Systems & Control Transactions

Research Article - Peer Reviewed Conference Proceeding
ESCAPE 35 - European Symposium on Computer Aided Process Engineering
Ghent, Belgium. 6-9 July 2025
Jan F.M. Van Impe, Grégoire Léonard, Satyajeet S. Bhonsale,
Monika E. Polańska, Filip Logist (Eds.)

PSE PRESS

Hybrid Modelling for Reaction Network Simulation in Syngas Methanol Production

Harry Kay[a], Fernando Vega-Ramon[a], and Dongda Zhang[a]*

[a] Department of Chemical Engineering, University of Manchester, Manchester, M13 9PL, UK.
* Corresponding Author: dongda.zhang@manchester.ac.uk.

ABSTRACT

Sustainability is a thriving global topic of concern and following the advancement of technological progress and increased standards of living, the demands for energy, fuels, chemicals and other requirements have increased significantly. Methanol is one such chemical which has seen increases in demand due to its importance as a precursor in the development of widely used chemicals such as formaldehyde. In order to gain insight into the reaction mechanisms driving the process, it is beneficial to develop kinetic models that accurately describe the system for several reasons: (i) to develop process understanding; (ii) to facilitate control and optimisation; (iii) to reduce experimental burdens; and (iv) to expedite scale up and scale down of processes. Two commonly used kinetic reaction rate models are the power law and Langmuir-Hinshelwood expressions, however the strong assumptions made when developing such models may limit their predictive performance through the introduction of inductive bias (i.e. model structural uncertainty). A solution to counter these drawbacks is known as hybrid modelling where, the inauguration of a data-driven component within the kinetic modelling framework allows for any complex, less understood kinetics to be instead learnt from historical data by a machine learning model. In order to identify the pros and cons associated with each kinetic and hybrid modelling strategy for chemical reaction network modelling, a thorough comparison was made using syngas methanol production as a case study. It was shown that hybrid models offered increased predictive accuracy, robust uncertainty quantifications, and improved generalisability under limited data availability.

Keywords: Kinetic modelling, Hybrid modelling, Uncertainty estimation

INTRODUCTION

Increases in global population and standards of living have intensified the demand for bulk and specialty chemicals, however growing sustainability concerns and economic pressures necessitate further process design innovations, and for optimisation of current production facilities. In order to combat such issues, advanced modelling techniques are required that describe bulk and specialty chemical reaction systems under industrial operation, allowing for the precise estimation of key process variables and reaction performance indicators. Typically, first principle based kinetic models are constructed in order to represent the reaction network of interest due to their inherent interpretability and capacity for proving deep kinetic insight when used to discriminate between reaction mechanisms with two of the most popular applied strategies being the Langmuir-Hinshelwood and power law models.

The Langmuir-Hinshelwood expression finds prominent application within heterogeneous catalysis, specifically for gas phase reactants on solid catalyst surfaces [1], however the model is also used for other catalytic reactions. In comparison, the power law model finds much broader application being employed within catalytic reactions [2], non-catalytic reactions [3], and within fields beyond kinetic modelling. Although kinetic models are interpretable and may be used to offer kinetic insight, it is often the case that many chemical reactions are governed by complex underlying physics, so the identification of a correct structure is challenging, typically required simplifications to be made which may be limiting and reduce generalisability. Because of this, model parameters will exhibit large uncertainties, preventing any

use for control and optimisation.

To address the challenges inherent to kinetic modeling, data driven models have gained significant attention within the literature in recent years. Machine learning based black-box models leverage the availability of extensive experimental datasets and do not require a detailed understanding of the underlying reaction mechanism, thus showing advantages in areas where complex phenomena are present. Machine learning has shown great potential in many fields such as (bio)chemical reaction kinetics predictions [4] and soft-sensing, however they are known to be inhibited by various disadvantages including the requirement for large volumes of data and a lack of interpretability, which hinders their ability to provide mechanistic insights.

In recent years, hybrid modelling has emerged as a potential solution to overcome the challenged associated with kinetic and data driven modelling through the combination of a mechanistic backbone with data-driven knowledge. The mechanistic backbone leverages available prior knowledge whilst the data-driven component accommodates for information not captured through the mechanistic structure. In this manner, the interpretability of the kinetic model is retained without the introduction of large inductive bias, and any available data can be exploited to quantify unknown kinetics that are not described by the kinetic structure. The advantages of hybrid modelling are shown clearly with its successful application in predicting complex reaction kinetics [5], boasting accurate predictions and robust uncertainty estimations.

Another crucial benefit of hybrid modelling stems from their reliable uncertainty quantification, which is especially important when encountering complex bio(chemical) reactions where large process variability or historical dependencies may be present. Typically, kinetic models may suffer with large parameter uncertainties and fail to account for historical dependencies, whereas data driven models may overfit when large batch to batch variability occurs, both of which prevent their application under such conditions. However, this is not true of hybrid modelling which has exhibited the capability to accurately describe process dynamics whilst maintaining narrow uncertainty ranges [6]. Even though the value of using hybrid modelling has been demonstrated, little research has focused on investigating its use within the field of chemical reaction kinetics nor on the effects of the kinetic component on model performance.

Therefore, in this work we present a study to explore the application of hybrid models for predictive modeling of chemical reaction kinetics. A key focus of this study is the role of the mechanistic backbone in hybrid model performance. Specifically, we investigate how the selection of different kinetic model structures influences the predictive accuracy and uncertainty quantification of hybrid models. To this end, two hybrid models were developed using the Langmuir-Hinshelwood and power-law rate expressions as their physical components, respectively. To demonstrate the potential of hybrid models for industrial applications, the methanol synthesis reaction from syngas, coupled with the water-gas shift reaction was chosen as a case study. The reliability of the hybrid models was evaluated from their uncertainty estimations and predictions across a broad range of operating conditions. Lastly, the study evaluated the potential of hybrid models to uncover new physical insights.

Case study

In this work, the case study is an industrial reaction system consisting of the water-gas shift reaction and syngas methanol synthesis reaction [7]. The reaction system was catalysed by a $Cu/ZnO/Al_2O_3$ catalyst, and the reactions were conducted within fixed-bed, integral-operation microreactors. In total, 38 experiments were carried out, each containing the measurements of 5 states (i.e. reactants and products) at 4 equidistant points along the reactor length. The experiments were conducted under varying feedstock conditions with CO to CO_2 ratios of 3:3, 3:1, and 3:0.3; reaction temperatures of 453 K, 463 K, 473 K, 483 K, and 493 K, and reaction pressures of 15 bar, 25 bar, 35 bar. Further details regarding the experimental setup and procedures are provided in [7]. The three possible reactions occurring under such conditions are shown in reactions 1, 2 and 3 below

$$CO_2 + \mathrm{H}_2 \rightleftharpoons CO + H_2O \quad (1)$$

$$CO_2 + 3\mathrm{H}_2 \rightleftharpoons CH_3OH + H_2O \quad (2)$$

$$CO + 2\mathrm{H}_2 \rightleftharpoons CH_3OH \quad (3)$$

where reaction 1 is the reverse water gas shift (RWGS) reaction, and reactions 2 and 3 are methanol synthesis from CO and CO_2 respectively.

In this study, it is assumed that CO serves as the primary source of carbon for methanol synthesis. Whilst it is also possible to assume CO_2 as the primary carbon source for producing methanol, this choice is not expected to impact model performance because the hybrid models, developed in this work, are designed to incorporate unknown process knowledge through their data-driven components, enabling them to adapt and accurately represent the underlying system dynamics.

METHODOLOGY

In this section, we will outline the development of two hybrid models based on different mechanistic backbones as shown in Figure 3. In previous literature, Langmuir-Hinshelwood kinetic structures have been proposed to describe reactions 1-3, however due to their

complexity, these often lead to high errors under broad operating conditions, and high parameter uncertainties [7]. Thus, to remove inductive bias, we simplify the equations by lumping complex kinetics into few parameters to be learnt and estimated by a data-driven component.

The second kinetic backbone adopts a power law structure as shown in Figure 3, however since the classical power law expression does not satisfy the thermodynamic constraint at equilibrium conditions [2], a modification has been made. This will ensure that, under equilibrium, the ratio of forward to reverse reaction rate constants are directly equal to the ratio of products to reactants raised to their stoichiometric coefficients.

Due to the complexity of the reaction network under a broad operating range, it is to be expected that some kinetic parameters may contain lumped information and therefore effectively be functions of the operating conditions and input states. Therefore, to construct an accurate and generalisable hybrid model, it is necessary to determine the minimum number of parameters that are dependent on reaction conditions which will be subsequently estimated by the data-driven component. To accomplish this, all parameters were initially assumed condition dependent and parameter estimation, using a modified least squares expression, was conducted to classify parameters as condition-dependent or condition-invariant.

A hybrid model was then constructed, using an artificial neural network (ANN), to estimate the identified condition dependent parameters given their respective operating conditions as shown in Figure 1. The input features to the ANN include $P_T, P_{CO}, P_{CO_2}, P_{H_2}, \frac{P_{CO}}{P_{CO_2}}, \frac{P_{CO}+P_{CO_2}}{P_{H_2}}$ and T where P_x refers to the partial pressure of species x, P_T is the total operating pressure, and T is the temperature of the experiment. To ensure high-quality solutions, Bayesian optimisation was employed for hyperparameter optimisation for the ANN.

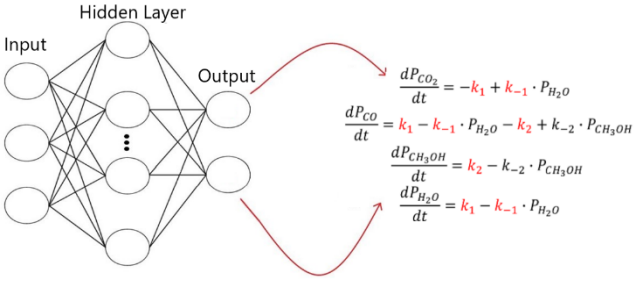


Figure 1. Schematic of the hybrid modelling framework using the Langmuir-Hinshelwood mechanistic backbone. The red highlights indicate condition dependent parameters estimated by the ANN.

In this study, the uncertainty of the hybrid model stems predominantly from the condition varying parameters, and hence the ANN's predictions. This is because, during parameter estimation, it was found that the condition invariant parameters show little change across all experiments. Thus, it is paramount that the uncertainty of the ANN is estimated. To accomplish this, a bootstrapping framework was employed to generate a number of ANN's which, when integrated with the mechanistic backbone to attain the corresponding hybrid models, can be used to generate a mean prediction and its respective standard deviation as shown in Figure 2.

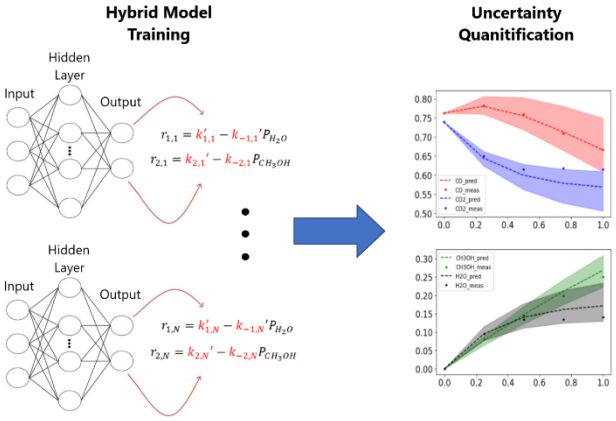


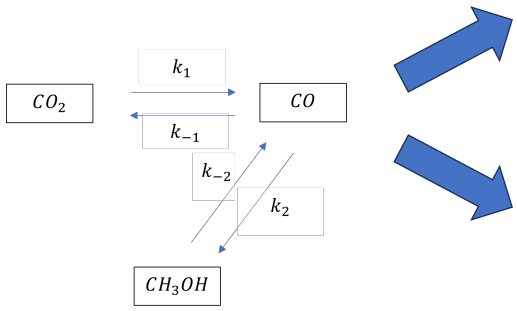
Figure 2. Schematic of the bootstrapping framework to estimate the hybrid models uncertainty where N refers to the number of hybrid models constructed. The red highlights indicate condition dependent parameters estimated by the ANN.

RESULTS AND DISCUSSION

Results of the Langmuir-Hinshelwood based hybrid model

In the methanol synthesis reaction case study, hydrogen was always provided in excess thus exhibiting hydrogen partial pressure 100-10000 times greater than those of other species. As a result of the excess hydrogen, its consumption and production throughout the process is negligible (i.e., there are little changes in P_{H_2}). Consequently, after parameter estimation was conducted, it was found that further model reduction is required from the Equations shown in Figure 3. More specifically, $K_{H_{2,1}}$ and $K_{H_{2,1}}$ were estimated to be zero, and the products $k_1 P_{H_2}$ and $k_2 P_{H_2}$ needed to be lumped as one apparent kinetic constant to overcome identifiability issues. Additionally, $K_{CH_3OH} \cdot P_{CH_3OH}$ and $K_{H_2O} \cdot P_{H_2O} \ll 1$, and $K_{CO_2} \cdot P_{CO_2}$, $K_{CO,1} \cdot P_{CO}$, $K_{CO,2} \cdot P_{CO} \gg 1$, thus either dominating or having a negligible effect on the denominator of the kinetic model.

After model reduction, the Langmuir-Hinshelwood



Langmuir-Hinshelwood model structure

$$r_1 = k_1 \frac{P_{CO_2} P_{H_2}}{1 + K_{CO_2} P_{CO_2} + K_{H_{2,1}} P_{H_2}}$$

$$r_{-1} = k_{-1} \frac{P_{CO} P_{H_2O}}{1 + K_{CO,1} P_{CO} + K_{H_2O} P_{H_2O}}$$

$$r_2 = k_2 \frac{P_{CO} P_{H_2}}{1 + K_{CO,2} P_{CO} + K_{H_{2,2}} P_{H_2}}$$

$$r_{-2} = k_{-2} \frac{P_{CH_3OH}}{1 + K_{CH_3OH} P_{CH_3OH}}$$

Power law model structure

$$r_1 = k_1 P_{CO_2}^{\alpha_{CO_2,1}} P_{H_2}^{\alpha_{H_2,1}} P_{CO}^{\alpha_{CO,1}} P_{H_2O}^{\alpha_{H_2O}}$$

$$r_{-1} = k_{-1} P_{CO_2}^{-1+\alpha_{CO_2,1}} P_{H_2}^{-1+\alpha_{H_2,1}} P_{CO}^{1+\alpha_{CO,1}} P_{H_2O}^{1+\alpha_{H_2O}}$$

$$r_2 = k_2 P_{CO}^{\alpha_{CO,2}} P_{H_2}^{\alpha_{H_2,2}} P_{CH_3OH}^{\alpha_{CH_3OH}}$$

$$r_{-2} = k_{-2} P_{CO}^{-1+\alpha_{CO,2}} P_{H_2}^{-2+\alpha_{H_2,2}} P_{CH_3OH}^{1+\alpha_{CH_3OH}}$$

Figure 3: Thermodynamics inspired power law and simplified Langmuir-Hinshelwood models for the given reaction network where P_j, K_j, and α are the partial pressure, adsorption equilibrium constants, and reaction orders of species j, and k_i and k_{-i} are the forward and reverse reaction rate constants for reaction i. The red highlights indicate key process parameters that are modelled using a data-driven component.

expression devolved into a basic power law structure as seen in Figure 1, where the apparent rate constants contain lumped kinetics, and will therefore be estimated by an ANN. As described in the methodology, the condition dependent parameters were identified as k_1, k_{-1}, and k_2 where k_1 and k_2 are functions of P_{H_2} because of any H_2 effects being combined into them, and k_{-1} is a function of P_{CO}:P_{CO_2} which is intuitive as it describes the forward water gas shift reaction where CO and CO_2 are influential.

The results of the parameter estimation were then used to train the ANN such that the hybrid model can make its predictions on unseen data. In order to quantify the uncertainty predictions of all models, we use the relative percentage uncertainty (RPU) calculated as shown in Equation 4:

$$RPU\% = \frac{100}{N} \cdot \sum_{n=1}^{N} \frac{X_{UB,n} - X_{LB,n}}{X_n} \quad (4)$$

where X_n, $X_{UB,n}$ and $X_{LB,n}$ are the nominal, upper bounds and lower bound predictions of species X at experiment n.

In total 16 training experiments were used and 22 testing experiments yielding percentage errors of 3.6% and 9.5%, respectively and an average testing RPU of 17%. The hybrid model prediction and uncertainty bounds are shown in Figure 4.

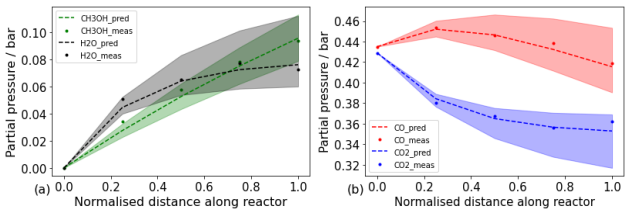


Figure 4. Plots showing the uncertainty and testing predictions of the Langmuir-Hinshelwood based hybrid model. (a) shows the predictions for CH_3OH and H_2O and (b) shows the predictions for CO and CO_2. The scatter points and dotted lines refer to measured data and predictions, respectively.

It can be seen that the data-driven component has correctly learnt missing information from the simplified mechanistic backbone leading to the hybrid model possessing high accuracies. Furthermore, the similar testing and training performance indicates a high level of generalisability displaying the potential of hybrid modelling for use in chemical reaction kinetics prediction.

Although the Langmuir-Hinshelwood based hybrid model exhibits good predictive accuracy and uncertainty quantifications, it is limited due to it not conforming to well known physics and equilibrium constraints. It is known that kinetic constants are solely functions of temperature typically expressed using an exponential relationship, however it was discovered that, because of the lumped kinetics in the apparent rate constants of the

simplified kinetic model, the exponential temperature dependency was not observed. An example can be seen in the values of k_1 where, at temperatures of 453K, 463K, 473K, 483K and 493K, the lumped kinetic constants values are 0.004, 0.0047, 0.0027, 0.012, and 0.005 bar kg^{-1}, respectively.

A second limitation derives from the definition of the reaction equilibrium constant where, under equilibrium conditions, the ratio of forward to reverse rate constants should be equal to the ratio of products to reactants raised to their stoichiometric coefficients. As can be seen in Equations 5 and 6, this constraint does not hold true. This inconsistency limits the interpretability and insight that the Langmuir-Hinshelwood based hybrid model may offer, especially under equilibrium conditions.

$$\frac{k_1}{k_{-1}} = \frac{P_{H_2O}}{1} \tag{5}$$

$$\frac{k_2}{k_{-2}} = \frac{P_{CH_3OH}}{1} \tag{6}$$

Results of the power law based hybrid model

Using the same methodology as for the Langmuir-Hinshelwood based hybrid model construction, parameter estimation was first used to classify condition dependent and condition invariant parameters using the thermodynamics inspired power law defined in Figure 3. Due to the excessive quantities of H_2 present in the system, $\alpha_{H_{2,1}}$and $\alpha_{H_{2,2}} = 0$ showing that H_2 has no effect on the forward reactions but instead inhibits the reverse reactions.

The exponents, $\alpha_{CO,1}$, $\alpha_{CO,2}$ and α_{CO_2} were found to be equal to 1, which can be reasoned by the large magnitudes of P_{CO} and P_{CO_2} in comparison to P_{CH_3OH} and P_{H_2O}, thus any changes are less influential. Therefore, the only condition dependent exponents were α_{CH_3OH} and α_{H_2O}. α_{H_2O} took negative values becoming less negative as temperature increased indicating that the role of water changes from inhibiting reaction r_1 (RWGS) to promoting the reverse reaction.

Further insight can be gained from the values of the reaction order of CO. Since the exponents, $\alpha_{CO,1}$, $\alpha_{CO,2} = 1$, from Figure 3 we can see that, as a product of the reverse water gas shift reaction, CO facilitates the forward reaction. This can be explained by the fact that the use of CO may remove adsorbed oxygen on the Cu catalyst which would ordinarily passivate it. Therefore, any metallic sites that would typically be unavailable for CO_2 activation, are now available. An alternative explanation is that the CO may act to remove water from the reaction network which is beneficial because, on a Cu based catalyst, water can inhibit the methanol synthesis reaction from CO_2 to CH_3OH.

All reaction rate constants within the power law model were found to be condition dependent being solely functions of temperature. Additionally, the trends of the rate constants with respect to temperature all followed the Arrhenius expression as shown in Figure 5. This provides additional confidence into the power law based hybrid model as the results appear to conform to the underlying physics present without any constraints to enforce this and thus, the results are more likely to hold physical significance.

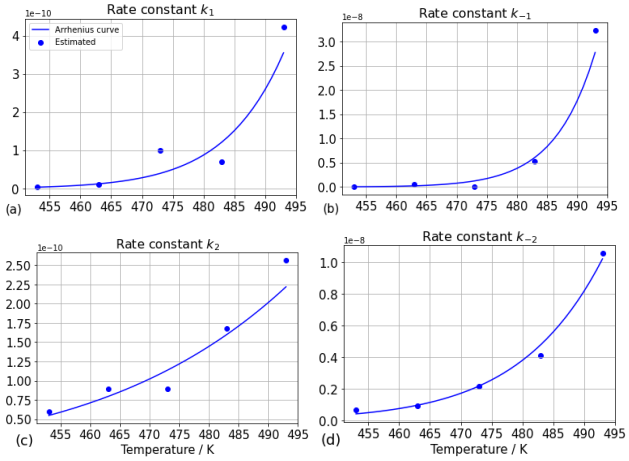


Figure 5. Plots showing the rate constants at different temperatures. The scatter points and lines represent the estimated parameter values and the values through fitting an arrhenius curve, respectively.

The results of parameter estimation were then used to train the ANN component in order to correlate the operating conditions and input states to the condition dependent parameters. The completed power law based hybrid model yielded average fitting and testing errors of 3.7% and 6.7%, respectively and an average relative percentage uncertainty of 13.0%. The similar training and testing performance proves the developed model is generalisable, whilst also boasting low prediction errors across a broad range of operating conditions. The hybrid model prediction and uncertainty bounds are shown in Figure 6.

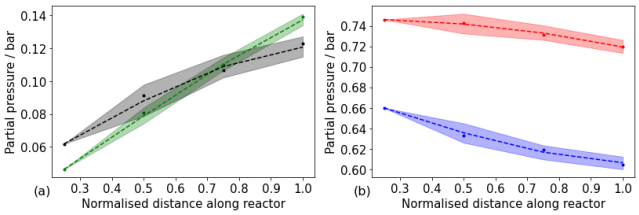


Figure 6. Plots showing the uncertainty and testing predictions of the power law based hybrid model. (a) shows the predictions for CH_3OH and H_2O and (b) shows the predictions for CO and CO_2.

The fitting and testing results for both the power law based and Langmuir-Hinshelwood based hybrid model are similar with a slight increase in performance for the

former as well as a slight decrease in overfitting. The same conclusion can be drawn about uncertainty where both hybrid structures provide robust estimations that cover the majority of measured datapoints. However, the Langmuir-Hinshelwood based hybrid model holds less interpretability as the apparent kinetic constants are functions of the operating conditions and states whereas, it is widely known that the rate constants should be functions solely of temperature. This phenomenon has been successfully described by the power law based hybrid model, meaning that the results are more likely to hold physical significance. Additionally, we can gain kinetic understanding from the reaction orders of the power law structure.

CONCLUSION

This study highlights the potential of hybrid modeling to address the limitations of traditional kinetic and data-driven approaches for complex chemical reaction systems. By combining mechanistic insights with the adaptability of data-driven models, the hybrid models developed provided interpretable structures with high predictive accuracy, and reliable uncertainty estimates without the requirement for large quantities of experimental data.

Two hybrid models based on Langmuir-Hinshelwood and power law rate expressions were constructed, demonstrating the flexibility of this approach, with the power law based model offering additional kinetic insights. These findings underscore the promise of hybrid modeling for industrial applications, particularly in process control and optimisation, where accuracy and uncertainty quantification are critical. Overall, this work emphasises the value of hybrid modeling in advancing our modelling capabilities and understanding in reaction engineering, especially for systems with incomplete mechanistic knowledge.

REFERENCES

1. E. A. Nakajima *et al.*, "Kinetics of dry reforming of methane catalyzed by Ni/Si-MCM-41," *Int J Hydrogen Energy*, vol. 48, no. 83, pp. 32331–32341, Oct. 2023, doi: 10.1016/J.IJHYDENE.2023.05.010.
2. M. Zhu and I. E. Wachs, "Iron-Based Catalysts for the High-Temperature Water-Gas Shift (HT-WGS) Reaction: A Review," *ACS Catal*, vol. 6, no. 2, pp. 722–732, Feb. 2016, doi: 10.1021/ACSCATAL.5B02594/ASSET/IMAGES/LARGE/CS-2015-02594J_0002.JPEG.
3. C. Mendoza-Covarrubias, C. E. Romero, F. Hernandez-Rosales, and H. Agarwal, "N 2 O Formation in Selective Non-Catalytic NO x Reduction Processes," *J Environ Prot (Irvine, Calif)*, vol. 2, pp. 1095–1100, 2011, doi: 10.4236/jep.2011.28126.
4. X. He and M. Yan, "GraphKM: machine and deep learning for KM prediction of wildtype and mutant enzymes," *BMC Bioinformatics*, vol. 25, no. 1, pp. 1–12, Dec. 2024, doi: 10.1186/S12859-024-05746-1/FIGURES/6.
5. A. W. Rogers, Z. Song, F. V. Ramon, K. Jing, and D. Zhang, "Investigating 'greyness' of hybrid model for bioprocess predictive modelling," *Biochem Eng J*, vol. 190, p. 108761, Jan. 2023, doi: 10.1016/J.BEJ.2022.108761.
6. G. Wang, H. Xing, Y. Chen, and Y. Liu, "A dynamic exploratory hybrid modelling framework for simulating complex and uncertain system," *Journal of Safety Science and Resilience*, vol. 5, no. 2, pp. 167–178, Jun. 2024, doi: 10.1016/J.JNLSSR.2024.03.001.
7. S. K. Wilkinson, L. G. A. Van De Water, B. Miller, M. J. H. Simmons, E. H. Stitt, and M. J. Watson, "Understanding the generation of methanol synthesis and water gas shift activity over copper-based catalysts – A spatially resolved experimental kinetic study using steady and non-steady state operation under CO/CO2/H2 feeds," *J Catal*, vol. 337, pp. 208–220, May 2016, doi: 10.1016/J.JCAT.2016.01.025.

© 2025 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and adaptations must be shared under the same terms. See https://creativecommons.org/licenses/by-sa/4.0/