Kinetic study of product distribution using various data-driven and statistical models for Fischer-Tropsch synthesis

Yixiao Wang
Jing Hu
Xiyue Zhang
Abubakar Yusuf
Binbin Qi

See next page for additional authors

Follow this and additional works at: https://ro.ecu.edu.au/ecuworkspost2013

Part of the Engineering Commons

10.1021/acsomega.1c03851
This Journal Article is posted at Research Online.
https://ro.ecu.edu.au/ecuworkspost2013/11360
Authors
Yixiao Wang, Jing Hu, Xiyue Zhang, Abubakar Yusuf, Binbin Qi, Huan Jin, Yiyang Liu, Jun He, Yunshan Wang, Gang Yang, and Yong Sun

This journal article is available at Research Online: https://ro.ecu.edu.au/ecuworkspost2013/11360
Kinetic Study of Product Distribution Using Various Data-Driven and Statistical Models for Fischer–Tropsch Synthesis

Yixiao Wang, Jing Hu, Xiyue Zhang, Abubakar Yusuf, Binbin Qi, Huan Jin, Yiyang Liu, Jun He, Yunshan Wang, Gang Yang, and Yong Sun

1. INTRODUCTION

The year 2020 witnessed how weak the supply chain is in how goods (the empty grocery-store shelves and the out of stock of surgical masks) are sourced, distributed, and where they are stored.1 These surely ring the alarm bell for the acceleration of long-term shifts in the supply chain on various aspects.2 Energy security issues are also increasing as crude oil prices are falling, the petrochemical production from crude oil is still in demand as the pendulum of daily life does not stop and the consumption of fast-moving consumer goods (FMCG) bounces back quickly.3,4 The challenging issues of importing crude oils from oil-rich countries through the weak supply chain still exist.5,6 Although the energy demand of using crude oil is significantly falling, the petrochemical production from crude oil is still in demand as the pendulum of daily life does not stop and the consumption of fast-moving consumer goods (FMCG) bounces back quickly.5,6 The challenging issues of importing crude oils from oil-rich countries through the weak supply chain still exist.

The alternative technology of using syngas to produce hydrocarbons (both for bulk/fine chemical manufacturing and energy consumption) via one of the most commercial viable process (Fischer–Tropsch, FT synthesis) shows its merit in tackling the thorny problems for many countries with poor oil reserve.7–10 The kinetic modeling of Fischer–Tropsch (FT) synthesis is the cornerstone for sizing and process design of gas to liquid (GTL) technology.11–13 Many scholars have proposed different modeling approaches to model the kinetics.14–16 These models include (i) data regression by formulating the minimization of the sum of the square of errors through a numerical Levenberg–Marquardt (LM) algorithm, (ii) a genetic algorithm (GA) to model the product distribution of FT synthesis such as using the Anderson–Schulz–Flory (ASF) method, (iii) an explicit model based on a second-order polynomial for correlating the critical operational parameters using a response surface methodology (RSM), and (iv) data-driven soft computation machine learning techniques such as using different artificial neuron networks (ANNs) in directly correlating the product distribution and critical operational parameters.17–24
With the advancement of computation techniques, computational power has become more economical and practical, and this looms new opportunities of combining different modeling techniques to analyze the complicated kinetic system such as FT synthesis.25–27 Until now, there have been very rare efforts in reporting the establishment of a robust paradigm that combines the advanced calculation systems for kinetic modeling of product distribution during FT synthesis.

To achieve this goal, large experimental data sets (a 29 × 36 matrix with total 1073 data sets, of which inputs were operational parameters such as temperature (T), pressure (P), gas hour space velocity (GHSV), and syngas ratio (SR), outputs were the rate of olefin/paraffin from C1−C15, respectively) were fed into different modeling techniques, namely, a radial basis function neural network (RBFNN), the comprehensive kinetics based on a genetic algorithm (CKGA) for the regional and global optimization using a genetic algorithm (GA) via the objective function (minimization of the sum of the square of errors) for mechanism selection using the Langmuir–Hinshelwood—Hougen—Watson (LHHW) approach, and the second-degree polynomial correlations through a response surface methodology (RSM), and explored and compared. Using additional validation data sets (15 × 36 matrix with total 540 data sets), the comparisons of these three techniques (RBFNN, CKGA, and RSM) were made, and a paradigm of accurate determination and prediction of reaction pathways and optimization of the system is proposed. To the best of the author’s knowledge, this new modeling strategy for product distribution during FT synthesis has been rarely reported before.

2. THEORETICAL BACKGROUND

2.1. Radial Basis Function Neural Network (RBFNN). Artificial neural networks (ANNs) are well-developed and applied soft computing technique.28 With the advancement of computation technology, these soft computing techniques have recently remerged and begun to attract attention in both academia and industries for complex system modeling.29 Inspired from the neurological system and its basic elements for handling the information, ANNs are principally constructed as an input function x of the formal neuron i corresponding to the incoming activity (e.g., synaptic input) of the biological neuron; weight wij represents the effective magnitude of information transmission between neurons (e.g., determined by synapses), activation function zij = f(xijwij) describes the main computation performed by a biological neuron (e.g., spike rates), and the output function yij = f(zij) corresponds to the overall activity transmitted to the next neuron in the processing stream.30 With the learning phase on prior representative data as a robust training and testing step, this constructed learning technique is recognized to yield good predictions in many complex systems. From configurations of neural topologies and learning procedure perspectives, different ANNs have been constructed and deployed. Among them, the multilayer perceptron (MLP) and radial basis function neural networks (RBFNNs) are one of the well-formulated and rigorous ANNs.31,32 In this work, we choose the RBFNN as the machine learning algorithm for product distribution predictions during FT synthesis. The rationale of using RBF lies in its relative simple neuron architecture and efficient computation. The schematic diagram of RBFNN is shown in Figure 1.

The RBFNN is unique in its neuron topology, which only possesses one hidden layer representing its nodes Nh

\[
\phi(z_c) = \exp \left[ \frac{\sum_{i=1}^{n} (x_i - c_{ik})^2}{2\sigma^2} \right]
\]

The cross-validation was employed for the evaluation of the statistical performance of the established model. The number of folds was set at 17, which results in exactly one data item being a test item during each fold iteration.33 The detailed procedures for cross-validation could be found from our previous literature reports.34 The mean square error (MSE) and mean average residual relation (MARR) can be calculated as follows

![Figure 1. Schematic diagram of RBFNN formulated from the Gaussian function.](https://doi.org/10.1021/acsomega.1c03851)
Table 1. Elementary Steps for Different Mechanisms,\textsuperscript{16,38,40–42} Where M1 Refers to the Carbide Mechanism, M2 Refers to the Enolic Mechanism, and M3 Refers to the CO Insertion Mechanism (with $n \geq 4$), A Refers to Alkylidene and S Refers to the Active Site on the Surface of the Catalyst

<table>
<thead>
<tr>
<th>Constant</th>
<th>M1</th>
<th>Constant</th>
<th>M2</th>
<th>Constant</th>
<th>M3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{H2}^{M1}$</td>
<td>$H_2 + 2S \leftrightarrow 2HS$</td>
<td>$K_{H2}^{M2}$</td>
<td>$CO + 2S \leftrightarrow COS$</td>
<td>$K_{H2}^{M3}$</td>
<td>$H_2 + 2S \leftrightarrow 2HS$</td>
</tr>
<tr>
<td>$K_{O2}^{M1}$</td>
<td>$CO + S \leftrightarrow COS$</td>
<td>$K_{CO}^{M2}$</td>
<td>$CO + H_2 \leftrightarrow H_2CO - S$</td>
<td>$K_{CO}^{M3}$</td>
<td>$CO + C, H_{2n+1} - S \rightarrow C, H_{2n+1}CO - S$</td>
</tr>
<tr>
<td>$K_{H2O}^{M1}$</td>
<td>$H_2O + S \leftrightarrow H_2O - S$</td>
<td>$K_{H2O}^{M2}$</td>
<td>$H_2 + 2S \leftrightarrow 2HS$</td>
<td>$K_{H2O}^{M3}$</td>
<td>$C, H_{2n+1}CO - S + H_2 \leftrightarrow C, H_{2n+1}C - S + H_2O$</td>
</tr>
<tr>
<td>$k_{O2}$</td>
<td>$O_2 + S \leftrightarrow O_2 - S$</td>
<td>$k_{CO}$</td>
<td>$H_2C - S + C, H_{2n+1} - S \rightarrow C, H_{2n+1}CH_2 - S + S$</td>
<td>$k_{CO}$</td>
<td>$C, H_{2n+1}C - S + H_2 \rightarrow C, H_{2n+1}CH_2 - S$</td>
</tr>
<tr>
<td>$k_{O}$</td>
<td>$CO - S + S \rightarrow C - S + O - S$</td>
<td>$k_{CH}$</td>
<td>$CH_3 - S + H - S \rightarrow CH_4 + 2S$</td>
<td>$k_{CH}$</td>
<td>$CH_4 - S + H_2 \rightarrow CH_4 + HS$</td>
</tr>
<tr>
<td>$k_{CH}$</td>
<td>$C - S + H - S \rightarrow CH - S + S$</td>
<td>$k_{C}$</td>
<td>$C, H_{2n+1} - S + H - S \rightarrow C, H_{2n+2} + S$</td>
<td>$k_{C}$</td>
<td>$C, H_{2n+1} - S + H_2 \rightarrow C, H_{2n+2} + HS$</td>
</tr>
<tr>
<td>$k_{CH}$</td>
<td>$CH - S + H - S \rightarrow CH_2 - S + S$</td>
<td>$k_{C}$</td>
<td>$C, H_{2n+1} - S \rightarrow C, H_{2n} + HS$</td>
<td>$k_{C}$</td>
<td>$C, H_{2n+1} - S \rightarrow C, H_{2n} + HS$</td>
</tr>
</tbody>
</table>
In this work, the procedure of model development involved the identification of carbide, enolic, and CO insertion mechanisms. The procedures were repeated twice to ensure statistical relevance to the experimental results. Once the rate constant can be expressed as:

\[
E_r^n = E_r^0 + nE_r
\]

where \(E_r^n\) refers to the adsorption energy that is independent of the chain length (kJ mol\(^{-1}\)), \(E_r\) refers to the part of activation energy that is dependent on the chain length (kJ mol\(^{-1}\)), and \(n\) refers to the carbon number (\(n \geq 4\)). Then, the rate constant can be expressed as:

\[
k_{\text{CH}}^n = k_w e^{-E_r^n/\theta RT}
\]

\[
k_{\text{CH}}^n = k_w e^{-E_r^n/\theta RT} = k_w e^{-E_r^n/\theta RT} e^{-\varepsilon_r^0}
\]

The rate constants of chain-length-dependent (CLD) re-adsorption and desorption can be expressed as follows:

\[
k_{\text{re}}^n = k_{\text{re}} e^{-E_r^n/\theta RT}
\]

\[
k_{\text{re}}^n = k_{\text{re}} e^{-E_r^n/\theta RT} = k_{\text{re}} e^{-E_r^n/\theta RT} e^{-\varepsilon_r^0}
\]

where \(\varepsilon_r^0\) and \(\varepsilon_r^\text{ad}\) are the constants reflecting the part of activation energy of olefin re-adsorption and desorption with carbon number dependency and \(\theta\) refers to the gas constant (8.314 J K\(^{-1}\) mol\(^{-1}\)). For FT synthesis, the surface fraction of CO, H\(_2\), and O\(_2\) (the olefin chain-length-dependent re-adsorption is considering eqs 13–15, with \(n \geq 4\)) can be expressed as follows:

\[
\theta_{\text{CHS}} = \frac{k_{\text{CH}}^n \theta_{\text{CHS}}}{k_{\text{CHS}}}
\]

Calling quasi-steady-state assumption (QSSA)

\[
\theta_{\text{CS}} = \frac{k_{\text{CSS}}^n \theta_{\text{CS}}}{k_{\text{CS}}^n} = \frac{\varepsilon_{\text{CS}} \varepsilon_{\text{CS}}}{\varepsilon_{\text{CS}}}
\]

where

\[
\varepsilon_{\text{CS}} = \frac{k_{\text{CSS}}^n \theta_{\text{CSS}}}{k_{\text{CSS}}^n}
\]

The QSSA is applied with a chain number longer than 4

\[
\sum_{n=1}^{\infty} \theta_{\text{A}-n-S} = \frac{k_{\text{IN}}^n \theta_{\text{CHS}}}{k_{\text{IN}}^n \theta_{\text{CHS}}} + \frac{k_{\text{CHS}}^n \theta_{\text{CSS}}}{k_{\text{CHS}}^n \theta_{\text{CSS}}} + \frac{k_{\text{CSS}}^n \theta_{\text{CSS}}}{k_{\text{CSS}}^n \theta_{\text{CSS}}} \geq 4
\]

The approximate second-order polynomial expression using \(\theta_{\text{CHS}}\) as variables can be solved as:

\[
A\theta_{\text{CHS}}^2 + B\theta_{\text{CHS}} + C = 0
\]

where

\[
A = \frac{k_{\text{IN}}^n k_{\text{A}}}{k_{\text{IN}}^n + k_{\text{A}}\theta_{\text{CSS}}}
\]

\[
B = k_{\text{IN}}^{n}\varepsilon + \frac{k_{\text{IN}}^n \theta_{\text{CSS}}}{k_{\text{IN}}^n + k_{\text{CSS}}^n \theta_{\text{CSS}}}
\]

\[
C = k_{\text{IN}}^n \varepsilon_{\text{CSS}}^n
\]

The corresponding root could be solved as
Using QSSA in alkylidene-1-S

\[ \theta_{A-1-S} = \frac{k_{\text{INCH}_{1-S}} \theta_{S}}{k_{\text{CH}_{1-S}} \theta_{S} + k_{\text{CH}_{1-S}} \theta_{S}} = \frac{k_{\text{INCH}_{1-S}} \theta_{S}}{k_{\text{CH}_{1-S}} \theta_{S} + k_{\text{CH}_{1-S}} \theta_{S}} \]

\[ \theta_{A-1-S} = \frac{k_{\text{INCH}_{1-S}} \theta_{S}}{k_{\text{CH}_{1-S}} \theta_{S} + k_{\text{CH}_{1-S}} \theta_{S}} = \frac{k_{\text{INCH}_{1-S}} \theta_{S}}{k_{\text{CH}_{1-S}} \theta_{S} + k_{\text{CH}_{1-S}} \theta_{S}} \]

where

\[ \tilde{\theta}_{1} = \frac{k_{\text{INCH}_{1-S}} \theta_{S}}{k_{\text{INCH}_{1-S}} \theta_{S} + k_{\text{CH}_{1-S}} \theta_{S}} \]

For the alkylidene-2-S intermediate

\[ \theta_{A-2-S} = \frac{k_{A} \theta_{A} \theta_{CH_{-1-S}} + k_{\text{ref}} \theta_{OS} \theta_{HS} \theta_{S}}{k_{\text{ASCH}_{2-S}} + k_{P} \theta_{S} + k_{O}} \]

\[ \theta_{A-2-S} = \frac{k_{A} \theta_{A} \theta_{CH_{-1-S}} + k_{\text{ref}} \theta_{OS} \theta_{HS} \theta_{S}}{k_{\text{ASCH}_{2-S}} + k_{P} \theta_{S} + k_{O}} \]

For alkylidene-3-S

\[ \theta_{A-3-S} = \frac{k_{A} \theta_{A} \theta_{CH_{-1-S}} + k_{\text{ref}} \theta_{OS} \theta_{HS} \theta_{S}}{k_{\text{ASCH}_{2-S}} + k_{P} \theta_{S} + k_{O}} \]
\[
\sigma_1 = \frac{k_{\text{A}^{	ext{CH},S}}}{k_{\text{A}^{	ext{CH},S}} + k_{\text{P}^{	ext{HS},S}} + k_{O}} \text{and} \gamma_3
\]
\[
= \frac{k_{\text{O}}^\text{rev} K_{O} \xi^\text{HS} \theta_S}{k_{\text{A}^{	ext{CH},S}} + k_{\text{P}^{	ext{HS},S}} + k_{O}}
\]

For the alkylidene-\(n\)-S intermediate
\[
\theta_{\lambda-n} = \frac{k_{\text{A}^{	ext{CH},S}} + k_{\text{P}^{	ext{HS},S}} + k_{O}}{k_{\text{A}^{	ext{CH},S}} + k_{\text{P}^{	ext{HS},S}} + k_{O}}
\]
\[
= \frac{k_{\text{A}^{	ext{CH},S}} + k_{\text{P}^{	ext{HS},S}} + k_{O}}{k_{\text{A}^{	ext{CH},S}} + k_{\text{P}^{	ext{HS},S}} + k_{O}} \text{within} \geq 4
\]

The following linear algebra will be casted
\[
\begin{align*}
\forall_1 &= \xi_1 \theta_S \\
\forall_2 &= \sigma_1 \forall_1 + \gamma_2 \theta_S \\
\forall_3 &= \sigma_1 \forall_2 \forall_1 + (\sigma_2 \gamma_3 + \gamma_3) \theta_S
\end{align*}
\]
\[
\forall_n = \sigma_1 \forall_{n-1} \forall_1 + (\sigma_n \gamma_n + \gamma_n) \theta_S
\]
\[
= \prod_{i=1}^{n} \sigma_i \forall_{i-1} \forall_1 + (\sum_{i=4}^{n} \prod_{i=4}^{n} \sigma_i \gamma_i) \theta_S
\]

With assumption of \(\sigma_1 = \sigma_2 = \ldots = \sigma_n \text{ and } \gamma_3 = \gamma_4 = \ldots = \gamma_n\)

\[
\theta_{\lambda-n} = \frac{k_{\text{A}^{	ext{CH},S}} + k_{\text{P}^{	ext{HS},S}} + k_{O}}{k_{\text{A}^{	ext{CH},S}} + k_{\text{P}^{	ext{HS},S}} + k_{O}}
\]
\[
= \frac{k_{\text{A}^{	ext{CH},S}} + k_{\text{P}^{	ext{HS},S}} + k_{O}}{k_{\text{A}^{	ext{CH},S}} + k_{\text{P}^{	ext{HS},S}} + k_{O}} \text{within} \geq 4
\]

Thus, the site balance is expressed as the following
\[
\theta_S = \left[ 1 + \xi_{\text{HS}} + \xi_{\text{S}^2} + \xi_{\text{CH},S} \right] \left[ 1 + \sum_{k=4}^{n} \prod_{k=4}^{n} \sigma_{i-1} \gamma_i \theta_S \right]^{\frac{1}{1}}
\]

The schematic diagram of model selection using a genetic algorithm for data regression is shown in Figure 2. Finally, the rate expressions of the reagent and different products are then expressed as follows with \(n \geq 4\)

\[
r_{\text{CH}_4} = k_{\text{D}^\text{CH}_4} \xi_{\text{HS}} \xi_{\text{CH},S} \theta_S
\]
\[
r_{\gamma} = k_{\gamma} (\sigma_1 \xi_1 + \gamma_1) \xi_{\text{HS}} \theta_S
\]
\[
r_{O} = k_{O} (\sigma_1 \xi_1 + \gamma_1) \xi_{\text{HS}} \theta_S
\]
\[
r_{\text{CO}} = k_{\text{CO}^\text{CO}} \xi_{\text{S}^2} \theta_S
\]

For the alkylidene-\(n\)-S intermediate
\[
\theta_{\lambda-n} = \frac{k_{\text{A}^{	ext{CH},S}} + k_{\text{P}^{	ext{HS},S}} + k_{O}}{k_{\text{A}^{	ext{CH},S}} + k_{\text{P}^{	ext{HS},S}} + k_{O}}
\]
\[
= \frac{k_{\text{A}^{	ext{CH},S}} + k_{\text{P}^{	ext{HS},S}} + k_{O}}{k_{\text{A}^{	ext{CH},S}} + k_{\text{P}^{	ext{HS},S}} + k_{O}} \text{within} \geq 4
\]

With assumption of \(\sigma_1 = \sigma_2 = \ldots = \sigma_n \text{ and } \gamma_3 = \gamma_4 = \ldots = \gamma_n\)

\[
\theta_{\lambda-n} = \frac{k_{\text{A}^{	ext{CH},S}} + k_{\text{P}^{	ext{HS},S}} + k_{O}}{k_{\text{A}^{	ext{CH},S}} + k_{\text{P}^{	ext{HS},S}} + k_{O}}
\]
\[
= \frac{k_{\text{A}^{	ext{CH},S}} + k_{\text{P}^{	ext{HS},S}} + k_{O}}{k_{\text{A}^{	ext{CH},S}} + k_{\text{P}^{	ext{HS},S}} + k_{O}} \text{within} \geq 4
\]
for i and adsorption constants, respectively, activation energy (kJ) were obtained from the model reaction rate constants. In addition, all of the parameters that optimization using the MATLAB toolbox to approximate are experimental and calculated molar rates for species (kJ).

The genetic algorithm (GA) was utilized for global optimization and Levenberg-Marquardt (LM) for local optimization for significance analysis, where NPR refers to a normal plot of residues, RVP refers to residues versus predictions, BCP refers to the Box–Cox plot for power transform, and CD refers to Cook’s distance.

\[
\xi_{\text{CH3}} = \left\{ \frac{k_{\text{in}}(k_{p,p}P_{R_H})^{1/2}}{k_R + (k_{R,k})^{1/2}} \right. \left. + \frac{k_{\text{in}}(k_{p,p}P_{R_H})^{1/2}}{k_R + (k_{R,k})^{1/2}} \right\} \frac{k_{\text{in}}(k_{p,p}P_{R_H})^{1/2}}{k_R + (k_{R,k})^{1/2}} \right\}
\]

where \( k_{\text{in}} \) and \( k_{p,p} \) are experimental and calculated molar rates for species \( i \) in \( j \).

The model parameters were calculated by minimizing the sum of squares of residuals defined by calling the objective function, as follows

\[
E_{\text{obj}} = \frac{1}{N_{\text{exp}}} \sum_{i=4}^{N_{\text{exp}}} \sum_{j=1}^{N_{\text{sam}}} \left| \frac{Y_{ij}^{\text{exp}} - Y_{ij}^{\text{calc}}}{r_{ij}^{\text{exp}}} \right| \times 100
\]

where \( N_{\text{exp}} \) is the number of experimental data and \( r_{ij}^{\text{exp}} \) and \( r_{ij}^{\text{calcd}} \) are experimental and calculated molar rates for species \( i \) in \( j \).

The genetic algorithm (GA) was utilized for global optimization and Levenberg-Marquardt (LM) for local optimization using the MATLAB toolbox to approximate reaction rate constants. In addition, all of the parameters that were obtained from the model fit should also be physically relevant to the Arrhenius and Van’t Hoff adsorption laws, as follows

\[
k_i(T) = k_i \exp \left( \frac{-E_{\text{cat}}}{\mathcal{R} T} \right)
\]

\[
K_i(T) = K_i \exp \left( \frac{-\Delta H_i}{\mathcal{R} T} \right)
\]

where \( k_i \) (mol·kg\text{-cat}⁻¹·h⁻¹) and \( K_i \) (mol·kg\text{-cat}⁻¹·h⁻¹) are reaction and adsorption constants, respectively, \( E_{\text{cat}} \) is apparent activation energy (kJ·mol⁻¹), and \( \Delta H_i \) is heat of adsorption for \( i \) species (kJ·mol⁻¹).

2.3. Response Surface Methodology (RSM). In statistics, response surface methodology explores the relationships between several explanatory variables and one or more response variables. The main idea of RSM is to use a sequence of designed experiments to obtain an optimal response. In addition, it can also be used to directly correlate the influential operational parameters with responses using the most widely deployed second-degree polynomial expressions to find out the offset, linear, quadratic, and interactive terms as follows

\[
Y_i = b_0 + \sum_{i=1}^{4} b_i X_i + \sum_{i=1}^{4} b_i^2 X_i^2 + \sum_{i<j=1}^{4} b_{ij} X_i X_j
\]

where \( Y_i \) is the responded value (in this work, there are 36 different responses including \( \text{X}_{\text{CO2}}, \text{S}_{\text{CH4}}, \text{r}_{\text{CD}} \), and \( \text{r}_{\text{Ox}} \) with \( n \leq 15 \)), \( X_i \) is the binary parameter of the investigated operational parameters (\( T, P, \text{GHSV}, \) and \( \text{SR} \)), and \( b_{ij}, b_i, b_{ii} \) are coefficients from the polynomial expression. Although it does not shed any insightful lights upon the kinetic mechanism, it does provide the useful statistical indications for the significances of responses that are caused by the variations of process variables (either singular or binary). With experimental results (Table 4 in the Experimental Part section), the RSM is capable of calculating the corresponding formation rate of olefin and paraffin with different carbon numbers (Figure 3).

3. RESULTS AND DISCUSSION

3.1. Model Comparisons. Once experimental data (Table 4 from the Experimental Part section) were fed into different models (RBFNN, CKGA, and RSM), model predictions against validation results (Table S1) were used to test the performances of constructed models. Because of a range of differences, the parameters, i.e., \( X_{\text{CO2}}, \text{S}_{\text{CH4}}, \text{r}_{\text{CD}} \), and \( \text{S}_{\text{C2-4}} \) and the formation rate of olefin and paraffin with different carbon numbers \( (\text{r}_{\text{Ox}} \text{P}_{\text{Ox}}) \), with \( n \leq 15 \)) are plotted in Figure 4a,b, respectively. Clearly, with different models, data present appreciably different aggregating patterns both in Figure 4a–c and a2–c2, respectively.

Taking the center aggregation (purple circle with the range from 0 to 0.0006 for both axis), for instance, the percentage of
Figure 4. Comparisons between the experiment and model prediction when the rate is in mole·h\(^{-1}\) for (a) RBFNN model, (b) comprehensive kinetics based on the genetic algorithm (CKGA) model, and (c) RSM-based model, where Exp\(_{rH2}\) and \(rCO\) refer to the experimental hydrogen and CO consumption rate (mole·h\(^{-1}\)), Calcd\(_{rH2}\) and \(rCO\) refer to the calculated hydrogen and CO consumption rate (mole·h\(^{-1}\)), Exp\(_{rPi}\) and \(rOi\) refer to the experimental paraffin and olefin rate with the carbon number \(i\) (mole·h\(^{-1}\)), and Calcd\(_{rPi}\) and \(rOi\) refer to calculated paraffin and olefin rates with the carbon number \(i\) (mole·h\(^{-1}\)).
data falling out of this range from RBFNN, CKGA, and RSM models is 14.2, 16.6, and 18.9%, respectively. Apparently, the patterns of data distribution using different modeling approaches are not same, which will lead to the variation of relative errors between the experiment and predictions. To further assess the deviations between the experimental results and model predictions, the corresponding standard deviations were calculated as follows

\[ \bar{x} = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2} \]  

where \( N \) represents the numbers of experimental runs, \( \bar{x} \) represents the averaged values of each parameters (\( X_{\text{CO}} \), \( S_{\text{H}_2} \), \( r_{\text{CO}} \), \( S_{\text{CO}_2} \), \( S_{\text{C}_2\text{C}_4} \), \( r_{\text{P}} \), \( r_{\text{O}} \) with \( n \leq 16 \)), and \( x_i \) refers to corresponding specific value. Then, the standard error (SE) of each experimental value is defined as follows

\[ SE = \frac{\bar{x}}{\sqrt{N}} \]  

The errors between experiment and model predictions using three different models are shown in Figure 5, with the color bar from blue to red indicating the variations of standard errors from ±5 to ±20% in validation tests. The standard errors for majority data are less than ±10% with the presence of some exceptionally high errors (over ±20%) for all of three models (RBFNN, CKGA, and RSM). The RBFNN model approach is found to have relatively fewer standard errors within an experimental data range of 0–0.001 (mol·h\(^{-1}\)), while the CKGA model presents the largest deviation among the three investigated models. The possible reasons could be more additional conditions needed to be met for constant regressions during the mechanism selection (both in statistical and physical relevance such as meeting Arrhenius and Van’t Hoff criteria).

In addition, regarding to those data leading to large discrepancies (SE > ±10%), the calculations in RBFNN are found to undershoot, while the calculations from CKGA and RSM overshoot with more larger errors being observed from CKGA model. The comparison of experimental data and model predictions for \( \text{C}_1-\text{C}_{15} \) at the baseline condition are shown in Figure 6 (the run baseline with best data fit was chosen). Different model approaches provide different accuracies during data regression. Because of direct computation (using training data for RBFNN and second-degree polynomial expression for RSM), both RBFNN and RSM tend to provide predictions with relatively higher accuracy and less computational time.

On the other hand, CKGA does provide elementary mechanism (using the Langmuir–Hinshelwood–Hougen–Watson approach for elementary reaction pathway derivation) that sheds insightful lights on the CO hydrogenation mechanism at the investigated experimental conditions, even though the accuracy of the CKGA model tends to be lower during predictions at some experimental runs (over ±30%). Therefore, RBFNN is found to be the most robust (computational time and accuracy) to make predictions among the investigated models, if sufficient training data were fed. The
relatively big uncertainties C5 in Figure 6a might lie in the vaporization and loss of C5 during trap swap. With the CKGA model, the kinetic expressions (constants in the kinetic expression Table 2) could be obtained through data regressions. The obtained constants can be compared with reported values from literature reports. For instance, the values of $k_p$ and $k_{p,\Lambda}$ are quite close to each other, suggesting the reactivity of alkylidene to paraffin being stable as the carbon number increases over 4. The FT activation energy is close to the typical cobalt-based catalyst indicating the validity of constructed reaction pathways. 38,39,45

One of major challenges of using CKGA lies in compromising the computational time and accuracy using GA for both regional and global optimization to converge the
objective functions. As a dimension of tensor increase, the demand for computational time and internal memories increases significantly. Clearly, it might not be the best choice of obtaining kinetic constants via direct data regression through the elementary mechanisms using quasi-steady state assumption (QSSA) if the goal is for process optimization. Instead, CKGA might be more meaningful in exploring possible reaction pathways during mechanism selections. With regard to optimizing and sizing the reaction system, the RSM surely shows its inherent advantage, even though it does not provide any insightful descriptions between process parameters and the responses. With RSM, the polynomial correlations are able to provide a very accurate correlation with those investigated critical process parameters. For instance, correlations are able to provide an accurate correlation with the responses. With RSM, the polynomial correlations are able to provide a very accurate correlation with those investigated critical process parameters. For instance, correlations are able to provide an accurate correlation with the responses. With RSM, the polynomial correlations are able to provide a very accurate correlation with those investigated critical process parameters. For instance, correlations are able to provide an accurate correlation with the responses. With RSM, the polynomial correlations are able to provide a very accurate correlation with those investigated critical process parameters. For instance, correlations are able to provide an accurate correlation with the responses. With RSM, the polynomial correlations are able to provide a very accurate correlation with those investigated critical process parameters.

The visualization of the C3 paraform formation rate owing to the variations of pressure and temperature is shown in Figure 7a. In addition, the residues (Figure 7b), residues versus predictions (Figure 7c), and Cook’s distance (Figure 7d) can be analyzed to find out the range that is out of normal values. For instance, experimental data of run 10 and run 28 from (from Table 1) tend to yield a relatively larger Cook’s distance (>0.4). The corresponding coefficients from a second-order quadratic expression for the propane (C3) formation rate together with the p-value are shown in Table 3.

For instance, the established coefficients for all (C1–C15) paraform formation rates can be found from Table S2 with less than ±10% experimental uncertainties.

### 3.2. Strategy for Kinetic Modeling and Process Optimization

For sizing and designing a reactor system, the key challenge still lies in constructing accurate and reliable kinetic expressions.\[17,19,50\] Albeit, the LHHW approach has been successfully and widely employed to derive kinetic models for FT synthesis; the inherent limitations such as yielding the debatable mechanistic elementary reactions and tedious computational times still need to be substantially improved in the practical applications.\[17,18,51–58\] Because of the flexibility for learning the information and a pertinence for pattern recognition in the complex system, the applied soft computing technique such as using RBFNN has been widely used for modeling and control tools for a complex system.\[50\]

However, quite a few research studies either focus on kinetic constant correlations by simply using ANNs or generate the central composite design (CCD) matrix through the RBFNN simulations, followed by statistical analysis through RSM.\[50–62\] The critical limitation lies in its lacking understanding of kinetic mechanisms.\[51\] As discussed before, the best strategy is

---

**Table 3. Coefficients of the Quadratic Polynomial Expression for the C3 Rate of Formation and the Corresponding p-Value**

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>$r_p$/mol·h$^{-1}$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>b0</td>
<td>$-7.53 \times 10^{-5}$</td>
<td>0.0427</td>
</tr>
<tr>
<td>b1</td>
<td>8.496 $\times 10^{-5}$</td>
<td>0.0080</td>
</tr>
<tr>
<td>b2</td>
<td>$-3.503 \times 10^{-4}$</td>
<td>0.5818</td>
</tr>
<tr>
<td>b3</td>
<td>$-8.224 \times 10^{-7}$</td>
<td>0.5908</td>
</tr>
<tr>
<td>b4</td>
<td>2.593 $\times 10^{-3}$</td>
<td>0.0099</td>
</tr>
<tr>
<td>b5</td>
<td>$-4.270 \times 10^{-7}$</td>
<td>0.8335</td>
</tr>
<tr>
<td>b6</td>
<td>2.074 $\times 10^{-8}$</td>
<td>0.5414</td>
</tr>
<tr>
<td>b7</td>
<td>$-9.799 \times 10^{-6}$</td>
<td>0.2917</td>
</tr>
<tr>
<td>b8</td>
<td>9.023 $\times 10^{-9}$</td>
<td>0.8609</td>
</tr>
<tr>
<td>b9</td>
<td>5.861 $\times 10^{-6}$</td>
<td>0.7056</td>
</tr>
<tr>
<td>b10</td>
<td>1.824 $\times 10^{-8}$</td>
<td>0.4259</td>
</tr>
<tr>
<td>b11</td>
<td>$-1.472 \times 10^{-7}$</td>
<td>0.4225</td>
</tr>
<tr>
<td>b12</td>
<td>1.235 $\times 10^{-8}$</td>
<td>0.0820</td>
</tr>
<tr>
<td>b13</td>
<td>4.499 $\times 10^{-11}$</td>
<td>0.9457</td>
</tr>
<tr>
<td>b14</td>
<td>$-1.456 \times 10^{-4}$</td>
<td>0.4094</td>
</tr>
</tbody>
</table>

---

**Figure 7.** RSM analysis of process parameters toward response, where P_3 refers to propane (C3 paraform). (a) Temperature and pressure versus the C3 paraform (propane) formation rate, (b) normal plot residues of all experimental runs for the C3 paraform (propane) formation rate, (c) residues versus predictions for all experimental runs of the C3 paraform (propane) formation rate, and (d) Cook’s distance analysis for all experimental runs of the C3 paraform (propane) formation rate.
to maximize the merits of each model (RBFNN, CKGA, and RSM); the hybrid approach of using a small number of data points to cost-effectively generate more representative data sets through RBFNN, followed by CKGA (for mechanism selection) and RSM (performance evaluation to scrutinize the effects of process parameters upon the investigated responses) will show more promising prospects. The schematic diagram of the strategy of kinetic modeling is shown in Figure 9. With the advancement of computation technologies, the consideration of a hybrid paradigm of using RBFNN, CKGA, and RSM has become more attractive. The necessity of combining inherent merits of RBFNN, CKGA (mechanistic kinetic), and RSM (statistical analysis) secures both accuracy of predictions and insightful understanding of the reactions (Table 4).

4. CONCLUSIONS

Three different types of modeling frameworks, namely, RBFNN, CKGA, and RSM, were explored and compared to model the product distribution for FT synthesis using a Ru-promoted Co/Al₂O₃ catalyst in a conventional packing bed reactor (PBR). A new strategic modeling paradigm for olefin/
Paraaffin rate expressions during FT synthesis was proposed by considering both accuracy and intrinsic understanding of FT kinetics. The proposed new strategy possesses appealing features of modeling the rates of different hydrocarbons (olefin, paraaffin with C₂−C₁₅), olefin to paraaffin ratio (OPR), and statistical analysis of the key operational parameters upon responses (selectivity, olefin, and paraaffin with different C₂−C₁₅ carbon numbers). The accuracy and data patterns of using
different models (RBFNN, CKGA, and RSM) were rigorously compared by an error visualization approach. By combining the inherent advantages such as fast converge and higher accuracy (a hybrid approach of coupling both semiempirical or empirical soft computation technique RBFNN), visualization, direct statistical analysis (RSM), and insightful understanding of mechanistic kinetics (kinetic modeling using genetic algorithm CKGA), a new modeling strategy of kinetic study for complex FT synthesis is proposed.

4.1. Experimental Part. The preparation of cobalt-based catalysts is followed by a conventional wet-impregnation approach with a constitution of 20 wt % Co, and 0.5 wt % Ru on γ-alumina.63,64 The catalyst with a sieve fraction of 37–75 μm was employed and loaded into the reactor (a fixed-bed stainless steel reactor with 12.7 mm ID and 400 mm in length). The detailed experimental operations for collecting kinetic data, catalyst reduction, and product analysis could be found from our previous reports.38,39,61 The CO conversion and alkene (C₂–C₄) selectivity were defined as follows

\[ X_{CO\%} = \frac{F_{CO_{in}} - F_{CO_{out}}}{F_{CO_{in}}} \times 100 \]  

(45)

\[ S_{C_2-C_4\%} = \sum_{i=2}^{4} \frac{n_i S_{A_i}}{F_{CO_{in}} - F_{CO_{out}}} \times 100 \]  

(46)

\[ S_{CH_4\%} = \frac{F_{CH_4}}{F_{CO_{in}} - F_{CO_{out}}} \times 100 \]  

(47)

where \( F_{CO_{in}} \) is the CO molar flow rate from the outlet (mol·h⁻¹), \( X_{CO} \) is the CO conversion (%), \( F_{CO_{out}} \) is the CO molar flow rate from the outlet (mol·h⁻¹), \( S_{A_i} \) is the molar flow rate of light olefin \( C_2-C_4 \) (mol·h⁻¹), and \( F_{CH_4} \) is the CO₂ molar flow rate from the outlet (mol·h⁻¹). The outlet concentrations were measured using an online refinery gas analyzer. The following differential equation is used as a reactor model

\[ \frac{dF_i}{dW} = \sum_{j=1}^{N_r} \alpha_j R_{ij} \]  

(48)

The boundary conditions (BCs) were set at

\[ W = 0; \quad F_i = F_i \text{ inlet} \]

\[ W = W_{\text{cat}}; \quad F_i = F_i \text{ exit} \]  

(49)

where \( N_r \) is the number of reactions involved; \( \alpha_j \) is the stoichiometric coefficient of the \( r \)th compound, these detailed
sets of coefficients can be found from previous reports.\textsuperscript{59} $R_0$ is the reaction rate of formation (mol·h\textsuperscript{-1}) of the $i$th component, refers to molar flow rate of species $i$ (mol·h\textsuperscript{-1}), and refers to the mass of catalyst (g). The partial pressure of the $i$th compound

$$
P_i = \frac{m_i}{\sum_{i=1}^{N_c} m_i} P_c (i = i - N_c)
$$

(50)

where $N_c$ refers to the total numbers of components, $P_i$ refers to partial pressure of the $i$th compound (bar), $m_i$ refers to the mass of components (g), and $P_c$ refers to the total pressure of the system (bar). The gases and liquid (both cold trap setting at 10 °C and hot trap setting at 110 °C) were collected and analyzed by liquid chromatography (GC) and gas chromatography-mass spectrometry (GC-MS), respectively. Before systematically exploring the different experimental conditions, the baseline condition (0°) was maintained for about 48 h. At the end of the experiment, the condition was adjusted back to the baseline condition (30°) for about 12 h and deactivation was relatively less than 6%. It must be noted that FT synthesis yields a wide range of products (C\textsubscript{i}−C\textsubscript{50+}); for the simplicity of kinetic modeling, C\textsubscript{3}−C\textsubscript{10}, olefin and n-paraffin (ignoring all branched hydrocarbons) were usually taken into account due to the fact that the rate of the hydrocarbons with a larger carbon number was not simply governed by the kinetics at the experimental conditions. To validate the established model, additional 15 runs were conducted, which is shown in Table S1. Therefore, the overall procedure for model establishment and validation are as follows: (i) data in Table 1 were used to establish the model (RBFNN—data train, CKGA—data regression, RSM—data design in the matrix) and (ii) data in Table S1 were used to validate three constructed models (RBFNN, CKGA, RSM).

# ASSOCIATED CONTENT

## Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/10.1021/acsomega.1c03851. Tables S1–S3 for additional information such as validation experimental results for kinetics during Fischer–Tropsch synthesis, kinetic expressions for detailed product distribution derived from the CKGA model, and coefficients of the quadratic polynomial expression for paraffin formation rate from C\textsubscript{1} to C\textsubscript{15}.

# AUTHOR INFORMATION

Corresponding Authors

Huan Jin — School of Computer Science, University of Nottingham Ningbo, Ningbo 315100, China; Email: huan.jin@nottingham.edu.cn

Yong Sun — Key Laboratory of Carbonaceous Wastes Processing and Process Intensification of Zhejiang Province, University of Nottingham Ningbo, Ningbo 315100, China; Edith Cowan University School of Engineering, Joondalup, WA 6027, Australia; orcid.org/0000-0002-4888-156X; Email: y.sun@ecu.edu.au

Authors

Yixiao Wang — Key Laboratory of Carbonaceous Wastes Processing and Process Intensification of Zhejiang Province, University of Nottingham Ningbo, Ningbo 315100, China

Jing Hu — Key Laboratory of Carbonaceous Wastes Processing and Process Intensification of Zhejiang Province, University of Nottingham Ningbo, Ningbo 315100, China

Xiuyue Zhang — Key Laboratory of Carbonaceous Wastes Processing and Process Intensification of Zhejiang Province, University of Nottingham Ningbo, Ningbo 315100, China

Abubakar Yusuf — Key Laboratory of Carbonaceous Wastes Processing and Process Intensification of Zhejiang Province, University of Nottingham Ningbo, Ningbo 315100, China

Binbin Qi — Department of Petroleum Engineering, China University of Petroleum—Beijing, Beijing 102249, China

Yiyang Liu — Department of Chemistry, University College London (UCL), London WC1H 0AJ, U.K.

Jun He — Key Laboratory of Carbonaceous Wastes Processing and Process Intensification of Zhejiang Province, University of Nottingham Ningbo, Ningbo 315100, China; orcid.org/0000-0001-8056-0347

Yunshan Wang — National Engineering Laboratory of Cleaner Hydrometallurgical Production Technology, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, China

Gang Yang — National Engineering Laboratory of Cleaner Hydrometallurgical Production Technology, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, China

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.1c03851

Author Contributions

Y.W. and J.H. contributed equally to this work.

Notes

The authors declare no competing financial interest.

# ACKNOWLEDGMENTS

This work was funded by the School of Engineering, Edith Cowan University, Australia, for open access publishing, the University of Nottingham Ningbo China (FoSE New Researchers Grant 101210100011), the Faculty Inpiration Grant of University of Nottingham (FIG. 2019), the Qianjiang Talent Scheme (QJD1803014), the National Key R&D Program of China (Grant: 2018YFC1903500), the Ningbo Science and Technology Innovation 2025 Key Project (Grant 2020Z100), the Ningbo Municipal Commonweal Key Program (Grant 2019C10033 and 2019C10104), and the Key Laboratory of Carbonaceous Wastes Processing and Process Intensification of Zhejiang Province (2020E10018).

# REFERENCES


