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

Automated Identification of Kinetic Models for Nucleophilic Aromatic Substitution Reaction via DoE-SINDy

Wenyao Lyu[a], Federico Galvanin[a]*

[a] University College London, Department of Chemical Engineering, London, United Kingdom

* Corresponding Author: f.galvanin@ucl.ac.uk.

ABSTRACT

Nucleophilic aromatic substitutions (SNAr) are key chemical transformations in pharmaceutical and agrochemical synthesis, yet their complex mechanisms (concerted or two-step) complicate kinetic model identification. Accurate kinetic models for SNAr are essential for scale-up, optimization, and control of the reaction process, but conventional methods struggle with mechanism uncertainty driven by substrates, nucleophiles, and reaction conditions, with data collection being difficult due to its source-intensive nature. We address this using DoE-SINDy, a data-driven framework for generative modelling without complete theoretical understanding. A benchmark study on the SNAr reaction of 2,4-difluoronitrobenzene with morpholine in ethanol was conducted, incorporating parallel and consecutive side-product formation. Ground-truth kinetic models validated in prior studies were used to generate in-silico data under varying noise levels and sampling intervals. DoE-SINDy successfully identified the true kinetic model with minimal runs, quantifying the impact of key design factors such as inlet concentrations, residence time, sample size and experimental budget on model identification.

Keywords: Modelling and Simulations, Reaction Engineering, System Identification, Machine Learning, Model Structure Generation, Design of Experiment

1. INTRODUCTION

Nucleophilic aromatic substitution (SNAr) reactions play a crucial role in synthesizing natural products, pharmaceuticals, and agrochemicals. Accurate kinetic model identification is essential for optimizing SNAr reaction conditions, improving process efficiency, and enabling predictive control in manufacturing [1]. SNAr reactions can proceed via a classical two-step pathway or a concerted mechanism (cSNAr), with the preferred pathway influenced by substrate, nucleophile, leaving group, and environmental conditions such as solvation and fluid dynamics. These factors complicate kinetic identification, especially at process scales [2].

While quantum chemical methods like DFT offer detailed mechanistic insights, they are computationally intensive. Molecular descriptors provide a user-friendly alternative but require broader datasets for generalization [3]. Integrating solvent, fluid dynamics, and operational factors is essential for process-scale simulations, which requires computational-experimental approaches [2]. Obtaining reliable experimental data is resource-intensive, especially when integrating multiple effects [4]. Though recent innovations in continuous-flow platforms and experimental design have improved data collection efficiency [2,4–6], their reliance on predefined kinetic models limits the application to systems with incomplete mechanistic understanding. Recently, automated model-building approaches, such as such as AI-DARWIN [7], PySR [8] and SINDy [9], have been developed to rapidly identify interpretable models with minimal data requirements and without requiring fully predefined model expressions.

Motivated by these challenges, a systematic model identification framework named DoE-SINDy is employed in this paper to simultaneously generate mechanistic model structures and estimate model parameters from a small dataset [10], with the ultimate objective to provide

a robust and accurate representations of concentration profiles in SNAr reactions.

2. METHODOLOGY

2.1 DoE-SINDy framework

A dynamic model of a chemical reaction system can be formulated as a set of differential-algebraic equations (DAEs) as:

$$\begin{cases} \mathbf{f}(\dot{\mathbf{x}}, \mathbf{x}, \mathbf{u}, t, \boldsymbol{\theta}) = 0 \\ \hat{\mathbf{x}} = \mathbf{h}(\mathbf{u}, t, \boldsymbol{\theta}) \end{cases} \quad (1)$$

where $\mathbf{f}$ and $\mathbf{h}$ are sets of model equations, $\mathbf{x}$ is the observable state variable vector, $\mathbf{u} \in \mathrm{U}$ the control variable vector, t time and $\boldsymbol{\theta}$ the parameter vector, and $\hat{\mathbf{x}}$ the predicted measurable state variables. This research focuses on simultaneously identifying the model structure $\mathbf{f}$, and parameters $\boldsymbol{\theta}$.

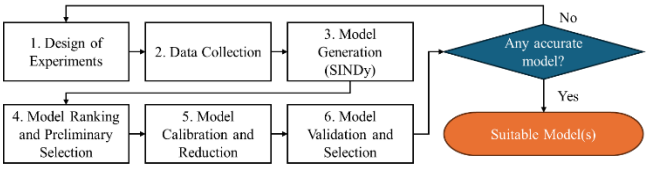


Figure 1. Framework of DoE-SINDy for identifying the most suitable models from experimental data [10].

Sparse identification of nonlinear dynamical systems (SINDy) [9] is a sparse regression-based approach that derives nonlinear models from a user-defined candidate term library and reaction profiles data. Assuming a process model is governed by a few key terms, SINDy ensures both accuracy and physical interpretability of the identified models. Building upon SINDy, Lyu & Galvanin proposed an iterative model identification framework named DoE-SINDy to address variability in model generation and enhance identifiability and reliability of identified models [10]. The procedure is articulated in steps illustrated in Figure 1.

In block (1), the DoE-SINDy framework begins with a preliminary design of experiments (DoE) using the minimum number of required runs to explore conditions underlying physical constraints. Step (2) requires the collection of time-varying state variable measurements, the numerical approximation of time derivatives and data splitting into training and testing subsets. In the model generation step (3), the SINDy algorithm, developed in [11,12], is called to build multiple candidate models from subsets of the training dataset. The resulting models are ranked by complexity, and the simplest candidates are selected (4). To improve parameter robustness, maximum likelihood estimation using the full training dataset is performed during the model calibration step (5), followed by model refinement to eliminate non-significant terms that minimally impact reaction profiles, further reducing unnecessary complexity. A sensitivity-based practical parameter identifiability analysis is integrated before and after calibration to ensure that only identifiable models are retained. Calibrated models then undergo model validation for statistical adequacy and check whether the models pass user-defined stopping criteria (6). If no model meets the stopping criteria, the framework iteratively returns to step (1), generating new candidate models using expanded datasets. This iteration continues until at least one model meets the stopping criteria or the experimental budget is exhausted. When the iteration stops, the output will be the statistically acceptable models or the models from the final iteration sorted in terms of Akaike Information Criterion (AIC).

The framework for DoE-SINDy is established in Python. ***PySINDy (v1.7.5)*** is the primary package used for model generation and numerical approximation of time derivatives [11,12]. Additionally, ***scipy*** is employed for ODE integration, parameter estimation, and statistical evaluations, including χ^2 test, normality test and t test.

2.2 Assessment criteria for DoE-SINDy identified models

The DoE-SINDy framework evaluates identified models using three main criteria. The first check is if the model is **statistically acceptable.** Goodness-of-fit test evaluates whether a model adequately represents experimental data considering measurement noise. A two-tailed χ^2 test assesses if residuals between predictions and measurements align with assumed noise distribution, detecting underfitting or overfitting cases [13]. However, since DoE-SINDy focuses on identifying promising model structures with limited data, precise parameter estimates are not guaranteed at this stage. A normality test, as a less strict alternative, checks if residuals follow a zero-mean Gaussian distribution [14], allowing inclusion of models that may subsequently improve through refined parameter estimation and additional data. Four options for user-defined stopping criteria are possible:1) passing both tests ('and') 2) passing only χ^2 test ('chi2'); 3) passing only normality test ('normality'), or 4) passing either ('or'). These are ranked from the most to the least strict, and determines when the iteration will stop, with models passing the chosen criterion labelled as "statistically acceptable".

To assess if the identified models successfully recover the ground-truth model structures, the second and the third criteria check if the model is **structurally promising,** as defined in Equation (2), or matches the **ground-truth**, as defined in Equation (3)**:**

$$\mathcal{I}(\boldsymbol{\theta}^{true}) \subset \mathcal{I}(\hat{\boldsymbol{\theta}}) \quad (2)$$

$$\mathcal{I}(\boldsymbol{\theta}^{true}) = \mathcal{I}(\hat{\boldsymbol{\theta}}) \quad (3)$$

where $\boldsymbol{\theta}^{true}$ and $\hat{\boldsymbol{\theta}}$ represent coefficient matrices of

Table 1: Potential scenarios for the models identified by DoE-SINDy.

Performance	Statistically acceptable	Structurally promising	Structurally Ground-truth	Circumstance
Best	T	T	T	Statistically acceptable and structurally ground-truth model identified
Good	T	T	F	Statistically acceptable and structurally promising model identified
Poor	T	F	F	Statistically acceptable but missing term(s) in the ground-truth model
	F	T	T	Structurally ground-truth but poor fit
	F	T	F	Structurally promising but poor fit
Worst	F	F	F	Poor fit and missing terms(s) in the ground-truth model

ground-truth and identified models, respectively. Each matrix has dimensions corresponding to the number of equations by the size of candidate term library. A non-zero entry in these matrices indicates the presence of a specific term in the model, with its value representing estimated parameters of that term. The notation $\mathcal{I}(\cdot)$ denotes the set of position indices of non-zero elements in the coefficient matrix of each identified model. For example, if the true and identified coefficients matrix of the differential equation for component A are $\theta_{c_A}^{true} = [2\ 0\ 0\ 3\ 0\ 0]$ and $\hat{\theta}_{c_A} = [2\ 0\ 0\ 3\ 1\ 0]$, the true and identified coefficient indices are $\mathcal{I}(\theta_{c_A}^{true}) = \{1,4\}$ and $\mathcal{I}(\hat{\theta}_{c_A}) = \{1,4,5\}$. The model is structurally promising since $\{1,4\} \subset \{1,4,5\}$, meaning that the identified structure contains all terms from the ground-truth model but with additional terms. It is structurally ground-truth if its structure exactly matches the true underlying model, even if parameter uncertainty is still present. For example, a model with the coefficient matrix $\hat{\theta}_{c_A} = [1.8\ 0\ 0\ 3.2\ 0\ 0]$ is considered structurally ground-truth because all the model terms are correctly identified.

2.3 Scenarios for DoE-SINDy identified models

Based on the assessment criteria defined in Section 2.2, labels are assigned to each identified model, as summarized in Table 1. The most desirable scenario is 'TTT,' where the identified model is both statistically acceptable and structurally ground-truth. Scenario 'TTF' is also acceptable, as the identified model contains all the ground-truth terms, with extra terms removable after additional data collection and parameter re-estimation. In practice, 'TFF' models are selected for their statistical adequacy when the ground-truth is unknown but may be eliminated through model discrimination techniques [15]. Scenarios 'FTT', 'FTF' and 'FFF' occur when no acceptable model is generated within the available experimental budget. In such cases, redefining the library of candidate terms or adjusting DoE-SINDy setting is recommended before conducting additional experiments.

3. CASE STUDY AND IMPLEMENTATION

3.1 In-silico data generation

A benchmark case study evaluated the performance of DoE-SINDy in identifying the rate models for the nucleophilic aromatic substitution (SNAr) of 2,4-difluoronitrobenzene 1 with morpholine in ethanol. The reaction produce a mixture of desired product *ortho*-substituted 3 and side products *para*-substituted 4 and *bis*-adduct 5 through parallel and consecutive elementary reaction steps as shown in Figure 2 [5,6].

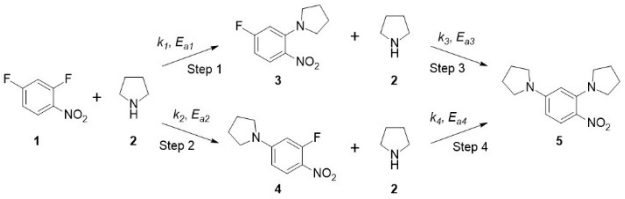


Figure 2. Assumed mechanism scheme for SNAr [5,6].

Data was generated through in-silico experiments simulating an automated continuous-flow reactor system with controlled linear flow ramps as in Hone *et al.* [5] and Agunloye *et al.* [6]. The residence time τ was manipulated through volumetric flow rate changes within a fixed-length reactor. The range of the time axis is equal to the residence time of the experiment with the longest operation time among the set of experiments in the ramp. The three experimental design variables in this system are: 1) initial concentrations; 2) sample size of the concentration-time profile; 3) experimental budget. The molar balance equations for the system, based on the reaction scheme are:

$$\frac{dc_1}{d\tau} = -(k_1 + k_2)c_1c_2 = \theta_1 c_1 c_2$$

$$\frac{dc_2}{d\tau} = -(k_1 + k_2)c_1c_2 - k_3c_2c_3 - k_4c_2c_4$$

$$= \theta_2 c_1 c_2 + \theta_3 c_2 c_3 + \theta_4 c_2 c_4 \quad (4)$$

$$\frac{dc_3}{d\tau} = k_1 c_1 c_2 - k_3 c_2 c_3 = \theta_5 c_1 c_2 + \theta_6 c_2 c_3$$

$$\frac{dc_4}{d\tau} = k_2 c_1 c_2 - k_4 c_2 c_4 = \theta_7 c_1 c_2 + \theta_8 c_2 c_4$$

Experiments were conducted for 2.4 minutes at 130°C, with ground-truth reaction rate constants were defined as $k_1, k_2, k_3, k_4 = [1.21, 0.21, 0.0, 0.057]$. Identifying dynamic models via DoE-SINDy requires multiple trajectories of time-varying states variables. Thus, design of experiment technique, Latin Hypercube Sampling, is applied to design multiple sets of initial concentrations within the experimental design space (Table 2). Notably, the initial concentration of $c_2(0)$ was defined as $c_2(0) = c_1(0) \cdot c_2^{equiv}(0)$.

Table 2: Experimental design space for control variables, initial concentrations, employed in the case study [6].

Limits	$c_1(0)$ (M)	$c_2^{equiv}(0)$	$c_3(0)$ (M)	$c_4(0)$ (M)
Lower	0.0967	0.2054	0	0
Upper	1.6917	2.5400	0.5000	0.5000

In-silico data were generated by adding random normally distributed noise with zero mean and given standard deviation to the noise-free simulated values of state variables. Subsequently, the time derivatives of the noisy state variables (concentrations) were numerically approximated using Kalman derivative as the measurement errors are assumed to follow a standard normal distributions [12].

The performance of DoE-SINDy in identifying models achieving specific target scenarios was evaluated using a metric, *Target Scenario Achievement Rate (TSAR)*, which is defined as:

$$\text{TSAR} = \frac{\text{Number Of Datasets with Models Meeting the Target Scenario}}{\text{Total Number Of Datasets}} \quad (5)$$

The robustness of DoE-SINDy was tested across three different experimental budgets (10, 15, and 20 experiments), three noise levels ($\sigma = [0.1\%, 1\%, 10\%]$), ten sample sizes (8, 12, 16, 18, 20, 24, 30, 36, 48, 60), and 50 sets of initial concentrations.

3.2 Application of DoE-SINDy for model identification

The DoE-SINDy implementation utilized the collected measurements and their time derivatives for model identification. In the model generation step (3), the candidate term library is constructed with the following set of features: $\boldsymbol{g}(\boldsymbol{c}) = [c_1c_2, c_1c_3, c_1c_4, c_2c_3, c_2c_4, c_3c_4]$. The Stepwise Sparse Regressor (SSR) is employed to generate models from data. In step 4 of the proposed procedure, the top three simplest models are selected for further analysis. four different stopping criteria were evaluated to assess their impact on model identification performance.

4. RESULTS

This section compares DoE-SINDy example models for the SNAr reaction rate, evaluating the TSAR for achieving 'TTT' and 'TTF' scenarios using four stopping

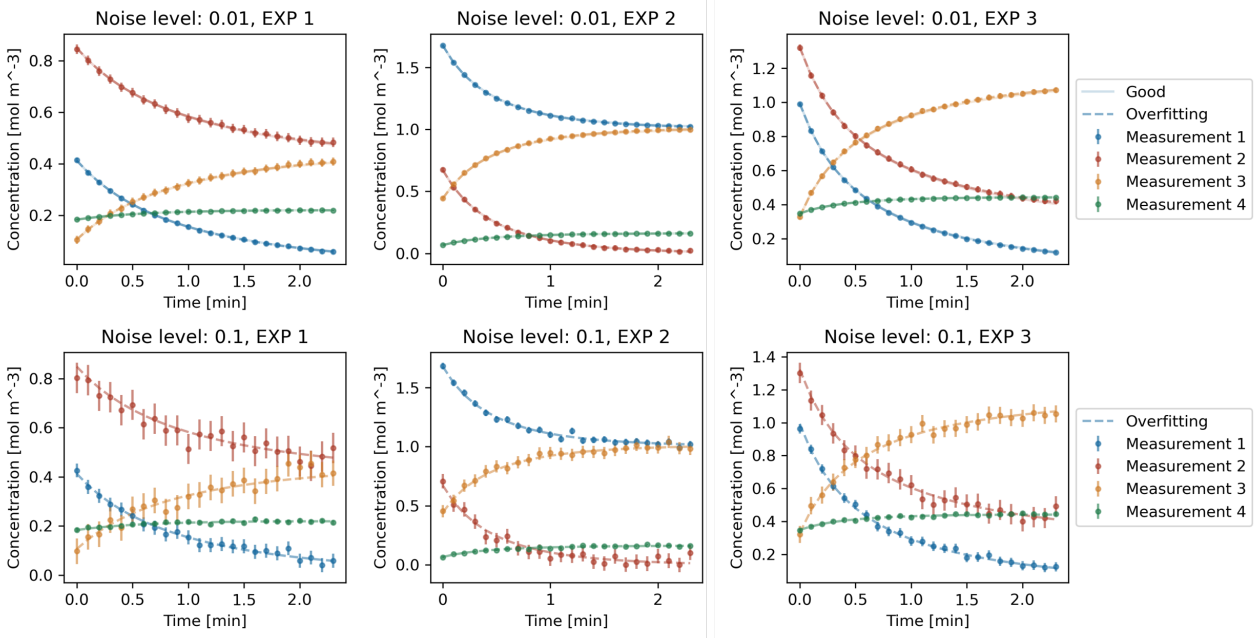


Figure 3: Simulations for example 'TTT' scenario models tested as good fit or overfitting based on the χ^2 test, under varying initial conditions and noise levels (σ=0.01 and 0.1, sample size=24, budget=20 experiments).

Table 3: Estimated coefficients ($\hat{\boldsymbol{\theta}}$) and corresponding t-values of models achieving 'TTT' scenario (sample size=24, budget=20). Results are presented for three noise levels ($\sigma = 0.1\%, 1\%\ and\ 10\%$).

	$\boldsymbol{\theta_{i,ref}}$ $(\mathbf{M^{-1}min^{-1}})$	$\hat{\boldsymbol{\theta}}_{i,\sigma=0.1\%}$	$\boldsymbol{t\ (t_{ref}(95\%) = 1.97)}$	$\hat{\boldsymbol{\theta}}_{i,\sigma=1\%}$	$\boldsymbol{t\ (t_{ref}(95\%) = 1.97)}$	$\hat{\boldsymbol{\theta}}_{i,\sigma=10\%}$	$\boldsymbol{t\ (t_{ref}(95\%) = 1.97)}$
$\boldsymbol{\theta_1}$	-1.42	-1.42	2.06×10^{-3}	-1.41	2.05×10^{-3}	-1.43	2.08×10^{-3}
$\boldsymbol{\theta_2}$	-1.42	-1.42	2.06×10^{-3}	-1.42	2.06×10^{-3}	-1.43	2.07×10^{-3}
$\boldsymbol{\theta_3}$	0.00	0.00	-	0.00	-	0.00	-
$\boldsymbol{\theta_4}$	-0.057	-0.056	1.07×10^{-4}	-0.071	1.29×10^{-4}	-0.064	1.23×10^{-4}
$\boldsymbol{\theta_5}$	1.21	1.21	1.75×10^{-3}	1.20	1.73×10^{-3}	1.21	1.76×10^{-3}
$\boldsymbol{\theta_6}$	0.00	0.00	-	0.00	-	0.00	-
$\boldsymbol{\theta_7}$	0.21	0.21	3.09×10^{-4}	0.21	3.10×10^{-4}	0.21	3.11×10^{-4}
$\boldsymbol{\theta_8}$	-0.057	-0.057	1.13×10^{-4}	-0.055	1.26×10^{-4}	-0.057	1.22×10^{-4}

criteria ('and', 'chi2', 'normality', 'or') under varying experimental designs and data conditions.

4.1 Identified SNAr reaction rate models

Figure 3 shows simulated reaction profiles for models achieving the 'TTT' scenario under varying noise levels. When $\sigma = 0.01$, the profiles of the good-fit and overfitting models overlap, indicating similar prediction performance. At $\sigma = 0.1$, no good-fit models are identified but the overfitting model aligns perfectly with the observations. These graphical results suggest that both good-fit and overfitting models perform well in representing the dynamic system when the 'TTT' scenario is achieved.

Scenario 'TTT' is the most desirable outcome, where the identified models are statistically acceptable and aligned with the ground-truth. DoE-SINDy successfully identifies 'TTT' models with at least 7 experiments.

We examine three 'TTT' models identified using data with a sample size of 24 and an experimental budget of 20 noise levels of σ=0.1%, 1% and 10%. As shown in Table 3, the estimated parameter values ($\hat{\theta}_i$) are generally consistent with reference values ($\theta_{i,ref}$), and close to the reference at higher noise levels (σ=1% and 10%). Figure 3 shows that these models reflect the system dynamics well, even when some parameters deviate slightly from the reference. Parameter θ_3 and θ_6 consistently yield zero estimates, indicating that Step 3 is inactivated in the reaction network, consistent with the assumed reaction mechanism as shown in Figure 2 [5,6].

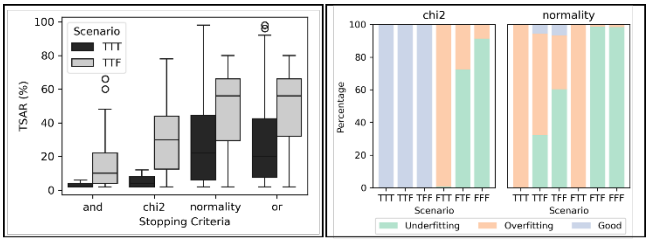
Although the models represent the system well, they might be statistically inadequate due to overfitting as indicated by the χ^2 test. Furthermore, small t values for all parameters suggest significant uncertainty in the estimates, pointing to the need for more informative data to improve parameter precision.

4.2 TSAR comparison across stopping criteria

The comparison of TSAR values across stopping criteria provides critical insights into DoE-SINDy's effectiveness in identifying suitable models for the SNAr reaction system. Higher TSAR indicates better performance in achieving specific scenarios.

Figure 4(a) shows that the strictness of stopping criteria significantly affects model identification. For the 'TTT' scenario, the 'normality' criterion achieves the highest average TSAR (over 20%), whereas 'chi2' achieves less than 5%, and 'and' nearly zero. This highlights 'normality' as the most effective criterion for identifying structurally and statistically adequate models. For the 'TTF' scenario, relaxing criteria from 'and' to 'or' improves TSAR from around 10% to 55%, with 'normality' and 'or' performing similarly. Less stringent criteria enable better identification of statistically adequate models, balancing efficiency and reliability.

The box plots reveal high dispersion in TSAR across datasets, reflecting the sensitivity of DoE-SINDy to data quality and experimental design. This underscores the importance of high-quality data and well-designed experiments for reliable model identification.



(a) (b)

Figure 4: (a) TSAR distribution for 'TTT' and 'TTF' scenarios under different stopping criteria across various data conditions; (b) Scenario distribution based on the χ^2 test using 'chi2' or 'normality' stopping criteria.

The distribution patterns observed in Figure 4(b) demonstrate the goodness-of-fit of the models categorised in each scenario. Notably, for the 'FTT' scenario under both the 'chi2' and 'normality' criterion, most models correctly identify the ground-truth structure but are excluded as overfitting. These models, with parameter re-estimation, are highly likely to achieve 'TTT'. Using 'normality' as the stopping criterion is more likely to retain such models compared to using 'chi2', explaining higher TSAR values of 'normality' shown in Figure 4(a). Its reliability is supported by Figure 3, where overfitting models passing the normality test align with experimental data under varying noise levels and experimental conditions.

5. CONCLUSION

This study demonstrates the effectiveness of DoE-SINDy in identifying kinetic models for SNAr reaction, successfully reconstructing statistically acceptable and structurally ground-truth models with as few as 7 experiments. Adopting the 'normality' stopping criterion allows to achieve the highest TSAR for identifying promising 'TTT' models, while stricter criteria, such as 'chi2' and 'and', tend to reject structurally accurate models due to overfitting.

However, the performance of DoE-SINDy varies under different experimental designs and data conditions. Further investigations into the impact of experimental design variables are necessary to optimize model identification and reduce experimental demands. The significant uncertainty in estimated parameters highlights the need for integrating techniques as model-based design of experiments (MBDoE) in the framework to improve parameter precision [16].Finally, as the ground-truth model is typically unknown in practice, further model discrimination is required to identify the most accurate representation of the system.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the funding support from UCL EPSRC DTP (EP/R513143/1 and EP/W524335/1).

REFERENCES

1. Rohrbach S, Smith AJ, Pang JH, Poole DL, Tuttle T, Chiba S, et al. Concerted Nucleophilic Aromatic Substitution Reactions *Angew Chem Int Ed* 58(46):16368–88 (2019) https://doi.org/10.1002/anie.201902216
2. Kim J, Hayashi Y, Badr S, Okamoto K, Hakogi T, Furukawa H, et al. Kinetic Study and Model-Based Design Space Determination for a Drug Substance Flow Synthesis Using an Amination Reaction via Nucleophilic Aromatic Substitution *Org Process Res Dev* 28(5):1793–805 (2024) https://doi.org/10.1021/acs.oprd.3c00380
3. Lu J, Paci I, Leitch DC. A broadly applicable quantitative relative reactivity model for nucleophilic aromatic substitution (SNAr) using simple descriptors *Chem Sci*. 13(43):12681–95 (2022) https://doi.org/10.1039/D2SC04041g
4. Hone CA, Boyd A, O'Kearney-McMullan A, Bourne RA, Muller FL. Definitive screening designs for multistep kinetic models in flow *React Chem Eng*. 4(9):1565–70 (2019) https://doi.org/10.1039/C9RE00180H
5. Hone CA, Holmes N, Akien GR, Bourne RA, Muller FL. Rapid multistep kinetic model generation from transient flow data *React Chem Eng*. 2(2):103–8 (2017) https://doi.org/10.1039/C6RE00109B
6. Agunloye E, Petsagkourakis P, Yusuf M, Labes R, Chamberlain T, Muller FL, et al. Automated kinetic model identification via cloud services using model-based design of experiments *React Chem Eng*. 9(7):1859–76 (2024) https://doi.org/10.1039/D4RE00047A
7. Chakraborty A, Sivaram A, Venkatasubramanian V. AI-DARWIN: A first principles-based model discovery engine using machine learning *Comput Chem Eng*. 154:107470 (2021) https://doi.org/10.1016/j.compchemeng.2021.107470.
8. Servia MÁ de C, Sandoval IO, Hii KK (Mimi), Hellgardt K, Zhang D, Chanona EA del R. The automated discovery of kinetic rate models – methodological frameworks *Digit Discov*. 3(5):954–68 (2024) https://doi.org/10.1039/D3DD00212H.
9. Brunton SL, Proctor JL, Kutz JN, Bialek W. Discovering governing equations from data by sparse identification of nonlinear dynamical systems *Proc Natl Acad Sci U S A*. 113(15):3932–7 (2016) https://doi.org/10.1073/pnas.1517384113
10. Lyu W, Galvanin F. DoE-integrated Sparse Identification of Nonlinear Dynamics for Automated Model Generation and Parameter Estimation in Kinetic Studies *Comput Chem Eng*. 53:169-174 (2024) https://doi.org/10.1016/B978-0-443-28824-1.50029-6
11. de Silva BM, Champion K, Quade M, Loiseau JC, Kutz JN, Brunton SL. PySINDy: A Python package for the Sparse Identification of Nonlinear Dynamics from Data *J Open Source Softw*. 5(49):2104–2104 (2020) https://doi.org/10.21105/joss.02104
12. Kaptanoglu AA, Silva BM de, Fasel U, Kaheman K, Goldschmidt AJ, Callaham J, et al. PySINDy: A comprehensive Python package for robust sparse system identification *J Open Source Softw*. 7(69):3994 (2022) https://doi.org/10.21105/joss.03994
13. Silvey SD. Statistical Inference. Routledge (1975)
14. D'Agostino R, Stephens MA. Goodness-of-Fit-Techniques. Routledge (1986)
15. Asprey SP, Macchietto S. Statistical tools for optimal dynamic model building *Comput Chem Eng*. 24(2):1261–7 (2000) https://doi.org/10.1016/S0098-1354(00)00328-8
16. Franceschini G, Macchietto S. Model-based design of experiments for parameter precision: State of the art *Chem Eng Sci*. 63(19):4846–72 (2008) https://doi.org/10.1016/j.ces.2007.11.034.