I&EC research
Industrial & Engineering Chemistry Research

pubs.acs.org/IECR

Article

Machine Learning-Based Operational Modeling of an Electrochemical Reactor: Handling Data Variability and Improving Empirical Models

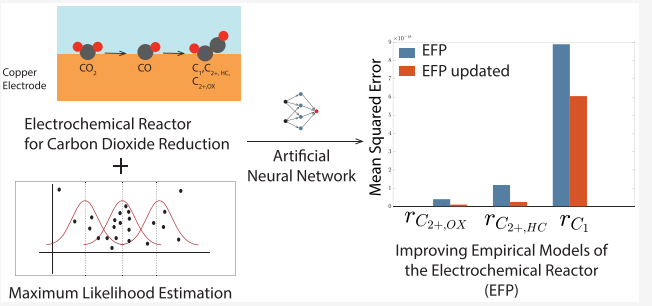
Junwei Luo, Vito Canuso, Joon Baek Jang, Zhe Wu, Carlos G. Morales-Guio, and Panagiotis D. Christofides*

Cite This: *Ind. Eng. Chem. Res.* 2022, 61, 8399–8410

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Electrochemical reduction of carbon dioxide (CO_2) has received increasing attention with the recent rise in awareness of climate change and the increase in electricity supply from clean energy sources. However, because of the complexity of its reaction mechanism and the largely unknown electron transfer pathways, the development of a first-principles-based operational model of a CO_2 electrocatalytic reactor is still in its infancy. This work proposes a methodology to develop a feed-forward neural network (FNN) model to capture the input−output relationship of an experimental electrochemical reactor from experimental data that are obtained from easy-to-implement sensors. This FNN model is computationally efficient and can be used in real-time to determine energy-optimal reactor operating conditions. To further account for the uncertainty of the experimental data, the maximum likelihood estimation (MLE) method is adopted to construct a statistical neural network, which is demonstrated to be able to address a usual overfitting problem that occurs in the standard FNN model. In addition, by comparing the neural network with an empirical first-principles-based model, it is demonstrated that the neural network model achieves improved prediction accuracy with respect to experimentally determined input−output operating conditions. Finally, the insights obtained from the FNN model and the limitations identified of the empirical, first-principles model (EFP model) are used to propose specific modifications to the EFP model to improve its prediction capability.



INTRODUCTION

The electrochemical transformation of carbon dioxide (CO_2) into carbon-based fuels and chemicals has received growing interest in this century, because of its potential to reduce CO_2 emissions and facilitate the production of energy from renewable sources.[1] The biggest challenge for research in this area is the difficulty in determining and quantifying the products that result from the reduction of CO_2. Specifically, the CO_2 reduction pathways constitute a complex web of reactions that result in the production of various alkanes, alkenes, and oxygenate species.[2] In addition, recent small-scale experiments on this process show varying levels of experimental uncertainty, because of the minimum measurable limit of the sensors and other inevitable experimental errors. This can introduce a level of uncertainty into the data that can increase the probability of overfitting.

Although developing mathematical models, such as first-principle models, is a classic and reliable way to describe and predict a physical process, the uncertainty and complexity of most engineering systems make it challenging to implement. To overcome this problem, various data-driven models, as well as artificial intelligence (AI) approaches, have been proposed historically. Early on, in the 1960s, an epochal AI logic, the Fuzzy logic, was proposed by Zadeh to approximate uncertain features.[3] From then on, techniques of machine learning for real-time process operation were investigated in 1990s,[4] such as the expert system.[5] In addition, the autoregressive model provided statistical strategies to develop data-driven models based on recorded observation. For example, the autoregressive-moving-average model (ARMA) proposed by Peter Whittle back in the early 1950s.[6]

With the development of open-source deep-learning libraries and availability of large datasets from experimental electrochemical reactors (as well as other chemical reactor systems), machine learning modeling of electrochemical reactors and other reactor systems has become a growing field of interest within chemical engineering. Specifically, various versions of

Special Issue: Machine Learning and Data Science in Chemical Engineering

Received: October 20, 2021
Revised: December 23, 2021
Accepted: December 23, 2021
Published: January 10, 2022

https://doi.org/10.1021/acs.iecr.1c04176
Ind. Eng. Chem. Res. 2022, 61, 8399−8410

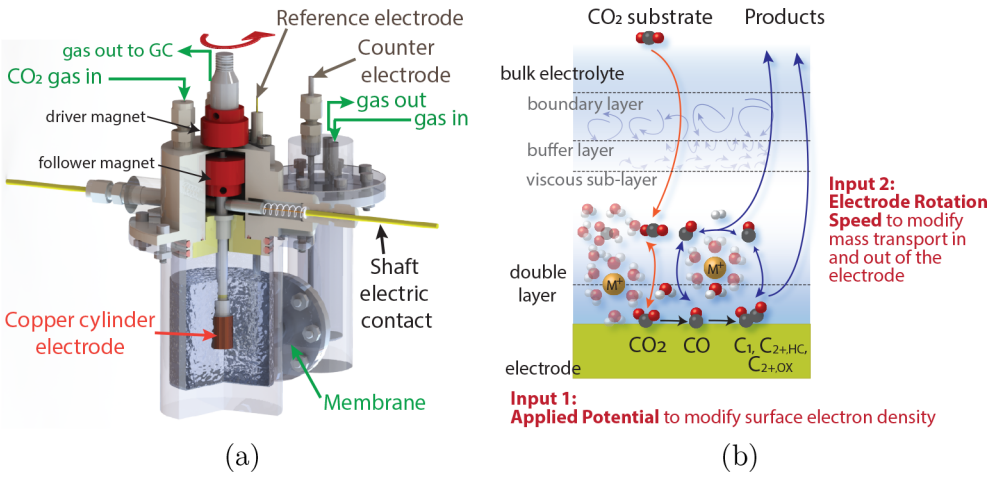


Figure 1. A diagram showing (a) the electrochemical reactor and (b) the multiple and complex reaction and mass-transfer processes involved in the transformation of CO_2 to CO and further reduced products on the polycrystalline copper cylinder electrode.

artificial neural network (ANN) models have demonstrated their ability to address regression and classification problems in the context of chemical process modeling.[7−12] An ANN has many degrees of freedom, which gives it the ability to capture the complex, nonlinear relationships between an electrochemical reactor system input and output variables. In addition, it is a customary approach to combine an ANN model with an empirical, first-principles model (that is a model that is based on chemical reaction engineering fundamentals, yet its parameters are fitted to experimental data) to investigate complex reaction mechanisms and reactor macroscopic input−output behavior.[13,14]

Over the past few years, ANNs have been used to model chemical engineering manufacturing processes in several studies. For example, in ref 15 a recurrent neural network (RNN) model was used to model a chemical reactor using the data from a large-scale process simulator (Aspen Plus Dynamics) and was used in a model predictive controller (MPC) to stabilize the process at a steady state. In ref 16, a feed-forward neural network (FNN) model was developed to correlate the input and output variables of a SiO_2 atomic layer deposition (ALD) process to calculate optimal half-cycle times to fully coverage, which is an important industrial parameter. Moreover, in ref 17, a methodology was discussed wherein neural networks were used for parameter estimation from experimental data. These research investigations provide a strong support for using neural network models as a reliable approximation to analyze complex nonlinear relationships from simulation/experimental data for electrochemical reactors.

Other works have applied deep learning methods to improve operational aspects of industrial chemical processes. ANNs have been used as process models to replace traditional models to further optimize the control and operation of chemical and industrial processes. In ref 18, a deep reinforcement learning controller was used to control a hydraulic fracturing process to improve the safety and optimization of system operation. In addition, an operational model was constructed for this process using a hybrid approach of a deep neural network and a first-principles model.[19] ANNs were further used to determine optimum operating conditions for chemical and industrial processes,[20−25] which contributed to maximizing the feasibility of novel processes from economic and safety perspectives.

Motivated by the above considerations, this work develops an FNN model using steady-state, input−output experimental electrochemical reactor data by solving a nonlinear regression problem, accounting for data variability. This FNN model is computationally efficient and can be used in real-time to determine safe and energy-optimal electrochemical reactor operating conditions. Specifically, the maximum likelihood estimation (MLE) method was integrated with the FNN model development algorithm to account for the uncertainty and variability of the experimental data by determining their confidence interval and weighing each point accordingly in the FNN model training process. Therefore, the FNN model is able to account for the data variability and provide the statistically most likely trajectory of the experiment output over a broad set of operating conditions. This probabilistic method decreases the chance that the model will be overfit to specific training points with large variation. The key novelty of this work is the development of an operational model for a state-of-the-art electrochemical reactor using a statistical machine learning method. In addition, the insights obtained from the FNN model are used to propose specific modifications to a classical, empirical first-principles model (EFP model) of electrochemical phenomena to improve its prediction capability, which can contribute to the investigation of the unknown first-principle chemical reactor equations.

The rest of this manuscript is organized as follows. In the section entitled "Preliminaries", the experimental reactor setup and the kinetics of the electrochemical reactions are described. In the next section, entitled "Development of Machine Learning Model", the formulation and the construction method of the FNN model are discussed. In the section entitled "Maximum Likelihood Estimation in ML Reactor Modeling", the methodology of the maximum likelihood estimation is integrated with the FNN modeling method. In the next section, entitled "Machine Learning Model Results and Analysis", the performance of the FNN models is evaluated, and the statistical FNN model predictions and insights are used to improve an EFP model for this reactor.

■ PRELIMINARIES

This section introduces the background of the experimental electrochemical reactor employed in this work. Specifically, the experimental setup and basic operating reactor mode are

https://doi.org/10.1021/acs.iecr.1c04176
Ind. Eng. Chem. Res. 2022, 61, 8399−8410

Table 1. Electrochemical Reactions To Reduce CO_2 to Various Products on Copper

index	reaction	classification
1	$CO_2 + 6H_2O + 8e^- \rightarrow CH_4 + 8OH^-$	C_1 hydrocarbon (HC)
2	$2CO_2 + 8H_2O + 12e^- \rightarrow C_2H_4 + 12OH^-$	C_{2+} hydrocarbon (HC)
3	$CO_2 + 5H_2O + 6e^- \rightarrow CH_3OH + 6OH^-$	C_1 oxygenate (OX)
4	$\mathbf{2CO_2 + 9H_2O + 12e^- \rightarrow C_2H_5OH + 12OH^-}$	**C_{2+} oxygenate (OX)**
5	$\mathbf{2CO_2 + 5H_2O + 8e^- \rightarrow CH_3COO^- + 7OH^-}$	**C_{2+} oxygenate (OX)**
6	$\mathbf{2CO_2 + 8H_2O + 10e^- \rightarrow (CH_2OH)_2 + 10OH^-}$	**C_{2+} oxygenate (OX)**
7	$\mathbf{2CO_2 + 6H_2O + 8e^- \rightarrow HOCH_2CHO + 8OH^-}$	**C_{2+} oxygenate (OX)**
8	$\mathbf{2CO_2 + 7H_2O + 10e^- \rightarrow CH_3CHO + 10OH^-}$	**C_{2+} oxygenate (OX)**
9	$\mathbf{3CO_2 + 13H_2O + 18e^- \rightarrow C_3H_7OH + 18OH^-}$	**C_{2+} oxygenate (OX)**
10	$\mathbf{3CO_2 + 11H_2O + 16e^- \rightarrow C_3H_5OH + 16OH^-}$	**C_{2+} oxygenate (OX)**
11	$3CO_2 + 11H_2O + 16e^- \rightarrow CH_3COCH_3 + 16OH^-$	**C_{2+} oxygenate (OX)**
12	$\mathbf{3CO_2 + 11H_2O + 16e^- \rightarrow C_2H_5CHO + 16OH^-}$	**C_{2+} oxygenate (OX)**
13	$CO_2 + H_2O + 2e^- \rightarrow CO + 2OH^-$	Intermediate

presented in this section. Then, an overview of the input−output behavior for this process is used to further explain the data structure used in the neural network model. The experimental reactor and microscopic transport diagrams are shown in Figure 1.

Experimental Electrochemical Reactor. The reactor was designed to study the effect of mass transport on electrochemical CO_2 reduction while keeping the electrochemical cell hermetically gastight for the online detection of gas products. This is allowed by magnetic coupling where the driver magnet outside, connected to the modulated speed rotator (MSR), transmits torque to the follower magnet inside the reactor (Figure 1). The reactor has two chambers separated by an ion-exchange membrane to prevent the crossover of products. One chamber contains the working electrode, which is the cathode in this case. The other chamber contains the counter electrode (anode). The CO_2 gas is directly bubbled into both chambers, where the electrodes are submerged in 0.2 M potassium bicarbonate buffer electrolyte. The cathode is a rotating cylinder electrode (RCE) made of polycrystalline copper. Copper is the only known single transition metal that can reduce CO_2 into hydrocarbons and oxygenates with more than two carbons (C_{2+}) at an appreciable rate, and it plays a critical role as the catalyst in the overall reaction scheme.[26] As the RCE shaft continuously stirs the electrolyte solution, hydrodynamics formed around the electrode can be systematically controlled by setting a rotation speed from the MSR. Finally, gas and liquid products are analyzed by gas chromatograph (GC) and nuclear magnetic resonance (NMR) spectroscopy, respectively, to determine the product composition under well-controlled mass transport characteristics. Further details on the reactor design and experimental setup are newly reported.[27]

The product compositions quantified using gas chromatography (GC) and nuclear magnetic resonance (NMR) are then used to determine the production rate of each species and the reaction selectivity, with respect to desired products. Polycrystalline copper produces various products as tabulated in Table 1 at a quantifiable level. Here, the competing hydrogen evolution reaction and the production of formate are excluded from the table and from the selectivity calculation, since they do not share the same reaction pathway as the products in Table 1. That is, although carbon monoxide and formate are two-electron reduction products, carbon monoxide is the main reaction intermediate toward further reduced products while formate cannot be further reduced. Among the products sharing carbon monoxide as a common intermediate, the desired products are the oxygenate species (labeled in Table 1), since they are of high value and are commonly used as liquid fuels and reagents. Therefore, the selectivity for this experiment is defined as the ratio of the rate of oxygenate production to the rate of hydrocarbon production.

With respect to the reactor mode of operation, the CO_2 gas dissolves into the buffer solution, and is carried to the electrode surface by convective mass transport caused by the rotating electrode. Subsequently, the CO_2 molecules are adsorbed onto the electrode surface and reduced to oxygenate and hydrocarbon products through consecutive proton-coupled electron injection steps. Therefore, the surface reaction rate is determined by the electron density on the Cu surface and the adsorption rate of CO_2 molecules to the Cu surface. The electron density is dictated by the applied potential, and the adsorption rate of CO_2 is the result of complex mass transport and electrode kinetics at the electrode/electrolyte interface (Figure 1).

DEVELOPMENT OF MACHINE LEARNING MODEL

In this section, a neural network model is constructed to capture the steady-state behavior of the reactor at varying applied potentials and electrode rotation speeds using experimental electrochemical reactor input−output data. The neural network model formulation, training process, and the data collection process are presented in the following subsections.

FNN Learning Algorithm. The general structure of an FNN model is shown in Figure 2 and can be mathematically represented by the following equations:

$$\mathbf{Y} = F_{\mathrm{NN}}(X) = \begin{cases} h_j^{[1]} = \sigma^{[1]}\left(\sum_{i=1}^{p} \omega_{ji}^{[1]} x_i + b^{[1]}\right) \\ h_j^{[2]} = \sigma^{[2]}\left(\sum_{i=1}^{p} \omega_{ji}^{[2]} h_i^{[1]} + b^{[2]}\right) \\ y_j = \sigma^{[l]}\left(\sum_{i=1}^{p} \omega_{ji}^{[l]} h_i^{[l-1]} + b^{[l]}\right) \end{cases} \quad (1)$$

where $\mathbf{X} = [x_1, \cdots, x_n] \in \mathbf{R}^n$ and $\mathbf{Y} = [\hat{y}_1, \cdots, \hat{y}_m] \in \mathbf{R}^m$ are the input and output vectors of the FNN model, respectively. The term $\omega_{ji}^{[k]}$ (where $i = 1, \cdots, p$, $j = 1, \cdots, p$, and $k = 1, \cdots, l$) stands for the weights connecting the ith input from the prior layer to the jth neuron in the kth layer, where l is the number of layers. p represents the number of neurons used in each layer. Therefore, $i = 1, \cdots, n$ for the first hidden layer, because there are n units in

https://doi.org/10.1021/acs.iecr.1c04176
Ind. Eng. Chem. Res. 2022, 61, 8399−8410

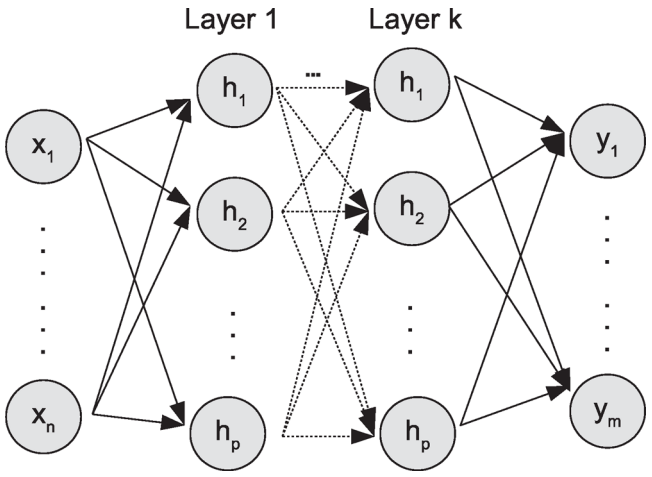


Figure 2. General structure of an FNN model, where subscript p is the index of neurons in the kth hidden layer.

the input layer. $b^{[k]}$ and $\sigma^{[k]}(\cdot)$ respectively denote the bias and activation function used in the kth layer.

In this study, a centralized two-input−multioutput FNN model is constructed to capture the nonlinear relationship between the two input states in Table 2 (i.e., rotation speed and applied potential) and the 14 outputs listed in Table 3. Specifically, the input and output training data are scaled by the maximum value of each respective state such that all the normalized states fall in the range between 0 and 1. The input layer is densely connected to 64 neurons in the hidden layer using the Rectified Linear Unit (ReLu) activation function, as defined in eq 2. The hidden layer is densely connected to the output layer using the Softplus activation function, $S(x) = \log(1 + e^x)$. Both the ReLu and Softplus functions are used to restrict the output predictions to be strictly non-negative and introduce nonlinearity to the model.

Table 2. Input States of the FNN Model

index	input states	units
1	applied potential	V vs the standard hydrogen electrode (V vs SHE)
2	rotation speed	rpm

Table 3. Output States of the FNN Model

index	output states	chemical formula
1	methane production rate	CH_4
2	ethylene production rate	C_2H_4
3	methanol production rate	CH_3OH
4	ethanol production rate	C_2H_5OH
5	acetate production rate	CH_3COO^-
6	ethylene glycol production rate	$(CH_2OH)_2$
7	glycolaldehyde production rate	$HOCH_2CHO$
8	acetaldehyde production rate	CH_3CHO
9	*n*-propanol production rate	C_3H_7OH
10	allyl alcohol production rate	C_3H_5OH
11	acetone production rate	CH_3COCH_3
12	propionaldehyde production rate	C_2H_5CHO
13	carbon monoxide production rate	CO
14	selectivity	

$$\mathrm{ReLu}(z) = \begin{cases} z & \text{for } z > 0 \\ 0 & \text{for } z \le 0 \end{cases} \quad (2)$$

Remark 1. *A single hidden layer is used for this model because it is the simplest structure to sufficiently capture the data trends. In addition, we apply a grid search for the number of neurons in the FNN, with 64 neurons having the best prediction. Specifically, neural networks with fewer than 64 neurons underfit the data, while networks with more neurons would overfit the data. In this work, both the prediction accuracy (in terms of mean-squared-error) and the output trajectories are considered to design the hyperparameters of the FNN model. Classical hyperparameter tuning algorithms did not perform effectively to capture reasonable trajectories, because of the difficulty of developing an explicit formula to evaluate the prediction trends. However, hyperparameter tuning algorithms, such as Bayesian optimization and random forest methods, are powerful tools to optimize the neural network structure.*[28] *We recommend that other users consider using those methods to develop their machine learning models.*

Data Generation and Dataset. As listed in Table 3, the oxygenates considered are Outputs 3−12, and the hydrocarbons are Outputs 1 and 2 (methane and ethylene). Therefore, the selectivity defined in the previous section is calculated as follows:

$$\text{selectivity} := \frac{\sum_{i=3}^{12} y_i}{\sum_{i=1}^{2} y_i} \quad (3)$$

where y_i refers to the production rate of species i, as defined in Table 3. Data are collected for the range of potential and rotation speed within which the reactor will operate. Specifically, the potential is varied from −1.2 to −1.47 (V vs the standard hydrogen electrode (SHE)), and the rotation speed is varied from 100 rpm to 800 rpm. For the data collection process, the potentiostat is set to be at a constant potential, and the electrode is rotated at a constant angular speed. The reactor is allowed to operate at steady state for 20 min prior to the data collection. The reactor then operates continuously with gas product samples taken every 20 min and liquid samples collected and analyzed after experiments to determine the concentration of the 13 relevant products, followed by the calculation of selectivity from the results of each sample.

The sampling process is an 80 min experiment that constitutes one data point for each input and output states. The sampling is repeated three to four times to ensure the data are consistent over time and to obtain the statistical information for the experimental results under the same operating conditions. Specifically, this 80 min experiment is repeated over 100 times to generate the data library that covers the specified range of operating conditions. Subsequently, the data are grouped into a single data vector based on the similarity of the operating conditions to compute the mean and standard deviations. As a result, 21 data points with mean and standard deviation information are each collected from two to five independent experiments, depending on the availability of experimental data.

Design of the Experiment. The range of potentials is limited by the overall resistance of the electrochemical cell (between the working and the counter electrode). This issue is resolved in the second generation of the cell reported in ref 27 by removing the channel that connects the two chambers to shorten the distance between the two electrodes and increase the surface area of the ion-exchange membranes. However, in this work, the first generation of the reactor is used which is not able to apply potentials more negative than −1.47 V vs SHE. The

https://doi.org/10.1021/acs.iecr.1c04176
Ind. Eng. Chem. Res. 2022, 61, 8399−8410

potential range is chosen to see appreciable rates of product generation considering the detection limits of the sensors (GC and NMR). The maximum rotation speed possible is 2000 rpm, as provided by the vendor of the RCE (Pine Research Instrumentation).

On the other hand, we have restricted the maximum electrode rotation speed to 800 rpm, mainly because of the mechanical instability of the custom-machined parts of the electrochemical cell. In addition, the chosen range of rotation speed is appropriate for studying mass transport effects from the perspective of mass transport characteristics around the RCE. As shown in ref 27, the film mass-transfer coefficient decreases as the electrode rotation speed increases with a 0.59 order dependency. The further increase in the rotation of the electrode beyond 800 rpm has a minimal effect on the mass-transport properties of the cell. The lower-bound of the rotation speed range is 100 rpm, below which the relationship between the film mass-transfer coefficient and the rotation speed starts to flatten out due to the convection created by the bubbling of CO_2 in the bulk of the electrochemical cell.

Standard FNN Training. The mean-squared-error (MSE) is used in the standard FNN training as the loss function that minimizes the difference between the experimental data value and the model predictive value. The MSE loss function is given below:

$$\text{loss} = \frac{1}{d}\frac{1}{m}\sum_{i=1}^{d}\sum_{j=1}^{m}\left|y_{i,j} - \hat{y}_{i,j}\right|^2 \tag{4}$$

where d and m are the number of data points in the training dataset and the number of output states. Specifically, from the original 21 data points, 4 are reserved for testing and the remaining 17 are used for training. Then, the 17 points are randomly split into training and validation sets with 80% used for training and 20% used for validation. The testing procedure compares the mean-squared difference between the FNN prediction and the testing data, using the loss function of eq 4 to evaluate the model performance. During this process, the parameter vector $\mathbf{W}$, which contains all the weights and bias of the neural network, is optimized using eq 5 to minimize the loss function.

$$\mathbf{W} = \mathbf{W} - \eta\frac{V_{dw}}{\sqrt{S_{dw}} + \epsilon} \tag{5}$$

where η is the learning rate, and ϵ is a small positive number to prevent the denominator being zero. V_{dw} and S_{dw} respectively introduce the momentum and root-mean-square factors of the parameters gradient to facilitate the optimization process. In practice, ϵ, V_{dw}, and S_{dw} can be set up by the machine learning API (e.g., Keras) automatically by specifying the optimizer. Tuning the value of ϵ will not have a significant impact on the model performance. In addition, the user can tune the learning rate (η) to improve the model performance. Usually, it is a small positive real value in the range between 0.0 and 1.0.

MAXIMUM LIKELIHOOD ESTIMATION IN ML REACTOR MODELING

Despite the standard FNN's capability of correlating the input and output variables of a complex nonlinear process, it treats all the data points equally, which might lead to overfitting when the data contains inconsistent levels of random error from the experimental data. To address this issue, the MLE method, originally developed by R. A. Fisher in the 1920s, is integrated in the FNN model to optimize the parameter set that maximizes the likelihood function of a probabilistic model.[29] Specifically, the likelihood function, $\mathcal{L}(\cdot)$, is used to correlate an unknown parameter vector (θ) with a random variable set (z) based on its probability-density function, $f(z,\theta)$. The maximum likelihood method can search for an optimum parameter set θ^* by maximizing the "likelihood of the sample", $\prod_{i=1}^{n} f(z, \theta)$, and it has been proved that this method can provide a solution to this optimization problem.[30] The MLE method assumes that the data are from a single population with the same standard deviation. However, this section proposes a modification that assumes each set of input parameters corresponds to a different population. Thus, each data point with its collected standard deviation is treated as an independent random variable.

To apply this method in our study, we first consider the experimental dataset to be a pseudoprobabilistic sample following the Gaussian distribution with an associated standard deviation. Therefore, the FNN outputs $\hat{y}_{i,j}$ must follow the same distribution as the reference data $y_{i,j}$, which means the joint likelihood of the neural network output is of Gaussian distribution, and can be expressed as follows:

$$\begin{aligned}\mathcal{L}(\mathbf{X};\mathbf{W},\sigma) &= \prod_{k=1}^{d\times m} f_{\mathbf{Y}}(y_k)\\ &= \prod_{k=1}^{d\times m}(2\pi\sigma_k^2)^{-0.5}\times\exp\left[-\frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{m}\left|\frac{y_{i,j}-\hat{y}_{i,j}(\mathbf{X},\mathbf{W})}{\sigma_{i,j}}\right|^2\right]\end{aligned} \tag{6}$$

where σ_i is the standard deviation for each data point. Subsequently, we find the optimum weight matrix $\mathbf{W}^*$ by maximizing the logarithm of the joint likelihood function:

$$\begin{aligned}\mathbf{W}^* &:= \arg\max_{\mathbf{W}} \log \mathcal{L}(\mathbf{X};\mathbf{W},\sigma)\\ &= \arg\max_{\mathbf{W}}\left(-\frac{1}{2}\sum_{k=1}^{d\times m}\log(2\pi\sigma_k^2) - \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{m}\left|\frac{y_{i,j}-\hat{y}_{i,j}(\mathbf{X},\mathbf{W})}{\sigma_{i,j}}\right|^2\right)\\ &= \arg\max_{\mathbf{W}}\left(-\sum_{k=1}^{d\times m}\log(2\pi\sigma_k^2) - \sum_{i=1}^{d}\sum_{j=1}^{m}\left|\frac{y_{i,j}-\hat{y}_{i,j}(\mathbf{X},\mathbf{W})}{\sigma_{i,j}}\right|^2\right)\end{aligned} \tag{7}$$

Since the first term of eq 7 is independent of $\mathbf{W}$, the maximum likelihood estimation of this model can be further simplified into eq 8.

$$\mathbf{W}^* = \arg\min_{\mathbf{W}}\left(\sum_{i=1}^{d}\sum_{j=1}^{m}\left|\frac{y_{i,j}-\hat{y}_{i,j}(\mathbf{X},\mathbf{W})}{\sigma_{i,j}}\right|^2\right) \tag{8}$$

The maximum likelihood estimation FNN model (MLE-FNN) is constructed using the same architecture and dataset as the standard FNN. However, the MLE-FNN model considers the standard deviation of each data point in its training process. Specifically, the sample standard deviation is calculated for each data point. The coefficient of variance (v) of each data point then is determined by the ratio of standard deviation and the respective output mean. This normalizes the data variability to

https://doi.org/10.1021/acs.iecr.1c04176
Ind. Eng. Chem. Res. 2022, 61, 8399−8410

allow for unbiased comparison between quantities of different magnitudes. The loss function, shown as eq 9, integrates eq 4 and eq 8. Thus, the weight matrix of the MLE-FNN is optimized to maximize both the accuracy of the prediction and the likelihood function during the training process.

$$\text{loss} = \frac{1}{d}\frac{1}{m}\sum_{i=1}^{d}\sum_{j=1}^{m}\frac{1}{v_{i,j}^{\ 2}}\left|y_{i,j} - \hat{y}_{i,j}\right|^2 \tag{9}$$

Remark 2. *In this study, error bars are constructed to represent the region of one standard deviation of uncertainty with respect to the mean, which is a ∼70% confidence interval for Gaussian distributed variables. Any statistic model can be used to develop a MLE-FNN model if it can provide reasonable statistical information on the experimental observations.*

Remark 3. *As shown in refs 31 and 32, the simplified log-likelihood function (eq 8) can be used directly as the loss function of a MLE-FNN model, since it contains the sum of squared error (SSE) in the loss function. We integrate it with eq 4 to demonstrate its similarity to the mean-squared error (MSE) loss function.*

Remark 4. *Bayesian optimization is another acknowledged method to develop a statistical machine learning model. Similar to the MLE method, the Bayesian optimization also considers the likelihood function model, which can account for data variance. Instead of focusing on the likelihood function, the Bayesian method implements optimization based on the posterior distribution of the machine learning model, which is defined by the Bayes' rule.*[33] *Therefore, the prior distribution of the parameter vector ($p(\theta)$) and the marginal likelihood of the observed data ($p(D)$) can be adopted to develop the statistical model.*

MACHINE LEARNING MODEL RESULTS AND ANALYSIS

In this section, we first compare the prediction performance of the standard FNN and MLE-FNN models. Subsequently, the MLE-FNN model's ability to capture the physical phenomenon behind the experiment is demonstrated through a comparison with a classical, EFP model. In addition, we propose an algorithm to improve the empirical model performance using the neural network model results and insights. Parameters used to generate the EFP models are described in this section and listed in Table 4.

FNN vs MLE-FNN. We first compare the performance of the MLE-FNN against the standard FNN. To account for the stochastic nature during the neural network training process, a Python script is used to train 100 FNN models in parallel with the structure discussed in the section entitled "Development of Machine Learning Model" and with randomly partitioned training and validation sets. The best FNN and MLE-FNN are chosen to minimize the MSE for the training dataset. This training method ensures that the selected models are trained consistently, following the same criteria. The selected FNN and MLE-FNN models then are evaluated with respect to the testing dataset, using the MSE between the normalized FNN outputs and the normalized testing set. The MSEs for the standard FNN and MLE-FNN are 0.0751 and 0.0791, respectively, which demonstrates a slight better performance of the standard FNN. Figure 3 shows that both models give accurate predictions across the majority of the data points, but the overall MSE for the MLE-FNN prediction increases, since it ignores the data points with high variance. However, the MSE of the two methods are sufficiently small, which implies that both models capture the input−output relationship well.

Table 4. Process Parameters for EFP Models with Units

parameter	value	units
EFP Model (Limiting Conditions)		
k^0	2.32×10^{-12}	cm s^{-1}
α	0.5	
F	96485	C mol^{-1}
R	8.314	J mol^{-1} K^{-1}
T	298	K
$E^{0\prime}$	−0.52	V
C_{CO_2}	3.40×10^{-5}	mol cm^{-3}
D_{CO_2}	1.91×10^{-5}	cm^2 s^{-1}
d_{RCE}	1.2	cm
ν_{H_2O}	1.03×10^{-2}	cm^2 s^{-1}
EFP Model		
$k_{0,5}$	2.02×10^{-28}	mol cm^{-1} s^{-2}
$k_{0,6}$	7.47×10^{-32}	mol cm^{-1} s^{-2}
$k_{0,7}$	2.61×10^{-13}	$mol^{1/2}$ cm $s^{-1/2}$
α_5	0.7	
α_6	0.85	
α_7	0.665	
EFP Model (Updated)		
$k_{0,5}$	7.2×10^{-22}	mol cm^{-1} s^{-2}
$k_{0,6}$	1.6×10^{-17}	$mol^{1/4}$ $cm^{3/2}$ $s^{5/4}$
$k_{0,7}$	9.5×10^{-23}	mol cm^{-1} s^{-2}
α_5	0.42	
α_6	0.53	
α_7	0.49	

To further compare the performance of the two models, the predictions for CO production rate are compared to some labeled outlier points, because of a slight drift in operating conditions in Figure 4. As shown in the figure, the MLE-FNN weighs the data points with critical experimental uncertainty less while the standard FNN overfits these points. This demonstrates the ability of the MLE-FNN model to improve its prediction by accounting for data variance. The goal of the MLE method is to generate models with a higher statistical significance that are suitable to be implemented with an experimental dataset. The MLE-FNN model demonstrates that it can provide an accurate approximation of the experimental data while outperforming the standard FNN in its ability to mitigate the impact of experimental uncertainty. Therefore, to simplify the discussion, the MLE-FNN will be used in the remainder of this section to provide comprehensive selectivity predictions for the electrochemical reactor, which is shown in Figure 5, and for comparison with other models (i.e., henceforth, the FNN will only refer to the MLE-FNN, and the standard FNN will not be included).

Remark 5. *The outlier points are included in the dataset, since they are valid but have higher variability. The predictive models are developed based on the experimental observation, even if some points are less likely to be reproduced. On the other hand, invalid data points from a failed experiment should not be included in the dataset.*

EFP Model vs MLE-FNN. First-principles models (FP models) are a fundamental approach to describing the operation of an electrochemical reactor, according to the energy and mass balances, as well as reaction kinetics. However, because of the complex mass transfer and reaction mechanisms of this process, it is challenging to obtain an accurate first-principles model. As a substitute, machine learning modeling provides an alternative approach to representing the physiochemical phenomena in the reactor with a desired prediction accuracy. In this subsection, an

https://doi.org/10.1021/acs.iecr.1c04176
Ind. Eng. Chem. Res. 2022, 61, 8399−8410

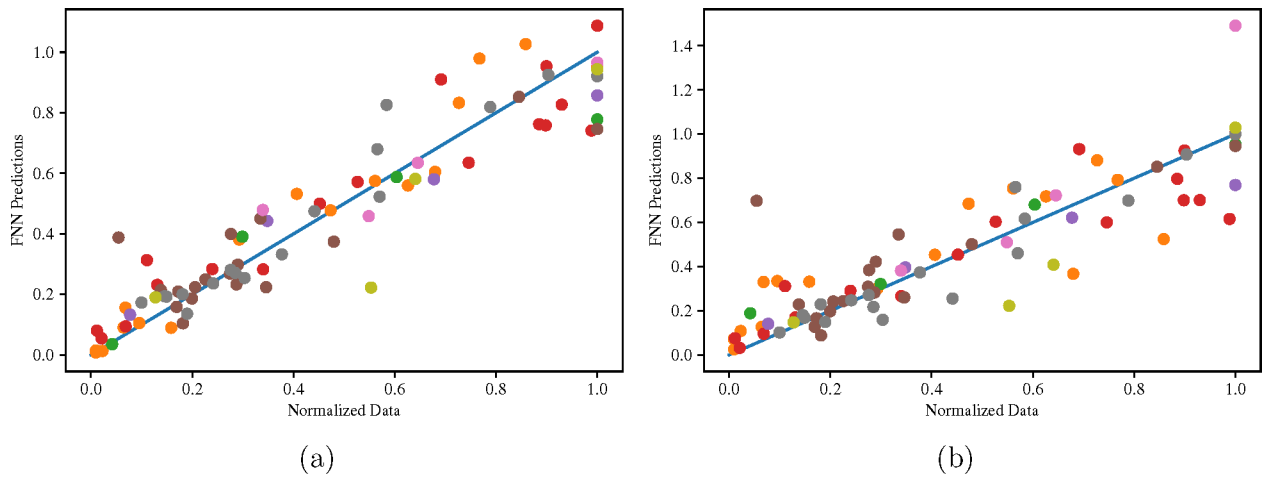


Figure 3. Comparison between the observed experimental outcome and the neural network predictions from (a) standard FNN and (b) MLE-FNN models.

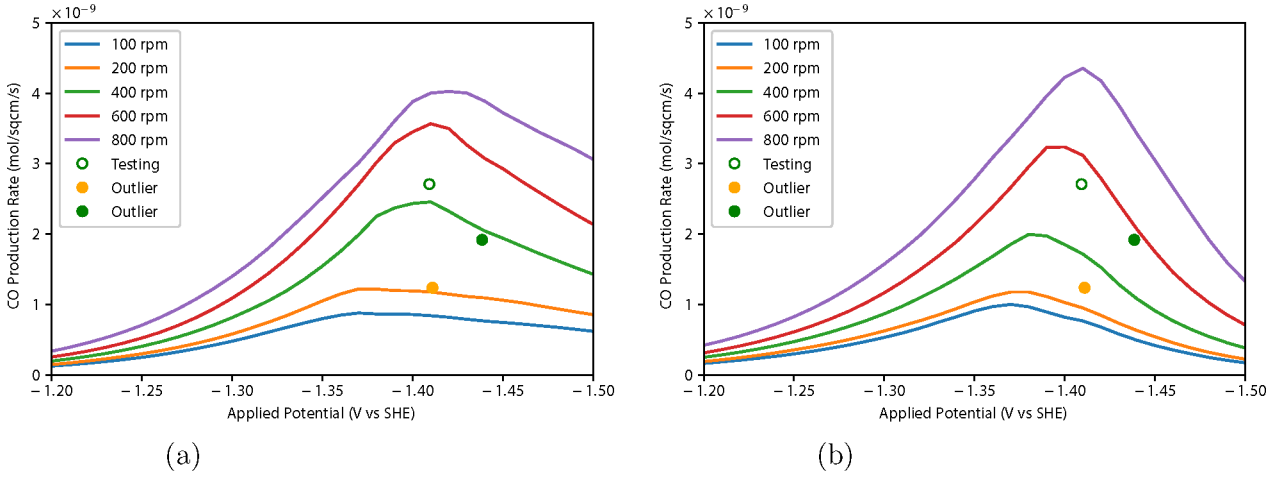


Figure 4. CO production rate predictions for various applied potentials, in units of V vs the standard hydrogen electrode (V vs SHE). The solid points are labeled uncertain data as having a drift in potential. The open point is from the testing set. (a) The CO prediction of standard FNN model overfitted the labeled uncertain data points. (b) The MLE-FNN model successfully learned the experimental uncertainty and provide prediction accordingly, but this feature introduces extra error to the testing results.

EFP model of a rotating electrode reactor is developed following the derivation in ref 34 to determine the rate of CO production under limiting conditions. Specifically, this model assumes that only a single, first-order reaction is occurring with no side reactions, and the reaction occurs only on the electrode surface following Butler−Volmer kinetics, which means it cannot capture the comprehensive kinetics of this experiment. The resulting equation is given as follows:

$$r_{CO} = \frac{k_f C^*_{CO_2}}{1 + \left(k_f / m_0\right)} \tag{10}$$

where $C^*_{CO_2}$ is the bulk concentration of CO_2, k_f is the kinetic rate constant, and m_0 is the convective mass-transfer coefficient. The rate constant k_f changes based on the applied potential, and is calculated as follows:

$$k_f = k^0 \exp\left[-\frac{\alpha F}{RT}(E - E^{0\prime})\right] \tag{11}$$

where k^0 is the standard rate constant, α the symmetry factor, F the Faraday constant, R the gas constant, T the temperature, E the applied potential, and $E^{0\prime}$ the standard reduction potential. The mass-transfer coefficient m_0 is determined based on the rotation speed of the electrode, but this correlation will change depending on the type of rotating electrode. For some simple rotating electrode geometries such as a flat disk, a mass-transfer coefficient is determined analytically, assuming a linear velocity profile in the boundary layer. However, the electrode used in this experiment has a cylindrical geometry, which is more complicated, so the mass-transfer coefficient is determined experimentally from the Sherwood number correlation as follows:[27]

$$m_0 = 0.204 {Re_{RCE}}^{0.59} {Sc}^{0.33} \frac{D_{CO_2}}{d_{RCE}} \tag{12}$$

where Re_{RCE} is the Reynolds number, Sc the Schmidt number, D the diffusion coefficient, and d_{RCE} the diameter of the RCE. The diffusion coefficient is assumed to be the same for the reactant and product species for simplicity. Since the Sherwood number is determined experimentally, this model will be referenced as an EFP model.

https://doi.org/10.1021/acs.iecr.1c04176
Ind. Eng. Chem. Res. 2022, 61, 8399−8410

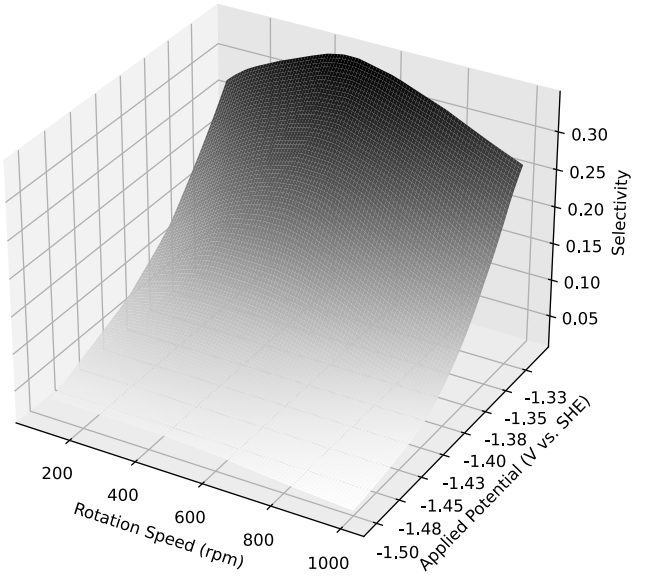


Figure 5. Selectivity of oxygenate species, with respect to rotation speed and applied potential (in the unit of V vs the standard hydrogen electrode (V vs SHE)), as predicted by the machine learning model.

The comparison between this EFP model and the FNN prediction is shown in Figure 6. As shown, the EFP model trajectory is similar to the FNN prediction under the operating conditions with less negative applied potentials and lower rotation speeds, because the side reactions are limited under these conditions. After switching to more-negative conditions, the EFP model's assumption becomes invalid. Therefore, these two models present different predictions after passing threshold conditions. This comparison demonstrates that the neural network can correctly capture the input−output relations from the experimental data.

EFP Model Improvement. EFP modeling is an efficient way to determine out how several experimental variables affect the experimental data without requiring complete knowledge of the underlying physical phenomena needed for large-scale FP models.[35] Although the neural network has demonstrated its ability to capture steady-state behavior of the electrochemical reactor, an empirical model with an explicit form is essential to improve the reactor phenomena understanding. However, parameter tuning and selection for an empirical model are challenging. Therefore, we propose an algorithm to use neural network model results to improve the EFP model structure.

Specifically, we first develop an EFP model consisting of several regression problems that predict the production rate of seven different classes of species produced in the reactor. This model is derived utilizing the same reaction kinetics and transport phenomena considerations mentioned in the section entitled "EFP Model vs. MLE-FNN". As shown in eq 13, the empirical regressions of interest are the production rate of C_1 products (FNN Outputs 1 and 3 listed in Table 3), C_{2+} hydrocarbons (FNN Output 2), and C_{2+} oxygenates (FNN Outputs 4, 5, 6, 7, 8, 9, 10, 11, and 12), denoted as r_{C_1}, $r_{C_{2+,HC}}$, and $r_{C_{2+,OX}}$, respectively.

$$r_{C_1} = k_{0,5} C_{CO} Sh_{RCE}^{-0.5} J_{HCO_3}^{-1} \exp\left(-\frac{\alpha_5 z_5 F}{RT} E\right)$$

$$r_{C_{2+,HC}} = k_{0,6} C_{CO} Sh_{RCE}^{-0.5} J_{HCO_3}^{-1} \exp\left(-\frac{\alpha_6 z_6 F}{RT} E\right)$$

$$r_{C_{2+,OX}} = k_{0,7} C_{CO} Sh_{RCE}^{-0.5} J_{HCO_3}^{0.5} \exp\left(-\frac{\alpha_7 z_7 F}{RT} E\right) \quad (13)$$

The notation C_{CO} is the concentration of carbon monoxide, and J_{HCO_3} is the flux of bicarbonate, both of which are calculated at the inner Helmholtz plane based on the bulk concentration, rotation speed, and applied potential. In addition, Sh_{RCE} is the Sherwood number of the rotating-cylinder electrode, which relates directly to the rotation speed. The rate constants $k_{0,i}$ and symmetry factors α_i are obtained by linearizing the equations for r_{C_1}, $r_{C_{2+,HC}}$, and $r_{C2+,OX}$, with respect to the applied potential. Furthermore, since these rate expressions each describe multiple products from different reaction steps, the number of electrons z_i

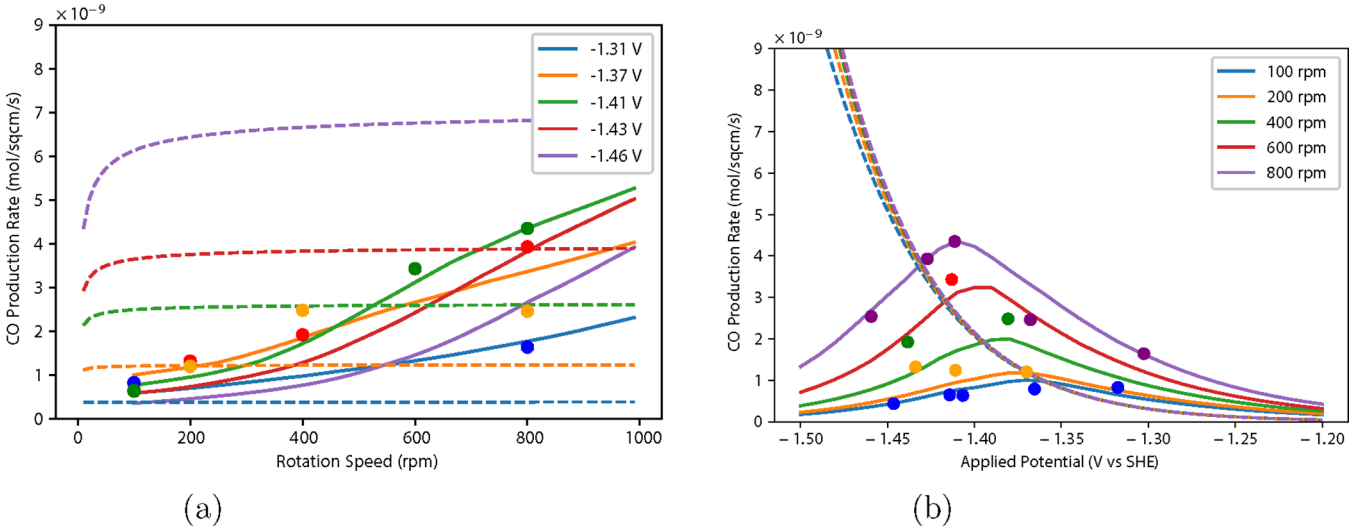


Figure 6. CO production rates for the first EFP model (dashed) and the MLE-FNN model predictions (solid), compared with the training data points over the range of (a) rotation speed and (b) applied potentials (given in units of V vs the standard hydrogen electrode (V vs SHE)). This EFP model can capture the general trend of the reactor for low applied potential and rotation speed. However, for the more negative potential and higher rotation speed, the initial assumption of the EFP model becomes invalid.

https://doi.org/10.1021/acs.iecr.1c04176
Ind. Eng. Chem. Res. 2022, 61, 8399−8410

is not a single value, and a modification is necessary for this case. Specifically, the parameter α_i is considered to be a fitting parameter for the exponential relationship between the potential and the rates. Thus, by fixing z_i to 1, α_i becomes an arbitrary positive value that can be optimized in the regression problem.

Subsequently, the EFP model is compared with the FNN model using the testing set as described in the section enetitled "Data Generation and Dataset". Table 5 shows that the FNN model outperforms the EFP model with significantly lower MSEs for all three rates, which implies that the accuracy of the EFP model can be improved by minimizing the deviation between FNN prediction and EFP model prediction. In other words, the FNN prediction can be considered as additional reference data to improve the EFP model performance for this reactor. In addition, by comparing the prediction trends from both models in Figure 7, the EFP model overestimates the effect of applied potential for higher rotation speed. Thus, the empirical model can be improved by modifying the existing terms. The process is summarized by Algorithm 1.

Table 5. Testing Data MSE Results of the FNN Model and the Empirical, First-Principles Model

rate index	FNN	EFP
$r_{C_{2+},OX}$	0.0239	0.042
$r_{C_{2+},HC}$	0.0098	0.042
r_{C_1}	0.003	0.021

Specifically, the regression parameters (θ_i) can be optimized to minimize the difference between the two models using a user-defined optimization algorithm in each iteration of the procedure. In addition, new system parameters, C_i, can be introduced to further develop the empirical model. For example, terms that describe the influence of gas pressure and flow rate on the electrochemical reaction can be included to further improve the current empirical model. Therefore, by following this procedure, the improvement of an empirical model can be represented as an optimization problem, which can be accomplished automatically by computers.

To demonstrate this procedure, we use the FNN to tune the parameters for the proposed EFP model. The difference between the EFP and FNN models is minimized in the range of −1.47 V to −1.30 V, since this is the range for the data collected to train the FNN model. The MSE between the EFP

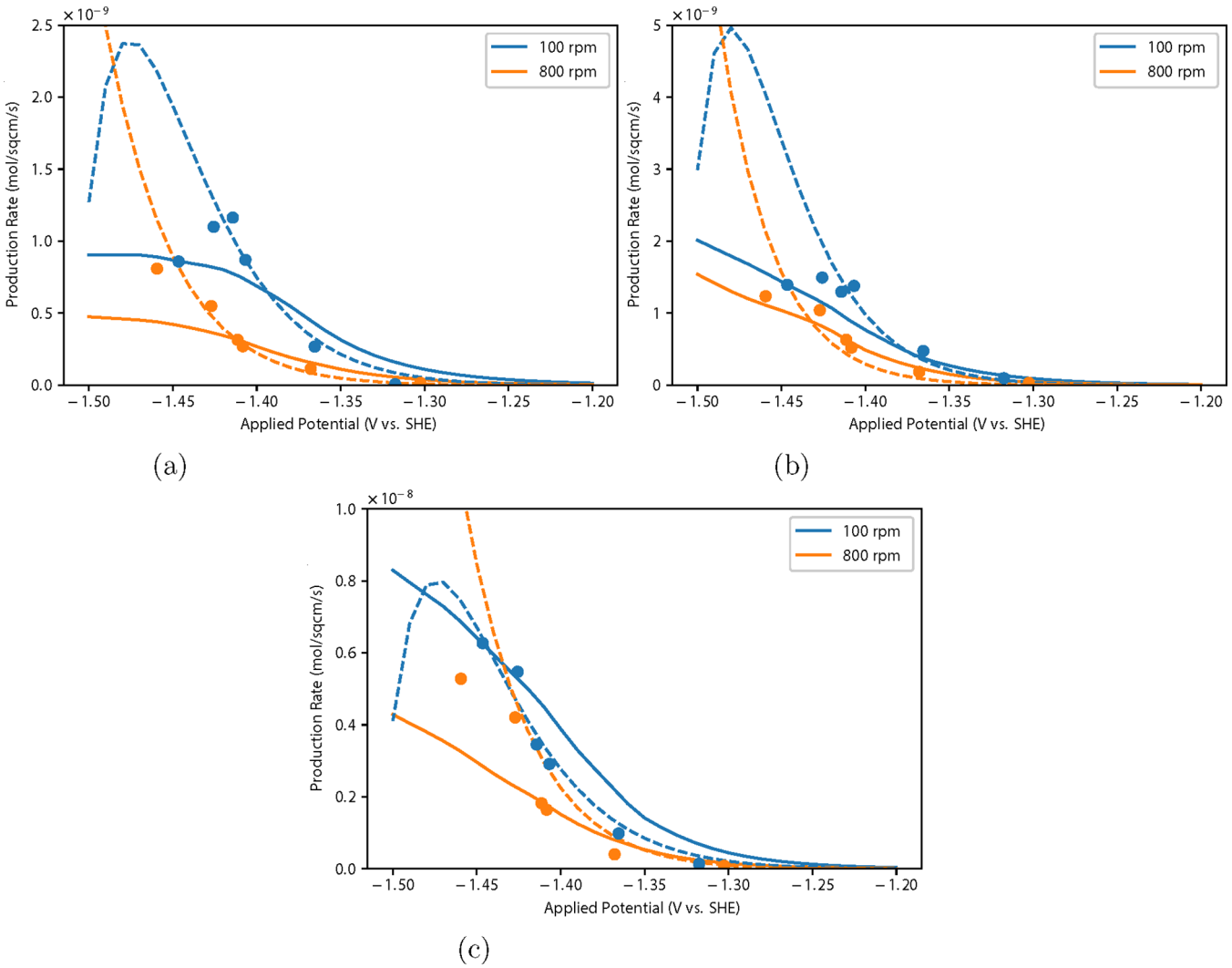


Figure 7. Production rates of (a) r_{C_1}, (b) $r_{C_{2+},HC}$, and (c) $r_{C_{2+},OX}$ from the EFP model (dashed) and the MLE-FNN model (solid), compared with the reference data points over the range of applied potentials (given in units of V vs the standard hydrogen electrode (V vs SHE)).

https://doi.org/10.1021/acs.iecr.1c04176
Ind. Eng. Chem. Res. 2022, 61, 8399−8410

Algorithm 1: Empirical, First-Principles Model Improvement Procedure

X is the input data sequence, Θ contains the regression parameters in the empirical model to be optimized, C represents any additional system parameters that can be added to the empirical model, E and F_{ANN} are the general equations for the empirical and FNN models respectively, $\mathbf{D}$ calculates the distance, and I_{max} is the maximum number of iterations.

for $i = 0$ **to** I_{max} **do**

Obtain the distance between the empirical model and the FNN model:

$$\mathbf{D}(\Theta, X, C) = \sum [E(\Theta, X, C) - F_{ANN}(X)]^2$$

if $\mathbf{D}(\Theta, X, C_{new}) < \mathbf{D}(\Theta, X, C)$ **then**

$C = C_{new}$

else

$C = C$

end

Optimize the regression parameters to minimize the distance:

$$\Theta_{new} = \{\theta_{i,new} \in \Theta \mid D(\Theta_{new}, X, C) < D(\Theta, X, C)\}$$

end

model and FNN is calculated at intervals of 0.01 V. For the model of eq 13, we apply a grid search to find the optimum values for the parameters k_i, α_i, and the exponent of the flux term J_{HCO_3}. Specifically, we search the α_i values from 0 to 1 with step sizes of 0.01, and the J_{HCO_3} exponents from −1 to 1 with step sizes of 0.25. However, small changes in the α_i and exponent of J_{HCO_3} can cause the k_i to change by several orders of magnitude, which caused problems when attempting to use nonlinear optimization packages. Therefore, the grid search for k_i must cover a large range of magnitudes from 0 to 1. To decrease the computational complexity, the grid search for k_i is split into two subsequent searches. Given that k_i can be expressed in the form of $A \times 10^{-B}$, we first search for the optimum order of magnitude B, and then search for the optimum number A. Specifically, the first search follows a geometric sequence from 10^{-25} to 10^{-10} with a geometric ratio of 10 (i.e., 10^{-25}, 10^{-24}, 10^{-23}, ..., 10^{-10}). Then, given the optimum order of magnitude B, we search for the optimal number from 1 to 10 with step size of 0.1 (i.e., 0.1×10^{-B}, 0.2×10^{-B}, 0.3×10^{-B}, ..., 10.0×10^{-B}) for the order of magnitude B and $B + 1$.

Following this procedure, the optimum parameters are determined and listed in Table 4. Then, the new EFP model is tested against the same reference data points, and its MSE results are compared with the original EFP model listed in Table 6. As a result, the MSE for r_{C_1}, $r_{C_{2+,HC}}$, and $r_{C_{2+,OX}}$ against the reference dataset decreases by 75%, 79%, and 32%, respectively. The new empirical, first-principles equations are updated in eq 14 to reflect the changes made from the procedure. Moreover, the new EFP model predictions are also shown in Figure 8, which shows that the overestimating problem is solved with new parameters.

Table 6. Nonscaled MSE for the Updated and Original EFP Models

rate index	EFP (original)	EFP (updated)
$r_{C_{2+,OX}}$	3.68×10^{-19}	9.20×10^{-20}
$r_{C_{2+,HC}}$	1.11×10^{-18}	2.32×10^{-19}
r_{C_1}	8.46×10^{-18}	5.76×10^{-18}

$$r_{C_1} = k_{0,5} C_{CO} Sh_{RCE}^{-0.5} J_{HCO_3}^{-1} \exp\left(-\frac{\alpha_5 z_5 F}{RT} E\right)$$

$$r_{C_{2+,HC}} = k_{0,6} C_{CO} Sh_{RCE}^{-0.5} J_{HCO_3}^{-0.25} \exp\left(-\frac{\alpha_6 z_6 F}{RT} E\right)$$

$$r_{C_{2+,OX}} = k_{0,7} C_{CO} Sh_{RCE}^{-0.5} J_{HCO_3}^{-1} \exp\left(-\frac{\alpha_7 z_7 F}{RT} E\right) \quad (14)$$

Remark 6. *Minimizing the difference between the two models will not result in exactly the same predictive model. During the optimization process, the empirical model structure derived from physical relations should remain unaltered. Furthermore, the additional terms C_i that have not been included in the previous empirical models should have physical meanings. In this way, the empirical model is modified toward a lower MSE for its prediction, while respecting the physics of the experiment.*

Remark 7 *The experimental data is used to calculate the original EFP parameters. Specifically, the original EFP model is determined using traditional methods to extract kinetic parameters of α_i and k_i. Since the reaction rate is proportional to the exponential of the applied potential, the relationship is linearized by plotting the natural logarithm of the reaction rate against the applied potential. A linear regression is then used to find the slope and intercept of the observed data. The value of α_i is then extracted from the slope, and k_i is extracted from the intercept. Furthermore, by using the FNN model to propose an updated EFP model, meaningful process parameters can be extracted from the neural network regression, which provides additional explicit values to evaluate the neural network performance.*

CONCLUSION

This work demonstrated the application of neural network modeling to an electrochemical reactor to capture input−output relationships of key operating variables, and developed a method to improve EFP models using neural network model results.

https://doi.org/10.1021/acs.iecr.1c04176
Ind. Eng. Chem. Res. 2022, 61, 8399−8410

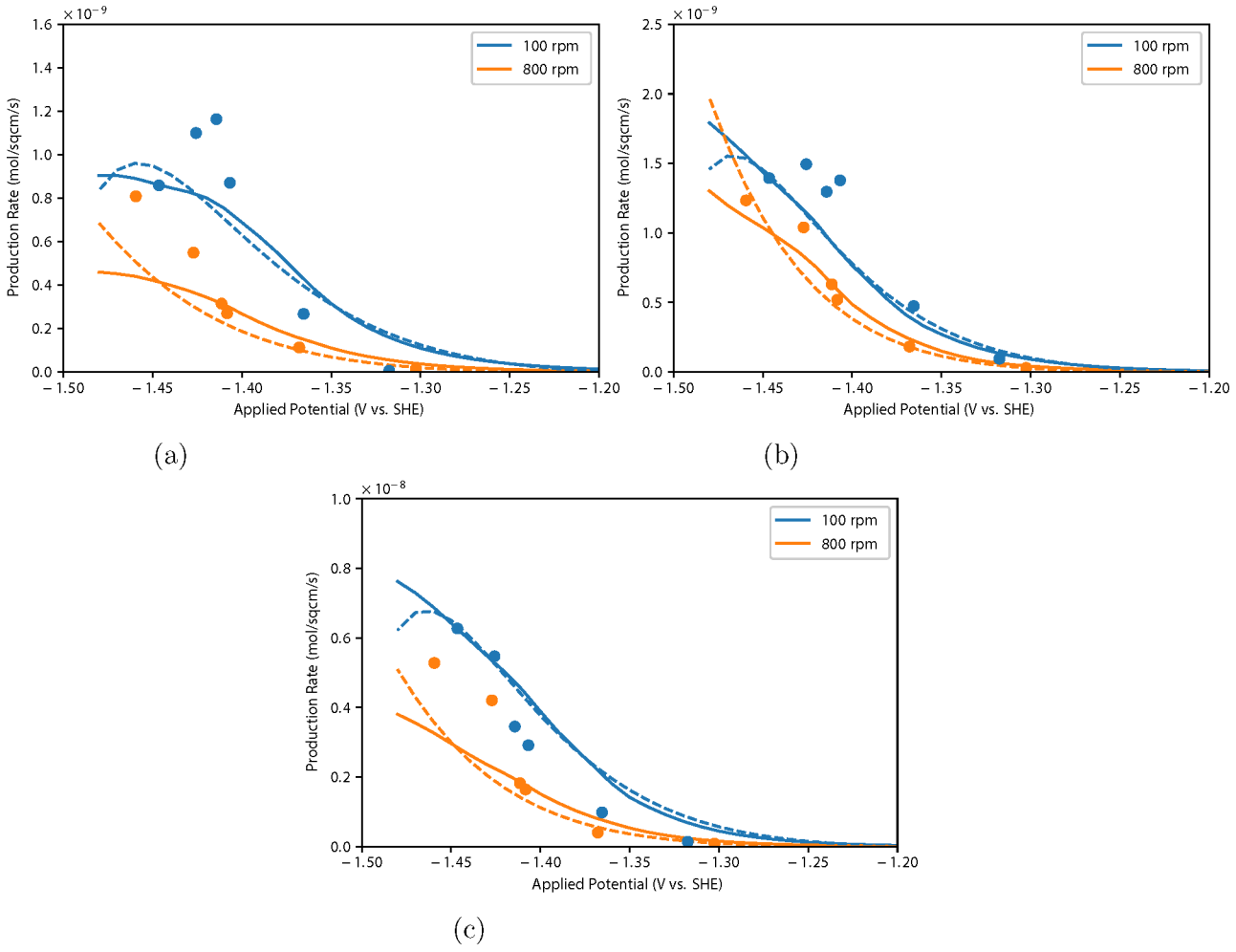


Figure 8. Production rates of (a) r_{C_1}, (b) $r_{C_{2+,HC}}$, and (c) $r_{C_{2+,OX}}$ from the updated EFP model (dashed) and the MLE-FNN model (solid), compared with the reference data points over the range of applied potentials (given in units of V vs the standard hydrogen electrode (V vs SHE)).

Since the electrochemical reduction of carbon dioxide and the overall chemical reaction pathways are not well-understood, the kinetic and empirical, first-principles models cannot fully capture the physiochemical phenomena of the reactor. To address this issue, an FNN model was developed to model the experimental reactor data over a broad range of operating conditions. In addition, statistical FNN models were developed by utilizing maximum likelihood estimation method to account for the variability of experimental data, and their predictive performance was demonstrated over a broad range of operating conditions. Lastly, we developed an algorithm to improve the EFP model by utilizing the neural network model results, which decreased the MSE of the EFP model prediction for three reaction rates, by 75%, 79%, and 32%, respectively.

AUTHOR INFORMATION

Corresponding Author

Panagiotis D. Christofides − *Department of Chemical and Biomolecular Engineering and Department of Electrical and Computer Engineering, University of California, Los Angeles, California 90095, United States;* orcid.org/0000-0002-8772-4348; Email: pdc@seas.ucla.edu

Authors

Junwei Luo − *Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, California 90095, United States*

Vito Canuso − *Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, California 90095, United States*

Joon Baek Jang − *Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, California 90095, United States*

Zhe Wu − *Department of Chemical and Biomolecular Engineering, National University of Singapore, 117585, Singapore*

Carlos G. Morales-Guio − *Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, California 90095, United States;* orcid.org/0000-0002-5840-5591

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.iecr.1c04176

Notes

The authors declare no competing financial interest.

https://doi.org/10.1021/acs.iecr.1c04176
Ind. Eng. Chem. Res. 2022, 61, 8399−8410

■ ACKNOWLEDGMENTS

Financial support from the Department of Energy is gratefully acknowledged.

■ REFERENCES

(1) Morales-Guio, C.; Cave, E.; Nitopi, S.; Feaster, J.; Wang, L.; Kuhl, K.; Jackson, A.; Johnson, N.; Abram, D.; Hatsukade, T.; et al. Improved CO_2 reduction activity towards C_{2+} alcohols on a tandem gold on copper electrocatalyst. *Nat. Catal.* **2018**, *1*, 764−771.

(2) Nitopi, S.; Bertheussen, E.; Scott, S.; Liu, X.; Engstfeld, A.; Horch, S.; Seger, B.; Stephens, I.; Chan, K.; Hahn, C.; et al. Progress and perspectives of electrochemical CO_2 reduction on copper in aqueous electrolyte. *Chem. Rev.* **2019**, *119*, 7610−7672.

(3) Zadeh, L. A. Probability measures of fuzzy events. *J. Math. Anal. Appl.* **1968**, *23*, 421−427.

(4) Vepa, R. A review of techniques for machine learning of real-time control strategies. *Intell. Syst. Eng.* **1993**, *2*, 77−90.

(5) Lee, C. Fuzzy logic in control systems: fuzzy logic controller. I. *IEEE Trans. Syst.s, Man., Cybernetics* **1990**, *20*, 404−418.

(6) Whittle, P. Tests of fit in time series. *Biometrika* **1952**, *39*, 309−318.

(7) Shen, Q.; Jiang, B.; Shi, P.; Lim, C. Novel neural networks-based fault tolerant control scheme with fault alarm. *IEEE Trans. Cybernetics* **2014**, *44*, 2190−2201.

(8) Hopfield, J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U. S. A.* **1982**, *79*, 2554−2558.

(9) Wu, Z.; Tran, A.; Ren, Y.; Barnes, C.; Chen, S.; Christofides, P. D. Model predictive control of phthalic anhydride synthesis in a fixed-bed catalytic reactor via machine learning modeling. *Chem. Eng. Res. Des.* **2019**, *145*, 173−183.

(10) Wu, Z.; Tran, A.; Rincon, D.; Christofides, P. D. Machine learning-based predictive control of nonlinear processes. Part I: Theory. *AIChE J.* **2019**, *65*, No. e16729.

(11) Raccuglia, P.; Elbert, K.; Adler, P.; Falk, C.; Wenny, M.; Mollo, A.; Zeller, M.; Friedler, S.; Schrier, J.; Norquist, A. Machine-learning-assisted materials discovery using failed experiments. *Nature* **2016**, *533*, 73−76.

(12) Malek, A.; Wang, Q.; Baumann, S.; Guillon, O.; Eikerling, M.; Malek, K. A Data-Driven Framework for the Accelerated Discovery of CO_2 Reduction Electrocatalysts. *Front. Energy Res.* **2021**, *9*, 52.

(13) Galvelis, R.; Sugita, Y. Neural network and nearest neighbor algorithms for enhancing sampling of molecular dynamics. *J. Chem. Theory Comput.* **2017**, *13*, 2489−2500.

(14) Natarajan, S.; Behler, J. Neural network molecular dynamics simulations of solid−liquid interfaces: water at low-index copper surfaces. *Phys. Chem. Chem. Phys.* **2016**, *18*, 28704−28725.

(15) Wu, Z.; Luo, J.; Rincon, D.; Christofides, P. D. Machine learning-based predictive control using noisy data: evaluating performance and robustness via a large-scale process simulator. *Chem. Eng. Res. Des.* **2021**, *168*, 275−287.

(16) Ding, Y.; Zhang, Y.; Ren, Y.; Orkoulas, G.; Christofides, P. D. Machine learning-based modeling and operation for ALD of SiO_2 thin-films using data from a multiscale CFD simulation. *Chem. Eng. Res. Des.* **2019**, *151*, 131−145.

(17) Jadid, M.; Fairbairn, D. Neural-network applications in predicting moment-curvature parameters from experimental data. *Eng. Appl. Artific. Intell.* **1996**, *9*, 309−319.

(18) Bangi, M.; Kwon, J. Deep reinforcement learning control of hydraulic fracturing. *Comput. Chem. Eng.* **2021**, *154*, 107489.

(19) Bangi, M.; Kwon, J. Deep hybrid modeling of chemical process: Application to hydraulic fracturing. *Comput. Chem. Eng.* **2020**, *134*, 106696.

(20) Kimaev, G.; Ricardez-Sandoval, L. A. Artificial Neural Network Discrimination for Parameter Estimation and Optimal Product Design of Thin Films Manufactured by Chemical Vapor Deposition. *J. Phys. Chem. C* **2020**, *124*, 18615−18627.

(21) Miriyala, S.; Mittal, P.; Majumdar, S.; Mitra, K. Comparative study of surrogate approaches while optimizing computationally expensive reaction networks. *Chem. Eng. Sci.* **2016**, *140*, 44−61.

(22) Inapakurthi, R.; Miriyala, S.; Mitra, K. Deep learning based dynamic behavior modelling and prediction of particulate matter in air. *Chem. Eng. J.* **2021**, *426*, 131221.

(23) Inapakurthi, R.; Miriyala, S.; Mitra, K. Recurrent neural networks based modelling of industrial grinding operation. *Chem. Eng. Sci.* **2020**, *219*, 115585.

(24) Rangel-Martinez, D.; Nigam, K.; Ricardez-Sandoval, L. A. Machine learning on sustainable energy: A review and outlook on renewable energy systems, catalysis, smart grid and energy storage. *Chem. Eng. Res. Des.* **2021**, *174*, 414−441.

(25) Kimaev, G.; Ricardez-Sandoval, L. A. Nonlinear model predictive control of a multiscale thin film deposition process using artificial neural networks. *Chem. Eng. Sci.* **2019**, *207*, 1230−1245.

(26) Nishimura, Y.; Peng, H.; Nitopi, S.; Bajdich, M.; Wang, L.; Morales-Guio, C.; Abild-Pedersen, F.; Jaramillo, T.; Hahn, C. Guiding the Catalytic Properties of Copper for Electrochemical CO_2 Reduction by Metal Atom Decoration. *ACS Appl. Mater. Interfaces* **2021**, *13*, 52044−52054.

(27) Jang, J.; Rüscher, M.; Winzely, M.; Morales-Guio, C. Gastight Rotating Cylinder Electrode: Towards Decoupling Mass Transport and Intrinsic Kinetics in Electrocatalysis. Submitted to *AIChE J.* **2021**.

(28) Wu, J.; Chen, X.; Zhang, H.; Xiong, L.; Lei, H.; Deng, S. Hyperparameter optimization for machine learning models based on Bayesian optimization. *J. Electron. Sci. Technol.* **2019**, *17*, 26−40.

(29) Myung, I. Tutorial on maximum likelihood estimation. *J. Math. Psychol.* **2003**, *47*, 90−100.

(30) Huzurbazar, V. The likelihood equation, consistency and the maxima of the likelihood function. *Ann. Eugenics* **1947**, *14*, 185−200.

(31) Dorling, S.; Foxall, R.; Mandic, D.; Cawley, G. Maximum likelihood cost functions for neural network models of air quality data. *Atmos. Environ.* **2003**, *37*, 3435−3443.

(32) Kumar, M.; Garg, D.; Zachery, R. Intelligent sensor modeling and data fusion via neural network and maximum likelihood estimation. *Proc. ASME Int. Mech. Eng. Congress Expos.* **2005**, *42169*, 1759−1768.

(33) Shahriari, B.; Swersky, K.; Wang, Z.; Adams, R.; De Freitas, N. Taking the human out of the loop: A review of Bayesian optimization. *Proc. IEEE* **2016**, *104*, 148−175.

(34) Bard, A.; Faulkner, L. *Electrochemical Methods: Fundamentals and Applications*, 2nd Edition; John Wiley & Sons: New York, 2001.

(35) Box, G.; Draper, N. *Empirical Model-Building and Response Surfaces*; John Wiley & Sons: New York, 1987.

https://doi.org/10.1021/acs.iecr.1c04176
Ind. Eng. Chem. Res. 2022, 61, 8399−8410