Title
Machine Learning-Based Operational Modeling of an Electrochemical Reactor: Handling Data Variability for Experimental Data

Author
Canuso, Vito Anthony

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Machine Learning-Based Operational Modeling of an Electrochemical Reactor: Handling Data Variability for Experimental Data

A thesis submitted in partial satisfaction of the
requirements for the degree Master of Science
in Chemical Engineering

by

Vito Anthony Canuso IV

2022

ABSTRACT OF THE THESIS

Machine Learning-Based Operational Modeling of an Electrochemical Reactor:
Handling Data Variability for Experimental Data

by

Vito Anthony Canuso IV
Master of Science in Chemical Engineering
University of California, Los Angeles, 2022
Professor Panagiotis D. Christofides, Chair

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, due to the complexity of its reaction mechanism and the largely unknown electron-transfer pathways, the development of a first-principles-based operational model of an electrocatalytic CO_2 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. Additionally, 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. The insights obtained from the FNN model are applied to propose specific modifications to the empirical, first-principles model (EFP model) to improve its prediction capability and to propose optimal set points for future experiments based on the FNN predictions of operating cost and profit. The FNN model is also used as the system model to perform relative gain array analysis to determine controllability for multi-input-multi-output control schemes.

The thesis of Vito Anthony Canuso IV is approved.

Dante A. Simonetti

Carlos G. Morales-Guio

Panagiotis D. Christofides, Committee Chair

University of California, Los Angeles

2022

Contents

List of Figures

List of Tables

ACKNOWLEDGMENTS

I would like to thank my advisor Professor Panagiotis D. Christofides for his guidance and support during the course of my research.

I would like to thank Professors Dante A. Simonetti and Carlos G. Morales-Guio for reviewing my thesis and contributing to my Master's thesis committee.

Financial support from the Department of Energy is gratefully acknowledged.

This work was submitted for publication in *Industrial & Engineering Chemistry Research, ie-2021-041768* and is co-authored by Junwei Luo, Joon Baek Jang, Zhe Wu, Professor Carlos G. Morales-Guio, and Professor Panagiotis D. Christofides. I would like to acknowledge their contributions and the contributions of Berkay Citmaçi. I am grateful for all of their help.

Chapter 1

Introduction

The electrochemical transformation of carbon dioxide (CO_2) into carbon-based fuels and chemicals has received growing interest in this century due to its potential to reduce CO_2 emissions and facilitate the production of energy from renewable sources [20]. 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 [23]. The complexity of the electrocatalytic reduction pathway poses challenges when attempting to quantify and model the reaction kinetics. In addition, recent small-scale experiments on this process show varying levels of experimental uncertainty caused by the minimum measurable limit of the sensors and other inevitable experimental errors. This can introduce a level of uncertainty into the data, which can increase the difficulty of data analysis.

Although developing mathematical models, such as first-principle models, is a classical 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 and artificial intelligence (AI) approaches have been proposed historically. In the 1960s, Zadeh proposed an epochal AI logic, fuzzy logic, to approximate uncertain features [38].

From then on, machine learning techniques for real-time process operation were investigated in the 1990s [30], such as the expert system [16]. Additionally, auto-regressive models—such as the autoregressive-moving-average model (ARMA) proposed by Peter Whittle in the early 1950s [32]—provided statistical strategies to develop data-driven models based on recorded observations.

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 artificial neural network (ANN) models have demonstrated their ability to address regression and classification problems in the context of chemical process modeling [7, 18, 24, 28, 36, 37].

In recent years, ANNs have been used to model chemical engineering manufacturing processes in several studies. For example, Wu et al. used a recurrent neural network (RNN) model to model a chemical reactor using data from a large-scale process simulator, Aspen Plus dynamics. This RNN was used in a model predictive controller (MPC) to stabilize the process under specific steady-state operating conditions [35]. Additionally, a feed-forward neural network (FNN) model was developed by Ding et al. to correlate the input and output variables of a SiO_2 atomic layer deposition (ALD) process to calculate optimal half-cycle times to achieve full surface coverage, which is an important parameter for industrial ALD processes [5]. Furthermore, a methodology was discussed by Jadid and Fairbairn, wherein neural networks were used for parameter estimation from experimental data. Neural-network applications in predicting moment-curvature parameters from experimental data [11]. These research investigations provide strong support for the use of neural network models as a reliable approximation for analyzing complex nonlinear relationships from experimental data and simulation 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. Bangi and Kwon developed a deep reinforcement learning controller to control a hydraulic fracturing process to improve the safety and optimization of system operation [2]. In addition, an operational model was constructed for this process using a hybrid approach of a deep neural network and a first-principle model [1]. ANNs were further used to determine the optimal operating conditions for chemical and industrial processes [9, 10, 13, 14, 19, 25], which contributed to maximizing the feasibility of novel processes from economic and safety perspectives.

Motivated by the aformentioned considerations, this work develops an FNN regression model using steady-state, input-output data generated from an experimental electrochemical reactor. The model is determined by solving a nonlinear regression problem that accounts for data variability. This FNN model is computationally-efficient and can be used in real time to determine safe and energy-optimal operating conditions for the electrochemical reactor. 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 each data points' respective confidence interval and weighing each point accordingly in the FNN 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 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-principle 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 chapter "Preliminaries", the experimental setup of the reactor and the kinetics of the electrochemical reactions are described.

In the chapter “Development of the Machine Learning Model”, the formulation and construction methods of the FNN model are discussed. In the chapter “Maximum Likelihood Estimation in ML Reactor Modeling”, the methodology of the maximum likelihood estimation is integrated with the FNN modeling method. In the chapter “Machine Learning Model Results and Analysis”, the performance of the FNN models is evaluated, and the statistical FNN model predictions and insights are demonstrated for various applications in further research.

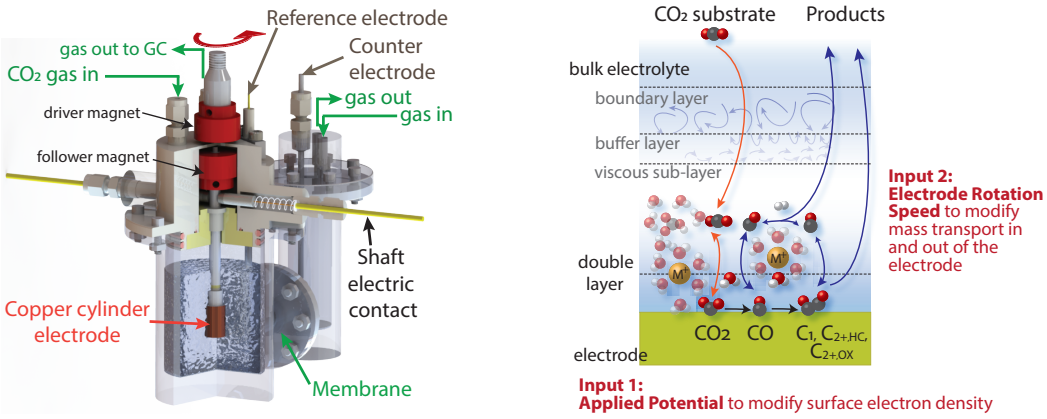
Chapter 2

Preliminaries

This chapter introduces the background of the electrochemical reactor used in this work and the background of the mathematical derivations of the analysis used in Chapter 5. Specifically, the experimental setup and the basic operating mode of the reactor are presented in this chapter. 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. Then, some traditional kinetic and mass transfer models are derived for the electrochemical reactor. Finally, the theory of relative gain array analysis is discussed for multi-input-multi-output control.

2.1 Experimental Electrochemical Reactor

The experimental reactor and microscopic transport diagrams are shown in Fig. 2.1. The reactor was designed to study the effect of mass transport on electrochemical CO_2 reduction while keeping the electrochemical cell hermetically gas-tight for the online detection of gas products. The reactor has two chambers separated by an ion-exchange membrane to prevent products from crossing between the chambers. The main chamber contains the working electrode (cathode), and the other chamber contains the counter electrode (anode). The CO_2 gas is directly bubbled



(a) (b)

Figure 2.1: A diagram showing (a) the electrochemical reactor and (b) the reaction and mass transfer processes involved in the transformation of CO_2 to further reduced products on the polycrystalline copper electrode surface.

into both chambers, where the electrodes are submerged in a 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 [22]. In order to keep the reactor airtight, the modulated speed rotator (MSR) uses magnetic coupling to drive the RCE. The driver magnet is attached to the MSR shaft, and it transmits torque to the follower magnet inside the reactor without requiring a direct shaft connection between the MSR and RCE (Fig. 2.1). As the RCE shaft continuously stirs the electrolyte solution, the 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 chromatography (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 have recently been reported [12].

The product compositions quantified using GC and NMR are then used to determine the production rate of each species and the reaction selectivity with respect to the desired products.

Polycrystalline copper produces various products at a quantifiable level, as tabulated in Table 2.1. Here, competing reactions—such as the hydrogen evolution reaction and the formate production reaction—are excluded from the table and from the selectivity calculation, as they do not share the same reaction pathway as the products in Table 2.1. Although carbon monoxide and formate are both 2-electron reduction products, carbon monoxide is the main reaction intermediate that leads to further reduced products, while formate cannot be reduced further. The desired products are the C_{2+} oxygenate species (labeled in Table 2.1), because 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 C_{2+} oxygenate production to the rate of hydrocarbon production.

Since the reactions are catalytic, the reactions only occur on the surface of the electrode. Therefore, the reactor operation is dictated by the balance between the surface kinetics and the fluid phase mass transfer. The CO_2 gas is bubbled in from the bottom of the reactor and dissolves in the buffer solution. The dissolved CO_2 is carried to the electrode surface by convective mass transport from 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 copper surface and the adsorption rate of CO_2 molecules on the copper surface. The electron density is dictated by the applied potential, and the CO_2 adsorption rate is the result of mass transport and reaction kinetics at the electrode/electrolyte interface (Fig. 2.1). Then, the products are carried away from the electrode surface by convective mass transfer. The liquid products remain in the solution until the end of the experiment, while the gaseous products diffuse into the gas phase and are carried out by the excess CO_2 carrier gas. The liquid phase is removed from the reactor at the end of the experiment, and samples are taken for NMR. The gas products are measured by the GC every 20 minutes throughout the experiment.

Table 2.1: Various reactions that are part of the CO_2 reduction mechanism on copper.

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

2.2 Electrochemical Kinetics and Mass Transport

This section will go into more detail on electrochemical kinetics and specific mass transfer models relevant to electrochemistry. In the context of electrochemistry, the rates of kinetics and mass transfer are defined in terms of the current because the current represents the measurable rate of electron transfer in the process. The current is defined as the rate of charge transfer in the system. The reactive species in electrochemical reactions are ions, so the current is proportional to the rate of mass transfer of these charged species in the system. For a given reaction, the rate of mass transfer is simply a unit conversion from the current

$$i = nF\frac{dN_j}{dt} \tag{2.1}$$

where N_j is the moles of species j, n is the number of electrons per molecule of species j, and F is the Faraday constant. Since electrochemical reactions are catalytic surface reactions, the overall reaction rate is a balance between the rate of kinetics and the rate of mass transfer. If the surface reaction is slow compared to the mass transfer, the system is said to be kinetically limited, so the

rate of reaction can be described by a kinetic model. If the mass transfer is slow compared to the surface reaction, the system is said to be mass-transfer-limited, and the measured current is called the limiting current. In either case, models exist to describe the behavior of the current in the system.

From a kinetics perspective, the Butler-Volmer equation describes the current for a single-step, single-electrode reaction [3]. The general form of the Butler-Volmer model works for reversible and irreversible reactions, and it is valid for systems with mass transfer if the surface concentration can be determined explicitly. For simplicity, the CO_2 reduction reactions are assumed to be irreversible, which gives the equation

$$i = FAk^0 C_O\left(t\right) e^{-\frac{\alpha F}{RT}\left(E-E^{0\prime}\right)} \tag{2.2}$$

where F is the Faraday constant, A is the electrode surface area, k^0 is the standard rate constant, C_O is the surface concentration of the reactant (i.e., CO_2), R is the gas constant, T is the temperature, E is the applied potential, α is the symmetry factor, and $E^{0\prime}$ is the standard reduction potential of the reaction. This equation holds for any surface concentration $C_O\left(t\right)$, even if it changes as a function of time. Under kinetically limited conditions, the surface concentration C_O is the same as the bulk concentration C_O^* and is assumed to be constant, since the mass transfer rate is much faster than the kinetic rate. When the reverse reaction is ignored, it can be helpful to simplify the analysis by defining the forward rate constant k_f as

$$k_f = k^0 e^{-\frac{\alpha F}{RT}\left(E-E^{0\prime}\right)} \tag{2.3}$$

Under mass-transfer-limited conditions, the surface reaction is assumed to be very fast relative to the mass transfer. Therefore, the current is completely described by the mass transfer of the system. A common mass transfer model for electrochemistry is the Levich equation [3] which describes the mass transfer around a rotating disk electrode (RDE) under steady-state and mass-

transfer-limited conditions. An RDE is a flat disk electrode covered with insulation so that the flat, circular face is the only surface exposed to the electrolyte solution. When rotated at sufficient speeds, the mass transport is dominated by convection, so the rate of mass transfer is determined by the rotation speed of the electrode. The mass transfer is determined by solving the convective-diffusion equation in cylindrical coordinates [3].

$$v_r\left(\frac{\partial C_O}{\partial r}\right)+\frac{v_\phi}{r}\left(\frac{\partial C_O}{\partial \phi}\right)+v_y\left(\frac{\partial C_O}{\partial y}\right)=D\left[\frac{\partial^2 C_O}{\partial y^2}+\frac{\partial^2 C_O}{\partial r^2}+\frac{1}{r}\frac{\partial C_O}{\partial r}+\frac{1}{r^2}\left(\frac{\partial^2 C_O}{\partial \phi^2}\right)\right] \quad (2.4)$$

Here, r, ϕ, and y refer to the radial, angular, and vertical coordinates; v_r, v_ϕ, and v_y refer to the velocity in each coordinate direction; C_O is the concentration; and D is the diffusivity. In this analysis, the concentration, diffusivity, etc. refer to the reactant species. Since the reactions are reduction reactions, the reactant species are the oxidized species (CO_2). This equation can be substantially simplified because C_O is not a function of r or ϕ at the surface of the electrode (based on the symmetry of the system), so Eq. 2.4 simplifies to

$$v_y\left(\frac{\partial C_O}{\partial y}\right)=D\frac{\partial^2 C_O}{\partial y^2} \quad (2.5)$$

This equation can be solved analytically, and the current is found from the mass transfer rate at the surface by

$$i_{l,c}=nFAD\left(\frac{\partial C_O}{\partial y}\right)_{y=0} \quad (2.6)$$

$i_{l,c}$ is defined as the current under completely mass-transfer-limited conditions (also called the limiting current). The Levich equation is the solution to Eqs. 2.5 and 2.6, and states

$$i_{l,c}=0.62nFAD^{2/3}\nu^{-1/6}\omega^{1/2}C_O^* \quad (2.7)$$

where ν is the kinematic viscosity of the electolyte, ω is the rotation speed of the RDE, C_O^* is the

bulk CO_2 concentration, and the other quantities are defined previously.

The Levich equation can also be combined with Butler-Volmer kinetics to model RDE reactions in more general situations, even when the reaction is not mass-transfer-limited. The mass transfer and kinetics effects are combined to give the Koutecký-Levich equation [3]

$$\frac{1}{i} = \frac{1}{i_K} + \frac{1}{i_{l,c}} \tag{2.8}$$

where i is the total current and i_K is the current in the absence of mass transfer effects. In this equation, i_K is given by combining Eqs. 2.2 and 2.3, and $i_{l,c}$ is given by Eq. 2.7 (the Levich equation). This gives the equation

$$\frac{1}{i} = \frac{1}{FAk_f C_O^*} + \frac{1}{0.62nFAD^{2/3}\nu^{-1/6}\omega^{1/2}C_O^*} \tag{2.9}$$

Solving for i in this equation gives the general current relationship for the RDE

$$i_{RDE} = \frac{FAk_f C_O^*}{1 + \frac{k_f}{0.62D^{2/3}\nu^{-1/6}\omega^{1/2}}} \tag{2.10}$$

Since the physical reactor uses a cylindrical electrode, rather than a disk electrode, the Levich model must be adapted to the cylindrical geometry to apply these results to the real system. Therefore, Sherwood analysis is used to nondimensionalize the RDE analysis so that it can be converted into an analogous solution for the RCE system. The Sherwood relation for the RDE can be found analytically from Eq. 2.7 (the Levich equation)

$$Sh_{RDE} = \frac{m_O}{D/r_{disk}} = 0.62Re_{RDE}^{0.5}Sc^{0.33} \tag{2.11}$$

where m_O is the mass transfer coefficient, r_{disk} is the radius of the disk, Re_{RDE} is the Reynolds number for the disk geometry, and Sc is the Schmidt number. Here, the Reynolds number is

defined as $Re_{RDE} = \frac{\omega r_{disk}^2}{\nu}$, with r_{disk} serving as the characteristic length scale. The Sherwood relation for the cylindrical geometry is found empirically by fitting parameters to data collected from experiments at varying rotation speed. The analysis is discussed in more detail by Jang et al. [12]. The empirical Sherwood relation is

$$Sh_{RCE} = \frac{m_O}{D/d_{cyl}} = 0.204 Re_{RCE}^{0.59} Sc^{0.33} \tag{2.12}$$

where d_{cyl} is the diameter of the cylinder and Re_{RCE} is the Reynolds number for cylindrical geometry. In this equation, the Reynolds number is defined as $Re_{RCE} = \frac{\omega d_{cyl}^2}{2\nu}$, with d_{cyl} serving as the characteristic length scale. This analysis is used to create the empirical models discussed in Chapter 5.

2.3 Relative Gain Array Analysis

Relative gain array (RGA) analysis is used to determine the controllability of a system under multi-input-multi-output control. Specifically, RGA analysis determines whether multiple PI controllers (multi-PI control) are sufficient to perform MIMO control on the given system, or if more complex nonlinear control schemes, such as model-predictive control, are required. The relative gain array is a square matrix that compares the relative effect of each manipulated variable on each output variable. Each column of the RGA corresponds to a manipulated variable m_i, and each row corresponds to an output variable y_i. The relative gain of each loop is defined as the ratio of two gains between the input and output variables: the gain when holding the other inputs constant and the gain when holding the other outputs constant. By this definition, the relative gain compares the relative effects of the manipulated variables on the specific output variable when all the other loops are open versus when all the other loops are closed. The relative gain of the output variable with

respect to the manipulated variable is defined as

$$\lambda_{ij} = \frac{(\partial y_i / \partial m_j)_m}{(\partial y_i / \partial m_j)_y} \tag{2.13}$$

where y_i is the output variable, m_j is the manipulated variable, and λ_{ij} is the relative gain of y_i with respect to m_j. The subscript m in the numerator denotes that all manipulated variables are held constant, except for m_j. Similarly, the subscript y in the denominator denotes that all output variables are held constant except for y_i [29]. The relative gain can be found either by conducting experiments or by calculating the gradients from a steady-state model. The experimental procedure is described by Stephanopoulos [29]. Riggs and Woolf explain how to use the steady-state model gradients to calculate the relative gain [26, 33]. For this system, the RGA analysis is carried out with the steady-state model approach with the FNN serving as the steady-state model.

For any size RGA, the first step of the steady-state model approach is to calculate the steady-state gain matrix G, which contains the gradient of each output variable with respect to each manipulated variable. The gain matrix is defined as

$$G_{ij} = \frac{\partial y_i}{\partial m_j} \tag{2.14}$$

where G_{ij} is the element in the ith row and jth column of the G matrix, and the partial derivatives are determined with all other manipulated variables held constant. This is the same as the numerator in Eq. 2.13. Then, the matrix R is defined as the inverse transpose of the steady-state gain matrix G

$$R = \left(G^{-1}\right)^T \tag{2.15}$$

Finally, the relative gain array is calculated by multiplying G and R elementwise, as in, each

element of the RGA is calculated by

$$\lambda_{ij} = G_{ij} \times R_{ij} \tag{2.16}$$

where λ_{ij} is the element in the ith row and jth column of the RGA. Given this definition of the RGA, the relative interactions of the various control loops can be determined based on the magnitude of the elements in the RGA.

As an example, the RGA calculations are shown for the case of a two-by-two RGA. First, the steady-state gain matrix is calculated at a given set point

$$G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} = \begin{bmatrix} \dfrac{\partial y_1}{\partial m_1} & \dfrac{\partial y_1}{\partial m_2} \\ \dfrac{\partial y_2}{\partial m_1} & \dfrac{\partial y_2}{\partial m_2} \end{bmatrix} \tag{2.17}$$

Then, the corresponding matrix R is given by

$$R = \left(G^{-1}\right)^T = \begin{bmatrix} \dfrac{G_{22}}{G_{11}G_{22} - G_{12}G_{21}} & \dfrac{-G_{21}}{G_{11}G_{22} - G_{12}G_{21}} \\ \dfrac{-G_{12}}{G_{11}G_{22} - G_{12}G_{21}} & \dfrac{G_{11}}{G_{11}G_{22} - G_{12}G_{21}} \end{bmatrix} \tag{2.18}$$

where G_{ij} is the element in the ith row and jth column of the G matrix (the parital derivative of output i with respect to input j). Finally, the RGA is calculated by the elementwise multiplication of G and R as follows

$$RGA = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} = \begin{bmatrix} \dfrac{G_{11}G_{22}}{G_{11}G_{22} - G_{12}G_{21}} & \dfrac{-G_{12}G_{21}}{G_{11}G_{22} - G_{12}G_{21}} \\ \dfrac{-G_{12}G_{21}}{G_{11}G_{22} - G_{12}G_{21}} & \dfrac{G_{11}G_{22}}{G_{11}G_{22} - G_{12}G_{21}} \end{bmatrix} \tag{2.19}$$

The results of the RGA are then analyzed based on the values of each element in the array. First, it is important to note some special properties of the RGA. All rows and all columns in the

RGA sum to 1. For a two-by-two RGA, this means that the diagonal elements are always equal, and the off diagonal elements are always equal. Additionally, only one element needs to be specified in order to know all values of a two-by-two RGA. Once the RGA is calculated at a specific set point, each element in the RGA can be interpreted by the following cases:

1. If $\lambda_{ij} = 0$, then y_i does not respond to m_j
2. If $\lambda_{ij} = 1$, then y_i is only affected by m_j. All other manipulated variables do not affect y_i.
3. If $\lambda_{ij} < 0$, then the system is unstable if y_i is paired with m_j.
4. If $0 < \lambda_{ij} < 1$ or $\lambda_{i,j} > 1$, then there are interactions between the control loops.

The interactions in Case 4 are further classified into different types of interactions.

- If $0 < \lambda_{ij} < 0.5$, then m_j has a weak effect on y_i. The combined effect of all the other control loops is larger than the influence of m_j on y_i. Avoid pairing y_i with m_j.
- If $0.5 < \lambda_{ij} < 1$, then m_j has a strong effect on y_i, but the other loops have some effect.
- If $\lambda_{ij} > 1$, then m_j has a strong effect on y_i, but some other control loops affect y_i in the opposite direction.

It is preferential to pair y_i with m_j in the latter two cases because m_j has a dominant effect on y_i. When $\lambda_{ij} < 0.5$, the influence of m_j is too small relative to the other control loops, so the interference from the other loops will overpower the control action of m_j. In the case where $\lambda_{ij} > 1$, the control will still work, but the other control loops will push y_i in the opposite direction of the control action of m_j, so the gain for m_j may need to be increased to overcome the other loops' interactions. The best control loop pairing will be the combination that gives the $\lambda_{i,j}$ value closest to 1.

Chapter 3

Development of the Machine Learning Model

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

3.1 FNN Learning Algorithm

The general structure of an FNN model is shown in Fig. 3.1 and can be mathematically represented by the following equations:

$$\mathbf{Y} = F_{NN}(X) = \begin{cases} h_j^{[1]} & = \sigma^{[1]}(\sum_{i=1}^{p} \omega_{ji}^{[1]} x_i + b^{[1]}) \\ h_j^{[2]} & = \sigma^{[2]}(\sum_{i=1}^{p} \omega_{ji}^{[2]} h_i^{[1]} + b^{[2]}) \\ y_j & = \sigma^{[l]}(\sum_{i=1}^{p} \omega_{ji}^{[l]} h_i^{[l]} + b^{[l]}) \end{cases} \tag{3.1}$$

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

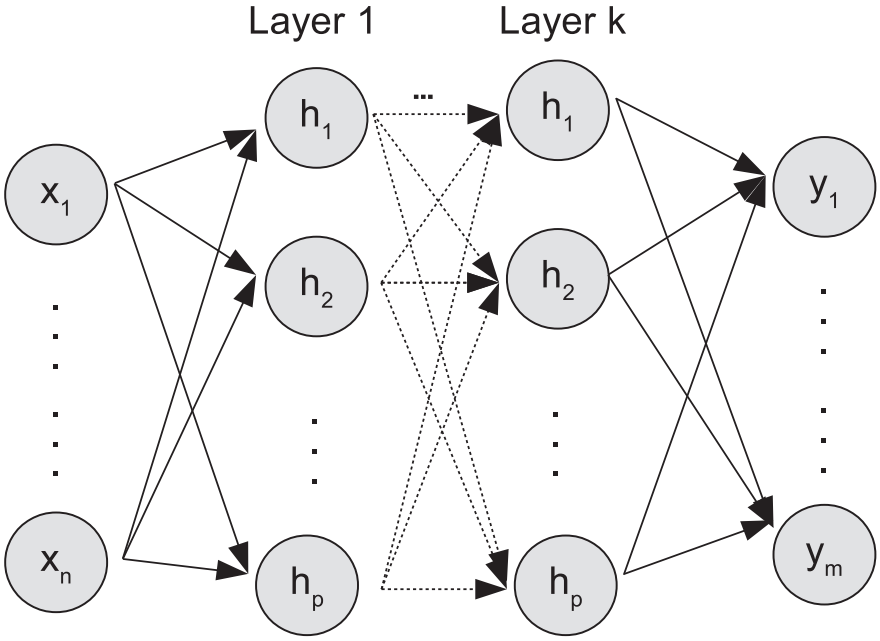


Figure 3.1: General structure of an FNN model, where subscript p is the index of neurons in the kth hidden layer.

In this study, a centralized two-input-multi-output FNN model is constructed to capture the nonlinear relationships between the two input states in Table 3.1 (i.e., rotation speed and applied potential) and the fourteen outputs listed in Table 3.2. 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 from 0 to 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. 3.2. 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.

$$ReLu(z) = \begin{cases} z & \text{for} \quad z > 0 \\ 0 & \text{for} \quad z \leq 0 \end{cases} \tag{3.2}$$

Remark 3.1 *A single hidden layer is used for this model because it is the simplest structure to sufficiently capture the data trends. Additionally, 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 due to 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 [34]. We recommend that other users consider using those methods to develop their machine learning models.*

3.2 Data Generation and Dataset

As listed in Table 3.2, the oxygenates considered are Outputs 3-12, and the hydrocarbons are Outputs 1 and 2 (methane and ethylene). Therefore, the selectivity defined in Section 2.1 is calculated as follows:

$$Selectivity := \frac{\sum_{i=3}^{12} y_i}{\sum_{i=1}^{2} y_i} \tag{3.3}$$

where y_i refers to the production rate of species i, as defined in Table 3.2. Data are collected for the range of potentials and rotation speeds within which the reactor will operate. Specifically, the potential is varied from -1.2 to -1.47 volts versus the standard hydrogen electrode (V vs. SHE),

and the rotation speed is varied from 100 to 800 revolutions per minute (rpm).

For the data collection process, the potentiostat is set to a constant potential, and the electrode is rotated at a constant angular speed. The reactor is allowed to equilibrate for twenty minutes prior to the data collection. Then, product samples are taken every twenty minutes to determine the concentration of the thirteen relevant products, and the selectivity is calculated from the results of each sample.

Each experiment takes 80 minutes to collect three to four GC measurements and one NMR sample. The GC is sampled three times per experiment to ensure that the measurements are consistent over time. Each experiment is repeated two to five times to ensure that the data are consistent and to obtain the statistical information for the experimental results under the same operating conditions. Subsequently, the data points 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 collected from approximately 100 experiments.

3.3 Design of the Experiment

The range of potentials is limited by the overall resistance of the electrochemical cell between the working and counter electrodes. This issue is resolved in the second generation of the cell by removing the channel that connects the two chambers to shorten the distance between the two electrodes and to increase the surface area of the ion-exchange membranes [12]. 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 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). We restricted the maximum electrode rotation speed to 800 rpm mainly due to the mechanical instability of the custom-machined parts of the electrochemical cell. Additionally, the chosen range of

rotation speed is appropriate for studying mass transport effects from the perspective of mass transport characteristics around the RCE. The film mass transfer coefficient decreases as the electrode rotation speed increases with a 0.59 order dependency—as shown by the Sherwood relationship in Section2.2—so increasing the rotation speed beyond 800 rpm has a negligible effect on the mass transport properties of the cell [12]. 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.

3.4 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 predicted value. The MSE loss function is given below:

$$Loss = \frac{1}{d}\frac{1}{m}\sum_{i=1}^{d}\sum_{j=1}^{m}|y_{i,j} - \hat{y}_{i,j}|^2 \tag{3.4}$$

where d is the number of data points in the training dataset, and m is the number of output states. To generate the training and testing sets, 4 points are reserved from the original 21 data points to be the testing dataset, and the remaining 17 data points 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 square difference between the FNN prediction and the testing data, using the loss function of Eq. 3.4 to evaluate the performance of the model. During this process, the parameter vector $\mathbf{W}$, which contains all the weights and bias

of the neural network, is optimized using Eq. 3.5 to minimize the loss function.

$$\mathbf{W} = \mathbf{W} - \eta \frac{V_{dw}}{\sqrt{S_{dw}} + \epsilon} \tag{3.5}$$

where η is the learning rate, and ϵ is a small positive number to prevent the denominator being zero. V_{dw} and S_{dw} 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 does not have significant impact to the model performance. Additionally, the user can tune the learning rate η to improve the model performance. η is usually a small positive number from 0 to 1.

Table 3.1: Input states of the FNN model.

Index	**Input State**	**Units**
1	Applied Potential	V vs. SHE
2	Rotation Speed	rpm

Table 3.2: Output states of the FNN model.

Index	**Output State**	**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	

Chapter 4

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 experimental error. 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 [21]. 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 shown that this method can provide a solution to this optimization problem [8]. The MLE method assumes that the data are from a single population with the same standard deviation. However, this chapter proposes a modification that assumes each set of input parameters corresponds to a different population. Thus, each data point with its associated 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 pseudo-

probabilistic sample following the Gaussian distribution with an associated standard deviation. Therefore, the FNN outputs $\hat{y}_{i,j}$ need to follow the same distribution as the reference data $y_{i,j}$, which means the joint likelihood of the neural network output is a Gaussian distribution and can be expressed as follows:

$$\begin{aligned}\mathcal{L}(\mathbf{X};\mathbf{W},\sigma) &= \prod_{k=1}^{d\times m} f_{\mathbf{Y}}(y_k) \\ &= \prod_{k=1}^{d\times m}(2\pi\sigma_k^2)^{-0.5} \times \mathbf{exp}\left[-\frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{m}\left|\frac{y_{i,j}-\hat{y}_{i,j}(\mathbf{X},\mathbf{W})}{\sigma_{i,j}}\right|^2\right]\end{aligned} \tag{4.1}$$

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:

$$\begin{aligned}\mathbf{W}^* :&= \arg\max_{\mathbf{W}} \mathbf{log}\mathcal{L}(\mathbf{X};\mathbf{W},\sigma) \\ &= \arg\max_{\mathbf{W}}\left(-\frac{1}{2}\sum_{k=1}^{d\times m}\mathbf{log}(2\pi\sigma_k^2) - \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{m}\left|\frac{y_{i,j}-\hat{y}_{i,j}(\mathbf{X},\mathbf{W})}{\sigma_{i,j}}\right|^2\right) \\ &= \arg\max_{\mathbf{W}}\left(-\sum_{k=1}^{d\times m}\mathbf{log}(2\pi\sigma_k^2) - \sum_{i=1}^{d}\sum_{j=1}^{m}\left|\frac{y_{i,j}-\hat{y}_{i,j}(\mathbf{X},\mathbf{W})}{\sigma_{i,j}}\right|^2\right)\end{aligned} \tag{4.2}$$

Since the first term of Eq. 4.2 is independent of $\mathbf{W}$, the maximum likelihood estimation of this model can be simplified further into Eq. 4.3.

$$\mathbf{W}^* = \arg\min_{\mathbf{W}}\left(\sum_{i=1}^{d}\sum_{j=1}^{m}\left|\frac{y_{i,j}-\hat{y}_{i,j}(\mathbf{X},\mathbf{W})}{\sigma_{i,j}}\right|^2\right) \tag{4.3}$$

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. Then, the coefficient of variance (v) of each data point is determined by the ratio of standard deviation and the respective output mean. This normalizes the data

variability to allow for unbiased comparison between quantities of different magnitudes. The loss function, shown as Eq. 4.4, integrates Eq. 3.4 and Eq. 4.3. Therefore, the MLE-FNN weight matrix is optimized to maximize both the accuracy of the prediction and the likelihood function during the training process.

$$Loss = \frac{1}{d}\frac{1}{m}\sum_{i=1}^{d}\sum_{j=1}^{m}\frac{1}{v_{i,j}^2}\left|y_{i,j} - \hat{y}_{i,j}\right|^2 \tag{4.4}$$

Remark 4.1 *In this study, error bars are constructed to represent the region of one standard deviation of uncertainty with respect to the mean, which is approximately the 70% confidence interval for Gaussian-distributed variables. Any statistical model can be used to develop an MLE-FNN model if it can provide reasonable statistical information of the experimental observations.*

Remark 4.2 *As shown by Dorling et al. and Kumar et al. [6, 15], the simplified log-likelihood function (Eq. 4.3) can be used directly as the loss function of an MLE-FNN model, since it contains the sum of squared error in the loss function. We integrate it with Eq. 3.4 to demonstrate its similarity to the mean-squared error (MSE) loss function.*

Remark 4.3 *Bayesian optimization is another acknowledged method to develop statistical machine learning model. Similarly to the MLE method, 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 Bayes' rule [27]. 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.*

Chapter 5

Machine Learning Model Results and Analysis

In this chapter, we demonstrate the MLE-FNN model's ability to provide accurate predictions of the reactor operation and demonstrate the future research areas that have made use of its predictions. First, we compare the prediction performance of the standard FNN and MLE-FNN models. Then, we compare the MLE-FNN model predictions with a classical EFP model to determine the MLE-FNN model's ability to capture the physical phenomenon behind the reactor operation. The results of the MLE-FNN model are then used to improve other empirical models, to determine economically optimal operating conditions, and to determine the reactor's controllability for multi-input-multi-output control.

5.1 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 Section 3.1 and with randomly partitioned

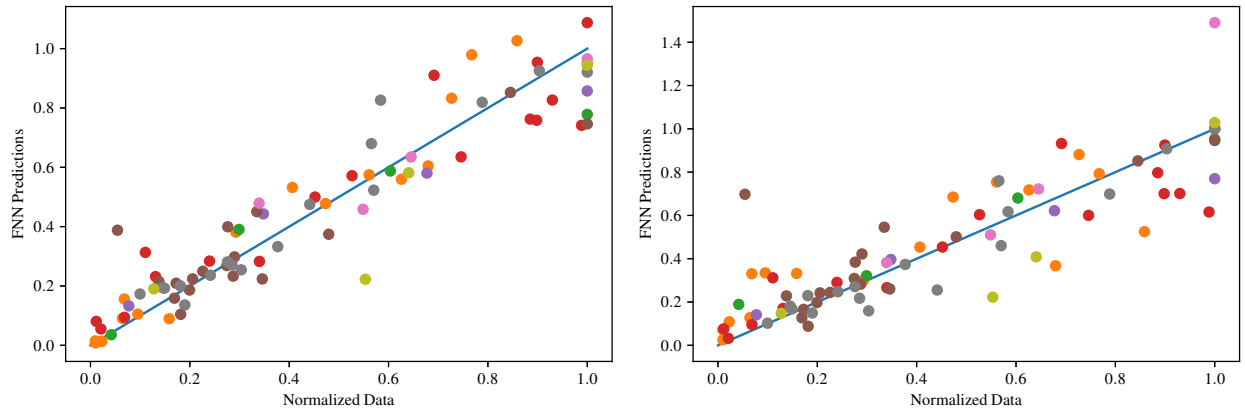
Table 5.1: Process parameters for EFP models with units.

Quantity	**Value**	**Units**
EFP model (limiting conditions)		
k_0	2.32×10^{-12}	$cm \cdot s^{-1}$
α	0.5	
F	96485	$C \cdot mol^{-1}$
R	8.314	$J \cdot mol^{-1} \cdot K^{-1}$
T	298	K
$E^{0\prime}$	-0.52	V
C_{CO_2}	3.40×10^{-5}	$mol \cdot cm^{-3}$
D_{CO_2}	1.91×10^{-5}	$cm^2 \cdot s^{-1}$
d_{RCE}	1.2	cm
ν_{H_2O}	1.03×10^{-2}	$cm^2 \cdot s^{-1}$
EFP model		
$k_{0,5}$	2.02×10^{-28}	$mol \cdot cm^{-1} \cdot s^{-2}$
$k_{0,6}$	7.47×10^{-32}	$mol \cdot cm^{-1} \cdot s^{-2}$
$k_{0,7}$	2.61×10^{-13}	$mol^{-\frac{1}{2}} \cdot cm \cdot s^{-\frac{1}{2}}$
α_5	0.7	
α_6	0.85	
α_7	0.665	
EFP model (updated)		
$k_{0,5}$	7.2×10^{-22}	$mol \cdot cm^{-1} s^{-2}$
$k_{0,6}$	1.6×10^{-17}	$mol^{\frac{1}{4}} \cdot cm^{\frac{3}{2}} \cdot s^{\frac{5}{4}}$
$k_{0,7}$	9.5×10^{-23}	$mol \cdot cm^{-1} \cdot s^{-2}$
α_5	0.42	
α_6	0.53	
α_7	0.49	

training and validation sets. The best FNN and MLE-FNN are chosen to minimize the MSE for the training dataset. This training method ensures the selected models are trained consistently following the same criteria. Then, the selected FNN and MLE-FNN models 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, demonstrating marginally better performance by the standard FNN. It is shown in Fig. 5.1 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 MSEs of the two methods are sufficiently small, which implies that both models capture the input-output relationship well.

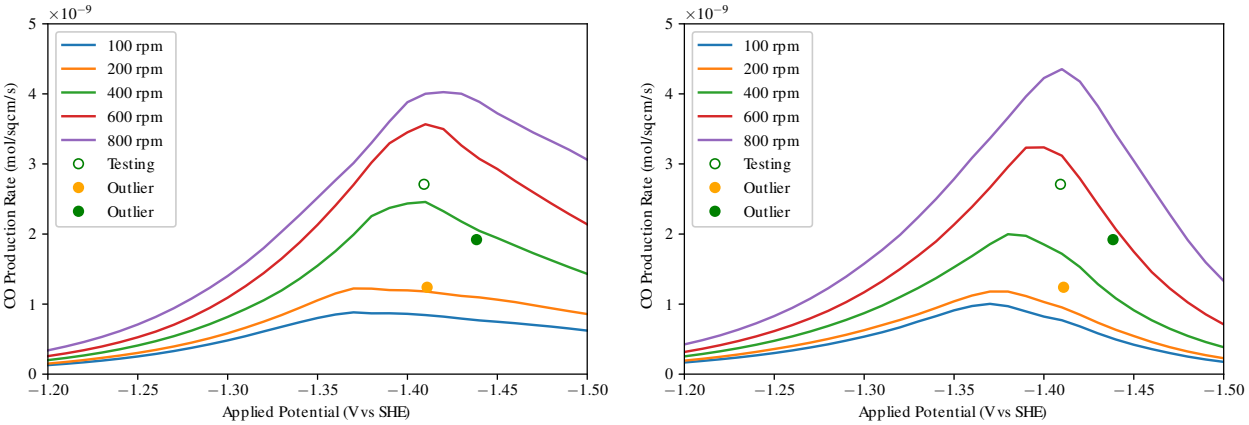
To further compare the performance of the two models, the predictions for CO production rate are plotted in Fig. 5.2. This figure shows the predicted CO production rate and known outlier points, which came from experiments with a slight drift in operating conditions. As shown in the figure, the MLE-FNN weighs the data points with higher 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 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. To simplify the discussion, only the MLE-FNN will be used in the remainder of this chapter (i.e., henceforth the FNN will only refer to the MLE-FNN, and the standard FNN will not be included).

In the following sections, the FNN model will be compared with other modeling techniques to further assess its performance. Additionally, various applications of this model are discussed to demonstrate the importance of accelerating the modeling process. One simple use case of this model is to provide comprehensive selectivity predictions over the range of realistic operating



(a) (b)

Figure 5.1: Comparison between the observed experimental outcome and the neural network predictions from (a) standard FNN and (b) MLE-FNN models.



(a) (b)

Figure 5.2: The CO production rate predictions are shown for various applied potentials in the unit of V vs. the standard hydrogen electrode (V vs. SHE). The solid points are known experiments that had a drift in potential. The open point is from the testing set. (a) The CO prediction of standard FNN model overfits the labeled uncertain data points. (b) The MLE-FNN model successfully accounts for the experimental uncertainty, even though this introduces additional error to the testing results.

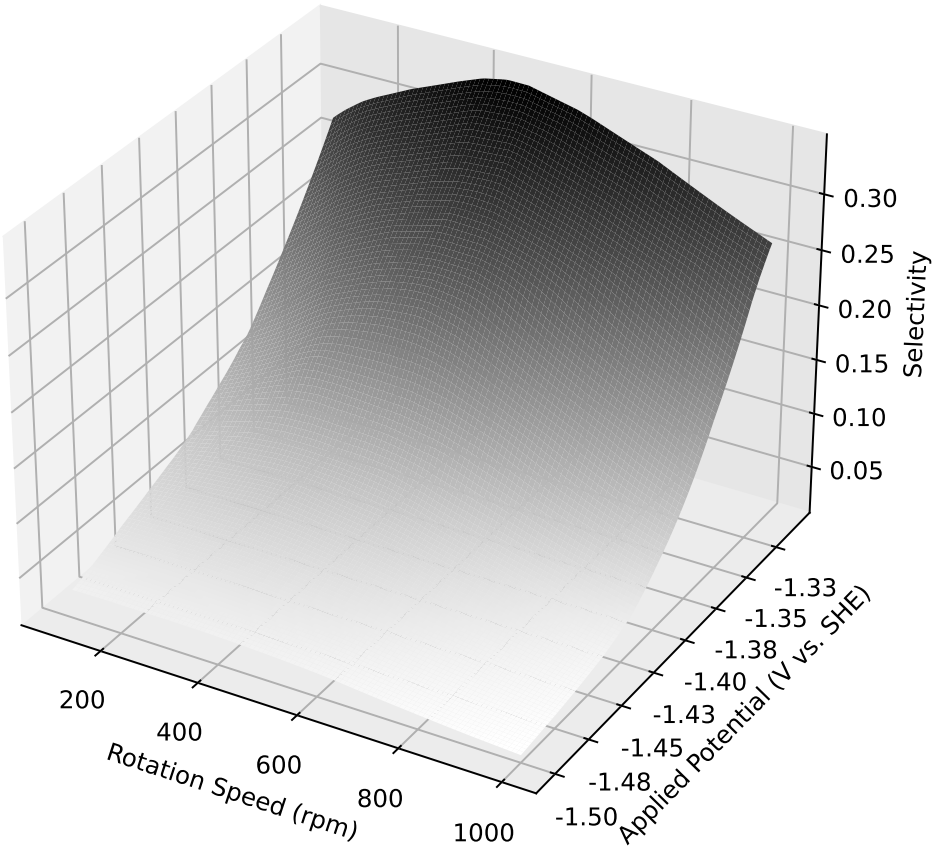


Figure 5.3: Selectivity of oxygenate species with respect to rotation speed and applied potential, in the units of V vs. the standard hydrogen electrode (V vs. SHE), as predicted by the machine learning model.

conditions, as shown in Fig. 5.3. This selectivity is defined in Section 3.2 with the purpose of maximizing the selectivity of oxygenate species. Maximizing the selectivity of the desired products can save energy for the operation of the reactor and minimize waste generated from disposal of the unwanted side products.

Remark 5.1 *The outlier points are included in the dataset since they are valid even if they have higher variability. The predictive models are developed on the basis of the experimental observations, even if some points are less likely to be reproduced than others. On the other hand, invalid data points from a failed experiment should not be included in the dataset.*

5.2 EFP model vs. MLE-FNN

First-principle models (FP models) are a fundamental approach to describe the operation of electrochemical reactors from the fundamental energy balance, mass balance, and reaction kinetics. However, it is challenging to obtain an accurate first-principle model for this specific system because of the complex mass transfer and reaction mechanisms of this process. As a substitute, machine learning modeling provides an alternative approach to representing the physicochemical phenomena in the reactor with a desired prediction accuracy. In this subsection, an empirical, first-principle model (EFP model) of a rotating electrode reactor is developed following the derivation by Bard [3], to determine the rate of CO production under limiting conditions. This derivation is showin in more detail in Section 2.2. Specifically, this model assumes that only a single, first-order reaction is occurring with no side reactions, which means it cannot capture the comprehensive kinetics of this experiment. The reaction is also assumed to occur only on the electrode surface following Butler-Volmer kinetics. The resulting equation is given as follows:

$$r_{CO} = \frac{k_f C^*_{CO_2}}{1 + k_f/m_O} \tag{5.1}$$

where $C^*_{CO_2}$ is the bulk concentration of CO_2, k_f is the forward rate constant, and m_O is the convective mass transfer coefficient. The forward rate constant k_f was defined in Section 2.2 as

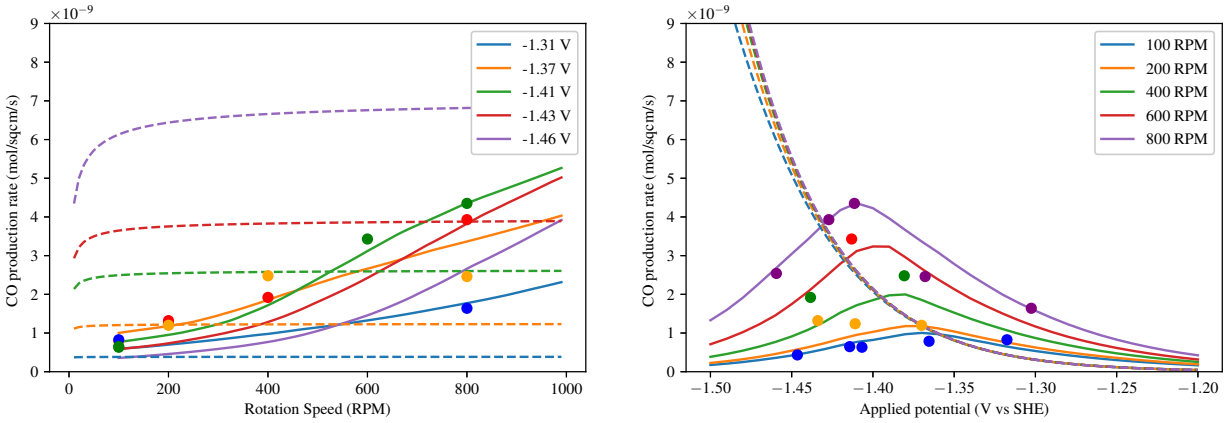
$$k_f = k^0 e^{-\frac{\alpha F}{RT}\left(E - E^{0\prime}\right)} \tag{5.2}$$

where k^0 is the standard rate constant, α is the symmetry factor, F is Faraday's constant, R is the gas constant, T is the temperature, E is the applied potential, and $E^{0\prime}$ is the standard reduction potential. The mass transfer coefficient m_O 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 [3]. 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 correlation as follows [12]:

$$m_O = 0.204 Re_{RCE}^{0.59} Sc^{0.33} \frac{D_{CO_2}}{d_{RCE}} \tag{5.3}$$

where Re_{RCE} is the Reynolds number, Sc is the Schmidt number, D is the diffusion coefficient, and d_{RCE} is the diameter of the RCE. This Sherwood correlation was previously discussed in Section 2.2. 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 referred to as an EFP model.

The comparison between this EFP model and the FNN prediction is shown in Fig. 5.4. 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 at these conditions. After switching to more negative conditions, the EFP model's assumptions become invalid. Therefore, these two models present different predictions after passing threshold conditions. This comparison demonstrates that the neural network can capture the input-output



(a) (b)

Figure 5.4: The 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 (rpm), and (b) applied potentials in the unit 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 more negative potentials and higher rotation speeds, the assumptions of the EFP model become invalid.

relations from the experimental data.

5.3 EFP Model Improvement

EFP modeling is an efficient way to find out how a number of experimental variables affect the experimental data without requiring complete knowledge of the underlying physical phenomena needed for large-scale FP models [4]. 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 understand the reactor phenomena. However, parameter tuning and selection for an empirical model are challenging. Therefore, we propose an algorithm to use the 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 Section 5.2. As shown in Eq. 5.4, the empirical regressions of interest are the production rate of C_1 products (FNN Outputs 1 and 3 listed in Table 3.2), 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_{C1}, $r_{C2+,HC}$, and $r_{C2+,OX}$ respectively.

$$\begin{aligned} r_{C_1} &= k_{0,5} C_{CO} Sh_{RCE}^{-0.5} J_{HCO_3}^{-1} \exp\left(-\frac{\alpha_5 z_5 F}{RT} E\right) \\ r_{C_{2+,HC}} &= k_{0,6} C_{CO} Sh_{RCE}^{-0.5} J_{HCO_3}^{-1} \exp\left(-\frac{\alpha_6 z_6 F}{RT} E\right) \\ r_{C_{2+,OX}} &= k_{0,7} C_{CO} Sh_{RCE}^{-0.5} J_{HCO_3}^{0.5} \exp\left(-\frac{\alpha_7 z_7 F}{RT} E\right) \end{aligned} \tag{5.4}$$

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. Additionally, 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_{C1}, $r_{C2+,HC}$, 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 is not a single value, and a modification is necessary for this case. Specifically the parameter α_i is considered as 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 to the FNN model using the testing set as described in Section 3.2. Table 5.2 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 the 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. Additionally, by comparing the prediction trends from both models in Fig. 5.5, 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 the following algorithm:

Algorithm 1: Empirical, First-Principle 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:

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

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

    O

  else

    p

  end

  timize 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

Table 5.2: Testing data MSE results of the FNN model and the empirical, first-principle 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

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. Additionally, 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 a computer.

Table 5.3: The non-scaled MSE for the updated and original EFP models.

Rate Index	**EFP (original)**	**EFP (updated)**
$r_{C_{2+,OX}}$	3.68E-19	9.20E-20
$r_{C_{2+,HC}}$	1.11E-18	2.32E-19
r_{C_1}	8.46E-18	5.76E-18

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 on the range of -1.47 V to -1.30 V as this is the range for the data collected to train the FNN model. The MSE between the EFP model and FNN is calculated at intervals of 0.01 V. For the model in Eq. 5.4, 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 value of k_i to change by several orders of magnitude which causes 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 0.1 to 10 with a 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 5.1. 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 5.3 . As a result, the MSE for r_{C1}, $r_{C2+,HC}$, and $r_{C2+,OX}$ are decreased by 75%, 79%, and 32%, respectively. The new empirical, first-principle equations are updated in Eq. 5.5 to reflect the changes made from the procedure. Moreover, the

new EFP model predictions are also shown in Fig. 5.6, which shows the overestimating problem is solved with the new parameters.

$$
\begin{aligned}
r_{C_1} &= k_{0,5} C_{CO} Sh_{RCE}^{-0.5} J_{HCO_3}^{-1} \exp\left(-\frac{\alpha_5 z_5 F}{RT} E\right) \\
r_{C_{2+,HC}} &= k_{0,6} C_{CO} Sh_{RCE}^{-0.5} J_{HCO_3}^{-0.25} \exp\left(-\frac{\alpha_6 z_6 F}{RT} E\right) \\
r_{C_{2+,OX}} &= k_{0,7} C_{CO} Sh_{RCE}^{-0.5} J_{HCO_3}^{-1} \exp\left(-\frac{\alpha_7 z_7 F}{RT} E\right)
\end{aligned}
\tag{5.5}
$$

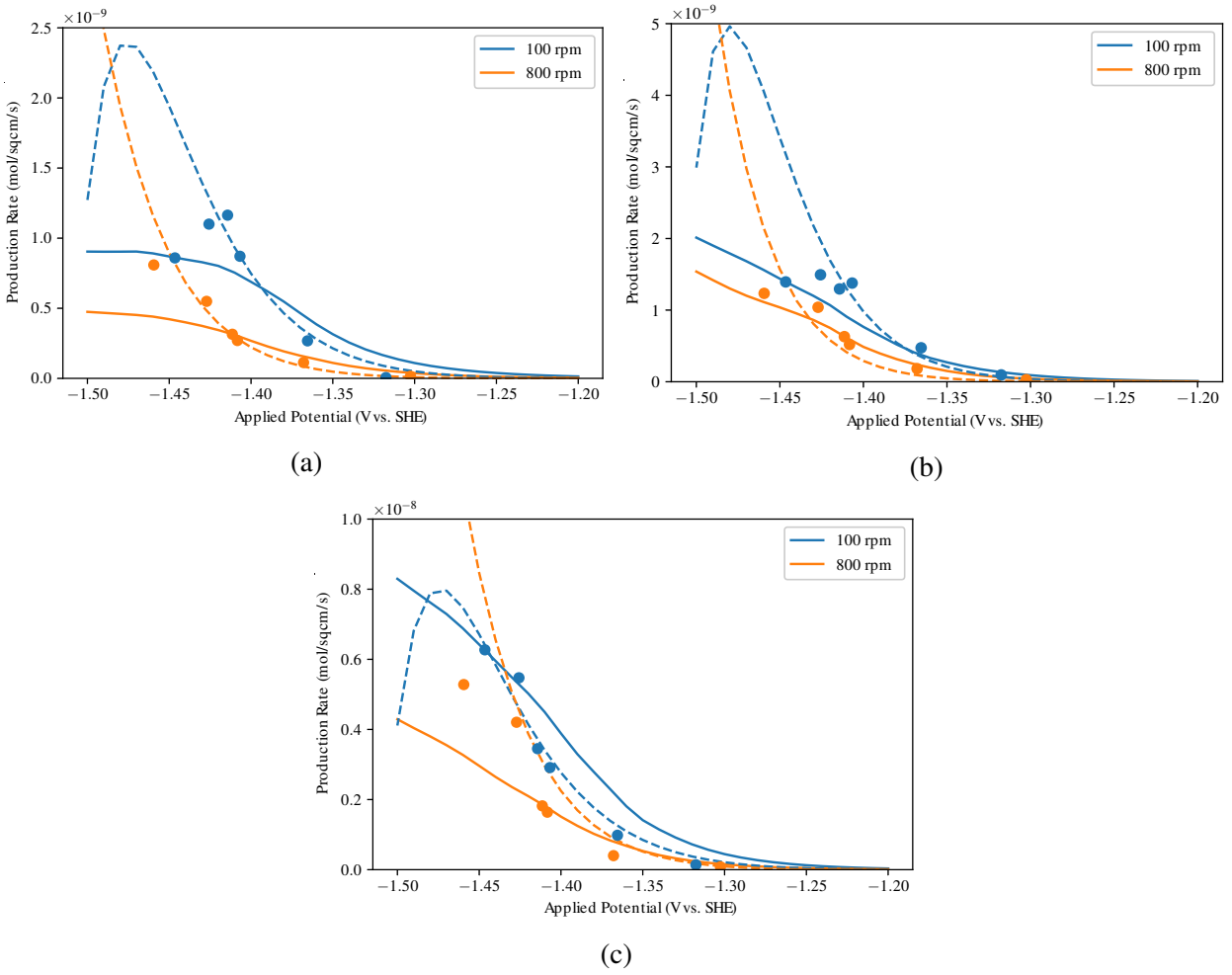


Figure 5.5: The production rates of (a) r_{C_1}, (b) $r_{C_{2+,HC}}$, and (c) $r_{C_{2+,OX}}$ from the EFP model (dashed) and the MLE-FNN model (solid) compared with the reference data points over the range of applied potentials in the unit of V vs. the standard hydrogen electrode (V vs. SHE).

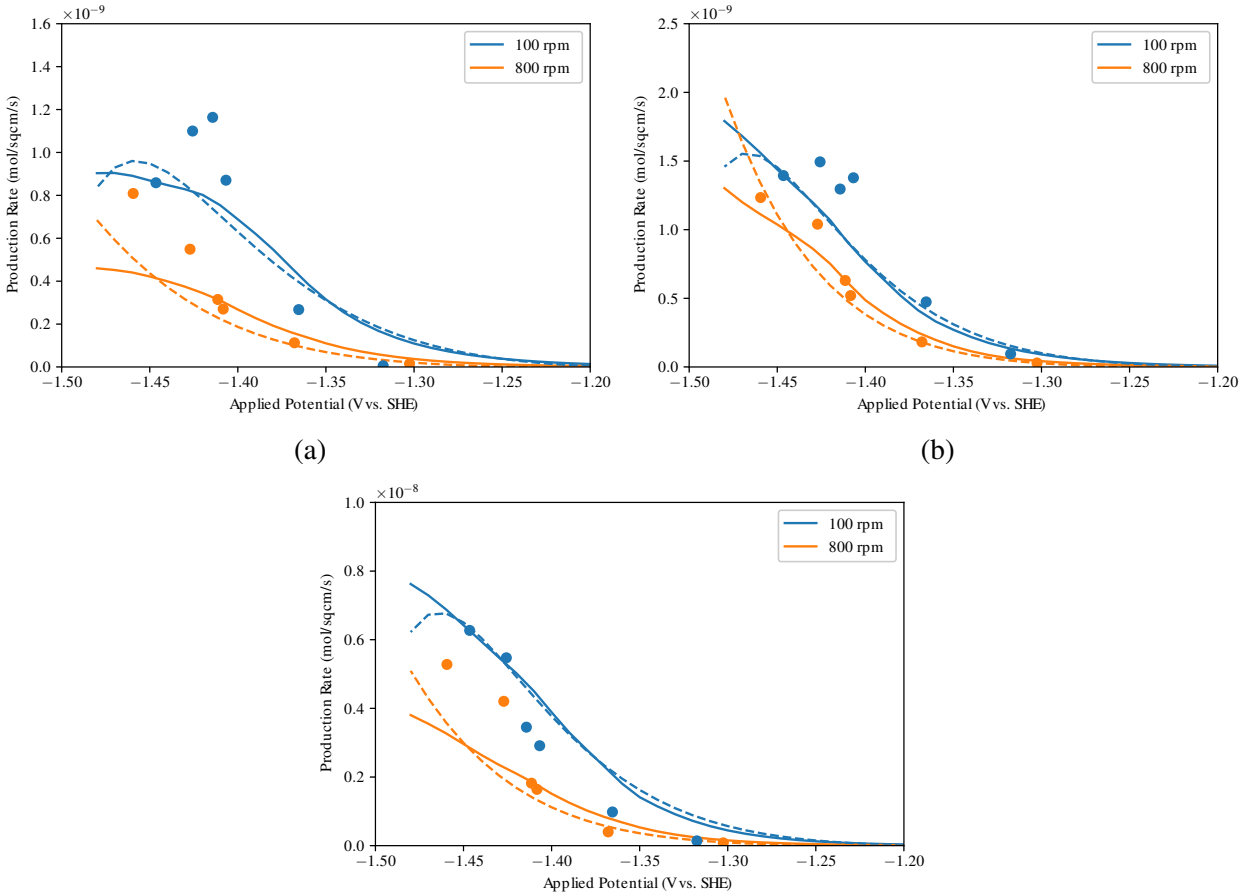


Figure 5.6: The production rates of (a) r_{C_1}, (b) $r_{C_{2+,HC}}$, and (c) $r_{C_{2+,OX}}$ from the updated EFP model (dashed) and the MLE-FNN model (solid) compared with the reference data points over the range of applied potentials in the unit of V vs. the standard hydrogen electrode (V vs. SHE).

Remark 5.2 *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. Therefore, the empirical model is modified to have a lower MSE for its prediction while retaining the physics of the experiment.*

Remark 5.3 *The experimental data is used to calculate the original EFP parameters. Specifically, the original EFP model is determined using traditional methods to extract the kinetic parameters 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 log of 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.*

5.4 Set Point Optimization

Set point optimization is critical for process operation as it can be used to maximize the benefits of a certain process. In this work, we demonstrate how the FNN results can be used to determine optimal operating conditions for the electrochemical reactor using results from an economic analysis of hypothetical operating costs. The results of the MLE-FNN described in Chapter 4 are used to predict the steady-state energy consumption and the steady-state production rate of profitable species. The energy consumption is used to calculate the electricity cost at a given set point, while the production rate is used to calculate the total profit at a given set point. The optimal set points found from the economic analysis are used to choose relevant operating conditions to test the con-

trollability of the reactor under PI control. To calculate the energy consumption, the total current is approximated by the total rate of consumption of electrons in the reduction reactions. Therefore, the current is given by the equation

$$i_{tot} = \sum_{j=1}^{m} F A e_j P_j \tag{5.6}$$

where i_{tot} is the total current, P_j is the molar production rate of species j, e_j is the number of electrons transferred per mole of species j produced (the values are given in Table 5.5), and m is the total number of species produced. F and A are the Faraday constant and electrode area, respectively, as defined in Section 2.2. Eq. 5.6 gives the total current used to produce the relevant products. Then, the energy consumption is simply the equation for electrical energy consumption which is

$$E = i_{tot} V t \tag{5.7}$$

where V is the applied potential and t is time. In this case, the time t is simply a basis for the operating duration, and it does not affect the results of the optimization problem. Therefore, the time is taken to be 24 hours (86,400 seconds) of continuous operation. Since these results are used for future control experiments, it is useful to express the results in terms of the manipulated variable and controlled variable that will be used in these experiments. In this case, the manipulated variable is the applied potential, and the controlled variable is the ethylene concentration in the reactor outlet stream, as measured by the GC. Since the flowrate is assumed to be constant in the reactor, the FNN-predicted production rate can be converted to outlet concentration using the unit conversion

$$M_j = \frac{P_j A V_g}{F_0} \tag{5.8}$$

where M_j is the molar concentration of species j, V_g is the standard molar volume for gases, and F_0 is the feed flow rate of CO_2.

Table 5.4: Parameters of economic evaluation.

notations	**value**	**unit**
A	3	cm^2
F	96485.3	$C \cdot mol^{-1}$
V_g	22.4	$L \cdot mol^{-1}$
F_0	0.02	$L \cdot min^{-1}$
V	variable	V
I	variable	A
E	variable	W
M	variable	ppm

To perform the economic analysis, it is assumed that electrical energy consumption is the only major operating cost. Similarly, we approximate the revenue of the reactor from the sale of the products without considering additional factors such as the cost of separation or purification. Therefore, the optimal set point for operating the electrochemical reactor can be determined by solving the following optimization problem:

$$\mathcal{J} = \arg\max_{\hat{x} \in \mathbf{D}} \; R(\hat{x}, V) - C(i_{tot}, V) \tag{5.9a}$$

$$s.t. \quad C(i_{tot}, V) = c_e \times E(i_{tot}, V) \tag{5.9b}$$

$$R(\hat{x}, V) = \sum_{j=1}^{m} c_j \times \hat{x_j} \tag{5.9c}$$

$$r = 100 \tag{5.9d}$$

$$-1.5 \leq V \leq -1.27 \tag{5.9e}$$

where $\mathbf{D} \subset \mathbb{R}^m$ is the bounded space of the production rates determined from the training dataset of the FNN model [17]. The vector $\hat{x}$ denotes the predicted production rates given by the FNN model. c_e is the electricity price (\$/kWh), and $c_j,\ (j = 1, \ldots, m)$ is the sale price for the jth product listed in Table 5.5. In this study, the rotation speed of the working electrode is set to

be constant at 100 rpm, and the surface potential is bounded from -1.5 V to -1.27 V shown as Eqs. 5.9d and 5.9e. Eqs. 5.9b and 5.9c use the FNN prediction to approximate the revenue and cost of operating this reactor. i_{tot} and E are given by Eq. 5.6 and Eq. 5.7, respectively.

The optimization problem is solved using Ipopt, an open-source software for large-scale optimization problems. In this work, we use the forward finite difference method to approximate the first-order derivatives of the optimization problems by adding small steps, Δu, on the optimized variables (i.e., potential and rotation speed). Additionally, the second-order derivatives are approximated with the Quasi-Newton method, to provide information for determining the search directions [31]. Subsequently, the derivatives and constant parameters (e.g. products and electricity prices) are provided to Ipopt to optimize the operating conditions in terms of surface potential and rotation speed. Finally, the result conditions are converted to the corresponding ethylene concentration set point (**c**) by using Eq. 5.10, where F_0 is the gas inlet flowrate (0.2 $L \cdot min^{-1}$) and a is a constant for unit conversion (1000000).

$$\mathbf{c} = \frac{aP_iAV_g}{F_0} \tag{5.10}$$

The optimization is performed for various electricity costs ranging from 0.02 to 0.03 $/kWh. As shown in Fig. 5.7a, the profit is plotted against the ethylene set point at various electricity prices. Fig. 5.7b shows the relationship between ethylene concentration and surface potential at a rotation speed of 100 rpm. In the plot, the optimal set point shifts to a lower ethylene concentration when the electricity price increases. This illustrates the balance between selectivity and production rate in this optimization. When electricity prices are cheap, the profit is maximized by maximizing the production rate of all products, which is shown by the optimum point appearing at more negative potentials. As the price of electricity increases, the profit is maximized by increasing the selectivity of the more expensive oxygenate species, which is shown by the optimum appearing at less negative potentials. From Fig. 5.7a, the profit decreases with increasing electricity price, but the

Table 5.5: Chemical information of products.

Index	Products	Number of Transferred Electrons	Chemical Formula
1	methane	8	CH_4
2	ethylene	12	C_2H_4
3	methanol	6	CH_3OH
4	ethanol	12	C_2H_5OH
5	acetate	8	CH_3COO^-
6	ethylene glycol	10	$(CH_2OH)_2$
7	glycolaldehyde	8	$HOCH_2CHO$
8	acetaldehyde	10	CH_3CHO
9	n-propanol	18	C_3H_7OH
10	allyl alcohol	16	C_3H_5OH
11	acetone	16	CH_3COCH_3
12	propionaldehyde	16	C_2H_5CHO
13	carbon monoxide	2	CO

optimization is still able to find the maximum profit for a given price, implying that the optimizer is capable of making intelligent decisions to reduce the production rate with higher operating costs. Based on the results of this optimization, these optimal operating conditions are utilized for future control experiments to demonstrate the controller performance when handling changes in the operating set point. This would reflect industrial scenarios in which the electricity price changes due to Time-of-Use (TOU) pricing or seasonal price changes.

5.5 RGA Analysis using the MLE-FNN

In addition to the determination of optimal set points, future control experiments will be conducted using multi-input-multi-output (MIMO) control schemes. For MIMO control, the number of controlled variables (outputs) must be the same as the number of manipulated variables (inputs) to ensure the control scheme has enough degrees of freedom to meet all of the set point specifications. The simplest MIMO control scheme is to use multiple PI control loops (multi-PI control),

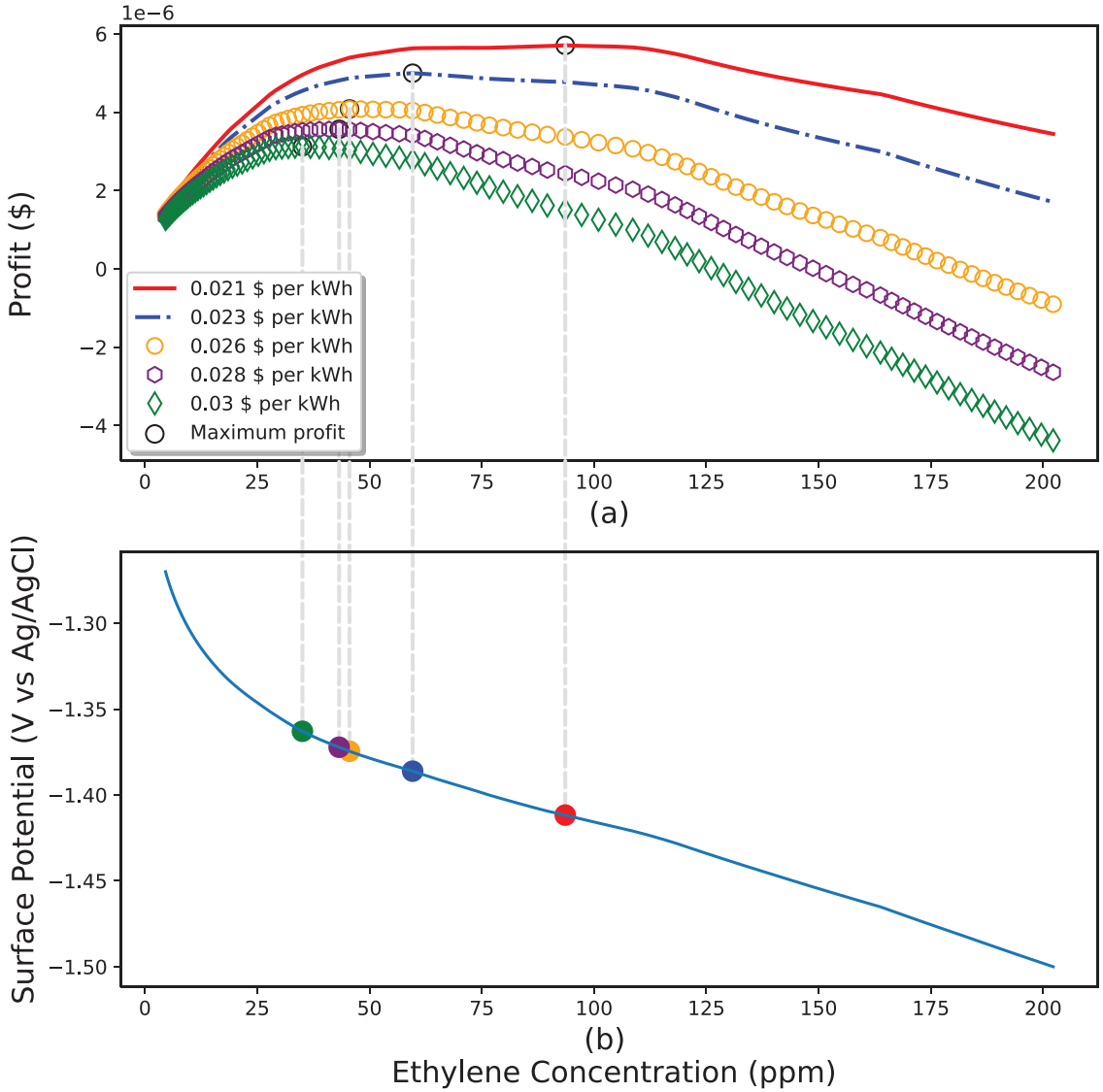


Figure 5.7: The approximated daily profit to operate the electrochemical reactor under various ethylene set points with changing electricity price. (a) The approximated daily profit profile to operate the reactor under different electricity costs (USD). The open black points are the maximum profits that can be obtained by operating the reactor under respective electricity prices. (b) The ethylene concentration profile under various surface potential conditions. The solid color points emphasize the optimum ethylene set points that give the maximum profit.

where each PI controller is assigned one manipulated variable to control one output variable, so this constitutes one input-output variable pair. Multi-PI control becomes difficult if the controlled variables are affected by more than one manipulated variable. This will cause some PI controllers to influence other output variables besides their assigned variable. Relative gain array analysis is used to determine the interaction between multiple control loops by comparing the relative effect that each manipulated variable has on the controlled variables. Specifically, the RGA will be calculated using the steady-state gain matrix method with the MLE-FNN serving as the steady-state model. The RGA analysis procedures are described in Section 2.3.

The goal of the RGA analysis is to determine the combination of controlled variables that have the least interaction. The RGA analysis will also determine whether multiple PI control loops are sufficient for control or if nonlinear control schemes, such as MPC, are required. In this case, two inputs will be used to control two outputs, so two control loops are required. For the RCE reactor, the two manipulated variables are the applied potential and the rotation speed, since both can be manipulated at the per-second timescale. The controlled variables can be any of the gas products measured by the GC during the experiment. These gas products are methane, ethylene, and carbon monoxide. As described by Stephanopoulos [29], an RGA is constructed for all three combinations of two controlled variables—specifically, methane and ethylene, methane and carbon monoxide, and ethylene and carbon monoxide. For each combination, the first output variable is paired with applied potential, and the second output variable is paired with rotation speed. Additionally, for nonlinear systems, the RGA values change at each set point, since the relative effect of each manipulated variable will change as a function of the other variables. Therefore, the three RGAs are also calculated at each set point within the range of applied potentials (-1.5 V to -1.25 V) and rotation speeds (100 rpm to 800 rpm). This gives a map of the controllable region for each combination of controlled variables. As discussed in Section 2.3, only one value is needed in order to specify all the values of a two-by-two RGA. Therefore, only the first element of each array is calculated for each control pair at each set point. Each set point is considered to have stable control

if the first element is near 1, specifically between 0.65 and 2. Alternatively, the set point is considered stable if the first element is near 0, specifically between -1 and 0.35, because this signifies that the control is stable for the reversed input-output pairs. The steady-state gain matrix is calculated using the MLE-FNN predictions. Since the FNN model does not have an analytical solution, the partial derivatives are calculated using the fourth-order centered difference approximation. The results are shown in Figs. 5.8, 5.9, and 5.10.

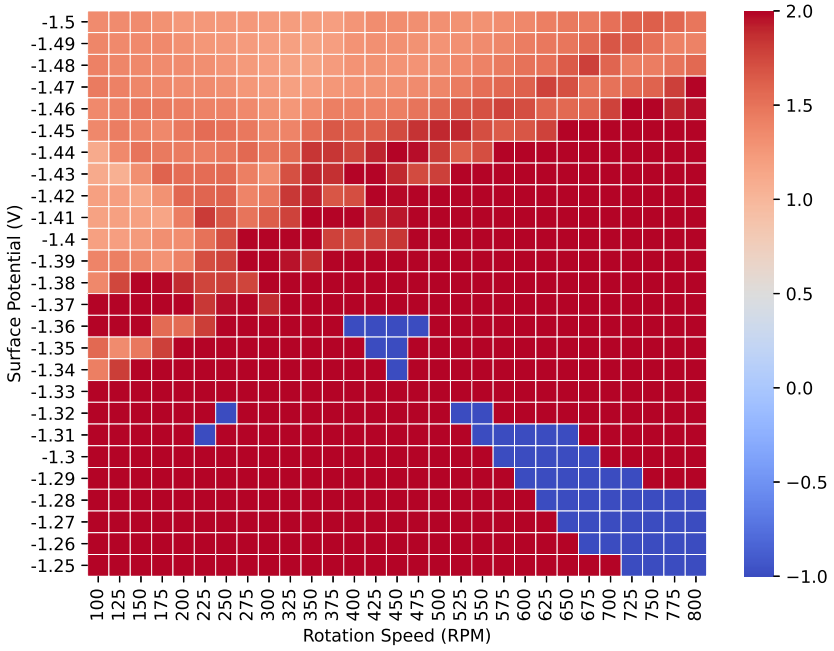


Figure 5.8: The relative gain of the control loop combination of methane and ethylene under various steady-state conditions. The relative gain describes the control stability when methane is paired with applied potential and when ethylene is paired with rotation speed.

From the figures, it can be seen that the controllable region for each combination is different. The controllable regions correspond to the areas where the RGA value is between 0.65 and 2 or between -1 and 0.35 (light red or light blue, respectively). Fig. 5.8 shows the interactions between the methane and ethylene control loops, Fig. 5.9 shows the interactions between the methane and carbon monoxide control loops, and Fig. 5.10 shows the interactions between the ethylene and

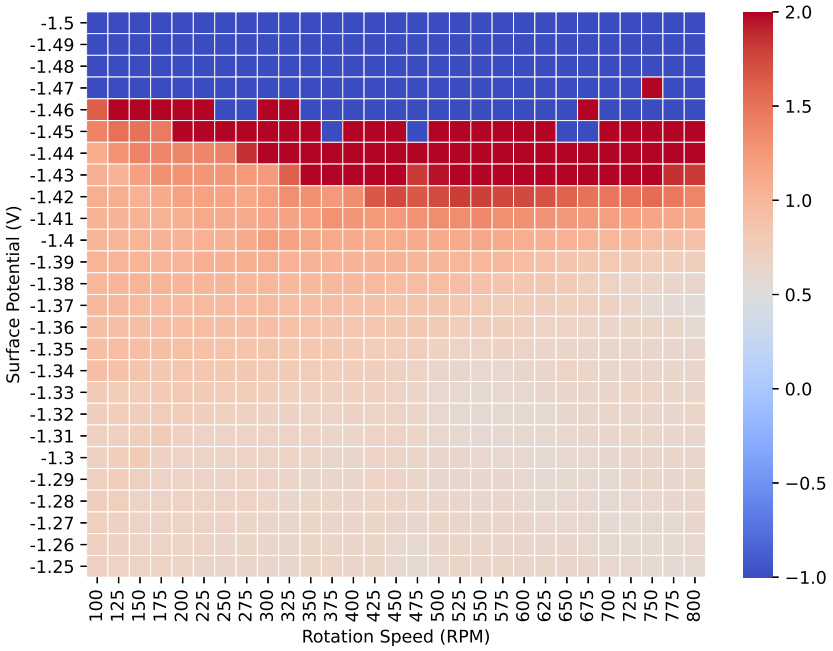


Figure 5.9: The relative gain of the control loop combination of methane and carbon monoxide under various steady-state conditions. The relative gain describes the control stability when methane is paired with applied potential and when carbon monoxide is paired with rotation speed.

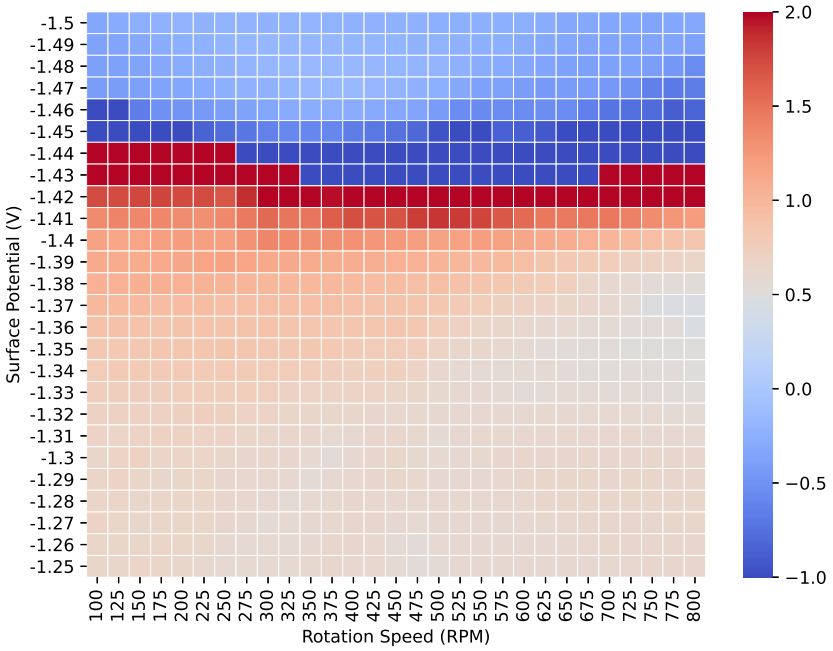


Figure 5.10: The relative gain of the control loop combination of ethylene and carbon monoxide under various steady-state conditions. The relative gain describes the control stability when ethylene is paired with applied potential and when carbon monoxide is paired with rotation speed.

carbon monoxide control loops. In Fig. 5.8, the controllable region only occurs at more negative potentials or lower rotation speeds. The lighter colors in the top left corner correspond to the value of the RGA being around 1. At less negative potentials and higher rotation speeds, the RGA value fluctuates dramatically as it takes on large positive and negative values, meaning that the control interactions are very sensitive to the potential and rotation speed in this regime. In Fig. 5.9, the pair is controllable at potentials from -1.25 V to -1.41 V and for all rotation speeds. At potentials more negative than -1.41 V, the RGA fluctuates from large positive to large negative values, so the control interactions are unstable. In Fig. 5.10, there are two controllable regions, but the control pairing is different for each region. Similarly to the combination of methane and carbon monoxide, the control is stable from -1.25 V to -1.41 V and for all rotation speeds. In addition, the control is stable at applied potentials between -1.46 V and -1.50 V, but the input-output control pairs must be swapped to be stable. At less negative potentials, the control is stable when ethylene is paired with the applied potential, but at more negative potentials, the control is stable when ethylene is paired with rotation speed.

The results of the RGA analysis show that this electrochemical reactor is highly nonlinear, so multi-input-multi-output control is complex. The nonlinear relationships between the manipulated variables and the output variables mean that the control loop interactions change as a function of the set point. The relative gain array analysis shows that multi-input-multi-output control is possible for all output variable combinations, but each combination restricts the operating conditions under which multi-loop control is stable. Therefore, it is possible to do MIMO control with any of the three combinations, so the combination should be selected based on the desired operating region. For the purpose of controlling this system under all possible operating conditions, it is recommended to use a nonlinear controller, such as model-predictive control, since each multi-PI control loop combination has regions of unstable control.

Chapter 6

Conclusion

This work demonstrated the application of neural network modeling to capture the steady-state input-output relationships of operating variables for an electrochemical reactor. Since the electrochemical reduction of carbon dioxide and the overall chemical reaction pathways are not well understood, the kinetic and empirical, first-principle models cannot fully capture the physico-chemical 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. Additionally, a statistical FNN model was developed by utilizing the maximum likelihood estimation method to account for the experimental data variability, and its predictive performance was demonstrated over a broad range of operating conditions. The FNN model was then used for various application such as for improving empirical models, optimizing operation parameters, and determining stable regions for control.

Bibliography

[1] M. Bangi and J. Kwon. Deep hybrid modeling of chemical process: Application to hydraulic fracturing. *Computers and Chemical Engineering*, 134:106696, 2020.

[2] M. Bangi and J. Kwon. Deep reinforcement learning control of hydraulic fracturing. *Computers and Chemical Engineering*, 154:107489, 2021.

[3] A.. Bard and L. Faulkner. *Electrochemical Methods: Fundamentals and Applications*. John Wiley & Sons, New York, 2nd edition, 2001.

[4] G. Box and N. Draper. *Empirical model-building and response surfaces.* John Wiley & Sons, New York, 1987.

[5] Y. Ding, Y. Zhang, Y. Ren, G. Orkoulas, and P. D. Christofides. Machine learning-based modeling and operation for ALD of SiO_2 thin-films using data from a multiscale CFD simulation. *Chemical Engineering Research and Design*, 151:131–145, 2019.

[6] S. Dorling, R. Foxall, D. Mandic, and G. Cawley. Maximum likelihood cost functions for neural network models of air quality data. *Atmospheric Environment*, 37(24):3435–3443, 2003.

[7] J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982.

[8] V. Huzurbazar. The likelihood equation, consistency and the maxima of the likelihood function. *Annals of Eugenics*, 14(1):185–200, 1947.

[9] R. Inapakurthi, S. Miriyala, and K. Mitra. Recurrent neural networks based modelling of industrial grinding operation. *Chemical engineering science*, 219:115585, 2020.

[10] R. Inapakurthi, S. Miriyala, and K. Mitra. Deep learning based dynamic behavior modelling and prediction of particulate matter in air. *Chemical Engineering Journal*, 426:131221, 2021.

[11] M. Jadid and D. Fairbairn. Neural-network applications in predicting moment-curvature parameters from experimental data. *Engineering Applications of Artificial Intelligence*, 9(3):309–319, 1996.

[12] J. Jang, M. Rüscher, M. Winzely, and C. Morales-Guio. Gastight rotating cylinder electrode: Towards decoupling mass transport and intrinsic kinetics in electrocatalysis. *AIChE Journal* , page submitted, 2021.

[13] G. Kimaev and L. A. Ricardez-Sandoval. Nonlinear model predictive control of a multiscale thin film deposition process using artificial neural networks. *Chemical Engineering Science*, 207:1230–1245, 2019.

[14] G. Kimaev and L. A. Ricardez-Sandoval. Artificial neural network discrimination for parameter estimation and optimal product design of thin films manufactured by chemical vapor deposition. *The Journal of Physical Chemistry C*, 124(34):18615–18627, 2020.

[15] M. Kumar, D. Garg, and R. Zachery. Intelligent sensor modeling and data fusion via neural network and maximum likelihood estimation. *Proceedings of ASME International Mechanical Engineering Congress and Exposition*, 42169:1759–1768, 2005.

[16] C. Lee. Fuzzy logic in control systems: fuzzy logic controller. I. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(2):404–418, 1990.

[17] J. Luo, V. Canuso, J. Jang, Z. Wu, C. Morales-Guio, and P. D. Christofides. Machine learning-based operational modeling of an electrochemical reactor: Handling data variability and improving empirical models. *Industrial & Engineering Chemistry Research, in press*, 2022.

[18] A. Malek, Q. Wang, S. Baumann, O. Guillon, M. Eikerling, and K. Malek. A data-driven framework for the accelerated discovery of CO_2 reduction electrocatalysts. *Frontiers in Energy Research*, 9:52, 2021.

[19] S. Miriyala, P. Mittal, S. Majumdar, and K. Mitra. Comparative study of surrogate approaches while optimizing computationally expensive reaction networks. *Chemical Engineering Science*, 140:44–61, 2016.

[20] C. Morales-Guio, E. Cave, S. Nitopi, J. Feaster, L. Wang, K. Kuhl, A. Jackson, N. Johnson, D. Abram, T. Hatsukade, and et al. Improved CO_2 reduction activity towards C_{2+} alcohols on a tandem gold on copper electrocatalyst. *Nature Catalysis*, 1(10):764–771, 2018.

[21] I. Myung. Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47(1):90–100, 2003.

[22] Y. Nishimura, H. Peng, S. Nitopi, M. Bajdich, L. Wang, C. Morales-Guio, F. Abild-Pedersen, T. Jaramillo, and C. Hahn. Guiding the catalytic properties of copper for electrochemical CO_2 reduction by metal atom decoration. *ACS Applied Materials & Interfaces*, 13(44):52044–52054, 2021.

[23] S. Nitopi, E. Bertheussen, S. Scott, X. Liu, A. Engstfeld, S. Horch, B. Seger, I. Stephens, K. Chan, C. Hahn, and et al. Progress and perspectives of electrochemical CO_2 reduction on copper in aqueous electrolyte. *Chemical reviews*, 119(12):7610–7672, 2019.

[24] P. Raccuglia, K. Elbert, P. Adler, C. Falk, M. Wenny, A. Mollo, M. Zeller, S. Friedler, J. Schrier, and A. Norquist. Machine-learning-assisted materials discovery using failed experiments. *Nature*, 533(7601):73–76, 2016.

[25] D. Rangel-Martinez, K. Nigam, and L. A. Ricardez-Sandoval. Machine learning on sustainable energy: A review and outlook on renewable energy systems, catalysis, smart grid and energy storage. *Chemical Engineering Research and Design*, 174:414–441, 2021.

[26] J. Riggs. *Chemical Process Control*. Ferret Publishing, Texas, 1999.

[27] B. Shahriari, K. Swersky, Z. Wang, R. Adams, and N. De Freitas. Taking the human out of the loop: A review of bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2015.

[28] Q. Shen, B. Jiang, P. Shi, and C. Lim. Novel neural networks-based fault tolerant control scheme with fault alarm. *IEEE Transactions on Cybernetics*, 44(11):2190–2201, 2014.

[29] G. Stephanopoulos. *Chemical Process Control: An Introduction to Theory and Practice*. Prentice-Hall Inc., New Jersey, 1984.

[30] R. Vepa. A review of techniques for machine learning of real-time control strategies. *Intelligent Systems Engineering*, 2(2):77–90, 1993.

[31] A. Wächter and L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1):25–57, 2006.

[32] P. Whittle. Tests of fit in time series. *Biometrika*, 39(3/4):309–318, 1952.

[33] P. Woolf and et al. *Chemical Process Dynamics and Controls*. LibreTexts Project, Michigan, 2022.

[34] J. Wu, X. Chen, H. Zhang, L. Xiong, H. Lei, and S. Deng. Hyperparameter optimization for machine learning models based on bayesian optimization. *Journal of Electronic Science and Technology*, 17(1):26–40, 2019.

[35] Z. Wu, J. Luo, D. Rincon, and P. D. Christofides. Machine learning-based predictive control using noisy data: evaluating performance and robustness via a large-scale process simulator. *Chemical Engineering Research and Design*, 168:275–287, 2021.

[36] Z. Wu, A. Tran, Y. Ren, C. Barnes, S. Chen, and P. D. Christofides. Model predictive control of phthalic anhydride synthesis in a fixed-bed catalytic reactor via machine learning modeling. *Chemical Engineering Research and Design*, 145:173–183, 2019.

[37] Z. Wu, A. Tran, D. Rincon, and P. D. Christofides. Machine learning-based predictive control of nonlinear processes. part i: Theory. *AIChE Journal*, 65(11):e16729, 2019.

[38] L. Zadeh. Probability measures of fuzzy events. *Journal of mathematical analysis and applications*, 23(2):421–427, 1968.