Comparison of machine learning based hybrid modelling methodologies for dynamic simulation of chemical reaction networks

Harry Kay, Fernando Vega-Ramon, Dongda Zhang\*

Department of Chemical Engineering, University of Manchester, Oxford Road, Manchester, M1 3AL, UK.

\*Corresponding author, email: dongda.zhang@manchester.ac.uk

Abstract

Accurate and reliable prediction of state variables and performance indicators within chemical reaction networks is essential for the control and optimisation of chemical process kinetics. Traditionally, kinetic modelling is employed to derive a set of ordinary differential equations. However, simplifications are often introduced via assumptions throughout the modelling process. These assumptions may reduce the model integrity and generalisability to unique conditions. Data driven modelling can overcome some of the aforementioned challenges but is characterised by further problems such as the lack of interpretability and data quality issues. A solution to overcome these challenges is the integration of kinetic and data-driven modelling, named hybrid modelling. Here, the kinetic component retains the underlying structure of the reaction network, and the data-driven component accommodates for any missing information within the kinetic model.

In order to identify the most promising hybrid model structure and construction strategy, different techniques were adopted and benchmarked against a traditional pseudo steady state kinetic model for a C-16 hydroisomerisation reaction network. It was found that hybrid models often outperform alternative methods, offering high accuracy, and overcoming a range of challenges associated with pure kinetic and data-driven modelling, highlighting their potential to be used for industrial chemical reaction applications.

**Keywords**: Machine learning, dynamic simulation, reaction networks, hybrid modelling.

* 1. Introduction

Mathematical models were historically derived to express complex relationships between predictor and response matrices in effort to describe any given physical phenomena of interest. Such models were developed for many reasons including predictive capabilities of current and future states, control, optimisation, reducing experimental burdens, accelerating system design (e.g., faster development of industrial processes, and rapid scale-up transitions) and investigations of underlying system behaviour (e.g., variable dependencies). In many fields, physical insight and knowledge can be introduced to derive equations through first-principles and experimental observations used to estimate model parameters or add empirical correlations to increase the predictive performance. It is, however, typical that certain parameters and variables within industrial processes possess too complex of interactions to adequately model via first principles. In such cases, the non-linearities can be better represented via the paradigm of machine learning, thus enabling the modelling of complicated dynamic systems. Machine learning lends itself naturally to Industry 4.0, with the large quantity of process information available through sensor recordings and experiments.

Recent works have incorporated black-box models within many fields including soft sensing (Kay et al., 2022), image classification (Rawat & Wang, 2017), model-based design of experiment (Greenhill et al., 2020), and chemical reaction engineering (Yang et al., 2020). Although capable of formulating accurate predictive models, black-box machine learning suffers from lack of interpretability and usually lacks the ability to generalise to other systems or sufficiently different conditions. Ideally, a middle ground exists that preserves the advantages of both physical and data-driven modelling yet does not retain any associated disadvantages. A solution that provides this is known as hybrid modelling where both mechanistic and data-driven information is incorporated into a single model.

In the literature, two major forms exist to characterise hybrid modelling:

* Using machine learning to estimate dynamic changes within key parameters of a defined kinetic model. Commonly, this method is applied to systems with complex phenomena where parameters may deviate through time such as in bioprocess modelling (Rogers et al., 2023; Zhang et al., 2019). This approach offers increased interpretability as no change to the structure is made. The complex non-linear dynamics of the parameters are instead evaluated by data-driven methods.
* Extending a defined kinetic structure with an unknown data driven error parameter to account for missing information (Quaghebeur et al., 2022). This is commonly referred to as a hybrid discrepancy model and is implemented to rectify an error with the model structure itself.
	1. Problem Statement

This study focuses on an in-depth analysis and comparison of the major forms of hybrid modelling and delves further into different methodologies of solving each system. More specifically, we will focus on simulating a dynamic C16 hydroisomerisation reaction. A rigorous microkinetic model (Figure 1) was derived in (Vega-Ramon et al., 2023) and reduced using optimisation to remove comparatively insignificant reactions. From this, the pseudo steady state hypothesis (PSSH) was applied to simplify the kinetic structure (Figure 1) and remove any dependencies on non-observable states (surface coverages). The model parameters were then lumped to ensure identifiability for better quality solutions upon optimisation and the resulting model used as a benchmark for traditional modelling techniques. The microkinetic model was exploited to generate 8 training datasets with varying initial mass fractions of , mono branch product , multi branch product , cracking byproducts , and catalyst concentrations . The training datasets were then used to determine the kinetic constants of the PSSH model. Three testing dataset were chosen to mimic the experimental data with mass fractions of , , *,* , and catalyst concentrations of 0.1 wt.%, 0.3 wt.% and 0.5 wt.%.



Figure 1 - Schematic showing the microkinetic model and the resulting reduced model after the pseudo steady state hypothesis was applied and parameters were lumped

* 1. Methodology

Here we will introduce two strategies for solving the hybrid modelling frameworks namely the one-step and two-step approaches. However, we must define an appropriate metric to ascertain parameters within the kinetic structure which are not adequately described through being limited to a constant value. Ideally, this is achieved through some determination of uncertainty, as the parameters holding the largest uncertainty are ultimately less confident in their assigned deterministic value. In this work, a measure of uncertainty is made through a leave-two-out optimisation of the pseudo steady state model, where two of the training datasets are withheld from the optimiser through each iteration and the variance of the optimal parameters taken. In this manner, a form of parametric uncertainty is calculated, and the highest variances can be associated with kinetic constants that require to be estimated at each individual timestep.

Firstly, we will formally define the two-step methodology as shown in Figure 2 where parameter has been identified as best suited to be time varying.



Figure 2 - Schematic of the two-step hybrid modelling approach for a single time varying kinetic constant .

This methodology is defined by initially discretising the set of time varying parameters as constants over set time intervals between each measured datapoint. This optimization was performed using interior point optimisation (IPOPT) in Pyomo. Once the optimal set of constants were obtained, a neural network can be constructed and employed to learn the correlations between the set of state parameters and Pd catalyst loading weight percentage, and the relevant set of time varying kinetic constants describing the system of differential equations between the current and next future states. This strategy was also applied to the development of the discrepancy model.

The second hybrid modelling strategy is the one-step approach where the network and physical parameters are optimised and evaluated simultaneously. This can be addressed in multiple manners such as in (Psichogios & Ungar, 1992) where the sensitivity approach

is taken. Using the forward sensitivities, however, requires the solving of excessive numbers of ODE’s scaling linearly with respect to the numbers of parameters. The adjoint sensitivity method can be used to alleviate this burden which scales constantly with number of parameters and is therefore far less computationally expensive. This approach was implemented for solving neural ODE’s in (Chen et al., 2018). In this work, the time varying and constant parameters are solved simultaneously through a combination of automatic differentiation and gradient descent. To evaluate and assess the performance of all strategies and models, two metrics were defined being the mean absolute percentage error (MAPE) (1), and the mean absolute error (MAE) (2).

|  |  |  |
| --- | --- | --- |
|  | (1) |  |
|  | (2) |  |

* 1. Results and Discussion

Through the analysis as described in the methodology, three parameters were determined to hold the largest uncertainty in the order of , , and . A multistep ahead approach is taken to obtain the predictions of the neural networks on the training sets by inputting the network outputs through the set of ODE’s that define the chemical reaction network and using these as inputs to ascertain the time varying constants that represent the kinetics until the next future state. The prediction results for the PSSH, discrepancy, one-step and two-step models are presented in Table 1 below where the discrepancy model has four time varying parameters (one per ODE) and the others have three.

Table 1 – Average MAPE and MAE testing results (over the 3 testing datasets) for the pseudo steady state benchmark model, discrepancy model, one-step time varying parameter model, and two-step time varying parameter model.

|  |  |  |
| --- | --- | --- |
| **Model** | **Mean average percentage error / %** | **Mean absolute error** |
|  |  |  |  |  |  |  |  |
| PSSH | 10.3 | 6.0 | 71.1 | 18.9 | 0.030 | 0.023 | 0.030 | 0.022 |
| Discrepancy (two-step) | 15.8 | 12.4 | 36.8 | 14.0 | 0.048 | 0.049 | 0.008 | 0.018 |
| One-step | 4.6 | 4.5 | 65.0 | 10.3 | 0.017 | 0.016 | 0.010 | 0.009 |
| Two-step | 7.9 | 7.2 | 16.7 | 10.0 | 0.023 | 0.030 | 0.006 | 0.010 |

From Table 1, it is evident that the discrepancy model performed the worst out of the hybrid configurations, yielding the largest mean absolute errors often being nearly 100% larger than the rest. This is reciprocated with the mean average percentage errors with the exception of the multi branch product predicting similarly to the other configurations. This can be reasoned due to the inherent structure of the ODE system not missing information but the parameters being misinformed instead. The PSSH benchmark model seemed to accurately represent the reactant and mono branch product mass fractions with high accuracy, however evidently lacked knowledge regarding the multi branch product and cracking byproduct mass fractions.

Both of the hybrid model frameworks that simulated time varying parameters improved significantly over the benchmark both in terms of MAPE and MAE. The simultaneous estimation of the network and physical parameters (i.e., one step approach) through automatic differentiation offered an average absolute improvement in MAPE and MAE of 5.5 % and 0.013 respectively. The large percentage errors associated with the prediction of the multi branch product result from mass fractions close to zero. Using the two-step approach provided average absolute improvements in MAPE and MAE of 11.7 % and 0.009. The mass fraction profiles estimated by the models for a single testing dataset can be seen in Figure 3.



Figure 3 - Predicted mass fraction profiles for initial conditions: = 1, = 0, = 0, = 0, C = 0.3 wt. % using the one and two-step hybrid models and the PSSH model (markers represent measured data and the solid curves represent the models predictions).

From Figure 3, it is clear that the PSSH model lacks necessary information regarding the trend of the multi branch product profile, and has an observable discrepancy for the cracking by-products, mono branch products and for the C16 profile. Both of the time varying parameter approaches provided the required missing information for the multi branch product profile and amended the predictions to exhibit the correct trend, reducing the MAE by 67 % and 80 % for the one and two step approaches respectively. In addition, the overshoot at the terminal of the C16 reaction is mitigated and the mean absolute errors corresponding to the cracking by-products reduced by 55 % and 59 %. In general, the models focused on minimising the MAE/MAPE accompanying the profiles closer to zero as these would generate the largest standardised errors upon a small change in predictions.

The two-step approach is limited by the parameter estimation in the first stage of its methodology as, the neural network is purposely trained to correlate the predictor matrix to these values, whereas the one-step approach is not. However, as the physical and data-driven parameters are simultaneously estimated, the network does not learn any soft constraints imposed through parameter estimation, leading to rapid changes in gradients in the mass fraction profiles. Hence, as shown in Figure 3, the curves are less smooth and

tend to oscillate around the measured values, which is not physically meaningful. Furthermore, additional penalties had to be imposed within the one-step method to prevent any kinetic constants becoming less than zero. If such penalties were not introduced, this constraint would often be broken, whereas this was never found in the two-step approach as the constraints were preserved from parameter estimation.

* 1. Conclusion

In conclusion, three hybrid models were constructed to simulate a dynamic chemical reaction network, proving to be useful for developing accurate, interpretable models for predicting reaction profiles. The discrepancy model performed the worst failing to improve upon the benchmark PSSH model, however, the one and two-step methodologies for time varying parameters improved upon the benchmark MAE by over 34%. Such hybrid models offer increased interpretability over traditional black box machine learning techniques as the kinetic structure that was defined through analysis and modelling is preserved and only the uncertain kinetic constants manipulated (typically these are estimated using parameter uncertainty estimation). Through the use of automatic differentiation, parameter estimation, and neural networks, it was shown possible to identify high quality solutions at low computational cost (network training was less than 5 minutes under all conditions). Upon testing of the optimised frameworks, good predictive performance and generalisation capabilities were exhibited. This study has presented the potential of integrating machine learning and mechanistic knowledge for use in chemical reaction networks and as a general modelling procedure.

References

Chen, R. T. Q., Rubanova, Y., Bettencourt, J., & Duvenaud, D. (2018). Neural Ordinary Differential Equations. *NIPs*, *109*(NeurIPS), 31–60. https://arxiv.org/abs/1806.07366v5

Greenhill, S., Rana, S., Gupta, S., Vellanki, P., & Venkatesh, S. (2020). Bayesian Optimization for Adaptive Experimental Design: A Review. *IEEE Access*, *8*, 13937–13948.

Kay, S., Kay, H., Mowbray, M., Lane, A., Mendoza, C., Martin, P., & Zhang, D. (2022). Integrating Autoencoder and Heteroscedastic Noise Neural Networks for the Batch Process Soft-Sensor Design. *Industrial and Engineering Chemistry Research*, *61*(36), 13559–13569.

Psichogios, D. C., & Ungar, L. H. (1992). A hybrid neural network‐first principles approach to process modeling. *AIChE Journal*, *38*(10), 1499–1511.

Quaghebeur, W., Torfs, E., De Baets, B., & Nopens, I. (2022). Hybrid differential equations: Integrating mechanistic and data-driven techniques for modelling of water systems. *Water Research*, *213*, 118166.

Rawat, W., & Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation*, *29*(9), 2352–2449.

Rogers, A. W., Song, Z., Ramon, F. V., Jing, K., & Zhang, D. (2023). Investigating ‘greyness’ of hybrid model for bioprocess predictive modelling. *Biochemical Engineering Journal*, *190*, 108761.

Vega-Ramon, F., Wang, W., Wu, W., & Zhang, D. (2023). Developing a rigorous chemical reaction network reduction strategy for n-hexadecane hydroisomerisation. *Computer Aided Chemical Engineering*, *52*, 157–162.

Yang, F., Dai, C., Tang, J., Xuan, J., & Cao, J. (2020). A hybrid deep learning and mechanistic kinetics model for the prediction of fluid catalytic cracking performance. *Chemical Engineering Research and Design*, *155*, 202–210.

Zhang, D., Del Rio-Chanona, E. A., Petsagkourakis, P., & Wagner, J. (2019). Hybrid physics-based and data-driven modeling for bioprocess online simulation and optimization. *Biotechnology and Bioengineering*, *116*(11), 2919–2930.