



Predictive Accuracy and Deep Ensemble Estimates of Epistemic Uncertainty in Constrained Neural ODEs for Kinetic Modeling

Downloaded from: <https://research.chalmers.se>, 2026-10-03 14:30 UTC

Citation for the original published paper (version of record):

Hajireza, K., Eklund, D., Olsson, L. et al (2026). Predictive Accuracy and Deep Ensemble Estimates of Epistemic Uncertainty in Constrained Neural ODEs for Kinetic Modeling. Industrial & Engineering Chemistry Research, In Press.
<http://dx.doi.org/10.1021/acs.iecr.6c03344>

N.B. When citing this work, cite the original published paper.

Predictive Accuracy and Deep Ensemble Estimates of Epistemic Uncertainty in Constrained Neural ODEs for Kinetic Modeling

Kian Hajireza, David Eklund,* Louise Olsson, Derek Creaser, and Ronnie Andersson

Cite This: <https://doi.org/10.1021/acs.iecr.6c03344>

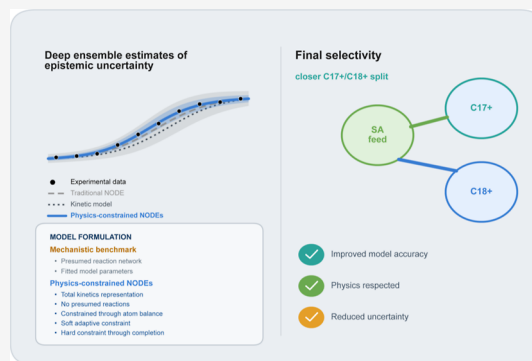
Read Online

ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: Robust modeling of complex reaction systems requires not only accurate predictions and physical consistency, but also reliable estimation of predictive uncertainty. Here, we investigate how mechanistic constraint type influences the deep ensemble estimates of epistemic uncertainty in neural ordinary differential equations for catalytic kinetics. We impose atom-conservation constraints either softly through a learnable, time-dependent regularization strategy or exactly through a completion-based architecture that guarantees conservation by construction. The approaches are evaluated on real experimental data for stearic acid hydrodeoxygenation under data-scarce conditions, with the deep ensemble estimates of epistemic uncertainty quantified using deep ensembles. Constrained neural ordinary differential equations improve predictive accuracy relative to a mechanistic kinetic benchmark from 80% up to 93% for the best constrained model and reduce the deep ensemble estimates of epistemic uncertainty compared with its corresponding unconstrained model. Although the hard-constrained model produces narrower predictive distributions, these do not necessarily better reflect the observed data. These findings demonstrate that embedding conservation laws into neural ordinary differential equations improves predictive performance and uncertainty behavior, while highlighting the need to assess uncertainty calibration alongside uncertainty magnitude.



1. INTRODUCTION

Reliable and accurate reactor models are essential for performing optimization¹ and process control.² Central to reactor modeling is the quantitative description of chemical kinetics, which classically incorporates first-principles mass and energy balances with kinetic rate expressions,³ typically parametrized using the Arrhenius equation. The exponential temperature dependence in the Arrhenius form introduces strong nonlinearity into the governing equations, complexifying parameter estimation, state reconstruction and process optimization. The full mathematical system description is a prerequisite for traditional reactor models.^{4–6} Such a formulation can prove to be cumbersome to construct in real-world applications where nonidealities may be present such as dead zones, axial dispersion and mass transfer limitations. Prior kinetic investigations related to biofuel production have provided valuable mechanistic insights^{7–12} however classical lumped or empirical kinetic models often rely on simplifying assumptions and can fail to capture the complexity of parallel and consecutive reactions, catalyst-dependent behavior and nonlinear dynamics across different operating conditions. These limitations highlight the challenges of relying solely on mechanistic models. Recently, hybrid modeling approaches have been proposed that integrate data-driven techniques with first-principles methods from chemical engineering.^{13,14} These approaches leverage the universal approximation capabilities of neural networks,^{15–18} while

constraining the model using domain knowledge through scientific machine learning frameworks.

The scientific machine learning frameworks have recently been advanced through the development of neural ordinary differential equations (ODEs),¹⁹ universal differential equations (UDEs)²⁰ and physics-informed neural networks (PINNs),^{21,22} among others. Neural ODEs¹⁹ offer an approach for learning continuous-time dynamics directly from irregularly sampled data. In this framework, a neural network represents the time derivative of the system state:

$$\frac{dy(t)}{dt} = \mathcal{NN}(y(t), t, \theta) \quad (1)$$

where $y(t) \in \mathbb{R}^d$ is the state vector, t represents time and $\mathcal{NN}: \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}^d$ is a neural network with learnable model parameters θ , where d is the number of state variables. Given the initial condition $y(t_0) = y_0$, the state at any future time t_1 is obtained by performing numerical integration on the ODE,

Received: June 26, 2026

Revised: September 7, 2026

Accepted: September 12, 2026

utilizing standard ODE solvers such as the Runge-Kutta methods. This formulation aligns naturally with first-principles models commonly used in chemical engineering in application areas within reaction kinetics and transport phenomena. Memory-efficient backpropagation is performed in neural ODEs by adopting the adjoint sensitivity method.²³ Neural ODEs in their traditional formulation are purely data-driven models exhibiting minimal inductive biases toward physical principles since they do not incorporate any physical knowledge and are considered weakly physics-informed with the underlying assumption that the training data follows an unknown differential equation. UDEs overcome this limitation by incorporating neural networks within first-principles models to represent unknown kinetics, constitutive relations or closure terms that are difficult to model from theory alone.^{24–26} UDEs enforce known physical relationships as hard constraints by directly embedding them into the model structure. Alternatively, neural ODEs can be trained with soft constraints by incorporating physics-based regularization terms into the loss function, which penalize deviations from expected physical behavior without strictly enforcing it. Such an approach is employed by Hajireza et al.²⁷ where mass balance constraints were integrated into the training objective, penalizing the model when predicted states violate mass conservation. Additional works have incorporated atom balance constraints into the physics-informed loss function, enforcing elemental mass conservation as a soft constraint during training.²⁸ Recent works employing neural ODEs in chemical engineering applications include modeling stiff chemical reaction systems,²⁹ integrating neural networks with discretized nonideal reactor models for mass transfer limited systems,³⁰ mechanistic parameter estimation through training of neural ODEs³¹ and last, encoding thermodynamic and stoichiometric constraints into neural ODEs.³²

This study extends the mechanistic kinetic model proposed by Arora et al.⁷ by replacing the kinetic model with a neural network, resulting in a neural ODE framework. Our approach utilizes flexible models such as neural networks, which do not require any assumptions of the specific reactions taking place. The objective of our proposed method is to serve as a surrogate for the kinetic model presented by Arora et al.,⁷ offering an alternative modeling approach. Consequently, employing an additive UDE where the neural network is intended to learn missing physical phenomena, is not suitable in this context. Such approaches typically require prior knowledge of kinetic parameters and model structure. In contrast, our method operates without these assumptions and unlike conventional mechanistic kinetic modeling, the neural ODE does not require specification of a reaction network followed by nonlinear regression of the associated kinetic parameters. The mechanistic kinetic model is used as a benchmark and is trained on all available data sets. In contrast, the neural network is trained using a leave-one-out strategy, where one data set is withheld and used as an independent test set to evaluate generalization performance. Three modeling strategies are explored:

- Traditional neural ODE: the entire right-hand side is modeled by a neural network scaled with the catalyst weight without incorporating any domain-specific knowledge.
- Self-adaptive soft-constrained neural ODE: identical to the traditional neural ODE with the exception of incorporating adaptive regularization weights in the loss

function, penalizing the model for atom balance deviations.

- Hard-constrained neural ODE: the neural network architecture is constrained by reducing the output dimensions to only produce output combinations that lie within the null space of the atom-molecule vector, resulting in physically invalid solutions becoming unreachable during training.

The self-adaptive soft-constrained approach is inspired by previous work in PINNs^{33,34} where gradient ascent is used to train the regularization weights alongside with the network parameters. However, this approach was applied to PINNs, utilizing information about the governing equations. To the best of our knowledge, adoption of the self-adaptive regularization approach utilizing gradient ascent for a vector of regularization weights has not been previously studied for neural ODEs, where previous studies have linearly increased the regularization weight for constraint violations during minimization.³⁵ In our approach, we are adopting neural ODEs as a form of representation discovery where we do not rely on the mechanistic model unlike PINNs. Instead, the aim of the study is to explore and investigate to what extent self-adaptive regularization can be adopted together with known conservation laws, integrating this method with a neural ODE framework and evaluating its potential for providing useful insights for reactor dynamics for estimating selectivity, beyond what mechanistic models can achieve for data sets that contain limited experimental information. For this reason we employ a real-world data set from literature,⁷ where previous attempts using mechanistic modeling for hydrodeoxygenation (HDO) of stearic acid, have not delivered fully satisfactory predictive performance. While neural ODEs and related physics-informed networks have shown promise in chemical engineering, most studies to date validate these methods on synthetic problems or well-controlled simulations (e.g., the Lorenz system, Robertson's kinetics).^{29,36,37} Although, notable exceptions include the work of but are not limited to Fedorov et al., Kircher and Votsmeier, Döppel et al.,^{31,38,39} applications to real-world experimental data sets nevertheless remain relatively limited, in part because data collection can be time-consuming and costly. In this work, we further explore the applicability of physics-informed neural ODEs to real-world experimental data by studying stearic acid hydrodeoxygenation, demonstrating efficient learning of the underlying kinetics despite the limited number of samples available per run.

In addition to the soft-constrained formulation, a hard-constrained neural ODE is investigated in which carbon atom conservation is enforced exactly at every ODE evaluation step through a completion approach. In this formulation, the neural network predicts only a subset of chemical species directly, and the remaining species are reconstructed analytically so as to satisfy the carbon atom balance exactly.^{39–43} Conceptually, this completion-based construction is closely related to the architecture-constrained networks introduced by Beucler et al.⁴¹ and to the completion-based framework for constrained optimization proposed by Donti et al.,⁴⁰ in which a subset of variables is predicted directly and the remaining variables are completed so as to satisfy equality constraints exactly. Döppel and Votsmeier⁴² showed atom conservation constraint can be enforced on the stoichiometric coefficients of a CRNN by representing them in a reduced-row-echelon null-space basis, referred to as the atom-conservation matrix and used this

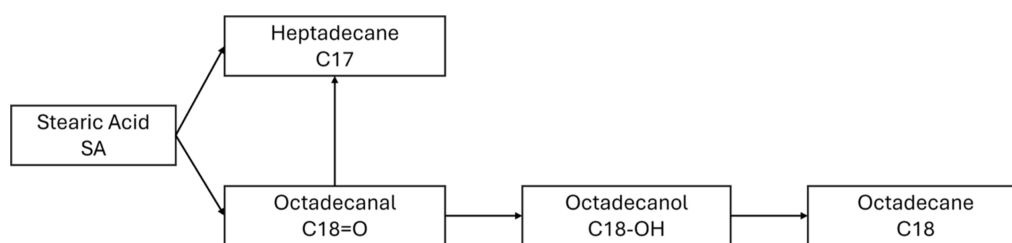


Figure 1. Key reaction pathways in the hydrodeoxygenation of stearic acid.

formulation within a neural ODE for mechanism discovery in combustion and biodiesel kinetics. Döppel et al.⁴³ subsequently applied completion to PINN surrogates of catalytic reactors with detailed surface kinetics, demonstrating improved accuracy, data efficiency, and robustness to noise relative to projection-based approaches. They also demonstrated completion at the source-term level in neural ODE surrogates for the dual-catalyst DME synthesis system.³⁹

This approach eliminates the need to incorporate an additional term in the loss function and guarantees constraint satisfaction by network architecture rather than through optimization. Similar to the self-adaptive approach, both methods circumvent the need to tune regularization weights, making them well suited to systems with only a small number of data sets.

Related work has incorporated atom conservation in neural ODEs through both soft constraints^{28,44–47} and hard constraints,^{39,42,48–50} while other studies have applied posthoc correction to trained models rather than enforcing conservation during training.⁵¹ Although these studies have examined the effect of atom-conservation constraints on predictive accuracy, their implications for epistemic uncertainty remain unexamined. While Üstün and Paykani⁵² demonstrated the incorporation of soft mass conservation constraints into deep ensemble feedforward neural networks for chemical kinetics integration and Nai et al.⁵³ highlighted a bias-variance trade-off induced by soft physics-informed constraints under epistemic uncertainty due to systematic model error in LSTM-based crystallization surrogates, the impact of physics-based constraints on epistemic uncertainty estimation in neural network surrogate models remains largely unexplored. Whereas Ott et al.⁵⁴ investigated uncertainty quality under a Laplace approximation with Hamiltonian conservation laws in mechanical systems,⁵⁵ our study focuses on atom conservation in chemical reactor systems and on how the choice of constraint formulation governs epistemic uncertainty in neural ODE ensembles. To our knowledge, no prior study has systematically examined how soft and hard atom-conservation constraints affect the epistemic uncertainty estimates in neural ODE ensembles. The primary contribution of this work is therefore a systematic comparison of soft and hard atom-conservation constraints in neural ODE ensembles, focusing on how they influence epistemic uncertainty alongside predictive accuracy in chemical reactor systems. As a secondary methodological contribution, we also introduce adaptive regularization for the soft-constraint setting, enabling the regularization weights to be adjusted automatically during training. Epistemic uncertainty is quantified using deep ensembles, whose implementation and motivation are described in Section 2.3. Therefore, the deep ensemble estimates of epistemic uncertainty represented by random initialization is studied, rather than the measure of total epistemic uncertainty.

2. MATERIALS AND METHODS

2.1. Reaction System

The reaction system studied is the catalytic hydrodeoxygenation of stearic acid (SA, $C_{18}H_{36}O_2$), which is a model compound for renewable oils, over a sulfided NiMo/ Al_2O_3 catalyst. The system consists of three competing deoxygenation routes: direct-HDO producing C18 hydrocarbons (octadecane) with water elimination, in addition to decarboxylation and decarbonylation which yield C17 hydrocarbons (heptadecane) with CO_2 or CO release, respectively as depicted in Figure 1. In addition to the main reactions illustrated in Figure 1, various side-reactions occur, such as water-gas shift, methanation, isomerization, hydrogenation of alkenes, esterification and cracking reactions.

The reaction mechanism in Figure 1 proceeds through sequential hydrogenation steps. Stearic acid is first reduced to octadecanal ($C18=O$, $C_{18}H_{36}O$), which serves as an intermediate. Octadecanal can either undergo further hydrogenation to octadecanol ($C18-OH$, $C_{18}H_{37}OH$), producing C18 products via direct-HDO or decarbonylation to form C17 products directly. Experiments with octadecanal and octadecanol feeds confirmed these pathways. For more details regarding the reaction system, the reader is encouraged to refer to the work by Arora et al.⁷ In addition to the C17 and C18 products, their corresponding isomers are also formed and are modeled separately in this work, whereas the kinetic model presented by Arora et al.⁷ treats them as lumped species. The lumped species for heptadecane and octadecane are in this work denoted as C17+ and C18+, respectively.

2.2. Training Procedure

Kinetic experiments were conducted for the temperature range 275–325 °C, pressure ranges 40–70 bar hydrogen pressure and feed concentration ranges 0–8 wt% SA feed concentration. For the experiments with 0 wt% SA feed concentration, the intermediates $C18-OH$ and $C18=O$ are used as the starting material to provide additional information on the reaction pathways. A total of eight data sets with different reaction conditions are used during the training process, employing the center point as the test data set, essentially adopting a leave-one-out strategy to test the trained models. Figure 2 illustrates the experimental design for the seven of the nine experiments that are incorporated into the model development procedure, where pressure is denoted as P , temperature as T and the initial number of moles of SA as $n_{0,SA}$ with Table 1 presenting the experimental conditions for each given experiment number. Notably, experiments 7 and 8 use octadecanal and octadecanol as their starting materials. The experimental measurements are treated as deterministic observations during model training and experimental uncertainty is not incorporated into the loss function. Arora et al.⁷ reported that repeated baseline experiments conducted at 300 °C, 50 bar, 1000 rpm and an initial stearic acid concentration of 5 wt% resulted in a relative standard deviation of approximately 5% for the hydrodeoxygenation conversion activity. However, no uncertainties were reported for the other experimental conditions. The number of temporal data points remains limited for each experiment, which makes identification of the underlying dynamics challenging. Additional experimental observations would generally be expected to provide more information to the model and reduce the epistemic uncertainty of the neural ODE models.⁵⁶ Nevertheless, species- and time-resolved experimental uncertainties

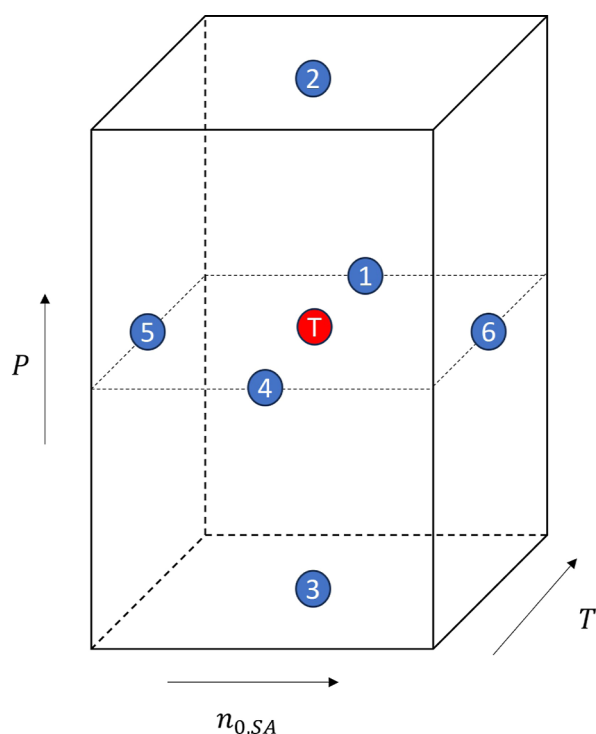


Figure 2. Experimental design where the blue and red scatter points represent data set(s) used for training and test, respectively.

Table 1. Experimental Conditions for the Various Experiment Numbers

experiment number	temperature [°C]	pressure [bar]	amount of SA [wt%]
1	325	50	5
2	300	70	5
3	300	40	5
4	275	50	5
5	300	50	2
6	300	50	8
7	275	50	0 ^a
8	275	50	0 ^b
Test	300	50	5

^aThis experiment employed C18=O as the starting material. ^bThis experiment employed C18–OH as the starting material.

were not reported and are therefore not explicitly incorporated in the present analysis.

Network parameter initialization is performed using Glorot initialization. For minimizing the loss function, we use the Adam optimizer.⁵⁷ To mitigate convergence to a saddle point, we implement a hybrid optimization strategy that combines several established techniques in a novel configuration. Specifically, the learning rate is adaptively reduced when the training loss plateaus, as detected by monitoring the variation in loss over a sliding window of epochs. If the loss remains stagnant at the minimum learning rate for a prolonged period, we trigger a warm restart by resetting the optimizer's internal state, reinitializing momentum buffers while retaining the network parameters from the previous epoch. This conditional restart increases the learning rate to a fraction of its original value, enabling the optimizer to escape local minima and resume effective exploration of the parameter space. While plateau-based learning rate decay is typically applied to test loss, we instead monitor training loss due to the limited available data. This approach in combination with periodic resets of the optimizer's momentum buffers, enables the model to escape flat regions of the loss landscape. While each component of this strategy is

individually known,⁵⁸ their integration in this form particularly for neural ODEs represents a novel and effective approach to improving training robustness and convergence efficiency. When selecting the network architecture, stability, predictive accuracy and convergence rate are considered as primary criteria. Accordingly, a feedforward neural network comprising four hidden layers with 20 neurons per layer is employed, using the Swish activation function.⁵⁹ The model inputs are standardized using Z-score normalization.

2.3. Uncertainty Estimation

To systematically assess epistemic uncertainty, we employ a deep ensemble approach.⁶⁰ Specifically, we train $M = 90$ independently initialized models using identical network architectures, hyperparameters, and data sets, differing only in the random seeds used for parameter initialization. This setup enables ensemble members to explore distinct local minima of the training landscape⁶¹ and thereby capture variability arising from model uncertainty.

Predictive uncertainty in machine learning models is commonly decomposed into aleatoric and epistemic components. Aleatoric uncertainty arises from irreducible noise in the data, such as experimental measurement error, whereas epistemic uncertainty reflects uncertainty in the model due to limited data or incomplete knowledge of the underlying process.^{56,62,63} In the present work, we focus specifically on epistemic uncertainty, which in our setting is associated with incomplete knowledge of the reaction kinetics and underlying reaction mechanisms. We therefore quantify the deep ensemble estimates of epistemic uncertainty through intermember disagreement within the ensemble, without explicitly introducing aleatoric uncertainty terms into the model. For conciseness, we refer to this deep ensemble estimate as epistemic uncertainty hereafter.

The ensemble mean prediction, $\hat{z}$, is defined as

$$\hat{z} = \frac{1}{M} \sum_{i=1}^M z_i \quad (2)$$

where z_i denotes the prediction of the i -th ensemble member. Epistemic uncertainty is then quantified as the standard deviation across ensemble predictions:

$$\sigma_{\text{epistemic}} = \sqrt{\frac{1}{M-1} \sum_{i=1}^M (z_i - \hat{z})^2} \quad (3)$$

All uncertainty metrics are computed in product distribution (%) for each species (SA, C17+, and C18+) at the experimental measurement time points, excluding $t = 0$.

Although Bayesian inference methods can also quantify epistemic uncertainty in neural ODEs,^{64–68} we adopt deep ensembles as the uncertainty estimation framework because they can be implemented within a standard training pipeline, do not require modification of the underlying model formulation or optimization objective and have shown strong practical performance for uncertainty estimation in UDE settings.⁶⁹ In addition, independently initialized ensemble members can explore functionally distinct solutions across multiple local minima.⁶¹ Since we are adopting deep ensembles through random initialization of network parameters, the deep ensemble estimate of epistemic uncertainty is being studied rather than the total epistemic uncertainty.

2.4. Neural ODE System

The number of inputs to the neural network corresponds to the number of chemical species being modeled, with two additional inputs representing temperature and pressure. The right-hand sides of the state equations for temperature and pressure are set to zero, as the experiments were conducted isothermally and the reaction are treated isobarically due to the small changes in pressure during the experiment, typically less than 5%. This results in the number of outputs in the neural network corresponding to the number of chemical species being modeled. The system of ODEs is defined as

$$\begin{aligned}\frac{dn}{dt} &= NN(\mathbf{n}(t), T, P, \theta) \cdot w, \\ \frac{dT}{dt} &= 0, \\ \frac{dP}{dt} &= 0\end{aligned}\quad (4)$$

where $\mathbf{n}(t) \in \mathbb{R}^S$ is the vector of molar amounts of the S chemical species at time t , T and P correspond to temperature and pressure, respectively, w is the catalyst weight, acting as a scaling factor on the output of the neural network. The temperature and pressure are held constant due to the isothermal and isobaric assumptions. The flexibility of the neural network circumvents explicit formulation for which reactions are taking place, thus enabling the model to capture side-reactions that are not explicitly included in mechanistic kinetic models.

The loss function minimizes the mean squared error (MSE) between the model predictions $\hat{n}$ and the experimental data for moles n , defined in eq 5 as

$$\mathcal{L}_{\text{data}}(\theta) = \frac{1}{S} \frac{1}{D} \sum_{i=1}^S \sum_{j=1}^D \sum_{k=1}^{N_j} \frac{1}{N_j} (n_{i,j,k} - \hat{n}_{i,j,k}(\theta))^2 \quad (5)$$

where D represents the number of different data sets and starting conditions used and N_j is the number of experimental time points, which is a function of the corresponding data set, j . All modeled species contribute to $\mathcal{L}_{\text{data}}$ with equal explicit weighting where no species is deliberately prioritized. Species-specific weighting could be useful when justified by measured species-dependent uncertainty. However, species- and time-resolved measurement uncertainties are unavailable for this data set.

2.5. Atom Balance-Regularized Neural ODE

As an alternative to the traditional neural ODE, we introduce a physics-based regularization approach that encourages conservation of moles of carbon atoms during training. This method retains the flexible structure of the neural ODE but incorporates fundamental conservation laws through incorporation into the loss function. The total loss function for this approach is

$$\mathcal{L}_{\text{total}}(\theta) = \mathcal{L}_{\text{data}}(\theta) + \lambda_C \mathcal{L}_{\text{carbon}}(\theta) \quad (6)$$

where λ_C is the regularization weight and $\mathcal{L}_{\text{carbon}}$ quantifies violations of carbon balance and is computed as

$$\mathcal{L}_{\text{carbon}} = \frac{1}{D} \frac{1}{K} \sum_{j=1}^D \sum_{l=1}^K (C_{\text{initial},j} - \hat{C}_{j,l}(\theta))^2 \quad (7)$$

where K is the number of collocation points, C_{initial} is the total moles of carbon atoms initially present in the liquid phase. As reported by Arora et al.,⁷ the mass balance in the studied system closes when only the liquid-phase components are considered. Accordingly, the atom-balance constraint was imposed using only the liquid-phase species, where

$$\hat{C}(t_l, \theta) = \sum_{i=1}^S \nu_i^C \hat{n}_i(t_l, \theta) \quad (8)$$

is the total moles of carbon in the liquid phase at time t_l , with ν_i^C being the number of carbon atoms in species i . This regularization term penalizes deviations from the initial number of moles of carbon. The regularization weight λ_C controls the trade-off between fitting experimental data and satisfying the carbon balance constraint. Furthermore, the gaseous components exhibit a near-total absence of observations within the transient regime relevant to the neural ODE framework. Across the entire total data set, one of the data sets contains two time points with dynamic measurements and only three of the five gaseous components are observed even at these sparse time points, while the remaining two species are entirely unmeasured. The remaining data sets contain no measurements in the transient regime of the gaseous species. Since the neural network in the neural ODE

framework learns the continuous-time dynamics from the structure of temporal trajectories; including these gaseous components would introduce additional inputs without sufficient information to identify their dynamics meaningfully. Their inclusion would expand the input space and increase model complexity, despite the available data being insufficient to inform the corresponding dynamics in a meaningful way. Moreover, in line with the curse of dimensionality, increasing the dimensionality of the learning problem generally raises the amount of data required to learn the dynamics robustly. Additionally, Arora et al.⁷ also attempted to model the gas-phase products but reported generally poor predictive performance, in part because the transient gas-phase measurements were sparse. The presently available gas-phase observations therefore did not provide sufficient temporal resolution and coverage both to train the continuous-time gas-phase dynamics and to reserve independent transient measurements for robust validation. Consequently, we model solely the liquid components.

The physical constraint is restricted to solely carbon conservation because the data set contains measurements of the main carbon-containing reactants, intermediates, and products in the liquid phase, allowing a meaningful carbon-chain balance to be formulated. Complete hydrogen and oxygen balances could not be constructed from the available measurements. In the data set from Arora et al.,⁷ hydrogen was supplied in excess and neither liquid water nor water vapor was measured. Since water is an important product of the hydrodeoxygenation pathways, imposing hydrogen or oxygen conservation would require assumptions regarding unobserved quantities. If sufficiently complete measurements become available, incorporating additional elemental conservation constraints would be a natural extension of the present method and could further improve the physical consistency of the learned dynamics. Additionally, Arora et al.⁷ reported that the liquid-product carbon-chain mole balance which was based on species with 17 or 18 carbon atoms, was within $100 \pm 10\%$, providing the basis for imposing the carbon-only conservation constraint.

2.6. Self-Adaptive Carbon Balance Regularization

The physics-informed penalty described in Section 2.5 introduces a regularization weight through eq 6 that must be specified during training. In conventional workflows, this weight is selected through manual hyperparameter tuning, which requires partitioning the available data sets into training, validation, and test subsets, with the validation set used exclusively for hyperparameter selection. While such a procedure is routine in data-rich settings, it is less suitable here because the total data set comprises a limited and fixed number of experiments reported in the literature,⁷ with no practical opportunity for augmentation through additional experiments. Under these conditions, allocating part of the data to validation would reduce the information available for learning the system dynamics. To avoid this loss, the regularization weight is treated as a trainable parameter and optimized jointly with the model parameters, thereby preserving the full data set for training.

Furthermore, a single fixed weight treats all temporal regions uniformly, failing to account for time points where carbon conservation is more difficult to enforce. To address these limitations, we extend the regularization framework by replacing the fixed scalar weight λ_C with a set of learnable point-wise weights $\lambda = (\lambda_1, \dots, \lambda_K)$ that adapt automatically during training, where K is the number of collocation points. Thus, the collocation points introduce additional model parameters into the optimization problem. This self-adaptive weighting scheme, inspired by recent advances in PINNs,^{33,34} allows the regularization strength to vary across the temporal domain.

The optimization problem for training the model parameters is reformulated from the traditional setup by incorporating the regularization parameters into the training process. Particularly, we introduce $\lambda = (\lambda_1, \dots, \lambda_K) \in \mathbb{R}^K$, which are self-adaptive weights applied to the carbon balance constraint and reformulate the objective function as

$$\mathcal{L}(\theta, \lambda) = \mathcal{L}_{\text{data}}(\theta) + \tilde{\mathcal{L}}_{\text{carbon}}(\theta, \lambda) \quad (9)$$

The carbon balance loss term is extended to incorporate learnable point-wise weights:

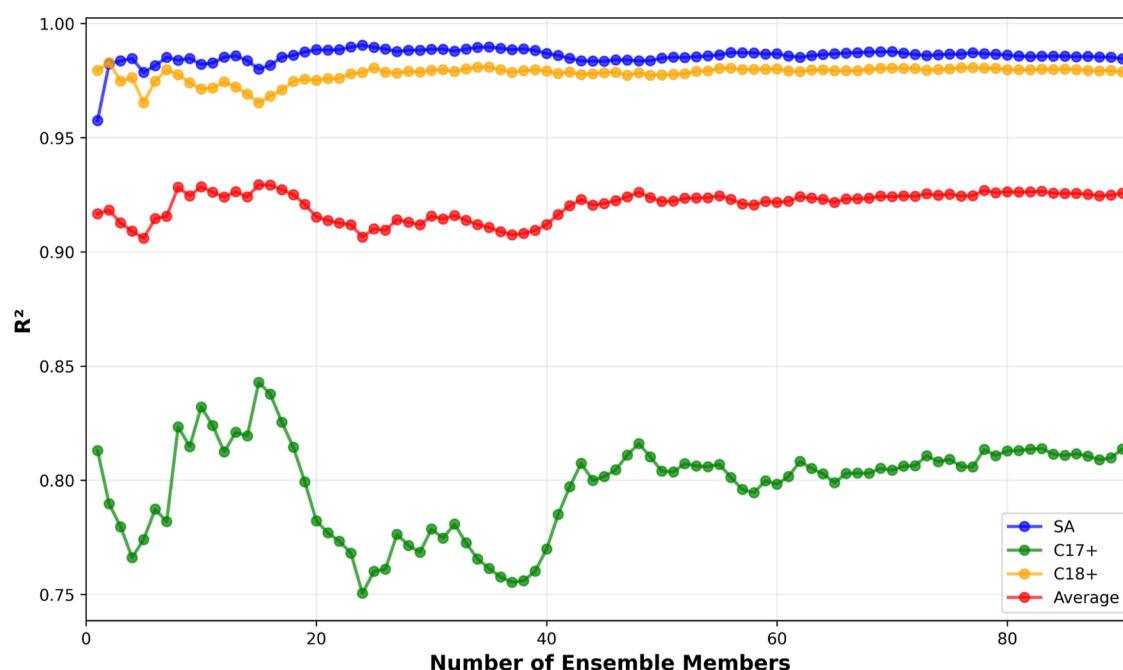


Figure 3. Mean and median model accuracy, R^2 for the test data set as a function of number of ensembles for the self-adaptive soft-constrained neural ODE.

$$\tilde{\mathcal{L}}_{\text{carbon}}(\theta, \lambda) = \frac{1}{D} \frac{1}{K} \sum_{j=1}^D \sum_{l=1}^K \lambda_l (C_{\text{initial},j} - \hat{C}_{j,l}(\theta))^2 \quad (10)$$

where $K = 500$. Unlike the conventional formulation in eq 7 that employs a single fixed scalar weight λ_C , this formulation allows each collocation point to have its own adaptive weight λ_l .

The optimization problem in eq 9 is solved by gradient descent with respect to θ and gradient ascent with respect to λ . The gradients are

$$\nabla_{\theta} \mathcal{L} = \nabla_{\theta} \mathcal{L}_{\text{data}} + \nabla_{\theta} \tilde{\mathcal{L}}_{\text{carbon}} \quad (11)$$

$$\nabla_{\lambda} \mathcal{L} = \frac{1}{DK} \sum_{j=1}^D \begin{bmatrix} (C_{\text{initial}} - \hat{C}_j(t_1, \theta))^2 \\ \vdots \\ (C_{\text{initial}} - \hat{C}_j(t_K, \theta))^2 \end{bmatrix} \quad (12)$$

By performing gradient ascent with respect to λ , the self-adaptive weights increase at time points where the carbon balance error is large, thereby penalizing the network for violating conservation laws. Conversely, weights remain small at time points where carbon balance is easily satisfied, allowing the network to prioritize fitting experimental data in those regions. Alternative approaches have given residual weights a statistical interpretation through maximum-likelihood formulations, where weighting is derived from estimated covariance or precision matrices.⁷⁰ However, in the present formulation, the adaptive weights are updated directly by gradient ascent and therefore act as adaptive penalty weights rather than variance or precision estimates.

The parameter updates for θ and λ have individual learning rates, where η_{θ} and η_{λ} are the learning rates for network parameters and self-adaptive weights, respectively. Note that we implement gradient ascent through the positive sign in front of the gradient term in the update formulas. Similar to the network parameters, θ , the regularization weights, λ , are independently initialized using different sets of random seeds for uncertainty estimation.

Both sets of parameters are optimized using the Adam optimizer with separate optimizer states to maintain independent adaptive learning rate schedules. The self-adaptive weights λ are initialized from a uniform distribution $\mathcal{U}(0, 10^{-4})$ to provide small positive starting values. The optimization is performed for up to 30,000 epochs using a

plateau-based termination criterion as described in more detail in Section 2.2.

2.7. Hard Constraint via Completion

Rather than penalising violations of the atom balance through a loss term, the atom conservation constraint can be embedded directly into the network architecture, ensuring exact satisfaction at every evaluation of the ODE right-hand side. Hard constraints for atom balances can be enforced exactly through for instance orthogonal projection or completion.^{39,41,42} For this work, the completion approach is adopted to enforce atom conservation as a hard architectural constraint while simultaneously reducing model complexity, by restricting the predicted outputs to the physically realizable null space of the atom-molecule vector and expressing the dependent species source term as fixed linear combinations of a reduced set of key species outputs. This approach was originally introduced for chemical kinetics surrogate modeling by Döppel and Votsmeier⁴² and generalized to source term enforcement by Döppel et al.³⁹ Regarding the choice for the dependent species, we follow the heuristic of Döppel et al.³⁹ and Döppel and Votsmeier,⁴² which states that the dependent species should exhibit the largest absolute source term magnitude, resulting in the choice of stearic acid as the dependent species. Computing a small rate as a linear combination of larger rates introduces relative errors through catastrophic cancellation and assigning the largest-magnitude species as dependent minimizes likelihood of this occurring.

A basis, B of the null space of the atom-molecule vector N is first computed using for instance singular value decomposition and subsequently transformed into reduced row echelon form to obtain the completion matrix $C \in \mathbb{R}^{S \times N_{\text{key}}}$, where $N_{\text{key}} = S - \text{rank}(N) = 6$. The neural network thus predicts a reduced set of unconstrained source terms $\mathcal{NN}_{\text{reduced}} \in \mathbb{R}^{N_{\text{key}}}$, corresponding to the key species. Specifically, C is incorporated as a final linear layer in the neural network with fixed, precomputed weights, consequently leading to a reduction in the number of network output nodes and thus circumventing any modification in the loss or the optimizer, unlike the integration of the soft constraint. The complete source term is reconstructed via the linear mapping shown in eq 13

$$\mathcal{NN} = C \cdot \mathcal{NN}_{\text{reduced}} \quad (13)$$

In the present system, SA is the most abundant species throughout the experiments and is therefore selected as the dependent species (as

Traditional neural ODE

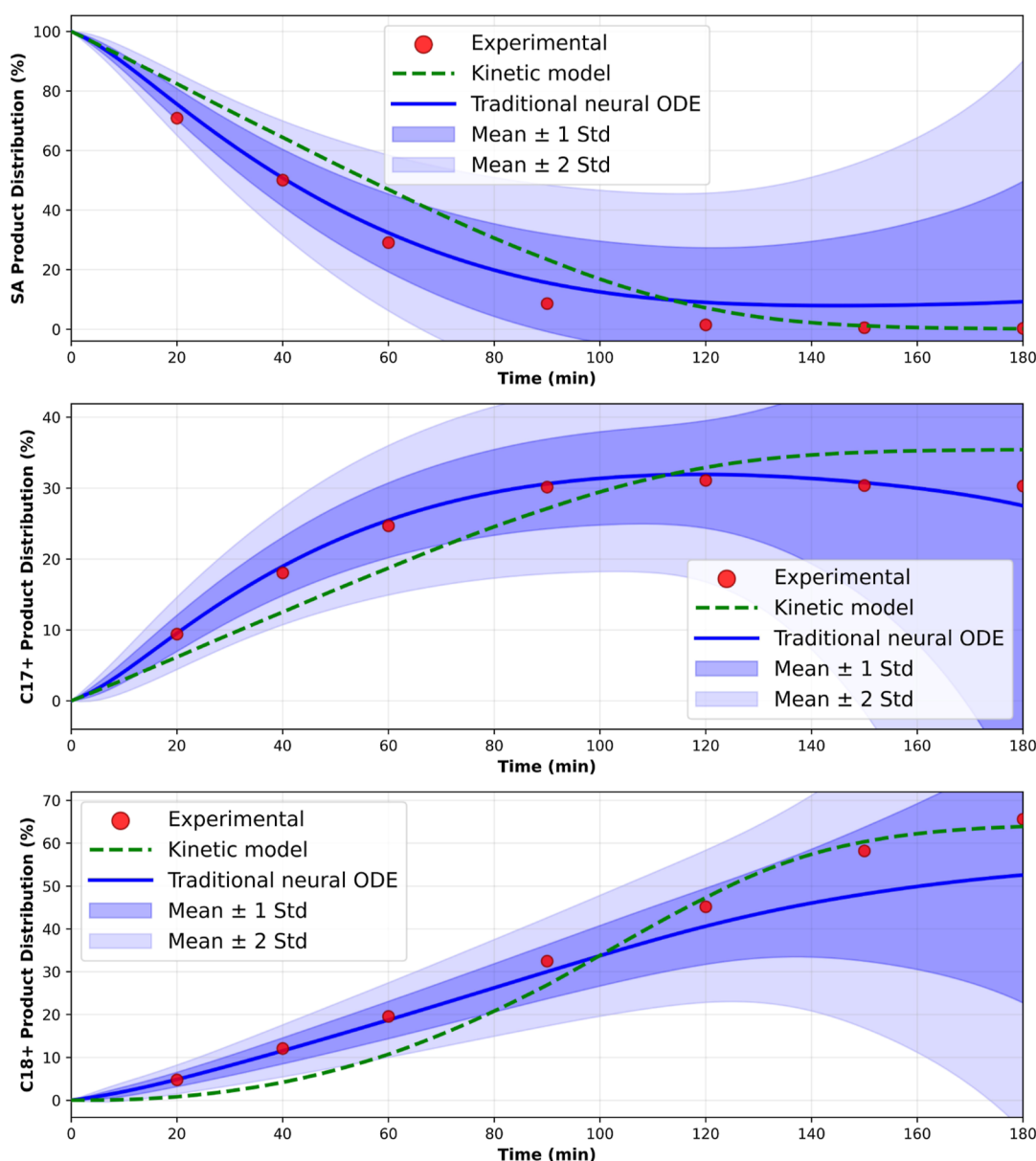


Figure 4. Ensemble-based uncertainty estimation for traditional neural ODE predictions without mechanistic priors predicting product distributions for the center point in the experimental design for time-resolved product distributions for SA, C17+ and C18+ species. Red circles represent experimental measurements. Solid blue lines show ensemble mean predictions from deep ensembles. Green dashed lines show predictions from the mechanistic kinetic model for comparison.

shown in row 0 of C), while the remaining six species are treated as key species predicted directly by the network. The resulting completion matrix is shown in eq 14

$$C = \begin{bmatrix} -1 & -1 & -\frac{17}{18} & -1 & -\frac{17}{18} & -1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (14)$$

where rows correspond to [SA, C18=O, C18-OH, C17=, C18=, C17, C18] and columns to the six key-species rates.

2.8. Software and Hardware

For this study, we utilize the Python programming language, employing JAX as the deep learning framework, providing automatic differentiation and just-in-time compilation capabilities. Numerical integration of the differential equations is performed using Diffrax, a JAX-based library offering adaptive step-size ODE solvers. Gradient-based optimization is conducted using Optax. NumPy is used for numerical operations and data preprocessing tasks, while Pandas is used for loading of experimental data sets. Visualization of trajectory predictions is performed using Matplotlib.

Training of the ensemble members is performed on the Leonardo supercomputer at Cineca with an Intel Xeon 8358 CPU and LUMI with

Soft-constrained neural ODE

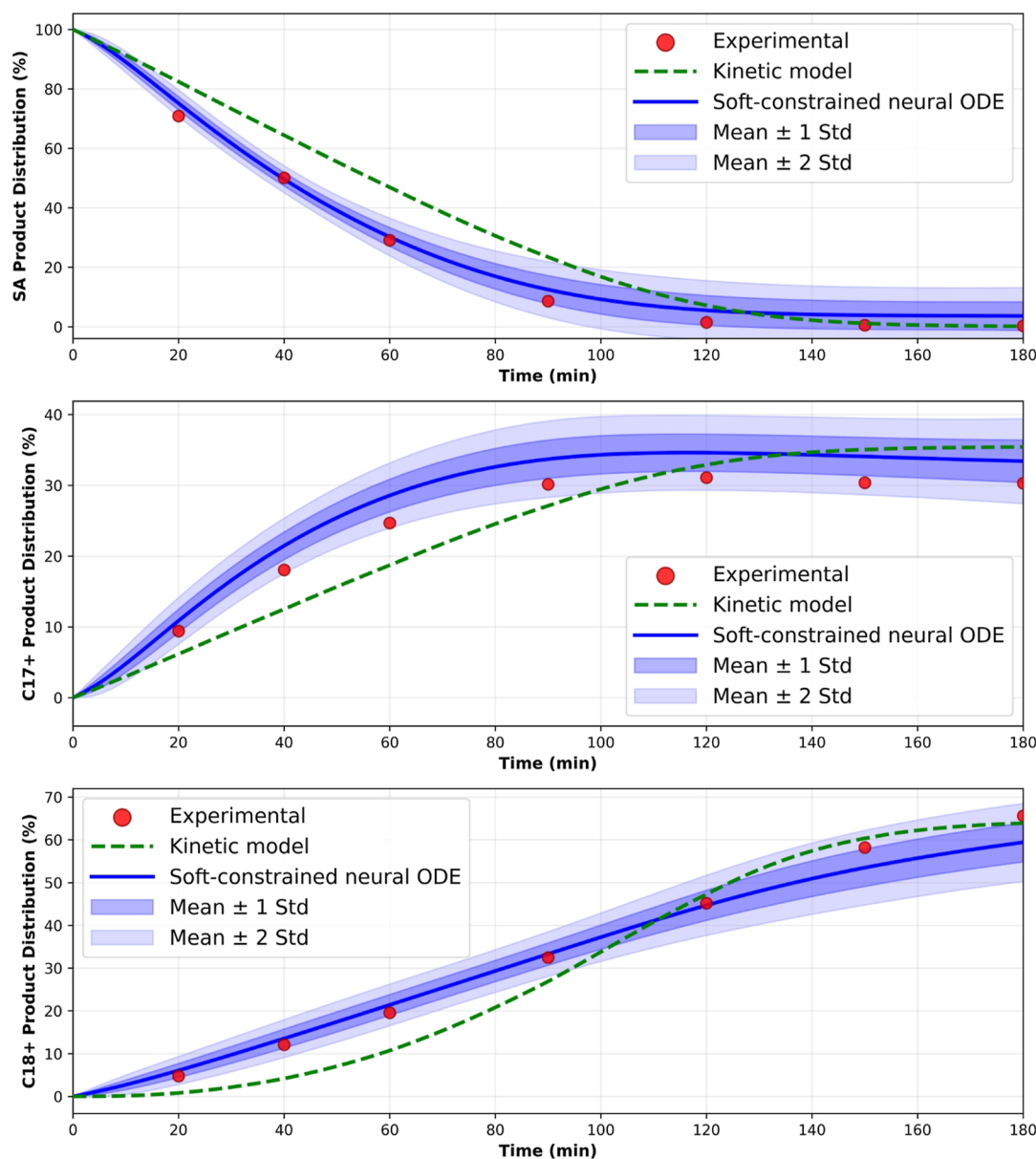


Figure 5. Ensemble-based uncertainty estimation for self-adaptive soft-constrained neural ODE models predicting product distributions for the center point in the experimental design for time-resolved product distributions for SA, C17+ and C18+ species. Red circles represent experimental measurements. Solid blue lines show ensemble mean predictions from deep ensembles. Shaded regions indicate model uncertainty. Green dashed lines show predictions from the mechanistic kinetic model for comparison.

an AMD EPYC 7763 CPU. Inference is conducted on 11th Gen Intel (R) Core (TM) i7–11800H @ 2.30 GHz. While individual runs are computationally tractable on a standard desktop workstation, parallel execution for the training of ensemble members benefit from high-performance computing resources.

3. RESULTS AND DISCUSSION

3.1. Ensemble Members

Following the recommendation of Hajireza et al.²⁷ for employing physics-based regularization for chemical reactor systems, we space the collocation points in time logarithmically to capture the key dynamics which typically take place in the earlier stages of sampling for chemical reactor systems.

Table 2. Prediction Accuracy, R^2 for the Mean Ensemble Predictions for the Center Point in the Experimental Design for Stearic Acid^a

model	SA	C17+	C18+	total
kinetic model	0.81	0.67	0.94	0.80
soft-constrained neural ODE	0.98	0.81	0.98	0.93
hard-constrained neural ODE	1.00	0.68	0.99	0.89

^aThe prediction accuracy of the kinetic model is also presented as a benchmark.

Therefore, utilizing logarithmic sampling, we populate the earlier stages of the temporal domain with a higher density of collocation points, ensuring accurate resolution of fast transients

Hard-constrained neural ODE

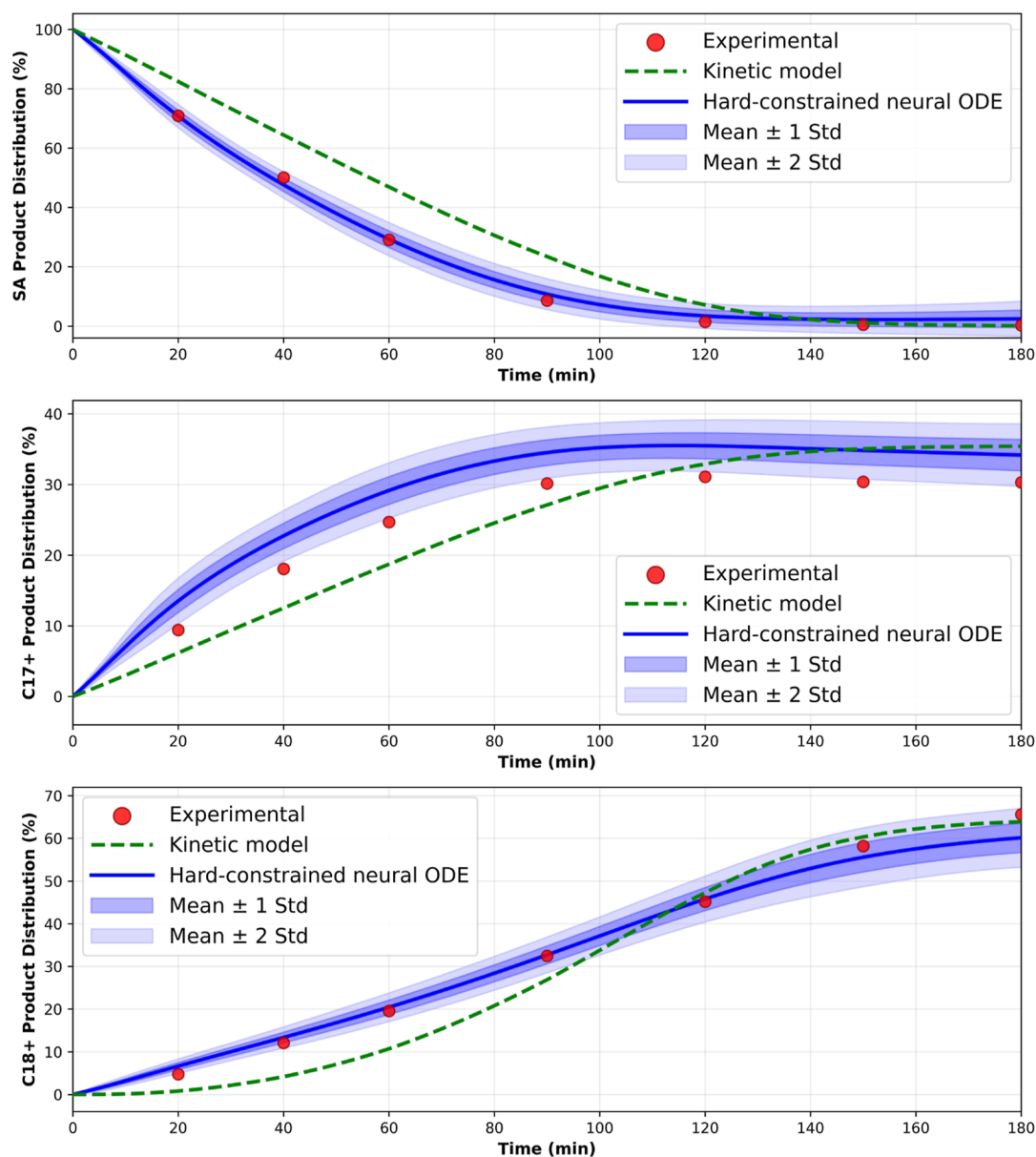


Figure 6. Ensemble-based uncertainty estimation for hard-constrained neural ODE models predicting product distributions for the center point in the experimental design for time-resolved product distributions for SA, C17+ and C18+ species. Red circles represent experimental measurements. Solid blue lines show ensemble mean predictions from deep ensembles. Shaded regions indicate model uncertainty. Green dashed lines show predictions from the mechanistic kinetic model for comparison.

and steep gradients that dominate the initial dynamics of chemical reactor systems. For the uncertainty estimation we utilize 90 ensemble members. The number of ensemble members is motivated by Figure 3, which demonstrates the test accuracy showing a clear plateau with decreasing amplitudes for the oscillations as the number of ensemble members increases for both the individual species and in aggregate.

3.2. Model Prediction and Uncertainty Estimation

In this study, we use the mechanistic model originally presented by Arora et al.⁷ as a benchmark for evaluating model performance. Arora et al.⁷ have lumped the modeling of product distributions of heptadecane and octadecane with their corresponding isomers. As explained in Section 2.1, we model

each isomer separately, however, to enable a direct comparison, the figures in this section present results by aggregating the predictions of the product distributions of heptadecane and octadecane together with their respective isomers. To assess generalization, we designate the center point in the experimental data set as the test set for both the data-driven and hybrid model. It is important to note that Arora et al.⁷ included the center point in their training data. Consequently, our comparison contrasts training accuracy for the mechanistic model with test accuracy for the data-driven and hybrid models. This setup allows us to evaluate the generalization capabilities of the data-driven approaches relative to the mechanistic benchmark.

3.2.1. Traditional Neural ODE. Figure 4 illustrates that the purely data-driven model exhibits substantial epistemic

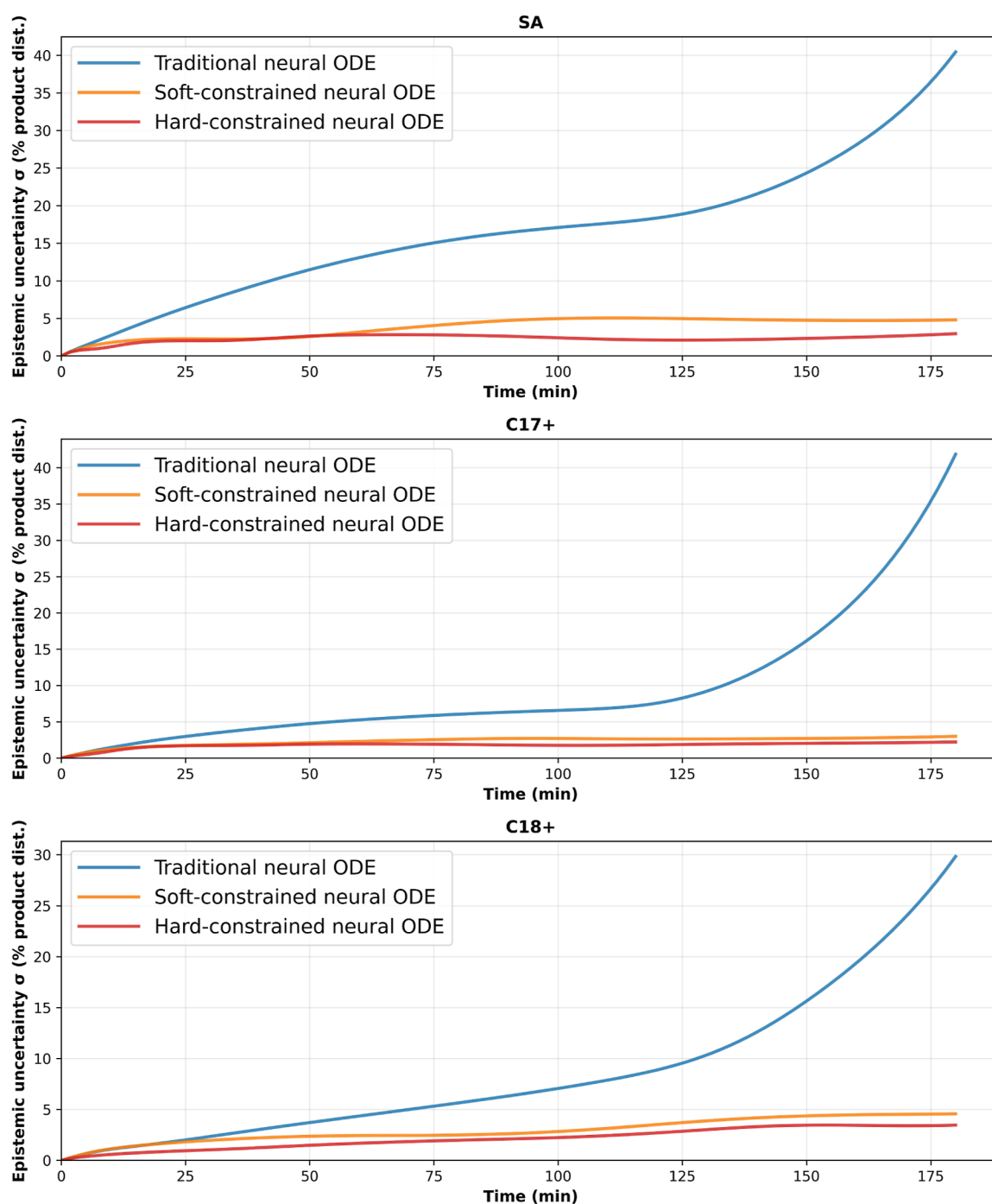


Figure 7. Epistemic uncertainty for all key chemical species as a function of time.

uncertainty, reflecting poor model reliability. Epistemic uncertainty here arises from insufficient prior knowledge embedded in the model architecture, leaving the parameter space unconstrained and highly sensitive to the limitations of the training data. Consequently, the model exhibits wide predictive intervals for the observed samples. To mitigate this deficiency, we introduce an atom balance as a soft constraint within the learning framework as shown in eq 9, thus avoiding rigid enforcement through hard constraints. Such a formulation preserves flexibility for data-driven learning while embedding fundamental physical principles, thereby reducing epistemic uncertainty by constraining the parameter space and improving generalization, with the carbon balance constraint acting as an

informative prior, adaptively penalizing the loss function for deviations from the inductive bias.

3.2.2. Constrained Neural ODE. As shown in Figure 5, the incorporation of physics-based regularization yields clear improvements in predictive reliability for the reasons explained above. This shows that incorporating mechanistic knowledge into the model effectively reduces the epistemic uncertainty. This alignment demonstrates that the model effectively combines data-driven inference with physical consistency.

Furthermore, Figure 5 benchmarks the self-adaptive soft-constrained model against the kinetic framework proposed by Arora et al.⁷ In comparison, the soft-constrained model obtains a higher accuracy for both the reactant; stearic acid and the end-

Soft-constrained neural ODE

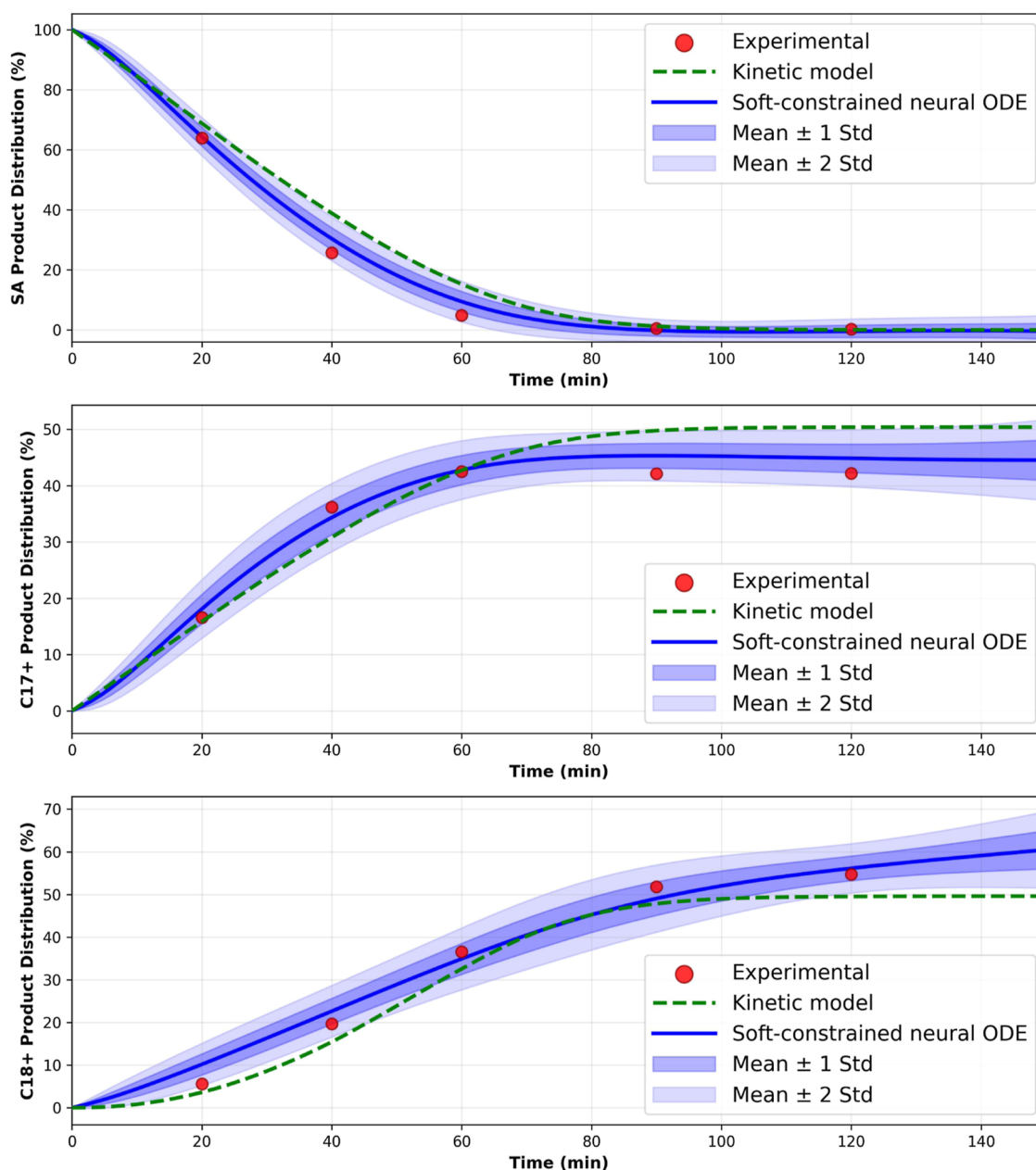


Figure 8. Ensemble-based uncertainty estimation for soft-constrained neural ODE models predicting product distributions for the high-temperature corner point in the experimental design for time-resolved product distributions for SA, C17+ and C18+ species. Red circles represent experimental measurements. Solid blue lines show ensemble mean predictions from deep ensembles. Shaded regions indicate model uncertainty. Green dashed lines show predictions from the mechanistic kinetic model for comparison.

products, heptadecane and its corresponding isomers and octadecane and its corresponding isomers which is clearly demonstrated in Table 2. Interestingly, the largest improvement is noted for heptadecane and its isomers, going from an R^2 of 0.67 in the kinetic model to a prediction accuracy for the ensemble mean of 0.81. A noteworthy improvement is also observed for stearic acid with a prediction accuracy for the ensemble mean of 0.98. Even for the chemical species where the kinetic model achieves its highest level of predictive accuracy, the soft-constrained model consistently delivers higher predictive accuracy, as evidenced by the prediction of C18+. Taken together, the total R^2 is presented in Table 2 where the

prediction accuracy for the soft-constrained neural ODE is proven to be higher than the kinetic model, demonstrating strong generalization capabilities. The median ensemble obtains similar R^2 values as the mean ensemble for the soft-constrained neural ODE as demonstrated in Table 2.

As Figure 6 and Table 2 illustrate, the hard-constrained neural ODE exhibits higher predictive accuracy for the hold-out data set for the species SA and C18+ compared to the other models. In addition, the predictive intervals are narrower than the soft-constrained model, thus indicating a reduced epistemic uncertainty. The epistemic uncertainties for all model types are more clearly presented in Figure 7, where the constrained

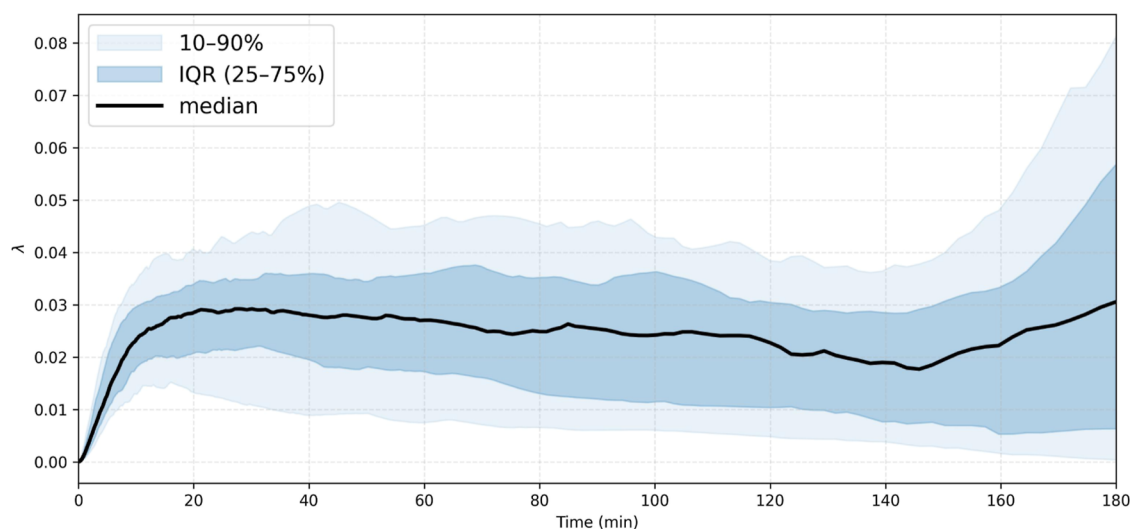


Figure 9. Temporal evolution of the learnable regularization weight, λ . The solid line represents the median trajectory across multiple realizations of λ , while the darker shaded region indicates the interquartile range of all realizations, reflecting variability in learned weights.

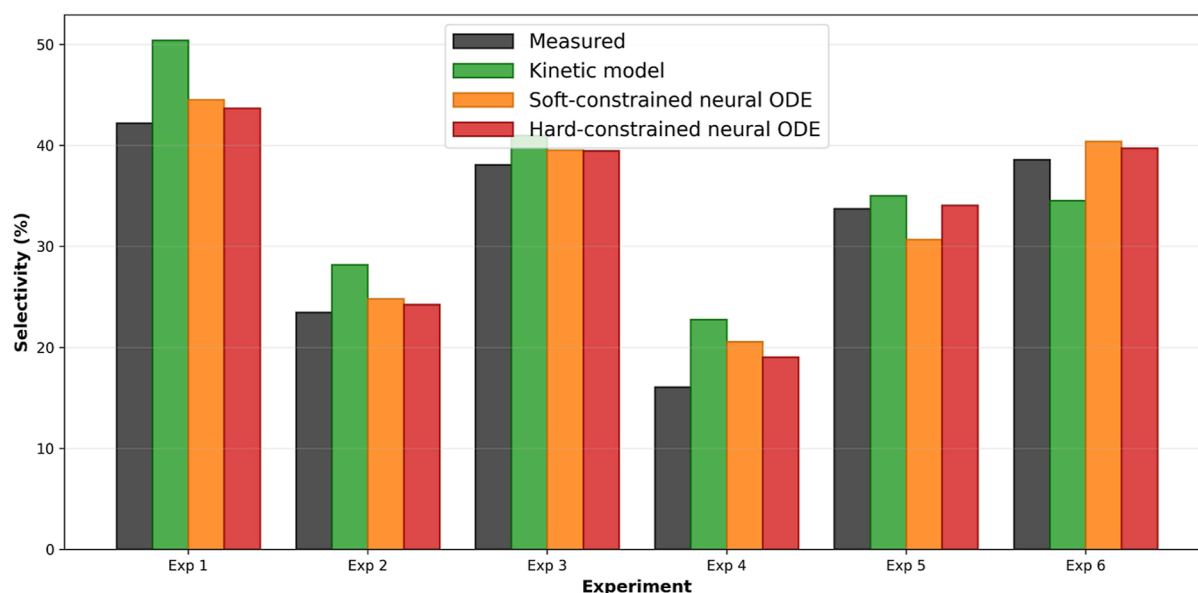
models display lower epistemic uncertainties than the traditional neural ODE, demonstrating that the incorporation of constraints to neural ODEs reduce epistemic uncertainties. Additionally, the hard-constrained model exhibits lower epistemic uncertainty than the soft-constrained model, likely because the constraint is enforced exactly rather than encouraged through a penalty term. Nevertheless, despite its lower epistemic uncertainty, the hard-constrained model appears overconfident for C17+, as several observations lie outside the interval defined by two standard deviations around the ensemble mean. By contrast, the soft-constrained model yields wider predictive intervals, within which all C17+ observations are contained, indicating better calibrated uncertainty for this species. This suggests that exact enforcement of the constraint may reduce ensemble spread without necessarily improving the reliability of the resulting uncertainty estimates.

Figures 5 and 6 indicate that the constrained neural ODEs introduce a systematic error for the prediction of the steady-state moles for SA and C17+. Since the test data set is a center point in the design space, the training data sets consist of the corner points in the design space partially illustrated in Figure 2. Convergence to steady-state was markedly slower for data sets corresponding to the corner points in the design space, particularly those with excess starting material or lower reaction temperatures. We hypothesize that for this reason, the absence of training data clearly conveying the steady-state product distribution with respect to its design parameters between the corner points leads to difficulties for the model to accurately interpolate the steady-state product distribution. Figure 8 presents the prediction of the soft-constrained neural ODE for a corner point, specifically for a training data set where the temperature is increased in the experimental design. The figure clearly illustrates that the soft-constrained neural ODE manages to capture the steady-state moles for SA and C17+ in the figure correctly. This supports our theory that the failure of the physics-informed models to correctly capture steady-state behavior is likely due to a lack of variation in the data set rather than a complete failure of the model to capture steady-state behavior. Particularly, Figure 5 displays wider uncertainty bands compared to Figure 8, which is expected since the model is

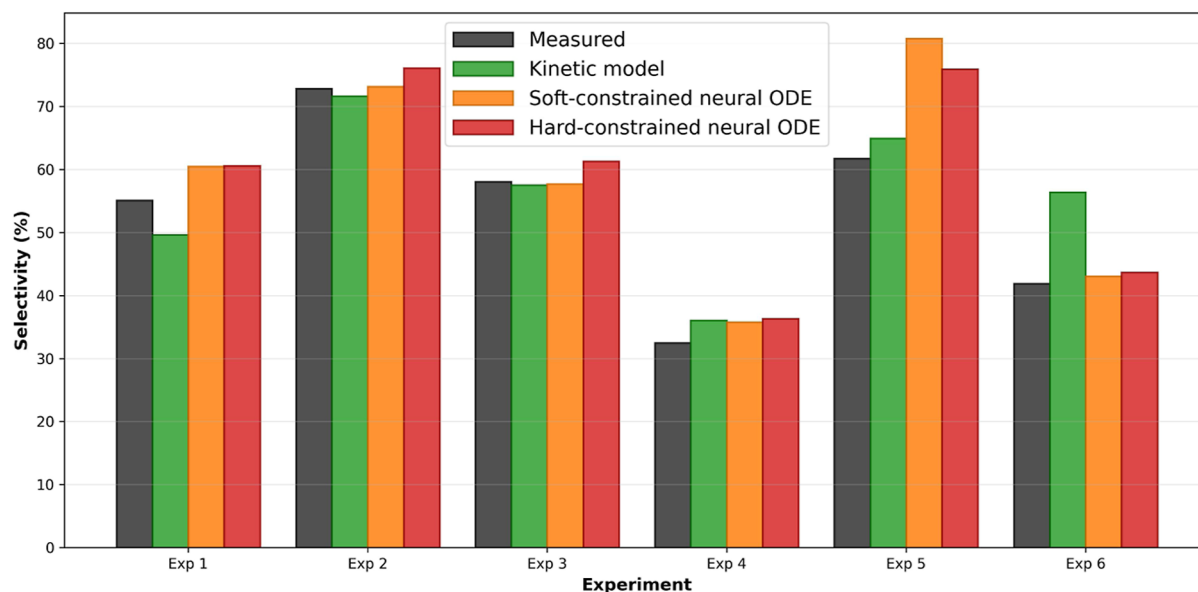
trained on the data in Figure 8, while the data presented in Figure 5 is unseen by the model, thus demonstrating greater confidence for its predictions for the training data set. While Figure 8 shows only the soft-constrained model trajectory, the hard-constrained model attains comparable predictive accuracy and produces tighter predictive intervals. Notably, experimental uncertainty, in addition to amount of available data and the prior knowledge introduced through physical constraints can all influence model uncertainty. The present study provides a detailed comparison of traditional, soft-constrained, and hard-constrained neural ODE formulations for a well-characterized experimental reaction system, with an established mechanistic kinetic model serving as a benchmark. For the system and data set investigated here, incorporating physical constraints improves predictive performance and reduces deep ensemble estimates of epistemic uncertainty. Evaluation on additional reaction systems would provide complementary evidence on whether these observed benefits extend to other chemistries, data sets, and experimental conditions.

3.3. Regularization Weights

The size of the regularization weights indicate regions where enforcing the mass balance is more important. We illustrate in Figure 9 the dynamic behavior of the regularization weight magnitudes. The observed trend reveals three distinct phases, namely an initial sharp increase within the first few minutes, a monotonically decreasing region in the midranges (approximately 30–145 min) and a secondary rise near the terminal phase. This pattern suggests that early time dynamics require stronger constraint enforcement, likely due to rapid transients in species concentrations, whereas midrange stability permits a more moderate penalty. Notably, for two of the chemical species presented in Figure 5, the kinetic model obtains higher accuracy at the final time point compared to the soft-constrained neural ODE. The sharp increase in the regularization weight near the terminal phase may reflect an adaptive response aimed at compensating for prediction discrepancies at this stage. This selective enforcement strategy mitigates overfitting by introducing domain knowledge in the temporal space, where it matters most, rather than uniformly penalizing all collocation points. This shows that traditional approaches in physics-informed models that uniformly penalize all collocation points equally,



(a) C17+



(b) C18+

Figure 10. Selectivity predictions for 10a) C17+ and 10b) C18+.

have a limitation as certain collocation points require greater regularization strengths.

3.4. Selectivity

Accurate selectivity prediction is essential for design and optimization of chemical processes. Figure 10 compares the selectivity predictions for the kinetic and neural ODE models to the measured selectivity. As illustrated in the figure, the hard-constrained neural ODE model gives a more accurate prediction of the selectivity in aggregate for C17+ than the other models. For C18+, the soft-constrained model generally exhibits a more accurate prediction of selectivity compared to the other models. Notably, for experiment 5 the kinetic model provides a more accurate selectivity prediction. This deviation is primarily

attributed to the higher epistemic uncertainty observed in the neural ODE models for that specific experiment.

4. CONCLUSION

This work has studied the effect of mechanistic constraint type on epistemic uncertainty in neural ODE ensembles, showing that hard constraints can reduce predictive spread relative to soft constraints, but do not necessarily yield better calibrated uncertainty estimates. In particular, the limited amount of available data made this problem well suited to a hard-constrained, self-adaptive approach, which avoids the need to tune regularization weights. For the soft-constrained model, to avoid uniformly penalizing the model in all collocation points, we introduced a vector for the regularization weights that

dynamically adapts the regularization strength at each collocation point based on the local loss contribution by reformulating the optimization problem by performing gradient ascent with respect to the regularization weights. This adaptive strategy ensures that the regularization is concentrated where physical violations are most pronounced, rather than applying a uniform penalty. Although we applied the self-adaptive regularization strategy to an atom balance, the approach is general and can be extended to other conservation laws and constraints.

Focusing on the hydrodeoxygenation of stearic acid, with stearic acid as the reactant and heptadecane and octadecane as end-products, we found that the traditional neural ODE exhibits a large epistemic uncertainty when applying deep ensembles. Incorporating an atom balance as either a soft or hard constraint restricts the model within the parameter space, guiding it toward a family of functions that inherently satisfy elemental conservation of carbon. In doing so, the constrained formulations retain predictive flexibility while substantially reducing epistemic uncertainty by embedding mechanistic knowledge through elemental conservation.

Adopting the center point as the test data set and employing the kinetic model as a benchmark, the soft- and hard-constrained neural ODE obtains a higher predictive accuracy compared to the kinetic model, demonstrating efficient model surrogate capabilities. The improvement in predictive accuracy is noted on both macro-level and individually, with the model prediction improving from 80% to 89–93% on the macro-level when comparing the kinetic model to the hard- and soft-constrained models, respectively. Moreover, all experimental data points fell within the model's uncertainty bands for the soft-constrained model, defined as two standard deviations around the predicted mean, evidencing reliable generalization. Further evaluation of the generalization capability of the models necessitates incorporation of additional experimental data sets within the experimental design presented in Figure 2. When working with synthetic data, this is readily established by performing a Monte Carlo simulation however this solution is evidently not viable with real-world data sets. Due to the lack of experimental data sets within the adopted experimental design from Arora et al.,⁷ the predictive accuracy is derived from one experimental data set. This limitation highlights the importance of including additional independent data sets within the experimental design, enabling a more rigorous assessment of model generalization.

This study examines a limited data regime with few experimental data sets that can be utilized for training. As it has been shown, employing a soft- and hard-constrained model serves as a strong candidate for generalization and reduction of epistemic uncertainty under data scarcity. Despite focusing on one reaction system, the approach is not limited to specifically hydrodeoxygenation of stearic acid. Instead it could be extended to other reactions where obtaining large data sets is impractical. This real-world application demonstrates the potential of the methodology, bridging a crucial gap between simulation and experiment. Thus, this framework should naturally extend to other systems governed by conservation laws in chemical reaction engineering, which possess limited number of observations.

AUTHOR INFORMATION

Corresponding Author

David Eklund – Industrial AI, RISE Research Institutes of Sweden, Stockholm SE-16440, Sweden; orcid.org/0000-0002-1954-760X; Email: david.eklund@ri.se

Authors

Kian Hajireza – Department of Chemistry and Chemical Engineering, Chalmers University of Technology, Gothenburg SE-41296, Sweden; Industrial AI, RISE Research Institutes of Sweden, Stockholm SE-16440, Sweden; orcid.org/0009-0008-1682-4975

Louise Olsson – Department of Chemistry and Chemical Engineering, Chalmers University of Technology, Gothenburg SE-41296, Sweden; orcid.org/0000-0002-8308-0784

Derek Creaser – Department of Chemistry and Chemical Engineering, Chalmers University of Technology, Gothenburg SE-41296, Sweden

Ronnie Andersson – Department of Chemistry and Chemical Engineering, Chalmers University of Technology, Gothenburg SE-41296, Sweden; orcid.org/0000-0002-2838-6115

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.iecr.6c03344>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Financial support from the Swedish Foundation for Strategic Research (Grant: FID20-0002) is gratefully acknowledged. We acknowledge the EuroHPC Joint Undertaking for awarding this project access to the EuroHPC supercomputer LEONARDO, hosted by CINECA (Italy) and the LEONARDO consortium through a EuroHPC Development Access call. We acknowledge the EuroHPC Joint Undertaking for awarding this project access to the EuroHPC supercomputer LUMI, hosted by CSC (Finland) and the LUMI consortium through a EuroHPC Development Access call.

REFERENCES

- (1) Wang, Z.; Parhi, S. S.; Rangaiah, G. P.; Jana, A. K. Analysis of weighting and selection methods for pareto-optimal solutions of multiobjective optimization in chemical engineering applications. *Ind. Eng. Chem. Res.* **2020**, *59*, 14850–14867.
- (2) Bequette, B. W. Nonlinear control of chemical processes: A review. *Ind. Eng. Chem. Res.* **1991**, *30*, 1391–1413.
- (3) Soltani, S.; Andersson, R.; Andersson, B. Enhancement of time resolution in transient kinetics. *Chem. Eng. J.* **2015**, *264*, 188–196.
- (4) Tatar, M.; Vashisth, V.; Iqbal, M.; Butterworth, T.; van Rooij, G.; Andersson, R. Analysis of a plasma reactor performance for direct nitrogen fixation by use of three-dimensional simulations and experiments. *Chem. Eng. J.* **2024**, *497*, 154756.
- (5) Zhou, X.; Nolte, M. W.; Mayes, H. B.; Shanks, B. H.; Broadbelt, L. J. Experimental and mechanistic modeling of fast pyrolysis of neat glucose-based carbohydrates. 1. Experiments and development of a detailed mechanistic model. *Ind. Eng. Chem. Res.* **2014**, *53*, 13274–13289.
- (6) Kumar, H.; Froment, G. F. A generalized mechanistic kinetic model for the hydroisomerization and hydrocracking of long-chain paraffins. *Ind. Eng. Chem. Res.* **2007**, *46*, 4075–4090.
- (7) Arora, P.; Grennfelt, E. L.; Olsson, L.; Creaser, D. Kinetic study of hydrodeoxygenation of stearic acid as model compound for renewable oils. *Chem. Eng. J.* **2019**, *364*, 376–389.

- (8) Jenišťová, K.; Hachemi, I.; Mäki-Arvela, P.; Kumar, N.; Peurla, M.; Capek, L.; Wärnå, J.; Murzin, D. Y. Hydrodeoxygenation of stearic acid and tall oil fatty acids over Ni-alumina catalysts: Influence of reaction parameters and kinetic modelling. *Chem. Eng. J.* **2017**, *316*, 401–409.
- (9) Hachemi, I.; Murzin, D. Y. Kinetic modeling of fatty acid methyl esters and triglycerides hydrodeoxygenation over nickel and palladium catalysts. *Chem. Eng. J.* **2018**, *334*, 2201–2207.
- (10) Kubička, D.; Kaluža, L. Deoxygenation of vegetable oils over sulfided Ni, Mo and NiMo catalysts. *Appl. Catal., A* **2010**, *372*, 199–208.
- (11) Donnis, B.; Egeberg, R. G.; Blom, P.; Knudsen, K. G. Hydroprocessing of bio-oils and oxygenates to hydrocarbons. Understanding the reaction routes. *Top. Catal.* **2009**, *52*, 229–240.
- (12) Pedroza, M. A. P.; Segtovich, I. S. V.; Da Silva, M. A. P. Oleic acid hydrodeoxygenation into renewable paraffins as biofuels: Kinetic study and parameter estimation. *Chem. Eng. Sci.* **2026**, *327*, 123623.
- (13) Venkatasubramanian, V. The promise of artificial intelligence in chemical engineering: Is it here, finally? *AIChE J.* **2019**, *65*, 466–478.
- (14) Mukherjee, A.; Zavala, V. M. Physics-constrained machine learning for chemical engineering. *Curr. Opin. Chem. Eng.* **2026**, *51*, 101228.
- (15) Cybenko, G. Approximation by superpositions of a sigmoidal function. *Math. Cont. Sig. Syst. (MCSS)* **1989**, *2*, 303–314.
- (16) Hornik, K.; Stinchcombe, M.; White, H. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Netw.* **1990**, *3*, 551–560.
- (17) Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Netw.* **1991**, *4*, 251–257.
- (18) Leshno, M.; Lin, V. Y.; Pinkus, A.; Schocken, S. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Netw.* **1993**, *6*, 861–867.
- (19) Chen, R. T. Q.; Rubanova, Y.; Bettencourt, J.; Duvenaud, D. K., Neural Ordinary Differential Equations In *Advances in Neural Information Processing Systems*, Bengio, S.; Wallach, H.; Larochelle, H.; Grauman, K.; Cesa-Bianchi, N.; Garnett, R., Eds., Curran Associates, Inc.: 2018; Vol. 31.
- (20) Rackauckas, C.; Ma, Y.; Martensen, J.; Warner, C.; Zubov, K.; Supekar, R.; Skinner, D.; Ramadhan, A.; Edelman, A. Universal differential equations for scientific machine learning. *arXiv* **2021**, arXiv:2001.04385.
- (21) Raissi, M.; Perdikaris, P.; Karniadakis, G. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **2019**, *378*, 686–707.
- (22) Karniadakis, G. E.; Kevrekidis, I. G.; Lu, L.; Perdikaris, P.; Wang, S.; Yang, L. Physics-informed machine learning. *Nat. Rev. Phys.* **2021**, *3*, 422–440.
- (23) Pontryagin, L. S.; Mishchenko, E.; Boltyanskii, V.; Gamkrelidze, R. *The Mathematical Theory of Optimal Processes*; Wiley-Interscience, 1962.
- (24) Santana, V. V.; Costa, E.; Rebello, C. M.; Ribeiro, A. M.; Rackauckas, C.; Nogueira, I. B. Efficient hybrid modeling and sorption model discovery for non-linear advection-diffusion-sorption systems: A systematic scientific machine learning approach. *Chem. Eng. Sci.* **2023**, *282*, 119223.
- (25) Chew, J. W.; Andersson, R.; Cocco, R. A. Understanding and Predicting Bubble Growth in Bubbling Fluidized Beds via Physics-Aware Data-Driven Models. *Ind. Eng. Chem. Res.* **2026**, *65*, 7273–7286.
- (26) Sorourifar, F.; Peng, Y.; Castillo, I.; Bui, L.; Venegas, J.; Paulson, J. Physics-Enhanced Neural Ordinary Differential Equations: Application to Industrial Chemical Reaction Systems. *Ind. Eng. Chem. Res.* **2023**, *62*, 15563–15577.
- (27) Hajireza, K.; Eklund, D.; Andersson, R. Enhancing generalization and training efficiency of neural ordinary differential equations for chemical reactor modeling. *Chem. Eng. J.* **2025**, *519*, 165183.
- (28) Kumar, T.; Kumar, A.; Pal, P., A physics-informed autoencoder-neuralode framework (phy-chemnode) for learning complex fuel combustion kinetics In *NeurIPS Machine Learning and the Physical Sciences Workshop*; Argonne National Laboratory, 2024; p 1.
- (29) Kim, S.; Ji, W.; Deng, S.; Ma, Y.; Rackauckas, C. Stiff neural ordinary differential equations. *J. Nonlinear Sci.* **2021**, *31*, 093122.
- (30) Kircher, T.; Votsmeier, M. Learning kinetics from non-ideal reactors by implicitly solved finite volumes and global reaction neural networks. *Chem. Eng. J.* **2025**, *522*, 166749.
- (31) Fedorov, A.; Perehodjuk, A.; Linke, D. Kinetics-constrained neural ordinary differential equations: Artificial neural network models tailored for small data to boost kinetic model development. *Chem. Eng. J.* **2023**, *477*, 146869.
- (32) Kircher, T.; Döppel, F. A.; Votsmeier, M. Global reaction neural networks with embedded stoichiometry and thermodynamics for learning kinetics from reactor data. *Chem. Eng. J.* **2024**, *485*, 149863.
- (33) McClenny, L. D.; Braga-Neto, U. M. Self-adaptive physics-informed neural networks. *J. Comput. Phys.* **2023**, *474*, 111722.
- (34) Liu, D.; Wang, Y. A dual-dimer method for training physics-constrained neural networks with minimax architecture. *Neural Netw.* **2021**, *136*, 112–125.
- (35) Coelho, C.; Costa, M. F. P.; Ferrás, L. L. A Self-Adaptive Penalty Method for Integrating Prior Knowledge Constraints into Neural ODEs. *arXiv* **2023**, arXiv:2307.14940.
- (36) Ji, W.; Qiu, W.; Shi, Z.; Pan, S.; Deng, S. Stiff-pinn: Physics-informed neural network for stiff chemical kinetics. *J. Phys. Chem. A* **2021**, *125*, 8098–8106.
- (37) Peng, W.; Liu, Z.-S.; Boy, M. SPIN-ODE: Stiff Physics-Informed Neural ODE for Chemical Reaction Rate Estimation. *ECAI 2025: 28th European Conference on Artificial Intelligence 25-30 October 2025, Bologna, Italy-Including 14th Conference on Prestigious Applications of Intelligent Systems (PAIS 2025) Proceedings*; SAGE Publications 1 Oliver's Yard, 2025.
- (38) Kircher, T.; Votsmeier, M. Robust extraction of thermodynamically consistent neural-network-based kinetic models from reactor data: Methanation, methanol synthesis, and dimethyl ether synthesis. *J. Catal.* **2026**, *461*, 116958.
- (39) Döppel, F.; Kircher, T.; Votsmeier, M. Formation–consumption neural networks for efficient chemical kinetics modeling. *Chem. Eng. J.* **2026**, *535*, 175101.
- (40) Donti, P.; Rolnick, D.; Kolter, J. Z. DC3: A learning method for optimization with hard constraints In *International Conference on Learning Representations. arXiv* **2021**, arXiv:2104.12225.
- (41) Beucler, T.; Pritchard, M.; Rasp, S.; Ott, J.; Baldi, P.; Gentine, P. Enforcing analytic constraints in neural networks emulating physical systems. *Phys. Rev. Lett.* **2021**, *126*, 098302.
- (42) Döppel, F. A.; Votsmeier, M. Robust mechanism discovery with atom conserving chemical reaction neural networks. *Proc. Combust. Inst.* **2024**, *40*, 105507.
- (43) Döppel, F. A.; Bracconi, M.; Maestri, M. Physically Consistent Neural Network Surrogates of Catalytic Reactors with Detailed Surface Kinetics. *ACS Eng. Au* **2026**, *6*, 359–371.
- (44) Chi, C.; Xu, X.; Thévenin, D. Efficient premixed turbulent combustion simulations using flamelet manifold neural networks: A priori and a posteriori assessment. *Combust. Flame* **2022**, *245*, 112325.
- (45) Almeldein, A.; Van Dam, N. Accelerating chemical kinetics calculations with physics informed neural networks. *J. Eng. Gas Turbines Power* **2023**, *145*, 091008.
- (46) Kumar, T.; Kumar, A.; Pal, P. Phy-ChemNODE: An End-to-End Physics-Constrained Autoencoder-NeuralODE Framework for Learning Stiff Chemical Kinetics of Hydrocarbon Fuels. *Frontiers Therm. Eng* **2025**, *5*, 1594443.
- (47) Luo, L.; Liu, Q.; Sun, J.; Huang, Y. Exploring surface reaction mechanism using a surface reaction neural network framework. *Chem. Eng. Sci.* **2025**, *306*, 121307.
- (48) Kircher, T.; Votsmeier, M. Machine learning surrogate models for mechanistic kinetics: Embedding atom balance and positivity. *J. Phys. Chem. Lett.* **2025**, *16*, 4715–4723.
- (49) Biermann, F.; Uglietti, R.; Döppel, F. A.; Kircher, T.; Bracconi, M.; Maestri, M.; Votsmeier, M. Enabling micro-kinetics based simulation of industrial packed-bed reactors by physics-enhanced neural networks. *Chem. Eng. J.* **2025**, *519*, 163598.

- (50) Sturm, P. O.; Wexler, A. S. Conservation laws in a neural network architecture: enforcing the atom balance of a Julia-based photochemical model (v0. 2.0). *Geosci. Model Dev.* **2022**, *15*, 3417–3431.
- (51) Sturm, P. O.; Silva, S. J. A nudge to the truth: atom conservation as a hard constraint in models of atmospheric composition using a species-weighted correction. *ACS ES&T Air* **2025**, *2*, 99–108.
- (52) Üstün, C. E.; Paykani, A. On-the-fly uncertainty-aware deep ensembles for stiff chemical kinetics. *Proc. Combust. Inst.* **2025**, *41*, 105896.
- (53) Nai, D.; Li, H.; Grover, M. A.; Medford, A. J. Modeling Batch Crystallization under Uncertainty Using Physics-Informed Machine Learning. *Ind. Eng. Chem. Res.* **2026**, *65*, 9563–9577.
- (54) Ott, K.; Tiemann, M.; Hennig, P. Uncertainty and structure in neural ordinary differential equations. *arXiv* **2023**, arXiv:2305.13290.
- (55) Greydanus, S.; Dzamba, M.; Yosinski, J. Hamiltonian neural networks. *Adv. Neural Inf. Process. Syst.* **2019**, *32*.
- (56) Kendall, A.; Gal, Y. What uncertainties do we need in bayesian deep learning for computer vision?. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.
- (57) Kingma, D. P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
- (58) Loshchilov, I.; Hutter, F. S. G. D. R. Stochastic gradient descent with warm restarts. *arXiv* **2016**, arXiv:1608.03983.
- (59) Ramachandran, P.; Zoph, B.; Le, Q. V. Searching for activation functions. *arXiv* **2017**, arXiv:1710.05941.
- (60) Lakshminarayanan, B.; Pritzel, A.; Blundell, C. Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles In *Advances in Neural Information Processing Systems*; Curran Associates, Inc., 2017; Vol. 30.
- (61) Fort, S.; Hu, H.; Lakshminarayanan, B. Deep ensembles: A loss landscape perspective. *arXiv* **2019**, arXiv:1912.02757.
- (62) Hüllermeier, E.; Waegeman, W. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Mach. Learn.* **2021**, *110*, 457–506.
- (63) Kiureghian, A. D.; Ditlevsen, O. Aleatory or epistemic? Does it matter? *Struct. Saf.* **2009**, *31*, 105–112.
- (64) Dandekar, R.; Chung, K.; Dixit, V.; Tarek, M.; García-Valadez, A.; Vemula, K.; Rackauckas, C. Bayesian Neural Ordinary Differential Equations. *arXiv* **2021**, arXiv:2012.07244.
- (65) Li, Q.; Chen, H.; Koenig, B. C.; Deng, S. Bayesian chemical reaction neural network for autonomous kinetic uncertainty quantification. *Phys. Chem. Chem. Phys.* **2023**, *25*, 3707–3717.
- (66) Koenig, B. C.; Chen, H.; Li, Q.; Zhao, P.; Deng, S. Uncertain lithium-ion cathode kinetic decomposition modeling via Bayesian chemical reaction neural networks. *Proc. Combust. Inst.* **2024**, *40*, 105243.
- (67) Nieves, E.; Dandekar, R.; Rackauckas, C. Uncertainty quantified discovery of chemical reaction systems via Bayesian scientific machine learning. *Front. Syst. Biol.* **2024**, *4*, 1338518.
- (68) Koenig, B. C.; Zhao, P.; Deng, S. Comprehensive thermal-kinetic uncertainty quantification of lithium-ion battery thermal runaway via bayesian chemical reaction neural networks. *Chem. Eng. J.* **2025**, *507*, 160402.
- (69) Schmid, N.; del Pozo, D. F.; Waegeman, W.; Hasenauer, J. Assessment of uncertainty quantification in universal differential equations. *Philos. Trans. A Math. Phys. Eng. Sci.* **2025**, *383*, 20240444.
- (70) Gusmão, G. S.; Medford, A. J. Maximum-likelihood estimators in physics-informed neural networks for high-dimensional inverse problems. *Comput. Chem. Eng.* **2024**, *181*, 108547.