierapetritou utilized a support vector machine (SVM) classifier to replace the feasibility constraints in the vertex formulation of the process design problem.[30] An artificial neural network (ANN) model with hyperbolic tangent activation function has also been embedded for compressors, fermentation, and chemical process operation optimization.[31] The ANN model with sigmoid activation function was used as the surrogate for different unit operation models based on Aspen Plus simulation.[32] Adaptive sampling strategy has been implemented using both ANN and Kriging surrogate models to optimize computationally expensive processes.[33,34] Kim and Boukouvala extended Gaussian Process and ANN to mixed-integer surrogate models by using one-hot encoding for optimization.[35] Rectified linear unit (ReLU) is another widely used ANN activation function that demonstrates the ability to capture the nonlinearity of the model with low model complexity.[36] Grimstad and Andersson demonstrated the MILP formulation of optimization problems using the ReLU deep neural network to surrogate the nonlinear well performance and pressure drop functions.[37]

A successful integrated process design is expected to strike a balance between high profit, low emission, and sufficient flexibility. A TSSP model of the biorefinery design and operation is first developed in this work to consider price and biomass supply fluctuation. The ϵ-constraint method is utilized to convert flexibility index and global warming potential (GWP) objectives into the model constraints.[25] We then incorporate the process flexibility requirement into the biorefinery design through surrogate-embedded TSSP and illustrated the LCA uncertainty through the performance in each sampled scenario. Since multiple plant configurations may have the same flexibility index, direct enumeration of all alternatives during the design stage is impossible. Three different data-driven surrogate models are built for the flexibility index constraints of the biorefinery, namely support vector regression (SVR) with radial basis function (RBF) features, third-order Lasso regression, and feed-forward neural network with ReLU activation function. The ANN model with ReLU activation function is able to achieve excellent approximation while retaining the mixed-linear characteristic of the model, which is well-suited for the biorefinery multiobjective optimization problem.

2. PROBLEM FORMULATION

2.1. Multiobjective TSSP of Multifeedstock/Multiproduct Processes. *2.1.1. Parameter Uncertainties.* The process flowsheets of each candidate conversion technology could be developed in Aspen Plus based on experimental data or literature.[38] The input and output flow rates, operating cost, and capital investment of each block are extracted and scaled to formulate the optimization objectives and constraints. The optimization framework could also include other promising biomass conversion routes and unit operations that are not modeled in the Aspen Plus. Rigorous kinetic and differential equation models or measurements on actual equipment could provide the raw material flow rate, utility usage, and product generation in those cases.[21] Both profitability and environmental benefits are essential considerations for the design of integrated processes with multiple feedstock and product choices, which requires a biobjective optimization model. More specifically, the decisions of the TSSP are made in two stages. The first-stage design decisions, such as the technology choices and capacities, are made before uncertainty realization. After the actual feedstock supply and prices are observed, the second-stage operational-level decisions are made for the actual production activity and stream flow rates, ensuring process feasibility and improving performance. On the one hand, feedstock supply (e.g., biomass availability) and prices are the uncertainty parameters considered in this TSSP problem since empirical data, historical trends, or predictive models could provide relatively reliable probability distribution information for the modeling.

On the other hand, future product demands and the novel technology conversion coefficients are challenging to predict since the market is still emerging, and the technologies are at the early stage of their development. Although the TSSP framework can include these demand and conversion uncertainties in the process design using arbitrary assumptions on the parameter's probability distributions (e.g., uniform distribution), it might lead to overly exaggerated LCA

https://doi.org/10.1021/acs.iecr.2c02968
Ind. Eng. Chem. Res. 2023, 62, 2067−2079

uncertainty. To avoid overestimating the magnitude of LCA uncertainty, the nominal values of demand and conversion coefficients are incorporated in the stochastic programming, and their uncertainties are considered in the flexibility index analysis.

2.1.2. Objective Functions. The superstructure optimization problem aims to maximize the expected profit and minimize the expected GWP while having a certain level of flexibility. The expected profit includes the revenue of selling products/byproducts minus both the first-stage annualized capital investments (CAPEX) and second-stage feedstock and operating costs (OPEX). The capital investment and fixed operating costs (such as administration and maintenance) demonstrate the famous "economy-of-scale" and follow the power-law relationship.[39]

$$\text{CAPEX} = \sum_{m\in M} C_0(m) \times \left(\frac{Q(m)}{Q_0(m)}\right)^{\alpha(m)} \tag{1}$$

$$\text{OPEX}(\omega) = \sum_{m\in M}\left[O_0(m) \times \left(\frac{Q(m)}{Q_0(m)}\right)^{\beta(m)} + R_0(m) \times \left(\frac{x(m,\omega)}{Q_0(m)}\right)\right] \forall\, \omega \in \Omega \tag{2}$$

where $C_0(m)$ is the known annualized capital cost basis of the process at a plant scale of $Q_0(m)$, $Q(m)$ is the actual designed capacity of the process, 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 $m \in M$. The actual production activity in the process unit m in scenario $\omega \in \Omega$ is denoted as $x(m,\omega)$. The economic objective, total annual profit, is calculated in eq 3.

$$\text{Profit}(\omega) = -\text{CAPEX} - \text{OPEX}(\omega) + \sum_{i\in I} \mu(i,\omega) \times \rho(i,\omega)\ \forall\, \omega \in \Omega \tag{3}$$

where $\mu(i,\omega)$ is the purchase or selling price and $\rho(i,\omega)$ is the flow rate of the feedstock, intermediate, or product $i \in I$ in scenario $\omega \in \Omega$. The profit objective is then piece-wise linearized using the function interpolation of grid points.[40]

The environmental impacts are evaluated by the LCA with a "cradle-to-gate" system boundary. Similar to the production cost, the total GWP in eq 4 includes the emissions associated with production activity, upstream raw material and utility extraction, credit from selling byproducts, and plant construction. This objective will focus on minimizing the overall greenhouse gas emission among all production activities. The GWP measures the contributions from all greenhouse gases, including (CO_2, CH_4, N_2O, and others).[41] The global warming effects of other greenhouse gases are normalized by that of CO_2. Next, the byproducts generated from the plant operation are treated by the "avoided burden" method, which deducts the emission of producing the byproducts (e.g., jet fuel,[42] surfactants, and lubricants)[43] from the standalone facility.[44] For instance, if biomass-based electricity or "green" hydrogen are generated as byproducts in the biorefinery, then customers could utilize them instead of purchasing from other suppliers that consume coal or natural gas to manufacture electricity or hydrogen. In this way, the total emission is reduced because the existing power plant will use less coal and has lower greenhouse gas emissions as compared to the case without the biorefinery. Thus, part of these suppliers' production activities and raw material usage is reduced, and environmental credit should be given to the biorefinery for avoiding high-emission activities.

$$\text{Total GWP}(\omega) = \sum_{m\in M} \pi_p(m) \times x(m,\omega) + \sum_{i\in I} \pi_u(i) \times \rho(i,\omega) + \sum_{s\in S} \pi_e(s) \times \gamma(s,\omega) + \sum_{m\in M} \pi_c(m) \times Q(m) \quad \forall\, \omega \in \Omega \tag{4}$$

where $\pi_p(m)$, $\pi_u(i)$, $\pi_e(s)$, and $\pi_c(m)$ are the greenhouse emissions during the production stage (e.g., flue gas from lignin combustion), upstream raw material $i \in I$ extraction, electricity generation from energy source $s \in S$ and plant construction, and $\gamma(s,\omega)$ is the electricity usage from energy source $s \in S$ in scenario $\omega \in \Omega$. The water consumption related to the process operation and biomass growth is considered for LCA, but it does not contribute significantly to GWP as compared to other chemicals and utilities (e.g., electricity and steam).

It is more common to normalize the total emissions by the amount produced in LCA so that the same product from different processes can be compared on the same basis. If the LCA functional unit is chosen as 1 kg of chemical $i \in I$ produced by the biorefinery, another bilinear constraint with both continuous variables is introduced in eq 5.

$$\text{Total GWP}(\omega) = \text{GWP}(\omega) \times \rho(i,\omega) \quad \forall\, \omega \in \Omega,\ i \in \text{Target Product} \tag{5}$$

2.1.3. Mass Balance. Based on the process simulation, conversion coefficients of each process unit could be collected to build the relationship between inlet and outlet flow rates for feedstocks, intermediates, and products $i \in I$.

$$\rho(i,\omega) = \sum_{m\in M} v(i,m) \times x(m,\omega) \quad \forall\, i \in I,\ \omega \in \Omega \tag{6}$$

when $v(i,m)$ is the conversion coefficient that represents the consumption/generation of products $i \in I$ using technology $m \in M$. 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. This formulation is similar to the matrix-based LCA calculation, which can readily expand and include more processes/technologies as they become available or of interest.

It is a common practice in integrated chemical plants, such as biorefineries, to incinerate residuals like lignin to generate heat and electricity for other units. Thus, different external electricity sources and the lignin-combustion unit are considered in the mass balance as the operational limits on electricity usage (eq 7).

$$-U_e \times b(\omega) \le \rho(\text{"Electricity"},\omega) \le U_e \times b(\omega) \quad \forall\, \omega \in \Omega \tag{7}$$

where U_e is the constant for the big-M constraints and $b(\omega)$ is the binary variable that decides whether to sell lignin-

https://doi.org/10.1021/acs.iecr.2c02968
Ind. Eng. Chem. Res. 2023, 62, 2067−2079

incinerated electricity or purchase electricity from different sources $s \in S$, including solar, wind, hydro, other biomass, and natural gas.

$$\sum_{s\in S} \gamma(s, \omega) + \rho(\text{“Electricity”}, \omega) - k(\omega) = 0 \quad \forall\, \omega \in \Omega \tag{8}$$

where $k(\omega)$ is the electricity sold to the market when excess electricity is generated from combusting lignin. It is forced to 0 when $b(\omega_s)$ is 0 by the logic constraints (eqs 9 and 10).

$$k(\omega) \le U_e \times b(\omega) \quad \forall\, \omega \in \Omega \tag{9}$$

$$\gamma(s, \omega) \le \gamma^{+}(s, \omega) \times (1 - b(\omega)) \quad \forall\, s \in S, \omega \in \Omega \tag{10}$$

where $\gamma^{+}(s, \omega)$ is the electricity availability from each source $s \in S$. Other energy sources and the future electricity mix could be incorporated by adding new elements and updating the $\gamma^{+}(s, \omega)$ values.

2.1.4. First-Stage Design Constraints. The binary variable $y(m)$ 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 $Q^{+}(m)$.

$$y(m) \times Q^{-}(m) \le Q(m) \le y(m) \times Q^{+}(m) \quad \forall\, m \in M \tag{11}$$

Additionally, some feedstocks (and technologies) may share the same facility, which is captured by equating their capacities in eq 12.

$$Q(m) = Q(m') \quad \forall\, m, m' \in M \tag{12}$$

and m, m' share the same facility.

2.1.5. Second-Stage Operating Constraints. There are also bounds on operational decisions. For example, the feedstocks flow rates could not exceed the maximum supply. The maximum production rate should be less than the market demands in each case. These requirements impose the lower and upper bounds ($\rho^{-}(i, \omega)$ and $\rho^{+}(i, \omega)$) on the material flow rates.

$$\rho^{-}(i, \omega) \le \rho(i, \omega) \le \rho^{+}(i, \omega) \quad \forall\, i \in I, \omega \in \Omega \tag{13}$$

Moreover, the actual production activity in each production unit should not exceed its designed capacity. This constraint (eq 14) connects the first- and second-stage decisions of the TSSP.

$$x(m, \omega) \le Q(m) \quad \forall\, m \in M, \omega \in \Omega \tag{14}$$

2.2. Flexibility Index Optimization Using the Active-Set Method. For a given plant configuration, a max-min-max formulation is adopted to calculate the flexibility index. Here, $f_j(d, z, \theta)$ is obtained after substituting equality constraints and eliminating the state variables in the inequalities. In this model, d is the design variable (e.g., choice of technologies $y(m)$ and capacities $Q(m)$), θ is the uncertain parameter that affects the process feasibility (such as supply or demands, $\rho^{-}(i, \theta)$ or $\rho^{+}(i, \theta)$, and conversion coefficients, $v(i, m, \theta)$), and z is the recourse/control actions (e.g., production activity $x(m, \theta)$ and material flow rates $\rho(i, \theta)$).

$$F = \max \delta$$
$$\text{s.t.} \max_{\theta\in T(\delta)} \psi(d, \theta) \le 0$$
$$\psi(d, \theta) = \min_{z} \max_{j} f_j(d, z, \theta)$$
$$T(\delta) = \{\theta: \theta^N - \delta\cdot\Delta\theta^{-} \le \theta \le \theta^N + \delta\cdot\Delta\theta^{+}\} \tag{15}$$

$\psi(d, \theta)$ is the function testing the system's feasibility when the parameter takes the value θ and the design is d. $T(\delta)$ is the hyperrectangular that allows uncertain parameters to vary without correlation and θ^N is the nominal value of process parameter. The interaction between different uncertain parameters could be captured by ellipsoidal or diamond shape uncertainty sets.[45] The active-set method is proposed to reformulate and efficiently solve the aforementioned bilevel optimization problem (eq 16), especially for nonconvex cases.[28,46]

$$F(d) = \min_{\theta, z, \delta, s_j, \lambda_j, y_j} \delta$$
$$\text{s.t. } s_j + f_j(d, z, \theta) = 0 \quad \forall\, j \in J$$
$$\sum_{j\in J} \lambda_j = 1, \sum_{j\in J} \lambda_j \cdot \frac{\partial f_i}{\partial z} = 0,$$
$$\lambda_j - y_j \le 0, s_j - U(1 - y_j) \le 0,$$
$$\sum_{j\in J} y_j = n_z + 1$$
$$T(\theta) = \{\theta: \theta^N - \delta\cdot\Delta\theta^{-} \le \theta \le \theta^N + \delta\cdot\Delta\theta^{+}\}$$
$$\delta \ge 0; y_j = 0, 1; s_j, \lambda_j \ge 0 \quad \forall\, j \in J \tag{16}$$

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

2.3. Multiobjective Integrated Process Design with Surrogate Flexibility Constraint. Different surrogate models for the flexibility index could be built based on sampled data, which will serve as a simplified constraint of the plant capacity design (eq 17). The surrogate model reduces the optimization problem complexity and enables the biorefinery design with flexibility requirements to be solved as a single-level optimization problem.

$$\hat{F}(d) \approx F(d) = \min_{\theta, z, \delta, s_j, \lambda_j, y_j} \delta \tag{17}$$

The ε-constraint method is applied to study the trade-off between the maximum profit and the minimum GWP for product $i \in I$ (or total annual GWP of the process) while maintaining the required flexibility level, δ_L.

$$\max_{y(m), Q(m), \rho(i,\omega), x(m,\omega)} \mathbb{E}_{\omega\in\Omega}[\text{Profit}(\omega)]$$
$$\text{s.t. Constraints}(1) - (14)$$
$$\mathbb{E}_{\omega\in\Omega}[\text{GWP}(\omega)] = \frac{\Sigma_{\omega\in\Omega}\text{Total GWP}(\omega)}{\Sigma_{\omega\in\Omega}\rho(i, \omega)} \le \varepsilon$$
$$\hat{F}(d) \ge \delta_L \tag{18a}$$

https://doi.org/10.1021/acs.iecr.2c02968
Ind. Eng. Chem. Res. 2023, 62, 2067−2079

After determining the highest and lowest GWP on the Pareto curve by solving the profit maximization and GWP minimization problem separately, the bilinear epsilon constraint (eq 18a) involving expected value among different scenarios could be converted to a linear one (eq 18b).

$$\Sigma_{\omega\in\Omega}\text{Total GWP}(\omega) \le \varepsilon \times \Sigma_{\omega\in\Omega}\rho(i,\omega) \tag{18b}$$

The average total GWP could be another linear objective function, which represents the lowest greenhouse gas emission of all production operations in a biorefinery.

$$\mathbb{E}_{\omega\in\Omega}\text{Total GWP}(\omega) \le \varepsilon \tag{19}$$

Overall, the multifeedstock/multiproduct process design with flexibility index requirement is shown in Figure 1. After the initial Pareto curve construction in the TSSP without any flexibility consideration, the capacities of each recommended design on the curve provide the reference ranges for capacity sampling. The Latin hypercube sampling (LHS) is then performed to create the data points for flexibility index evaluation (active-set problem). The data sets are then split into training and testing sets to build the data-driven surrogate model that approximates the flexibility index response as a function of each technology's capacity. The surrogate model is finally included in the original TSSP model as the flexibility constraint to ensure the process has enough resilience.

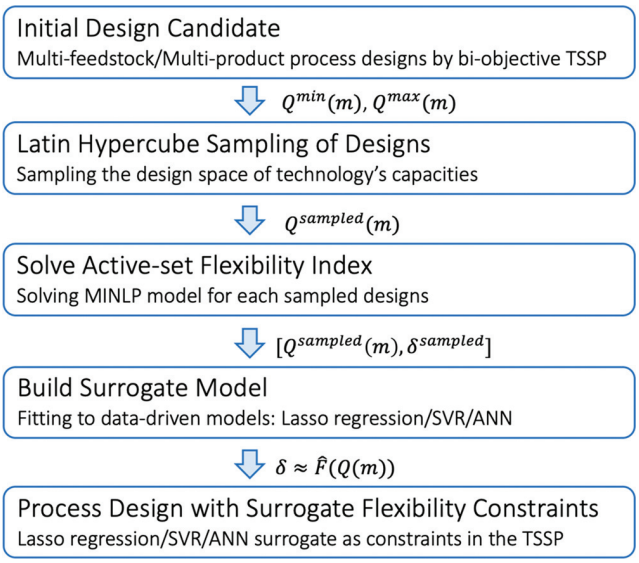


Figure 1. Proposed strategy for the surrogate-embedded process design with flexibility requirement.

This surrogate-embedded TSSP formulation could be applied to a variety of process design problems with feedstock/product/technology choices, including oil refinery operation,[47] chemical conversion of waste plastic,[48] biorefineries,[6,7] and other process superstructures.[49]

In particular, the integrated biorefinery design involves emerging technologies in its superstructure with uncertainties from various sources. The design of profitable, environmentally friendly, and flexible biorefineries is also of practical importance for the implementation of viable technologies toward circular economy.[50] Consequently, a biorefinery case study is selected in section 4 to demonstrate the effectiveness of the proposed method in handling supply, demand, conversion, and price uncertainties.

3. SURROGATE MODEL CONSTRUCTION FOR FLEXIBILITY INDEX

The following three types of data-driven surrogate models are deployed to replace the flexibility index calculation in the superstructure optimization problem. The model fitting performance is next compared based on the R^2 of both data sets. The Scikit-learn[51] and TensorFlow[52] Python machine learning packages are used for the model training, and the fitted parameters are then implemented in GAMS for optimization. Cross-validation by grid search is utilized to select the best performing hyperparameters for the surrogate model.

3.1. Support Vector Regression (SVR) with Radial Basis Function (RBF) Kernel. Similar to the SVM used in classification problems, SVR models also have a simple functional form and are easy to train. The RBF kernel is based on the Euclidean distance between the point of interest and the support vectors, as illustrated in eq 20.

$$\hat{F}^{\text{RBF}}(x) = \sum_{s\in SV} a_s\, e^{-\gamma\|x-v_s\|^2} + b \tag{20}$$

where v_s is the sth support vector with a weight of a_s, γ is the kernel parameter, and b is a constant term. The hyperparameters of the kernelized soft-margin SVR, including the tube size of the insensitive region (ε) and the regularization parameter of error outside the insensitive region (C), control the trade-off between model fitting accuracy and the risk of overfitting. The SVR model typically demonstrates good performance for training on small-sized high-dimensional data.[53] Nevertheless, the general RBF kernel is nonconvex, which adds computational complexity to the surrogate-embedded optimization problem.

3.2. Lasso Regression Up to Third-Order Terms. As one of the simplest surrogate models, linear regression with polynomial features could be readily encoded in the optimization software (eq 21a). However, the number of terms grows rapidly as the variable dimension increases, which is prone to overfitting.

$$\hat{F}^{\text{3rd}}(x) = \sum_i a_i x_i + \sum_i b_i x_i^2 + \sum_i c_i x_i^3 + \sum_i \sum_{j>i} a_{i,j} x_i x_j + \sum_i \sum_{j\neq i} e_{i,j} x_i x_j^2 + \sum_i \sum_{j>i} \sum_{k>j} a_{i,j,k} x_i x_j x_k + d \tag{21a}$$

However, the bilinear and trilinear terms in $\hat{F}^{\text{3rd}}(x)$ are also nonconvex and hinder the convergence of the final biorefinery design MINLP. Thus, a simplified polynomial surrogate (eq 21b) without some nonconvex terms ($x_i x_j^2$ and $x_i x_j x_k$) is considered to improve the computational efficiency at the expense of model accuracy.

$$\hat{F}^{3'}(x) = \sum_i a_i x_i + \sum_i b_i x_i^2 + \sum_i c_i x_i^3 + \sum_i \sum_{j>i} a_{i,j} x_i x_j + d \tag{21b}$$

Moreover, the Lasso (L1) regularization is adopted to avoid overfitting by penalizing the sum of absolute values of coefficients. This approach typically sets the coefficients of irrelevant terms to zero, which further reduces the model complexity.

3.3. Neural Network with ReLU Activation Function. Neural network models demonstrate excellent fitting perform-

https://doi.org/10.1021/acs.iecr.2c02968
Ind. Eng. Chem. Res. 2023, 62, 2067−2079

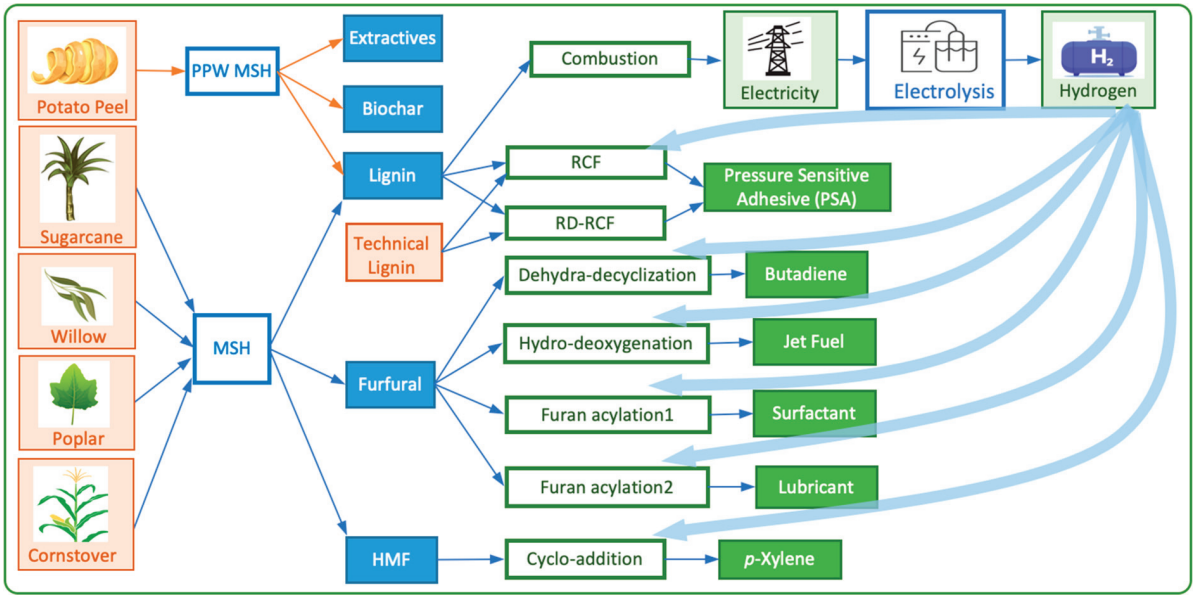


Figure 2. Biorefinery superstructure for the case study.

ance for various data types and applications, especially when data are abundant. For instance, the feed-forward ANN with only one hidden layer has been proven to approximate any smooth function with high accuracy on a compact domain given sufficient neurons.[54] In the feed-forward ANN model, the output vector of layer k is denoted as $z^{(k)}$. The first layer is the input layer with $z^{(1)}$ as its input vector, and the last layer is the output layer with a scalar $z^{(N)}$ as its output. For other layers $k \geq 2$, $z^{(k)}$ is calculated by applying the activation function to the linear combination (with weights $W^{(k-1)}$ and constant terms $b^{(k-1)}$) of the outputs ($z^{(k-1)}$) from the previous layer (eq 22).

$$z^{(k)} = h_k(W^{(k-1)}z^{(k-1)} + b^{(k-1)}) \tag{22}$$

The ReLU activation function, $h_k(x) = \max(0, x)$, is selected for the surrogate model. This piece-wise linear function could be converted into mixed-integer linear constraints with the big-M formulation (eq 23a−23e). Let $\eta \in \{0,1\}$ be a binary variable, and U be the big-M constant:

$$-U\cdot(1-\eta) \leq x \leq U\cdot\eta \tag{23a}$$

$$h_k(x) \leq U\cdot\eta \tag{23b}$$

$$0 \leq h_k(x) \tag{23c}$$

$$x \leq h_k(x) \tag{23d}$$

$$h_k(x) \leq x + U\cdot(1-\eta) \tag{23e}$$

Finally, the prediction from the output layer is just the linear combination without any activation function (eq 24):

$$z^{(N)} = W^{(N)}z^{(N)} + b^{(N)} \tag{24}$$

4. CASE STUDY: INTEGRATED BIOREFINERY DESIGN WITH FLEXIBILITY CONSTRAINTS

Biorefineries are designed to utilize various alternative technologies and convert different feedstocks into valuable products. The case study is modified from the aforementioned process design, and simulation work as well as the biorefinery superstructure presented in Bhosekar et al.[25] As illustrated by the superstructure in Figure 2, sugar cane bagasse, corn stover, as well as the poplar and willow wood are fed into the biorefinery as the biomass feedstock. The willow is chosen as a biomass feedstock instead of red oak because its yield uncertainty data are available in the literature, which better reflects the actual variability of biomass supply and the LCA uncertainty.[55] The biomass feedstocks are treated by the molten salt hydrate hydrolysis (MSH) technologies to produce important platform chemicals: furfural, hydroxymethylfurfural (HMF), and lignin.[55] Biochar and antioxidant extractive are also manufactured in the MSH process under different reaction and separation conditions using the potato peel feedstock.[56] Two reductive catalytic fractionation (RCF) processes, namely the traditional RCF with methanol as the solvent and the intensified reactive distillation-RCF (RD-RCF) with glycerin as the solvent, could convert lignin into the targeted main product, pressure-sensitive adhesive (PSA).[57] Technical lignin from other upstream treatments such as paper pulping could be utilized in the same RCF units. If the lignin byproduct exceeds RCF capacities, then it could also be burned to generate electricity for the plant or sent to the solid waste treatment facility at a higher cost. The cost of the power plant is calculated based on the simulation in Athaley et al.,[2] and the parameters are summarized in Table S1. Moreover, HMF and furfural could further go through hydro-deoxygenation, acylation, and cyclo-addition processes to synthesize a variety of value-added products, including butadiene, jet fuel, surfactant, lubricant, and *p*-xylene.[2,4] It is noteworthy that this superstructure does not contain all potential biomass conversion routes. Many other HMF and furfural conversion technologies could be considered in this framework when the necessary input-output relationship, cost, and emission data are compiled.[58−60]

Hydrogen gas is widely used in highly selective biomass technologies. But the hydrogen gas transportation requires pressurization or pipe construction, which could become a limiting factor in remote areas where biomass feedstocks are abundant. Thus, a new electrolysis process to produce "green hydrogen" is added as an alternative. The on-site electricity

https://doi.org/10.1021/acs.iecr.2c02968
Ind. Eng. Chem. Res. 2023, 62, 2067−2079

and hydrogen generation units are expected to provide more robustness to the plant such that the deficient electricity or hydrogen supply in extreme cases still has its recourse action. This biorefinery superstructure could easily include more processes from simulation or literature as only the input−output relationship, and the economic/environmental performance are required. It is similar to the matrix-based LCA framework, which shows the potential to be extended to more general LCA studies.[61,62]

In the early stage process design, the chemical prices and their uncertainties could be tracked through the historical trend.[63] However, it is challenging to measure and quantify the biomass feedstock supply before the plant operation. Although the regional biomass availability map at county scale resolution provides valuable information for supply chain design, the centralized large-scale biorefinery operation will rely on the supply from specific districts.[64] Consequently, the temporal correlation of the biomass supply is more relevant to the biorefinery operation. In this work, the yearly biomass yields per acre were selected to provide a reasonable estimate for the variability of raw material supply (Figure S1).[65,66] Recently, different biogeochemical models[67,68] and integrated biomass supply and logistic models[69] have been developed to more accurately predict feedstock availability. However, the stochastic programming framework is capable of incorporating more reliable uncertainty information in the future by sampling representative scenarios from the models above or historical data. The biomass composition variability was implicitly considered in the changes of HMF/furfural/lignin conversion coefficients. The biorefinery is designed to treat different biomass feedstock as a mixed stream. Thus, the model also considers the composition variability in the ratio of different feedstocks in the inlet stream. More sample collection, characterization, and data acquisition are required to quantify the magnitude of the composition variability in order to construct uncertain scenarios.[11]

The demands for the commodity chemical products traditionally from petroleum feedstocks (e.g., jet fuel, lubricant, and butadiene) are relatively stable and large due to the sheer volume of the market.[70] More valuable specialty chemicals such as extractives and pressure-sensitive adhesives have niche markets whose production may be severely affected by demand.[71,72] Since biomass-based products are still emerging and the market is developing, historical data or accurate models to describe the demand uncertainties are lacking. Although many biomass-based products are developed to substitute existing petrochemicals (same or similar molecules), there are opportunities to create new products with better performance or functions.[73] Currently, most chemicals are produced using several different routes and intermediates. It is difficult to predict which biomass-based intermediate molecule will be favored (with higher demand) in this competitive and volatile market. For the specialty chemical end-products, the customers' behavior and preference are more challenging to model, especially at the early stage. Product segmentation and market penetration will affect how much biomass-based products will take up in the existing market, which is also evolving over the years.[74] Another source of potential uncertainty is the reaction yields and conversions as the technology is still developing and the experimental measurement is performed at bench-scale experiments. The mean values and deviations (±15%) are thus assigned for the flexibility index analysis to capture these sources of uncertainty.

The summary of parameters used for the biorefinery optimization is listed in Table S1−S4 and Figure S1. All optimization models are implemented in GAMS 33.1 on a computer with Intel Xeon E-2274G CPU @ 4.00 GHz 32 GB RAM.

4.1. Data-Driven Surrogate Model for Flexibility Index Evaluation. In this section, SVR, Lasso regression, and ANN with ReLU are fitted to the same sampled data set of flexibility index as a function of different design capacities. The collected data set is shuffled and randomly split into the training and validation sets to identify underfitting and overfitting. Cross-validation is performed to determine the optimal polynomial degree for the model fitting (Figure S2). As the polynomial degree increases, the polynomial model fits the training data with higher accuracy, while the testing data has its peak at the third-order model. After that, overfitting occurs, and the model has a worsening performance at predicting unseen data. Thus, the third-order polynomial model is chosen for the regression. Feature selection is also a vital step in improving the data-driven model accuracy and reducing the model complexity. Especially in the polynomial regression and SVR models, the number of nonconvex terms increases rapidly as the input dimension grows, rendering the embedded optimization problem unsolvable in a reasonable time. The univariate test with F-statistic in the Scikit-learn Python package is implemented to remove three features that have minimal contributions to the model performance: the capacity of potato peel's MSH process, surfactant production, and waste combustion. As shown in Figures S3, S4 and S6, S7, the R^2 values in the training data set slightly decrease as these three features are removed, but the testing data set's R^2 values increases, meaning both Lasso regression and SVR model achieve better out-of-sample forecasting with fewer model parameters. More details on different surrogate models' fitting performance could be found in the parity plots (Figures S3−S8).

As compared in Table 1, the SVR model can capture the nonlinearity of the flexibility index (FI) as a function of the design capacities well (R^2 values higher than 0.93) due to the

Table 1. Summary of Surrogate Model Fitting and Optimization Performance

surrogate model	SVR (RBF kernel)	third-order Lasso (no $x_i x_j x_k$ and $x_i x_j^2$ terms)	ANN (ReLU activation function)
training R^2	93.7%	78.7%	99.7%
testing R^2	93.4%	75.2%	99.7%
optimization model type	MINLP	MINLP	MILP
no. of constraints	12341	12341	12443
no. of continuous variables	6171	6171	6203
no. of discrete variables	196	196	210
solver	BARON	BARON	CPLEX
max profit design (δ_L = 0.5) solution time	3600 s	1835 s	6.5 s
profit LB (M$/year)	47.3	45.8	47.9
profit UB (M$/year)	50.2	46.3	47.9
optimality gap (after 3600 s)	5.7%	1.0%	0%
FI prediction error (profit LB case)	4.0%	15.6%	−3.3%

https://doi.org/10.1021/acs.iecr.2c02968
Ind. Eng. Chem. Res. 2023, 62, 2067−2079

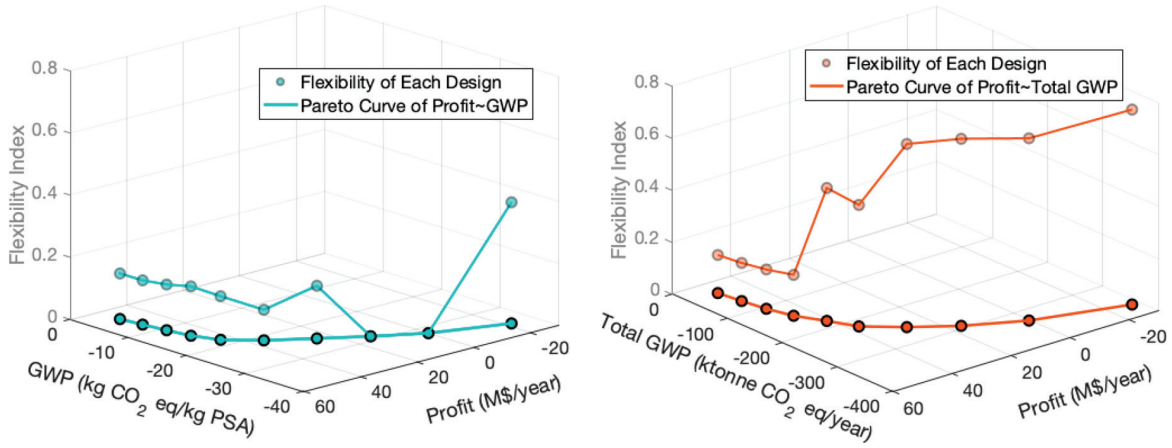


Figure 3. Pareto curve of the profit-PSA GWP (left) and profit-total GWP trade-offs (right) and the flexibility index evaluated for each recommended design.

excellent fitting ability of the RBF kernel.[75] However, a total of 293 support vectors are needed to fit the model with such accuracy. Therefore, the biorefinery design optimization with SVR surrogate flexibility constraint (FI $\geq \delta_L = 0.5$) encounters difficulties to find tighter upper bounds as this highly nonconvex function is challenging to relax.

The third-order Lasso regression is capable of fitting the general trend of the flexibility index but is less effective for the validation data set even with all polynomial terms up to the third-order (Figure S3). Since both bilinear and trilinear terms are nonconvex, the optimization problem will run into a similar convergence issue as the SVR model. Using third-order Lasso regression without $x_i x_j^2$ and $x_i x_j x_k$ terms, the surrogate's fitting performance drops, but the flexibility constraints become easier to relax and provide tighter upper bounds. The optimality gap is thus lower than the SVR flexibility surrogate, even though the SVR model predicts the highest profit design's flexibility index more accurately than the Lasso regression model. The solution quality of using third-order Lasso surrogated constraint is also the worst ($ 45.8M/year profit) as it overdesigns plant capacities, especially the MSH unit.

The feed-forward neural network surrogate with ReLU activation function demonstrates very accurate model fitting capability with R^2 of more than 99.7% with only two hidden layers (10 and 4 neurons in each layer). Such high surrogate model accuracy provides more confidence in the model when substituted as the flexibility index constraints in the overall process design optimization problem. When embedded into the superstructure optimization model, each neuron only introduces one extra integer variable in the ReLU function. Negative or near-zero values in each ReLU layer do not cause numerical errors or infeasibility, even in deep learning models.[36] Other activation functions, such as exponential linear unit (ELU), scaled exponential linear unit (SELU), and Gaussian error linear unit (GELU), could also be used.[76] Since the rest of the biorefinery design problem is linear with integer variables, the ReLU ANN model does not change the MILP model type and can be solved efficiently with the CPLEX solver, unlike the SVR- or Lasso-embedded nonconvex models that require global MINLP solver, such as BARON.

4.2. Profit-Emission Trade-Off of Biorefinery by TSSP.

The stochastic programming with environmental and economic objectives is performed using the ε-constraint method, and its Pareto curve is shown in Figure 3. After solving the maximum profit and the minimum emission optimization problems separately, the upper and lower bounds of ε are determined. Ten instances with gradually increased equidistant ε are then incorporated as constraints (eq 18b or 19) to obtain the Pareto optimum. The biorefinery configurations suggested by each case are then sent to the active-set optimization problem as a design variable, the flexibility index of which is evaluated and shown in the z-axis of Figure 3.

Without flexibility constraints, the design with the lowest total GWP emission (eq 19) has very large capacities since this configuration tries to utilize as much renewable biomass feedstock as possible, as indicated by the high flexibility index of all Pareto points. Nevertheless, the design that minimizes GWP of the PSA (total greenhouse gas emission divided by the total PSA production) has reduced capacities in each process as it tries to avoid the unnecessary usage of biomass feedstock that does not have the lowest greenhouse gas emission. On average, the designs with PSA's GWP as the second objective produce 8−8.47 kt PSA/year, and the designs with total GWP as the objective produce 8.45−15.56 kt PSA/year. The transition from high-profit to low-emission designs on both Pareto curves is achieved through selecting more expensive but less carbon-intensive feedstocks (e.g., from poplar to willow) and technologies/products (e.g., from RCF to RD-RCF).

Among different raw material supplies, the most restricting resources of the biorefinery operation are electricity and hydrogen gas, despite their relatively insignificant contribution to the cost. Thus, even if a bigger biorefinery is constructed with added capacities in some unit processes, the flexibility may not increase further as the operation at a higher flow rate does not have enough electricity or hydrogen. Moreover, the demand for most products is typically too high to make a difference in the flexibility index. Nonetheless, the supply and demand of the intermediates (HMF and furfural) are the main limiting factors in many cases. For instance, if the furfural production exceeds the market demand, treating the remaining furfural after the sale will consume electricity and hydrogen gas, which are also crucial in other processes.

Next, the flexibility constraint is added by using a 2-layer ANN surrogate with the ReLU activation function. The ANN prediction of the flexibility index based on the design capacities is very close to the actual value before and after imposing the flexibility constraint, as illustrated in Figure 4, which highlights

https://doi.org/10.1021/acs.iecr.2c02968
Ind. Eng. Chem. Res. 2023, 62, 2067−2079

the effectiveness of the ANN model in forecasting unseen flexibility data.

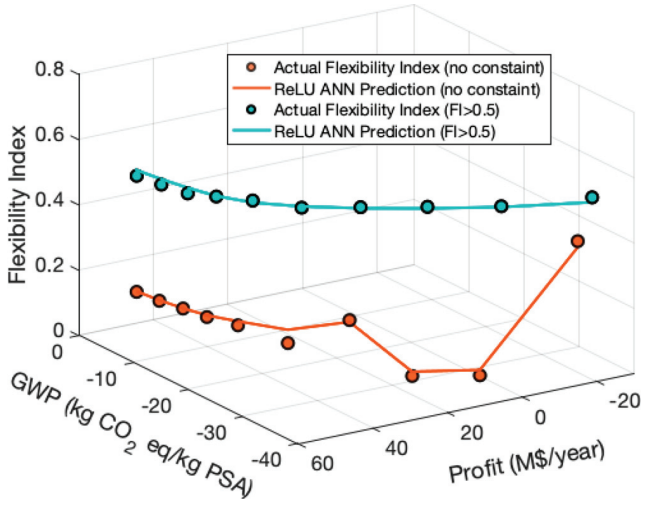


Figure 4. Pareto curve of the profit-PSA GWP trade-offs before and after implementing the flexibility constraints.

Apart from providing robust biorefinery designs that handle different sources of uncertainty, the profit and emission profile of each sampled scenario in the two-stage stochastic programming could also demonstrate the uncertainties of the biorefinery's economic and environmental performance. As illustrated in Figure 5, the variability in LCA results is the

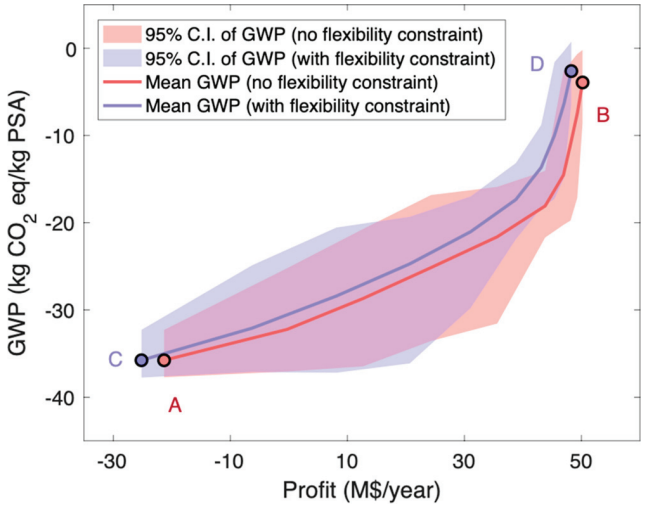


Figure 5. Biorefinery's GWP uncertainty ranges before and after the flexibility constraint (FI ≥ 0.5).

largest for designs in the middle of the profit-PSA GWP trade-off curve, as a mix of different technologies and feedstocks are used. The Pareto curve also helps to illustrate the marginal LCA results (i.e., the nonlinear response of emission when different amounts of product are generated).[77] It is challenging to quantify the LCA uncertainty and obtain such insights from the conventional linear LCA framework that is primarily based on matrix multiplication. Traditionally, the Pedigree method is utilized to rate the LCA data reliability, completeness, temporal, and geometric correlation, and technological correlation from 1 to 5. These scores are then transformed into uncertainty factors between 1 to 2 and assigned as geometric standard deviations for flow rate variability in Monte Carlo simulation.[17] Although this method could qualitatively describe the quality of data used in LCA, the Pedigree geometric standard deviations lack physical meanings and are not representative of the actual flow rate/emission variability. The optimization framework presented in this work, however, emphasizes the technology/feedstock choices and actual flow rate in each unit under various economic and operational circumstances with an empirical basis.

A detailed comparison of design capacities on the Pareto curves is illustrated in Table S5. As expected, designs C and D have higher capacities than their profit- or emission-prioritized counterparts (A and B). As the lowest GWP design (A) already has relatively high capacities, the relative changes in each unit are insignificant. The overdesigned capacities are mainly on the biomass combustion unit that generates extra electricity for the biorefinery when the electricity supply is tight and the *p*-xylene production process reduces the risk of producing too much HMF that exceeds the demand. However, design D has a relatively large increase from design B in almost all capacities, especially the MSH, RD-RCF, and *p*-xylene production units. Noticeably, the electrolysis process that generates H_2 on site is suggested for this maximum profit case as hydrogen gas is expensive. As the RD-RCF process does not consume hydrogen gas, the extra hydrogen could be utilized to convert the HMF/furfural into more value-added chemicals. It is a win-win situation where the intermediates' demand limits are avoided to maintain feasibility, and more valuable products are generated to increase the profit. Figure S9 further demonstrates that the electrolysis unit is not idle but actually generates H_2 more than 60% of the time to provide a better capability of intermediate conversion. The electrolysis process not only increases the flexibility of the biorefinery in extreme cases but also is economically favored in normal operation once it is installed.

After increasing the capacities for a higher flexibility requirement, the feedstock flow rates and production activity in each unit are still quite close to the base-case operation without flexibility constraints due to the unchanged background economic conditions (i.e., prices, supply, and demand). Consequently, the performance change mainly comes from the capital investment and emissions during the plant construction stage. This is the price paid for the increased flexibility by overdesigning the process capacities. Nevertheless, the installed electrolysis facility in design D allows the production of cheap on-site hydrogen gas and valuable furfural-based downstream products, which offsets part of the capital cost and causes slight changes in the shape of the Pareto curves. It is also possible that different technologies are chosen after the constraints are imposed due to their lower conversion/product demand deviations or the capability of treating intermediates to avoid producing more than the biorefinery could sell.

4.3. Multiobjective Optimization Using Surrogate Flexibility Index Model. The ε-constraint optimization method is applied in a nested way to solve the triobjective problem (profit−GWP−flexibility index). In the outer loop, the lower bound for flexibility index, δ_L, is gradually increased to determine the ranges of profit and GWP for the inner loop iteration. Consequently, only one MINLP needs to be solved in the outer loop when the lowest PSA GWP value is determined. Then, all other ε-constraint cases are solved efficiently as MILP by using eq 18b. The 3D Pareto curve

https://doi.org/10.1021/acs.iecr.2c02968
Ind. Eng. Chem. Res. 2023, 62, 2067−2079

(Figure 6) is obtained by solving the surrogate-embedded flexibility constraint model in 18 min. Each design's flexibility

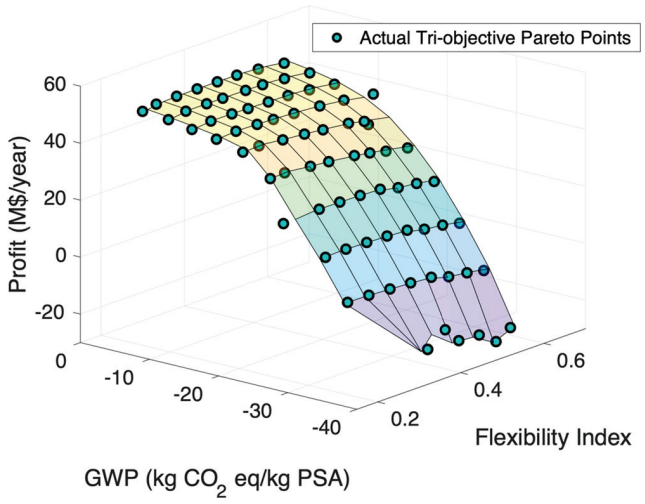


Figure 6. GWP−profit−flexibility trade-off surface constructed by surrogate-embedded optimization.

on the curve is then evaluated by the active-set method, which closely matches the ANN prediction. This performance again demonstrates the ANN surrogate model's excellent prediction ability over a relatively wide range.

The sampling and flexibility index evaluation steps are decoupled from the surrogate-embedded optimization. Hence, more efficient flexibility index solution algorithms or parallel computing will improve computational time significantly. Additionally, uncertainty sets that represents the parameter interactions in flexibility index analysis or other measures of the system flexibility could be included.[45,78] Although the ANN model fits the flexibility index with satisfactory accuracy throughout the variable range in this case study, it is possible to develop separate models for low and high flexibility index regions when an overall model does not capture the entire region.

Another implication from this work is that some high-dimensional and nonlinear functions, such as flexibility index, could be approximated effectively by a mixed-integer surrogate model (ANN with ReLU activation function) without introducing too much complexity. Misener and Floudas demonstrated the approximation of nonlinear functions based on the interpolation of grid points, but the formulation is complicated for input variables with dimensions higher than three.[40] As illustrated in eqs 23a−e, the ANN surrogate with ReLU activation function is relatively easy to encode in optimization software, which shows the potential for other applications.

5. CONCLUSIONS

Overall, a centralized biorefinery superstructure optimization is formulated to simultaneously minimize the GWP based on life cycle assessment (LCA) and maximize profit. The price and biomass supply uncertainties are utilized to generate scenarios in the two-stage stochastic programming. At the same time, the resilience of biorefinery in the face of biomass-based product demands and reaction conversion variability is quantified by the flexibility index analysis. This analysis is suited for an early stage biorefinery evaluation with mixed forms of uncertain information, including probability distributions, historical data, and ranges. Moreover, the ANN with ReLU activation function is shown as an appropriate surrogate model for flexibility index due to its accurate fitting ability and the mixed-integer linear formulation. It could then be incorporated into the biorefinery superstructure optimization to solve the design with flexibility constraints and even the profit−GWP−flexibility triobjective optimization.

This superstructure optimization is also compatible with the current matrix-based LCA framework consisting of the technological matrix and emission matrix. This model is not only able to guide the optimal biorefinery design and operation under different levels of price, supply, and demand uncertainties but also quantitatively translates them into the uncertainty of product emissions. This optimization-based LCA uncertainty quantification method provides the empirical foundation for material flow rate variability that the traditional Pedigree and Monte Carlo method lacks.[17]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.iecr.2c02968.

Economic parameters, nominal conversion coefficients, and process emissions for the biorefinery case study (Table S1−S4); relative yields of biomass feedstocks (Figure S1); Lasso regression fitting plots (Figures S2−S5); SVR fitting plots (Figure S6−S7); ANN fitting plots (Figure S8); process comparison for designs A−D (Table S5); electrolysis production activities in design D (Figure S9) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Marianthi Ierapetritou − *University of Delaware, Department of Chemical and Biomolecular Engineering, Newark, Delaware 19716, United States;* orcid.org/0000-0002-1758-9777; Email: mgi@udel.edu

Author

Yuqing Luo − *University of Delaware, Department of Chemical and Biomolecular Engineering, Newark, Delaware 19716, United States;* orcid.org/0000-0002-4174-946X

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.iecr.2c02968

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work is financially supported by the U.S. Department of Energy's RAPID Manufacturing Institute for Process Intensification (DE-EE000788-7.6), the National Science Foundation's Growing Convergence Research program (NSF GCR CMMI 1934887) and the Grant No. OIA-2119754.

■ NOMENCLATURE

Sets/Indices

$i \in I$ raw material, intermediate or products
$j \in J$ inequality constraints in the flexibility index analysis
$m \in M$ technologies
$\omega \in \Omega$ scenarios

https://doi.org/10.1021/acs.iecr.2c02968
Ind. Eng. Chem. Res. 2023, 62, 2067−2079

T hyperrectangle set of parameters in the flexibility index analysis

Parameters

$\alpha(m)$ capital cost scaling exponent for technology $m \in M$
$\beta(m)$ operating cost scaling exponent for technology $m \in M$
$\gamma^+(s, \omega)$ electricity availability from each source $s \in S$ in scenario $\omega \in \Omega$
δ_L lower bound of flexibility index requirement
ε constant used in the ε-constraint method
θ^N nominal point of parameter
$\Delta\theta^+, \Delta\theta^-$ positive and negative expected deviation of parameter
$\mu(i, \omega)$ purchase or selling price of $i \in I$ in scenario $\omega \in \Omega$
$\nu(i, m)$ conversion coefficient of products $i \in I$ using technology $m \in M$
$\pi_c(m)$ greenhouse emissions of plant construction (technology $m \in M$)
$\pi_e(S)$ greenhouse emissions of electricity generation from energy source $s \in S$
$\pi_P(m)$ greenhouse emissions during the production stage (technology $m \in M$)
$\pi_u(i)$ greenhouse emissions upstream raw material $i \in I$ extraction
$\rho^-(i, \omega)$ lower bounds of flow rate; supply of raw material $i \in I$
$\rho^+(i, \omega)$ upper bounds of flow rate; demand of product $i \in I$
n_z number of control variables in the flexibility index analysis
$C_0(m)$ annualized capital cost basis at a plant scale of $Q_0(m)$
$O_0(m)$ fixed operational cost basis at a plant scale of $Q_0(m)$
$Q_0(m)$ plant scale as the basis for costing
$Q^-(m)$ minimum capacities of technology $m \in M$
$Q^+(m)$ maximum capacities of technology $m \in M$
$R_0(m)$ variable operational basis at a plant scale of $Q_0(m)$
U_e, U big-M constants

Variables

$b(\omega)$ zero if electricity is sold and one if purchased in scenario $\omega \in \Omega$
d design variables in the flexibility index analysis
$k(\omega)$ amount of electricity sold to the market in scenario $\omega \in \Omega$
s_j slack variable for each constraint f_j
$x(m, \omega)$ the actual production activity of technology $m \in M$ in scenario $\omega \in \Omega$
$y(m)$ zero when technology m is not chosen and one when chosen
y_j zero when inequality constraint f_j is inactive and one when active
z recourse/control actions in the flexibility index analysis
$\gamma(s, \omega)$ electricity usage from energy source $s \in S$ in scenario $\omega \in \Omega$
δ flexibility index
θ uncertain parameters in the flexibility index analysis
λ_j Lagrange multiplier for each constraint f_j
$\rho(i, \omega)$ flow rate of $i \in I$ in scenario $\omega \in \Omega$
CAPEX annualized capital cost
GWP(ω) global warming potential of producing 1 kg product $i \in I$
OPEX(ω) total operating cost in scenario $\omega \in \Omega$
Profit(ω) total annual profit in scenario $\omega \in \Omega$
$Q(m)$ actual designed capacity in the biorefinery for technology $m \in M$
Total GWP(ω) total global warming potential of biorefinery in scenario $\omega \in \Omega$

Expressions

$f_j(d, z, \theta)$ inequality constraints after eliminating state variables by substituting equality constraints
ψ function to indicate constraint violation in the original model
F flexibility index as a function of design variables
$\hat{F}$ surrogate model for the flexibility index

Abbreviations

ANN artificial neural network
BARON branch and reduce optimization navigator
CPLEX IBM ILOG CPLEX solver
GWP global warming potential
HMF hydroxymethylfurfural
LCA life cycle assessment
LHS Latin hypercube sampling
MILP mixed integer linear programming
MINLP mixed integer nonlinear programming
MSH molten salt hydrate
PSA pressure-sensitive adhesive
RBF radial basis function
RCF reductive catalytic fractionation
RD - RCF reactive distillation−reductive catalytic fractionation
ReLU rectified linear unit
SVM support vector machine
SVR support vector regression
TSSP two-stage stochastic programming

REFERENCES

(1) Kaler, E. W.; Alger, M. M.; Barabino, G. A.; Beckham, G. T.; Collias, D. I.; Pablo, J. J. D.; Glotzer, S. C.; Hammond, P. T.; Iglesia, E.; Kim, S.et al.*New Directions for Chemical Engineering*; National Academy of Engineering, 2022.

(2) Athaley, A.; Annam, P.; Saha, B.; Ierapetritou, M. Techno-economic and life cycle analysis of different types of hydrolysis process for the production of p-Xylene. *Comput. Chem. Eng.* **2019**, *121*, 685−695.

(3) Alonso, D. M.; Hakim, S. H.; Zhou, S.; Won, W.; Hosseinaei, O.; Tao, J.; Garcia-Negron, V.; Motagamwala, A. H.; Mellmer, M. A.; Huang, K.; et al. Increasing the revenue from lignocellulosic biomass: Maximizing feedstock utilization. *Sci. Adv.* **2017**, *3* (5), No. e1603301.

(4) Athaley, A.; Saha, B.; Ierapetritou, M. Biomass-based chemical production using techno-economic and life cycle analysis. *AlChE J.* **2019**, *65* (9), No. e16660.

(5) Liao, Y.; Koelewijn, S.-F.; Van den Bossche, G.; Van Aelst, J.; Van den Bosch, S.; Renders, T.; Navare, K.; Nicolaï, T.; Van Aelst, K.; Maesen, M.; Matsushima, H.; Thevelein, J.; Van Acker, K.; Lagrain, B.; Verboekend, D.; Sels, B. F. A sustainable wood biorefinery for low-carbon footprint chemicals production. *Science* **2020**, *367* (6484), 1385−1390.

(6) Yuan, Z.; Eden, M. R. Superstructure optimization of integrated fast pyrolysis-gasification for production of liquid fuels and propylene. *AlChE J.* **2016**, *62* (9), 3155−3176.

(7) Gong, J.; You, F. Optimal Design and Synthesis of Algal Biorefinery Processes for Biological Carbon Sequestration and Utilization with Zero Direct Greenhouse Gas Emissions: MINLP

https://doi.org/10.1021/acs.iecr.2c02968
Ind. Eng. Chem. Res. 2023, 62, 2067−2079

Model and Global Optimization Algorithm. *Ind. Eng. Chem. Res.* **2014**, *53* (4), 1563−1579.
(8) Bartling, A. W.; Stone, M. L.; Hanes, R. J.; Bhatt, A.; Zhang, Y.; Biddy, M. J.; Davis, R.; Kruger, J. S.; Thornburg, N. E.; Luterbacher, J. S.; Rinaldi, R.; Samec, J. S. M.; Sels, B. F.; Román-Leshkov, Y.; Beckham, G. T. Techno-economic analysis and life cycle assessment of a biorefinery utilizing reductive catalytic fractionation. *Energy Environ. Sci.* **2021**, *14* (8), 4147−4168.
(9) Restrepo-Flórez, J.-M.; Ryu, J.; Witkowski, D.; Rothamer, D. A.; Maravelias, C. T. A systems level analysis of ethanol upgrading strategies to middle distillates. *Energy Environ. Sci.* **2022**, *15* (10), 4376−4388.
(10) Restrepo-Flórez, J. M.; Maravelias, C. T. Advanced fuels from ethanol - a superstructure optimization approach. *Energy Environ. Sci.* **2021**, *14* (1), 493−506.
(11) Luo, Y.; O'Dea, R. M.; Gupta, Y.; Chang, J.; Sadula, S.; Soh, L. P.; Robbins, A. M.; Levia, D. F.; Vlachos, D. G.; Epps, T. H.; Ierapetritou, M. A Life Cycle Greenhouse Gas Model of a Yellow Poplar Forest Residue Reductive Catalytic Fractionation Biorefinery. *Environ. Eng. Sci.* **2022**, *39* (10), 821−833.
(12) Dickson, R.; Brigljevic, B.; Lim, H.; Liu, J. Maximizing the sustainability of a macroalgae biorefinery: a superstructure optimization of a volatile fatty acid platform. *Green Chem.* **2020**, *22* (13), 4174−4186.
(13) Sahinidis, N. V. Optimization under uncertainty: state-of-the-art and opportunities. *Comput. Chem. Eng.* **2004**, *28* (6), 971−983.
(14) Lin, Z.; Nikolakis, V.; Ierapetritou, M. Life cycle assessment of biobased p-xylene production. *Ind. Eng. Chem. Res.* **2015**, *54* (8), 2366−2378.
(15) Heijungs, R. On the number of Monte Carlo runs in comparative probabilistic LCA. *Int. J. Life Cycle Assess.* **2020**, *25* (2), 394−402.
(16) Henriksson, P. J. G.; Heijungs, R.; Dao, H. M.; Phan, L. T.; de Snoo, G. R.; Guinée, J. B. Product Carbon Footprints and Their Uncertainties in Comparative Decision Contexts. *PLoS One* **2015**, *10* (3), No. e0121221.
(17) Ciroth, A.; Muller, S.; Weidema, B.; Lesage, P. Empirically based uncertainty factors for the pedigree matrix in ecoinvent. *Int. J. Life Cycle Assess.* **2016**, *21* (9), 1338−1348.
(18) Kätelhön, A.; Bardow, A.; Suh, S. Stochastic Technology Choice Model for Consequential Life Cycle Assessment. *Environ. Sci. Technol.* **2016**, *50* (23), 12575−12583.
(19) Yue, D.; You, F. Optimal supply chain design and operations under multi-scale uncertainties: Nested stochastic robust optimization modeling framework and solution algorithm. *AlChE J.* **2016**, *62* (9), 3041−3055.
(20) Ochoa, M. P.; García-Muñoz, S.; Stamatis, S.; Grossmann, I. E. Novel flexibility index formulations for the selection of the operating range within a design space. *Comput. Chem. Eng.* **2021**, *149*, 107284.
(21) Gulcan, B.; Eksioglu, S. D.; Song, Y.; Roni, M.; Chen, Q. Optimization models for integrated biorefinery operations. *Optim. Lett.* **2022**, *16*, 909−951.
(22) Li, Z.; Ding, R.; Floudas, C. A. A Comparative Theoretical and Computational Study on Robust Counterpart Optimization: I. Robust Linear Optimization and Robust Mixed Integer Linear Optimization. *Ind. Eng. Chem. Res.* **2011**, *50* (18), 10567−10603.
(23) Ostrovsky, G. M.; Datskov, I. V.; Achenie, L. E. K.; Volin, Y. M. Process uncertainty: Case of insufficient process data at the operation stage. *AlChE J.* **2003**, *49* (5), 1216−1232.
(24) Martín, M.; Martínez, A. Addressing Uncertainty in Formulated Products and Process Design. *Ind. Eng. Chem. Res.* **2015**, *54* (22), 5990−6001.
(25) Bhosekar, A.; Athaley, A.; Ierapetritou, M. Multiobjective modular biorefinery configuration under uncertainty. *Ind. Eng. Chem. Res.* **2021**, *60* (35), 12956−12969.
(26) Zhang, Q.; Grossmann, I. E.; Lima, R. M. On the relation between flexibility analysis and robust optimization for linear systems. *AlChE J.* **2016**, *62* (9), 3109−3123.
(27) Swaney, R. E.; Grossmann, I. E. An index for operational flexibility in chemical process design. Part I: Formulation and theory. *AlChE J.* **1985**, *31* (4), 621−630.
(28) Grossmann, I. E.; Floudas, C. A. Active constraint strategy for flexibility analysis in chemical processes. *Comput. Chem. Eng.* **1987**, *11* (6), 675−693.
(29) Wang, H.; Mastragostino, R.; Swartz, C. L. E. Flexibility analysis of process supply chain networks. *Comput. Chem. Eng.* **2016**, *84*, 409−421.
(30) Bhosekar, A.; Ierapetritou, M. Modular Design Optimization using Machine Learning-based Flexibility Analysis. *J. Process Control* **2020**, *90*, 18−34.
(31) Schweidtmann, A. M.; Mitsos, A. Deterministic Global Optimization with Artificial Neural Networks Embedded. *J. Optim. Theory Appl.* **2019**, *180* (3), 925−948.
(32) Henao, C. A.; Maravelias, C. T. Surrogate-based superstructure optimization framework. *AlChE J.* **2011**, *57* (5), 1216−1232.
(33) Metta, N.; Ramachandran, R.; Ierapetritou, M. A novel adaptive sampling based methodology for feasible region identification of compute intensive models using artificial neural network. *AlChE J.* **2021**, *67* (2), No. e17095.
(34) Rogers, A.; Ierapetritou, M. Feasibility and flexibility analysis of black-box processes part 2: Surrogate-based flexibility analysis. *Chem. Eng. Sci.* **2015**, *137*, 1005−1013.
(35) Kim, S. H.; Boukouvala, F. Surrogate-based optimization for mixed-integer nonlinear problems. *Comput. Chem. Eng.* **2020**, *140*, 106847.
(36) Katz, J.; Pappas, I.; Avraamidou, S.; Pistikopoulos, E. N. Integrating deep learning models and multiparametric programming. *Comput. Chem. Eng.* **2020**, *136*, 106801.
(37) Grimstad, B.; Andersson, H. ReLU networks as surrogate models in mixed-integer linear programs. *Comput. Chem. Eng.* **2019**, *131*, 106580.
(38) Ibarra-Gonzalez, P.; Rong, B.-G. Integrated Methodology for the Optimal Synthesis of Lignocellulosic Biomass-to-Liquid Fuel Production Processes: 1. Simulation-Based Superstructure Synthesis and Development. *Ind. Eng. Chem. Res.* **2020**, *59* (33), 14881−14897.
(39) Tribe, M. A.; Alpine, R. L. W. Scale economies and the "0.6 rule". *Eng. Costs Prod. Econ.* **1986**, *10* (1), 271−278.
(40) Misener, R.; Floudas, C. A. Piecewise-Linear Approximations of Multidimensional Functions. *J. Optim. Theory Appl.* **2010**, *145* (1), 120−147.
(41) Harvey, L. D. D. A guide to global warming potentials (GWPs). *Energy Policy* **1993**, *21* (1), 24−34.
(42) de Jong, S.; Antonissen, K.; Hoefnagels, R.; Lonza, L.; Wang, M.; Faaij, A.; Junginger, M. Life-cycle analysis of greenhouse gas emissions from renewable jet fuel production. *Biotechnol. Biofuels* **2017**, *10* (1), 64.
(43) Raimondi, A.; Girotti, G.; Blengini, G. A.; Fino, D. LCA of petroleum-based lubricants: state of art and inclusion of additives. *Int. J. Life Cycle Assess.* **2012**, *17* (8), 987−996.
(44) Azapagic, A.; Clift, R. Allocation of environmental burdens in multiple-function systems. *J. Cleaner Prod.* **1999**, *7* (2), 101−119.
(45) Pulsipher, J. L.; Zavala, V. M. A mixed-integer conic programming formulation for computing the flexibility index under multivariate gaussian uncertainty. *Comput. Chem. Eng.* **2018**, *119*, 302−308.
(46) Floudas, C. A.; Grossmann, I. E. Synthesis of flexible heat exchanger networks with uncertain flowrates and temperatures. *Comput. Chem. Eng.* **1987**, *11* (4), 319−336.
(47) Chen, Y.; Lin, M.; Jiang, H.; Yuan, Z.; Chen, B. Optimal design and operation of refinery hydrogen systems under multi-scale uncertainties. *Comput. Chem. Eng.* **2020**, *138*, 106822.
(48) Zhao, X.; You, F. Consequential Life Cycle Assessment and Optimization of High-Density Polyethylene Plastic Waste Chemical Recycling. *ACS Sustainable Chem. Eng.* **2021**, *9*, 12167−12184.
(49) Mencarelli, L.; Chen, Q.; Pagot, A.; Grossmann, I. E. A review on superstructure optimization approaches in process system engineering. *Comput. Chem. Eng.* **2020**, *136*, 106808.

https://doi.org/10.1021/acs.iecr.2c02968
Ind. Eng. Chem. Res. 2023, 62, 2067−2079

(50) Batchu, S. P.; Hernandez, B.; Malhotra, A.; Fang, H.; Ierapetritou, M.; Vlachos, D. G. Accelerating manufacturing for biomass conversion via integrated process and bench digitalization: a perspective. *React. Chem. Eng.* **2022**, *7* (4), 813−832.
(51) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(52) Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.TensorFlow: a system for Large-Scale machine learning; *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation*, Savannah, GA, November 2−4; 2016; pp 265−283.
(53) Serfidan, A. C.; Uzman, F.; Türkay, M. Optimal estimation of physical properties of the products of an atmospheric distillation column using support vector regression. *Comput. Chem. Eng.* **2020**, *134*, 106711.
(54) Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural Networks* **1989**, *2* (5), 359−366.
(55) Ghoreishi, S.; Løhre, C.; Hermundsgård, D. H.; Molnes, J. L.; Tanase-Opedal, M.; Brusletto, R.; Barth, T., Identification and quantification of valuable platform chemicals in aqueous product streams from a preliminary study of a large pilot-scale steam explosion of woody biomass using quantitative nuclear magnetic resonance spectroscopy. *Biomass Conv. Bioref.***2022**, DOI: 10.1007/s13399-022-02712-w (accessed 2022−10−03).
(56) Ebikade, E.; Athaley, A.; Fisher, B.; Yang, K.; Wu, C.; Ierapetritou, M. G.; Vlachos, D. G. The Future is Garbage: Repurposing of Food Waste to an Integrated Biorefinery. *ACS Sustainable Chem. Eng.* **2020**, *8* (22), 8124−8136.
(57) O'Dea, R. M.; Pranda, P. A.; Luo, Y.; Amitrano, A.; Ebikade, E. O.; Gottlieb, E. R.; Ajao, O.; Benali, M.; Vlachos, D. G.; Ierapetritou, M.; Epps, T. H., III Ambient-pressure lignin valorization to high-performance polymers by intensified reductive catalytic deconstruction. *Sci. Adv.* **2022**, *8* (3), No. eabj7523.
(58) Luo, Y.; Kuo, M. J.; Ye, M.; Lobo, R.; Ierapetritou, M. Comparison of 4,4′-Dimethylbiphenyl from Biomass-Derived Furfural and Oil-Based Resource: Technoeconomic Analysis and Life-Cycle Assessment. *Ind. Eng. Chem. Res.* **2022**, *61* (25), 8963−8972.
(59) Bozell, J. J.; Petersen, G. R. Technology development for the production of biobased products from biorefinery carbohydrates—the US Department of Energy's "Top 10" revisited. *Green Chem.* **2010**, *12* (4), 539−554.
(60) Davidson, M. G.; Elgie, S.; Parsons, S.; Young, T. J. Production of HMF, FDCA and their derived products: a review of life cycle assessment (LCA) and techno-economic analysis (TEA) studies. *Green Chem.* **2021**, *23* (9), 3154−3171.
(61) Heijungs, R.; Settanni, E.; Guinée, J. Toward a computational structure for life cycle sustainability analysis: unifying LCA and LCC. *Int. J. Life Cycle Assess.* **2013**, *18* (9), 1722−1733.
(62) Luo, Y.; Ierapetritou, M. Comparison between Different Hybrid Life Cycle Assessment Methodologies: A Review and Case Study of Biomass-based p-Xylene Production. *Ind. Eng. Chem. Res.* **2020**, *59* (52), 22313−22329.
(63) Jet fuel demand hit by Omicron; recovery delayed until 2023/2024. *Oil Energy Trends***2022**, *47* (2), 3−5,.
(64) O'Neill, E. G.; Martinez-Feria, R. A.; Basso, B.; Maravelias, C. T. Integrated spatially explicit landscape and cellulosic biofuel supply chain optimization under biomass yield uncertainty. *Comput. Chem. Eng.* **2022**, *160*, 107724.
(65) Volk, T. A.; Berguson, B.; Daly, C.; Halbleib, M. D.; Miller, R.; Rials, T. G.; Abrahamson, L. P.; Buchman, D.; Buford, M.; Cunningham, M. W.; Eisenbies, M.; Fabio, E. S.; Hallen, K.; Heavey, J.; Johnson, G. A.; Kuzovkina, Y. A.; Liu, B.; McMahon, B.; Rousseau, R.; Shi, S.; Shuren, R.; Smart, L. B.; Stanosz, G.; Stanton, B.; Stokes, B.; Wright, J. Poplar and shrub willow energy crops in the United States: field trial results from the multiyear regional feedstock partnership and yield potential maps based on the PRISM-ELM model. *GCB Bioenergy* **2018**, *10* (10), 735−751.
(66) Schmitz, A.; Kennedy, P. L.; Zhang, F. Sugarcane and sugar yields in Louisiana (1911−2018): Varietal development and mechanization. *Crop Sci.* **2020**, *60* (3), 1303−1312.
(67) Huntington, T.; Cui, X.; Mishra, U.; Scown, C. D. Machine learning to predict biomass sorghum yields under future climate scenarios. *Biofuels, Bioprod. Bioref.* **2020**, *14* (3), 566−577.
(68) Dzotsi, K. A.; Basso, B.; Jones, J. W. Development, uncertainty and sensitivity analysis of the simple SALUS crop model in DSSAT. *Ecol. Modell.* **2013**, *260*, 62−76.
(69) Sokhansanj, S.; Turhollow, A.; Wilkerson, E.*Development of the Integrated Biomass Supply Analysis and Logistics Model (IBSAL)*; Report No. ORNL/TM-2006/57; Oak Ridge National Laboratory, 2008.
(70) Barnett, J.Demand for jet fuel in the U.S. is recovering faster than in many other markets. https://www.eia.gov/todayinenergy/detail.php?id=44996 (accessed 2022-07−26).
(71) Park, D. S.; Joseph, K. E.; Koehle, M.; Krumm, C.; Ren, L.; Damen, J. N.; Shete, M. H.; Lee, H. S.; Zuo, X.; Lee, B.; Fan, W.; Vlachos, D. G.; Lobo, R. F.; Tsapatsis, M.; Dauenhauer, P. J. Tunable Oleo-Furan Surfactants by Acylation of Renewable Furans. *ACS Cent. Sci.* **2016**, *2* (11), 820−824.
(72) Mante, O. D.*Building Blocks from Biocrude: High-Value Methoxyphenols*; Report No. DOE-RTI-7730-1; RTI International: Research Triangle Park, NC, 2020.
(73) Bass, G. F.; Epps, T. H., III Recent developments towards performance-enhancing lignin-based polymers. *Polym. Chem.* **2021**, *12* (29), 4130−4158.
(74) Buyle, M.; Audenaert, A.; Billen, P.; Boonen, K.; Van Passel, S. The Future of Ex-Ante LCA? Lessons Learned and Practical Recommendations. *Sustainability* **2019**, *11* (19), 5456.
(75) Wang, Z.; Ierapetritou, M. A novel feasibility analysis method for black-box processes using a radial basis function adaptive sampling approach. *AlChE J.* **2017**, *63* (2), 532−550.
(76) Nguyen, A.; Pham, K.; Ngo, D.; Ngo, T.; Pham, L. In An Analysis of State-of-the-art Activation Functions For Supervised Deep Neural Network. In *Proceedings of the 2021 International Conference on System Science and Engineering (ICSSE)*, 26−28 Aug. 2021; pp 215−220.
(77) Qin, Y.; Yang, Y.; Cucurachi, S.; Suh, S. Non-linearity in Marginal LCA: Application of a Spatial Optimization Model. *Front. Sustain.* **2021**, *2* (28), 631080.
(78) Goyal, V.; Ierapetritou, M. G. Framework for evaluating the feasibility/operability of nonconvex processes. *AlChE J.* **2003**, *49* (5), 1233−1240.

https://doi.org/10.1021/acs.iecr.2c02968
Ind. Eng. Chem. Res. 2023, 62, 2067−2079