
KINETICS-INFORMED NEURAL NETWORKS

Gabriel S. Gusmão^a, Adhika P. Retnanto^{a,1}, Shashwati C. da Cunha^{a,1}, Andrew J. Medford^a

^aSchool of Chemical & Biomolecular Engineering
Georgia Institute of Technology
Atlanta, GA 30332

{gusmaogabriels, aretnanto6, shashwatidc, ajm}@gatech.edu

¹*These authors contributed equally to this work.*

March 21, 2022

ABSTRACT

Chemical kinetics and reaction engineering consists of the phenomenological framework for the disentanglement of reaction mechanisms, optimization of reaction performance and the rational design of chemical processes. Here, we utilize feed-forward artificial neural networks as basis functions to solve ordinary differential equations (ODEs) constrained by differential algebraic equations (DAEs) that describe microkinetic models (MKMs). We present an algebraic framework for the mathematical description and classification of reaction networks, types of elementary reaction, and chemical species. Under this framework, we demonstrate that the simultaneous training of neural nets and kinetic model parameters in a regularized multi-objective optimization setting leads to the solution of the inverse problem through the estimation of kinetic parameters from synthetic experimental data. We analyze a set of scenarios to establish the extent to which kinetic parameters can be retrieved from transient kinetic data, and assess the robustness of the methodology with respect to statistical noise. This approach to inverse kinetic ODEs can assist in the elucidation of reaction mechanisms based on transient data.

Keywords physics-informed neural network, surrogate approximator, physically informed neural network, catalysis, transient, chemical kinetics

1 Introduction

In recent decades, considerable attention has been devoted to chemical process intensification through the optimal design of unit operations based on rigorous modeling of thermodynamics, heat, mass and momentum transport, and chemical reactions [1]. Selectivity and yield of a chemical transformation, which are key factors in economic evaluation and design of associated unit operations, are determined by the intrinsic kinetics, transport resistances and thermodynamics constraints. The detailed description of chemical kinetics, especially in heterogeneous catalysis, is hence of utmost importance in the design and operation of chemical processes. Kinetic models provide the phenomenological framework for the elucidation of reaction mechanism, optimization of reaction performance, and process design [2]. Structurally, the detailed description of a chemical reaction network consists of the set of elementary steps involving bond breaking and/or formation from reactants to products [3,4]. Chemical kinetics can be described in terms of the transition-state barriers between thermodynamically stable intermediate states. Despite the high complexity of chemical kinetics networks, only few elementary steps typically control catalytic activity [5,6,7].

The mathematical description of chemical process models has departed from kinetic models that assume a single rate-limiting step and become more comprehensive [8]. The analysis and elucidation of reaction mechanism have benefited from the increasing accuracy of computational chemistry methods such as density functional theory (DFT). These advances in computational capabilities have led to an increased prevalence of “bottom-up” kinetic models. In the bottom-up approach, DFT-derived free energies from proposed reaction mechanisms give rise to free energy diagrams. The ensuing microkinetic models (MKMs) have provided insights into rate-controlling steps [7] and trends

in catalytic activity in terms of reaction conditions and catalyst compositions [9, 10]. However, bottom-up modelling of complex reaction networks requires in-depth understanding of the detailed atomic structure of active sites as well as the molecular structure of the thermodynamically stable intermediate species and dominant elementary reactions. Moreover, assumptions are often required to connect the molecular-scale models with measurable catalytic quantities such as turnover frequencies, yields, and selectivities. The mean-field approximation (MFA) is typically invoked to avoid the need for a specific spatial description of catalytic active domains [11]. Transport limitations on the surface are regarded as negligible or averaged out uniformly across the available catalyst surface. This requires adsorbate-adsorbate interactions to be either neglected or included as coverage-dependent properties [12, 13, 14]. For more heterogeneous or non-ideal systems, rigorous kinetic Monte Carlo methods can be used to build coarse-grained representations of catalyst active sites, allowing for an explicit description of diffusive processes and surface interactions [15]. Models of intermediate complexity, combining the MFA hierarchically in clusters, which benefits from the mathematical simplicity of the former and the ability to capture adlayer non-ideality of the latter, have also been considered [16]. However, these models are more difficult to parameterize, and in the absence of diffusion limitations and adsorbate-adsorbate interaction, studies by Hoffmann and Bliggard [17] and by Andersen et. al. [18] have shown that results derived from mean-field models approximate those of complex kinetic Monte Carlo schemes. Moreover, Reuter et. al. demonstrated that MFA models can be re-parameterized to reflect the results of complex kinetic Monte Carlo with great generalization properties [19]. These results suggest that the mean-field framework is promising for even complex catalytically active domains, though it may be necessary to depart from a direct mapping back to the molecular-scale system to obtain reliable results with the MFA.

Despite advances in first-principles and empirical mechanism elucidation, obtaining a bottom-up description of dynamic catalyst states over time-on-stream under operating conditions remains inherently challenging and computationally expensive. Uncertainties from first-principles calculated thermochemistry propagate exponentially to kinetic models turn-over frequencies, and their estimation has been an active research both of conventional and Bayesian parameter estimation fields [20]. It is also dependent on the specific system under study, since changes in catalyst state under different operating conditions may translate into shifts in the predominant chemical pathway, altering the selectivity and yield [21]. For instance, the carbide-promoted Fischer-Tropsch reaction is a complex mechanism where the presence of carbide-, oxygen- and metallic-rich states exhibit strong dependence on reaction conditions [22, 23]. Adsorbate-adsorbate interactions may also play a significant role in determining the preferential chemical pathway and the related rate-controlling steps [14, 24]. Furthermore, catalytic systems often encompass multiple active-site domains with different prevalent chemical pathways in which diffusive processes and adsorbate-adsorbate interaction may become relevant, as exemplified by Li et. al. in their comparison between mean-field approximation and kinetic Monte Carlo approaches for CO oxidation MKMs [25]. These nuances pose significant challenges in devising ab initio bottom-up descriptions of kinetic systems, particularly in the case of complex multi-component catalysts or large product/reactant molecules.

An alternative approach to kinetic model development is the “top-down” approach, where kinetic model parameters are fitted to the results of kinetic experiments. The top-down approach avoids the need for an explicit description of the catalyst active site, and can accommodate “lumped” reactions that include multiple elementary steps. However, the resulting models provide less detailed insight, and may not generalize well to other reaction conditions. In the top-down approach to chemical kinetics, the acquisition of steady-state data has traditionally been the foremost empirical procedure for the analysis and validation of reaction mechanism hypotheses [4]. However, acquiring intrinsic kinetic steady-state data is inherently laborious [26]. The design of experiments must ensure that mass- and heat-transfer limitations are negligible to detach transport processes from kinetics, and experiments must be run for long periods of times to reach steady state. More importantly, steady-state data is only controlled by the rate-limiting step at a given condition, overshadowing the detailed description of other elementary steps that may be relevant under other operational scenarios [27, 28, 29, 30, 31].

In contrast, dynamic or transient techniques to acquire kinetic data have been applied for nearly half a century [32, 33, 34, 35, 36], and have recently received renewed interest due to advances in numerical methods and computational processing capabilities. Transient kinetics methods such as temporal analysis of products [37, 38, 39] (TAP) associated with temperature-programmed reaction spectroscopy [40, 41, 42, 43] (TPRS), step-response experiments, steady-state isotopic transient kinetic analysis [44, 45, 46] (SSITKA), and tapered element oscillating microbalance (TEOM) [47] are so information-dense that a single series of states, gathered at their respective time points from one experiment, conveys kinetic information that would require numerous steady-state experiments. However, this information is convoluted with various other effects, making it difficult to extract meaningful intrinsic kinetic parameters. The temporal evolution of chemical system states can be mathematically represented as sets of coupled ordinary or partial differential equations encompassing the various state-changing steps, such as the elementary reactions and diffusive processes involved in a chemical reaction. In the past decades intensive effort has been put into constructing ab-initio microkinetic models, but recently there has been an increased focus on the related issue of catalyst optimization, parameter estimation or

model selection based on experimental data [48, 49, 50, 51, 52, 53, 54]. Such techniques simultaneously probe multiple elementary steps, potentially enabling more robust top-down microkinetic or “mesokinetic” models to be developed from less experimental data. Notably, both top-down and bottom-up kinetic models are often informed by additional experimental techniques such as operando [55] and in-situ photoelectron (XPS, LEED, TEM) and infrared (FT-IR) spectroscopy, temperature programmed reduction (TRP) and desorption (TPD), as well as numerous other material science characterization techniques including NMR, XAFS and XRD [56, 57, 2]. The coupling of these approaches with kinetic models is generally qualitative or semi-quantitative, and there are limited examples where bottom-up and top-down models are quantitatively combined or compared [50, 58, 59, 60, 61]. As the volume and variety of catalysis data increases, it is necessary to explore new frameworks that facilitate the quantitative fusion of information from various sources. One promising approach is the application of optimization techniques to minimize a loss function that contains terms penalizing the deviation between the model and various types of catalytic data [50, 58].

Optimization of kinetic models requires the model to be evaluated many times, and relies on computing derivatives of the model with respect to the kinetic parameters. To balance model accuracy with the computational cost of successive evaluation of high-fidelity models, surrogate models (SMs), also referred to as reduced-order, meta- or black-box models, have become an alternative functional mapping between model inputs and outputs [8]. SMs are models of models for which, given data, parameters are estimated from the minimization of some error metric or merit function. The construction of SMs necessarily relies on experimental data or data generated by mathematically complex phenomenological or first-principles models. SMs also require a proper choice of sampling strategies for the input space, since the computational cost of randomly sampling high dimensional spaces increases exponentially due to the “curse of dimensionality”. Uniform, random, Latin-hypercube, central composite design and quasi-random low-discrepancy are the most commonly adopted strategies for input space exploration [62], although more advanced strategies have recently been devised [63]. Surrogate models can be based on physical models, flexible data-driven models, or a combination thereof. In flexible model architectures, such as neural nets (NNs), model fitting generally depends on the ability to calculate gradients across the multiple layers through the chain-rule in the back-propagation algorithm [64]. However, to avoid the time-consuming process of estimating gradients or building the back-propagation structure, procedural algorithms for automatic or algorithmic differentiation (AD) [65] are used to evaluate gradients across NN layers.

Several authors have proposed methods that rely on NNs for the solution of scientific problems involving ODEs and PDEs either as approximators for their solutions or as surrogate models. Formative work regarding the utilization of NNs as approximators for the solution of differential equations was introduced by Lagaris et. al. [66], which set the groundwork for the recent development of the so called physics-informed neural networks, PINNs, by Karniadakis et al. [67, 68]. In the context of surrogate model construction, other endeavors have demonstrated that particular NN structures, such as residual networks (ResNets), resemble Euler-type discretization through successive time-steps [69, 70, 71]. Subsequently, Neural ODE (NODE) surrogate models have emerged from the analogy between the discrete time-domain in ResNet ODEs at their depth limits, which consists of an application of NNs to ODE that contiguously describes the physical model in its domain independent of discretization [72], i.e. in the hypothetical case of infinite recursive application of ResNets, where the discretization converges to a continuum. Following works have corroborated that ResNet-type deep NNs converge to Neural ODEs in the deep limit under stochastic gradient descent minimization [73]. The fundamental difference between PINNs and NODEs lies in their respective mappings. A PINN acts as a surrogate “approximator” for an ODE or PDE solution or integral, where NNs constitute the functional mapping between the independent variables and the states and their derivatives, and the physical model provides the phenomenological connection or restriction between states and their respective gradients/derivatives that NNs must satisfy. In NODEs, NNs instead constitute as mapping from states to their gradients/derivatives, in which case it acts as a true surrogate model. While NODEs can replace unknown parts of a differential equation, PINNs are a basis set for the solution of known differential equations.

In this work, we focus on the utilization of modified PINNs as a Surrogate Approximator (SA) for the solution of microkinetic models (MKMs) under the MFA, which are fully represented by the chemical system’s stoichiometric matrix [74, 75]. The MFA acts as a general surrogate for detailed kinetics, which could include not only the backbone elementary reactions but also diffusive steps. MKMs consist of a set of elementary or irreducible reactions. The elementary reactions involve the interaction of at most two individual species, in which the probability of occurrence is proportional to their concentrations as in power-law kinetics, and whose probability of success is a function of temperature, given by the Arrhenius law. We use this general mathematical representation to approach the solution of known MKMs and address the issue of parameter fitting for kinetically relevant steps in complete MKMs.

By relying on PINNs, we resort to the universal approximation theorem for artificial neural networks (NNs) [76, 77, 78, 79] to propose strategies towards the design of SAs that can be utilized for solving (integrating through collocation) kinetics ODEs, where differential algebraic equations (DAE) constraints must be satisfied. We introduce a framework for the classification of reaction network types according to the nature of involved elementary reactions (e.g. homogeneous systems, ad/desorption, reactions between surface intermediates). We further create reaction-type examples and apply

our PINN-based approach, demonstrating the need for semi-quantitative information on intermediate composition to fully recover ground-truth information on the MKM parameters. The results provide new tools for analysis of transient kinetic data and insight into top-down analysis of complex reaction networks. The framework has the potential to enable improved design of transient kinetic experiments and fusion of multi-modal transient experimental data in the future.

2 Methods

Given the methodological focus of this work, this section contains detailed mathematical formulations of the frameworks used. More applied readers may wish to proceed directly to Sec. 3

2.1 Mean-Field Chemical Kinetics Framework

Power-law kinetic models encompass a linear combination of chemical transformation events, whose sampling frequency scales linearly with the concentration of each participant, and whose probability of success is given by the Arrhenius law. The resulting set of intertwined molecular events is represented as a stack of ODEs, assigning the rate of change of each species to the power-law kinetic expression arising from stoichiometry. Let \mathbf{c} represent an array of concentrations or concentration-related state variables like partial pressures, concentrations and coverage fractions, for species that are unbound (like gases) or bound (adsorbed molecules or radicals), at time t . Then their rates of change, $\dot{\mathbf{c}}$, or the MKM, is written as (1.1):

$$\dot{\mathbf{c}} = \mathbf{M} \mathbf{r}(\mathbf{c}, \theta) = \mathbf{M} (\mathbf{k}(\theta) \circ f(\mathbf{c})) \quad (1.1)$$

Where $\mathbf{M} \in \mathbb{Z}^{n \times m}$ is the corresponding stoichiometry matrix and $f(\cdot) : \mathbb{R}_+^n \rightarrow \mathbb{R}_+^m$ maps the concentrations $\mathbf{c} := \{\mathbf{c} \mid \mathbf{c} \in \mathbb{R}_+^n\}$ to the concentration-based terms of power-law kinetics, and $\mathbf{k} := \{\mathbf{k}(\theta) \in \mathbb{R}_+^m, \theta \in \mathbb{R}_+\}$ is the temperature- and binding-energy-dependent, Arrhenius-like rate constant term. In this work, θ corresponds to temperature (rather than coverage) to keep a consistent standard of bold uppercase variables corresponding to matrices, normal uppercase variables corresponding to elements of a matrix, bold lowercase variables corresponding to vectors, and normal lowercase variables corresponding to scalars.

2.1.1 Complex Reaction Networks - *gdacs* representation

We denote the subset of \mathbf{c} for gas-phase species as $\mathbf{c}_g := \{c_i \mid i \in \mathcal{C}_g\}$, for bounded gas-counterpart coverage fractions as $\mathbf{c}_a := \{c_i \mid i \in \mathcal{C}_a\}$, and for bounded intermediates/radicals on the catalyst surface as $\mathbf{c}_s := \{c_i \mid i \in \mathcal{C}_s\}$ (the distinction between a and s is whether or not the adsorbed species is also sufficiently stable in the gas phase to be detected). An exemplary reaction network is shown in Fig. 1. Under this notation, $\cup_{i \in \{g, a, s\}} \mathcal{C}_i = \{1, 2, \dots, m \mid m \in \mathbb{N}\}$ and $\cap_{i \in \{g, a, s\}} \mathcal{C}_i = \emptyset$ (i.e. each species can only be in a single subset, and all species must be assigned to a subset). The state-dependent reaction rate vector, denoted as $\mathbf{r}(\mathbf{c}, \theta) := \{\mathbf{r} \in \mathbb{R}^m \mid \mathbf{r} = \mathbf{k}(\theta) \circ f(\mathbf{c})\}$, comprises rates of reaction (frequencies) associated with different kinds of elementary reactions: reactions in homogeneous phase, $\mathbf{r}_g := \{r_i \mid i \in \mathcal{G}_g\}$, those involving adsorption-desorption, $\mathbf{r}_d := \{r_i \mid i \in \mathcal{G}_d\}$, reactions between adsorbed molecules, $\mathbf{r}_a := \{r_i \mid i \in \mathcal{G}_a\}$, reactions involving adsorbed molecules and radicals/intermediates on the surface, $\mathbf{r}_c := \{r_i \mid i \in \mathcal{G}_c\}$, and reactions between intermediates on the surface, $\mathbf{r}_s := \{r_i \mid i \in \mathcal{G}_s\}$, such that $\cup_{i \in \{g, d, a, c, s\}} \mathcal{G}_i = \{1, 2, \dots, n \mid n \in \mathbb{N}\}$ and $\cap_{i \in \{g, d, a, c, s\}} \mathcal{G}_i = \emptyset$.

A full stoichiometry matrix in (1.2) embodies different types of reaction-type submatrices, where $\{\mathbf{i}\} \subset \{\mathbf{g}, \mathbf{d}, \mathbf{a}, \mathbf{c}, \mathbf{s}\}$ are the corresponding reaction types \mathbf{i} as columns of \mathbf{M} . One can, hence, classify the submatrices in the full stoichiometry matrix in terms of reaction- and species-types, as summarized in Table 1

Table 1: Stoichiometry Matrix Depth Connectivity - Reaction and Species Types

Species Types		<i>Gas phase</i>	<i>Adsorbed Molecules</i>	<i>Surface Radicals</i>
Reaction Types		g	a	s
<i>Homogeneous</i>	g	\mathbf{M}_{gg}	-	-
<i>Ad/desorption</i>	d	\mathbf{M}_{gd}	\mathbf{M}_{ad}	-
<i>Between Adsorbed Molecules</i>	a	-	\mathbf{M}_{aa}	-
<i>Between Adsorbed Molecules and Radicals</i>	c	-	\mathbf{M}_{ac}	\mathbf{M}_{sc}
<i>Between Radicals</i>	s	-	-	\mathbf{M}_{ss}

The stoichiometry matrix is then a composition of reaction-type submatrices, as follows in (1.2).

$$\begin{bmatrix} \dot{c}_g \\ \dot{c}_a \\ \dot{c}_s \end{bmatrix} = \begin{bmatrix} M_{gg} & M_{gd} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & M_{ad} & M_{aa} & M_{ac} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & M_{sc} & M_{ss} \end{bmatrix} \mathbf{r}(\mathbf{c}, \theta) \quad (1.2)$$

Complex reaction paths with multiple reacting surface intermediates may yield matrices of type **gdacs**, encompassing all possible reaction types. These networks may be seen under circumstances where alternate reaction pathways are possible depending on operating conditions, or when multiple possible pathways are being simultaneously considered in the solution of an inverse kinetics problem. Hence, multiple potential mechanisms may be regarded as feasible, and their parameters estimated together for the elucidation of the prevailing pathways.

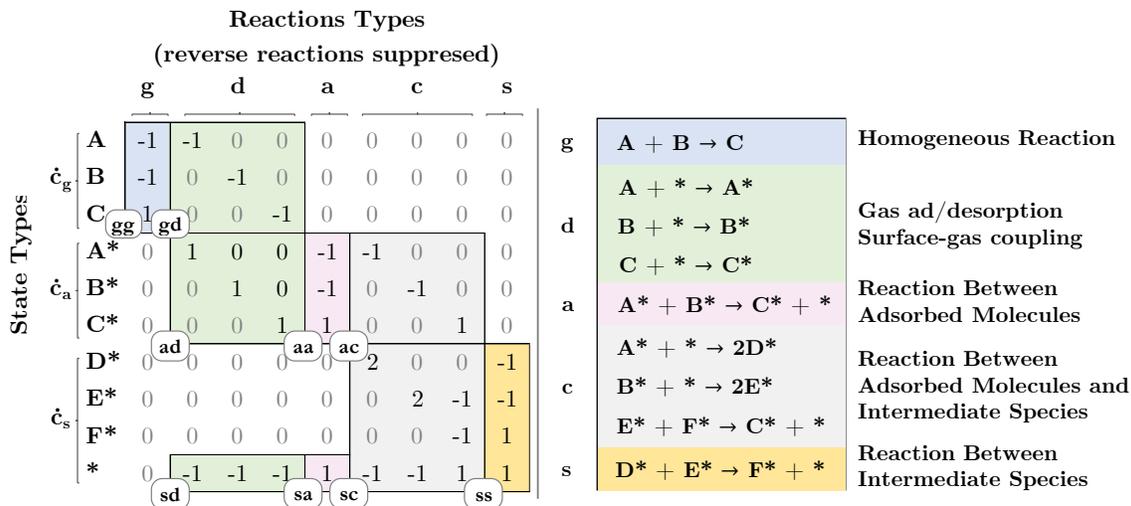


Figure 1: Complex Reaction Network Stoichiometry Matrix Example. A reaction-type is classified according to the columns of the stoichiometry matrix it encompasses.

For instance, combining the *g*, *da* and *dcs* reaction types listed above yields the matrix in Fig. 1 (inverse reactions are suppressed). The *gdacs* reaction network thus represents the maximum variety and depth in complexity, involving all types of species, either directly measurable through analytical methods or mathematically inferred from underlying material balances. This framework allows for the classification of types of elementary reactions based solely on the chemical reaction stoichiometry matrix. In section 3 we build anecdotal examples of reaction network subtypes to guide the discussion on the need for operando analyses for the data-driven estimation of kinetic parameters, and serve as scaffolds for the construction of chemical kinetics PINNs.

2.1.2 Reaction Network Examples

Reaction Network Type *g*: In the simplest homogeneous case conveyed in (g.1) species *A* and *B* reversibly react to form *C* in gas phase or in solution. To simplify the problem solution, the rate constants *k* are considered state- and spatially-independent, which implies an isothermal concentrated-parameter closed system. Various chemical reactions can be represented in terms of (g.1), such as dissociation, dimerization, and lumped-kinetics surrogates derived from complex kinetics data. A simple lumped-kinetics that could be represented as in (g.1) would be ethylene dimerization to 1-butene, where *A* and *B* would represent ethylene, C_2H_4 and *C* would denote 1-butene, C_4H_8 .



Reaction Network Type *da*: For the *da* example, we transpose the reactive step from the homogeneous phase in Eq. (g.1) to the heterogeneous phase, by including the adsorption and desorption elementary steps, **d**.(1-3), and a surface reaction between adsorbed molecules, Eq. (a.1). The set of elementary reactions may be a simple representation of catalyzed dimerization reaction, in the same spirit of Eq. (g.1), except that the reaction occurs in a single step on a

catalyst surface.



Reaction Network Type *dc*: The *a*-type reaction is replaced by two *d*-type reactions, i.e. reaction involving one intermediate species (D^*), which does not have a corresponding gas phase species. The *d*-type adsorption and desorption steps, **d**.(1-3), are suppressed, i.e. implicitly included in the set of elementary reactions comprising the *dc*-example below. An example of a reaction that could be represented overall in such terms would be ethylene epoxidation, where *A* would represent ethylene, *B* molecular oxygen, O_2 , *D* atomic oxygen, O , and *C* would represent ethylene epoxide.



Reaction Network Type *dcs*: Elementary reactions between surface intermediates exclusively are embodied by *s*-type reactions. In the following example, D^* , E^* and F^* are reaction intermediates with no stable desorbed counterpart. In terms of heterogeneous kinetics complexity depth, *s*-type reactions entail intermediates that cannot be directly measured or inferred from ordinary gas/liquid phase effluent/heads-space analytical chemistry techniques, such as gas/liquid chromatography and mass spectroscopy. Butene cracking followed by hydrogenation to ethane could be represented by such a reaction mechanism, where *A* would entail butene, *D* ethylene, *B* molecular hydrogen, H_2 , *E* atomic hydrogen, H , *F* is the ethyl intermediate, and *C* ethane.



2.2 Physics-Informed Neural Networks as Surrogate Approximators

The universal approximation theorem confers to feed-forward (FF) non-polynomial NNs the ability to approximate any continuous function given an appropriate number of neurons [76, 80, 78, 79]. The FFNN structures have proven to serve as appropriate basis functions for ODEs, which may also be extended to PDEs [81, 82, 66]. FFNNs can hence be used to approximate ODE initial value problems associated with chemical kinetics. We further show their use in a collocation framework to solve properly scaled inverse problems, under suitable choice of activation function and NN architecture necessary for the solution (integration) of stiff kinetic ODEs.

Let $\omega_s = \{\mathbf{W}^j \mid \mathbf{W}^j \in \mathbb{R}^{n_j \times m_j}, j \in \{1, 2, \dots, n_l + 1\}\}$ denote a parameter tensor for an n_l -layer feed-forward neural network (FFNN). Let $\mathbf{x} \in \mathbb{R}^n$ be a state vector and $\mathbf{t} \in \mathbb{R}^p$ be collocation points, and $\phi_k(\cdot) : \mathbb{R}^m \rightarrow \mathbb{R}^m$ be the activation function associated with layer j . The input-output structure for the FFNN can be represented by the recurrence relation in (2.1).

$$\begin{aligned} \mathbf{y}_j &= \{\phi_j(\mathbf{W}^j \mathbf{y}_{j-1}) \mid j \in \{1, 2, \dots, n_l\}, y_0 = t\} \\ \mathbf{x}(t, \omega_s) &= \mathbf{y}_{n_l+1} = \mathbf{W}^{n_l+1} \mathbf{y}_{n_l} \end{aligned} \quad (\text{2.1})$$

We refer to the general use of neural-networks methods as surrogate approximators to solve kinetic ODEs as KINNs, kinetics-informed neural networks, using PINNs as a scaffold with suitable modifications to enforce structural initial conditions (ICs) or other phenomenological constraints associated with the kinetics of heterogeneous catalytic systems. The general first-order non-homogeneous coupled ODE implicit representation of an MKM in terms of the KINN is shown in (2.2), which can be generalized to p^{th} order if \mathbf{x} is of class \mathcal{C}^p with respect to t .

$$g[t, \mathbf{x}, \dot{\mathbf{x}}] = g(t, \mathbf{x}(t, \boldsymbol{\omega}_s), \dot{\mathbf{x}}(t, \boldsymbol{\omega}_s)) = g(t, \boldsymbol{\omega}_s) = 0 \quad (2.2)$$

Where \mathbf{x} is the output of the SA encompassing one or more NNs and suitable output transformation layer, and $\dot{\mathbf{x}}$ is evaluated through AD of the former with respect to the single-input input layer. The degrees of freedom in the algebraic equation are all contained in the SA parameters, $\boldsymbol{\omega}_s$, which therefore determines the solution of (2.2) given proper ICs. The time derivative of the SA, $\dot{\mathbf{x}}$, is evaluated with respect to the single-input input layer through AD coupled with just-in-time (JIT) compilation through the JAX library [83][84]. The objective function of the neural network is trained with the *Adam* optimization algorithm. The MKM implicit ODE can be represented in terms of (2.2) as shown in (2.3).

$$\varepsilon_{\dot{\mathbf{x}}_i} = g(t_i, \boldsymbol{\omega}_s) = \dot{\mathbf{x}}(t_i, \boldsymbol{\omega}_s) - \mathbf{M}(\mathbf{k}(\theta) \circ \Pi(\mathbf{x}(t_i, \boldsymbol{\omega}_s))) \quad (2.3)$$

The implicit form, g , and its short-hand $\varepsilon_{\dot{\mathbf{x}}_i}$ are indiscriminately utilized in the following sections to denote residuals of the SA-based formulation to be minimized over collocation point in integration domains. In addition, fitting residuals associated with a particular data-point i are denoted by $\varepsilon_{\mathbf{x}_i}$ as follows in Eq. (2.4), where $\tilde{\mathbf{x}}_i$ are measured state values at t_i .

$$\varepsilon_{\mathbf{x}_i}(\boldsymbol{\omega}_s) = \mathbf{x}(t_i, \boldsymbol{\omega}_s) - \tilde{\mathbf{x}}_i \quad (2.4)$$

In the discussions that follow in the context of the utilization of FFNN as basis for the solution, or integration, of kinetic ODE, we generally denote by \mathbf{x} , the outputs of the FFNN that approximate the ground-truth solution of the related ODE. Although other state variables, such as temperature and pressure, could in principle also be represented in \mathbf{x} , for the isothermal-isobaric cases considered here

$$\mathbf{x} \approx \mathbf{c}.$$

2.2.1 Structural Boundary Conditions

Structural Dirichlet ICs are implemented by designing operators that impose certain behavior onto the NN in regions or points in the integration domain. In general, an operator $\mathbf{C}_{\text{bc}}^{\text{D}}[\cdot]$ must be devised to enforce Dirichlet-type initial conditions, i.e. $\mathbf{x}(\boldsymbol{\omega}_s, t_0) = \mathbf{x}_0 \forall \boldsymbol{\omega}_s$. A natural choice to satisfy a Dirichlet IC is the hyperbolic tangent function $\phi(t) = \tanh(\kappa t) \in \mathcal{C}^1$, such that $\tanh(0) = 0$ and $\partial_t \tanh(0) = 1$. Here $\kappa \in \mathbb{R}_+$ is a characteristic time constant that must be determined for each chemical system or also modelled in terms of independent variables, as shown in section 2.2.2. Here we utilize $\hat{\mathbf{x}}$ and $\mathbf{x}[\hat{\mathbf{x}}]$ as the NN and SA outputs, respectively, with parameters $\boldsymbol{\omega}_s$ and independent variable t . The IC operator is defined as $\mathbf{C}_{\text{bc}}^{\text{D}}[\hat{\mathbf{x}}(\boldsymbol{\omega}_s, t), \mathbf{x}_0] = \hat{\mathbf{x}}(\boldsymbol{\omega}_s, t)\phi(\kappa(t - t_0)) + \mathbf{x}_0(1 - \phi(\kappa(t - t_0)))$.

From simple inspection, the IC is universally enforced in (2.5) and gradients are a function of SA parameters in (2.6).

$$\begin{aligned} \hat{\mathbf{x}}(\boldsymbol{\omega}_s, t)\phi(\kappa(t - t_0))|_{t=t_0} + \mathbf{x}_0(1 - \phi(\kappa(t - t_0)))|_{t=t_0} &= \hat{\mathbf{x}}(\boldsymbol{\omega}_s, t_0)\phi(0) + \mathbf{x}_0(1 - \phi(0)) = \mathbf{x}_0 \quad (2.5) \\ \partial_t \hat{\mathbf{x}}(\boldsymbol{\omega}_s, t)\phi(\kappa(t - t_0))|_{t=t_0} + \partial_t \mathbf{x}_0(1 - \phi(\kappa(t - t_0)))|_{t=t_0} &= \\ &= \dot{\hat{\mathbf{x}}}(\boldsymbol{\omega}_s, t_0)\phi(0) + \mathbf{x}(\boldsymbol{\omega}_s, t_0)\partial_t \phi(0) - \mathbf{x}_0\partial_t \phi(0) \quad (2.6) \\ &= \hat{\mathbf{x}}(\boldsymbol{\omega}_s, t_0) - \mathbf{x}_0 \end{aligned}$$

Such that the SA vanishes at $t = t_0$, (2.5), with the continuous time-derivative of \mathbf{x} equal to the NN output, $\hat{\mathbf{x}}$, (2.6). Therefore, any Dirichlet IC can be automatically satisfied in this form for KINNs. The same approach can be extended to Neumann ICs with proper choice of ϕ function. The application of the operator $\mathbf{C}_{\text{bc}}^{\text{D}}$ on MKMs follows naturally by making $\mathbf{x} \approx \mathbf{c}$ in (1.1) as depicted by (2.3).

2.2.2 Characteristic Time

In the case of *tanh* as choice of ϕ for $\mathbf{C}_{\text{ic}}^{\text{D}}$, a proper choice for characteristic time constant can be circumvented by replacing κ by an ancillary shallow NN that can be learned along with the underlying SA NNs, with a monotonically increasing output layer. In this work we adopted an exponential form for κ whenever solving the forward ODEs, as in (2.7).

$$\kappa = e^{\mathbf{u}(\boldsymbol{\omega}_t, t - t_0)} \quad (2.7)$$

Where \mathbf{u} is a shallow FFNN with parameters $\boldsymbol{\omega}_t$. The IC operator on ϕ and κ , $\mathbf{C}_{\text{ic}}^{\text{D}}$, can thus be represented in terms of ϕ and \mathbf{u} as follows in (2.8), where parameter dependencies were suppressed. For the sake of simplicity, we further assume κ is embedded into the SA structure \mathbf{x} so its weights, $\boldsymbol{\omega}_t$, are included in $\boldsymbol{\omega}_s$.

$$\mathbf{C}_{\text{bc}}^{\text{D}}[\hat{\mathbf{x}}(t), \mathbf{x}_0] = \hat{\mathbf{x}}(t) \tanh\left(e^{\mathbf{u}(t - t_0)}(t - t_0)\right) + \mathbf{x}_0\left(1 - \tanh\left(e^{\mathbf{u}(t - t_0)}(t - t_0)\right)\right) \quad (2.8)$$

2.2.3 Normalization DAE Constraints

MKMs typically treat the concentration of adsorbed and intermediate species in terms of fractions of the total number of available active sites. It is hence desirable for SAs to structurally enforce normalization, eliminating a degree of freedom. We propose the projection of radius-one hypersphere onto the natural basis to enforce inherent normalization. Let $\mathbf{x}(\boldsymbol{\omega}_s, t) \in \mathbb{R}^k$ be the output values of the constrained surrogate related to the neural network $\hat{\mathbf{x}}(\boldsymbol{\omega}_s, t) \in \mathbb{R}^{k-1}$, such that

$$\begin{aligned} x_i(\boldsymbol{\omega}_s, t) &= (1 - \sin^2(\hat{x}_i(\boldsymbol{\omega}_s, t))) \prod_{j < i} \sin^2(\hat{x}_j(\boldsymbol{\omega}_s, t)) \quad \forall i < p; i, j \in \mathbb{N} \\ x_p(\boldsymbol{\omega}_s, t) &= \prod_{j < p} \sin^2(\hat{x}_j(\boldsymbol{\omega}_s, t)) \end{aligned} \quad (2.9)$$

Such a trigonometric transformation enforces bounds to \mathbf{x} , such that $0 \leq \mathbf{x} \leq 1$ and $\sum \mathbf{x} = 1 \quad \forall \hat{\mathbf{x}} \in \mathbb{R}^{p-1}$. The SA output can hence be denoted as the result of the normalization-constraint operator C_N over the NN output $\hat{\mathbf{x}}$, i.e. $\mathbf{x} = C_N[\hat{\mathbf{x}}]$. The normalization-constraint operator implicitly embeds DAE constraints to the ODE solutions, since it encodes $\sum \dot{\mathbf{x}} = 0$ for species coverage.

2.3 KINN Forward Problem

The surrogate approach to solving the forward problem involves training the neural network to satisfy the ICs, as well as the differential form specifying the reaction kinetics in (2.2), over a defined time grid. We constrain the analysis to non-spatially dependent closed systems, formulating the problem for an ideal, uniform (perfectly stirred), pseudohomogeneous batch reactor model, where the underlying ODE describing the state-variables evolution over time is analogous to (1.1). The approach would also hold for the steady-state solution over space of fixed or packed-bed reactors. The general solution for the forward problem, i.e. solving the ODE, consists of minimizing the implicit form in (2.2) given consistent ICs for the states, where the real-valued function g corresponds to (1.1) in standard form given fixed model parameters, \mathbf{p} , which is outlined by (2.3). The initial value is represented by \mathbf{x}_0 , and the residual is evaluated over the collocation points, \mathbf{t} , as shown in (3.1). This represents the minimization of the residual j_m , under some L2-norm, between the Jacobians of the states from the derivative-based rate laws and the surrogate approximator. AD provides the Jacobians, i.e. $\dot{\mathbf{x}}$ and $\partial_{\boldsymbol{\omega}_s} j_m(\mathbf{t}, \boldsymbol{\omega}_s)$, for the objective function. The *Adam* optimization algorithm is used to minimize the residual, j_m .

$$\begin{aligned} \min_{\boldsymbol{\omega}_s} \quad j_m(\boldsymbol{\omega}_s) &= \|g(\mathbf{t}, \boldsymbol{\omega}_s)\|_2^2 \\ &= \frac{1}{n} \sum_i^n \boldsymbol{\varepsilon}_{\dot{\mathbf{x}}_i}^T \boldsymbol{\varepsilon}_{\dot{\mathbf{x}}_i} \\ \text{s.t.} \quad \mathbf{x}(t_0, \boldsymbol{\omega}_s) &= \mathbf{x}_0 \end{aligned} \quad (3.1)$$

Importantly, in the case of MKMs for which surface coverage fractions are in play, the normalization of elements of g can be structurally enforced as shown in 2.2.3 by having $g(\mathbf{x}(\hat{\mathbf{x}}))$ as in (2.9), which imposes a DAE constraint. Furthermore, when dealing with mixed normalized (coverage) and non-normalized (bulk-phase) chemical species, the problem can be split into two separate underlying NNs which are concatenated in the SA output layer: a non-normalized one representing the mapping between time and gas-phase concentrations and a normalized one mapping between time and surface coverages. The implicit representation of MKM KINN is given in (2.3). When structural ICs are applied and normalization constraints are included, the surrogate model consists of the following combinations of underlying NNs and operators, (3.2), using the same nomenclature as defined in 2.1

$$\mathbf{x} = C_{ic}^D \left[\begin{array}{c} \hat{\mathbf{x}}_g \\ C_N[\hat{\mathbf{x}}_{as}] \end{array} \right] \quad (3.2)$$

Where $\hat{\mathbf{x}}_g$ corresponds to the output layer of the NN associated with gas (unbound) species and $\hat{\mathbf{x}}_{as}$ encompasses the output layer of subjacent NN that conveys the $p - 1$ angular components that, once mapped through the operator C_N , outputs p normalized coverage fractions. Finally, the initial-value IC operator C_{ic}^D is applied onto the combined (concatenated) NNs.

2.4 KINN Inverse Problem

Under the same model assumptions of the forward MKM problem, we extend the PINN-based KINN approach to MKM inverse ODE problems by letting $\tilde{\mathbf{x}} \in \mathbb{R}_+^{n \times d}$ represent observed (measured) concentrations of n chemical species over the respective d time-points $\mathbf{t} \in \mathbb{R}_+^d$, and including it in the objective function. As in the forward problem,

Algorithm 1 Forward KINN

-
- 1: **Input:** Boundary conditions (t_0, \mathbf{x}_0) , \mathbf{p} , maximum iterations n_{max} , tolerance tol
 - 2: Initialize NN, SA parameters ω_s , residual $j_m(\omega_s)$
 - 3: **while** iter $< M$ **or** $j_m(\omega_s) < tol$ **do**
 - 4: Predict states \mathbf{x} from SA over \mathbf{t}
 - 5: Predict states derivatives $\dot{\mathbf{x}}$ from SA over \mathbf{t}
 - 6: Compute objective function $j_m(\omega_s)$ over \mathbf{t}
 - 7: Compute gradients of $j_m(\omega_s)$ with AD; $\partial_{\omega_s} j_m(\omega_s)$
 - 8: Update SA parameters ω_s using *Adam*
 - 9: **Output:** Concentration profile $\mathbf{x}(t, \omega_s)$
-

state variables are represented by surrogates from separate NNs and applied operators, as in (3.2), i.e. $\mathbf{x}(t, \omega_s)$. The inverse problem consists of the forward KINN model with the inclusion of interpolation of observed datapoints, $\tilde{\mathbf{x}}$ as a regularization term in the objective function. In the regularized setup, the error to be minimized consists of the combination of the residual between the SA and the observed data (interpolation), j_d , and that between SA and the kinetic model describing the behavior (physics regularization), j_m , in a least-squares sense, similarly to the approach by Raissi *et al.* [67]. As in the forward problem, the combined error is minimized using the *Adam* optimization algorithm.

$$j_d(\omega_s) = \frac{1}{n} \sum_i^n \epsilon_{\mathbf{x}_i}^T \epsilon_{\mathbf{x}_i}$$

$$j_m(\omega_s, \mathbf{p}) = \frac{1}{n} \sum_i^n \epsilon_{\dot{\mathbf{x}}_i}^T \epsilon_{\dot{\mathbf{x}}_i}$$

Finally, let $\omega = \{\omega_s, \mathbf{p}\}$ combine the residual from the forward-KINN as regularization hyperparameter with weight α into the inverse ODE.

$$\begin{aligned} \min_{\omega} \quad & j_t(\omega) = j_m(\omega_s, \mathbf{p}) + \alpha j_d(\omega_s) \\ \text{s.t.} \quad & \mathbf{p} \in \mathbb{R}^{\dim(\mathbf{p})}, \omega_s \in \mathbb{R}^{\dim(\omega_s)}; \alpha \in \mathbb{R}_+ \end{aligned} \quad (4.1)$$

The regularization hyperparameter α conveys the ratio between the variances of the data (states) and model (state derivatives) under the hypothesis of homoscedasticity, and it needs to be optimized given an MKM postulate and the observed data. Furthermore, the regularization intensity plays a decisive role in the case of noisy data, since initial training steps with excessive regularization potentially lead to local convergence far from the global optimum in the model parameter space. A conservative strategy implemented in this work is to perform initial training steps at low values of α and then proceed with increments in α until an inflection point between $\log(j_d)$ and $\log(j_m)$ is observed. In section 3.2, a sensitivity analysis is performed over α values for an anecdotal inverse problem. Also, the learning of model parameters must take place in a similar domain and scale as the SA-associated NN weights. As kinetic parameters have non-negative values and may substantially vary in order of magnitude, we choose to train an exponential mapping, i.e., set $\mathbf{k} = \exp(\mathbf{p})$. Furthermore, for the *tanh* activation function, a saturation occurs at arguments of ± 5 , and for *swish* with a similar pattern as *exp*, such that \mathbf{p} and ω_s are of the same order of magnitude. This allows that they be grouped as a lumped variable and trained simultaneously as described in the following pseudo-algorithm. Different non-stiff anecdotal kinetic models are solved in the forward sense to justify the use of FFNNs as bases for the

Algorithm 2 Regularized Inverse KINN

-
- Input:** Observed data $(t, \tilde{\mathbf{x}})$, maximum iterations n_{max} , tolerance tol
 - 2: Build NN and initialize SA parameters ω_s , and kinetic model parameters \mathbf{p}
 - while** iter $< n_{max}$ **or** $j_t(\omega) < tol$ **do**
 - 4: Predict states $\mathbf{x}(t, \omega_s)$ from SA over \mathbf{t}
 - Predict states derivatives $\dot{\mathbf{x}}(t, \omega_s)$ from SA over \mathbf{t}
 - 6: Compute objective function $j_t(\omega)$ over \mathbf{t}
 - Compute gradients of $j_t(\omega)$ with AD; $\partial_{\omega} j_t(\omega_s, \mathbf{p})$
 - 8: Update SA parameters ω_s and kinetic model parameters \mathbf{p} using *Adam*
 - Output:** Kinetic model parameters \mathbf{p}
-

solution of the inverse problem. We increase reaction network complexity, moving through (i) an example of a simple homogeneous case (type *g*), (ii) an example involving reaction between adsorbed molecular species (type *da*), (iii) an example including reactions between adsorbed molecules and reaction intermediates (type *dc*) and (iv) an example

including a reaction between intermediates (type *dcs*). Abstract species are used in the following examples, which may suit a myriad of actual chemical reactions.

2.5 Synthetic Data Generation

The scheme for solving the forward KINNs problem, and the inverse problem in three different scenarios, is shown in Fig. 2. The scenarios, which vary in the completeness of knowledge about the chemical systems and the presence of noise representing actual measurements, as listed below.

- **Quantitative Unbound Species (*Q*):** no information about chemical intermediates and adsorbed molecules; the only measured states are related to unbound chemical species; e.g. gas, liquid species that can be measured downstream (flow-type reactors) or sampled from the reactor headspace (batch) with standard analytical equipment: gas/liquid chromatography, mass spectroscopy. For the *Q* scenarios, only data generated for the observable variables i.e. unbound chemical species are fed to the training loop, such that j_d , or ε_x , has the length of the observable variables. Meanwhile, j_m , or ε_x , in (4.1), encompasses derivatives for all present (or assumed present) chemical species and intermediates that are measured and unmeasured.
- ***Q* + Semiquantitative Bound Species (*SQ*):** it is assumed that resolved *SQ* data can be obtained, e.g. with operando analytical techniques, for all adsorbed molecules and stable reaction intermediates, and that the *SQs* are linearly related to the concentration of such species. For the *SQ* scenario, the surface coverages generated from the numerical solution of the forward problem are rescaled by dividing the time series values by their standard deviation (an effectively arbitrary number that mimics a calibration factor in spectroscopy). In the inverse KINNs problem, the scaling factors between fractional coverages and the *SQ* intensities are also included as parameters to be learned in ω_s .
- ***SQ* + noise (*SQ+n*):** a normally distributed homoscedastic error (white noise) with an arbitrary standard deviation of 0.025 is added to the scaled data for the *SQ* scenarios. This setup intends to represent a more realistic scenario that evaluates the robustness of the physics regularized setup in the presence of noise.

To obtain the ground truth set for the comparison of the forward KINN results, the ensuing ODEs of all combinations of MKMs and ICs were solved numerically with the stiff-nonstiff algorithm LSODA from the FORTRAN ODEPACK [85, 86]. Similarly, the generation of abstract observed data and further per-scenario processing was carried out using the same numerical method, as shown in Fig. 2(A). For the solution of forward KINNs, the MKM stoichiometry matrix, the kinetic parameters, \mathbf{p} , and ICs is provided, Fig. 2(D). Conversely, for the inverse problem, \mathbf{p} is adjusted and synthetic data is generated and provided as if experimentally assessed, Fig. 2(B). The KINN SA-based solution flow diagram is portrayed in Fig. 2(C), exhibiting the interplay between the SA and the physical models, and the final cost function, j_m .

Table 2: Rate Constants and NN architecture

Type	Rate Constants											Architecture (layers)	
	$\frac{k_1^g}{k_{-1}^g}$	$\frac{k_1^d}{k_{-1}^d}$	$\frac{k_2^d}{k_{-2}^d}$	$\frac{k_3^d}{k_{-3}^d}$	$\frac{k_1^a}{k_{-1}^a}$	$\frac{k_1^c}{k_{-1}^c}$	$\frac{k_2^c}{k_{-2}^c}$	$\frac{k_3^c}{k_{-3}^c}$	$\frac{k_4^c}{k_{-4}^c}$	$\frac{k_5^c}{k_{-5}^c}$	$\frac{k_1^s}{k_{-1}^s}$	Hidden [<i>tanh,swish,tanh</i>]	Output [linear,C _N [·]]
<i>g</i>	$\frac{10}{1}$	-	-	-	-	-	-	-	-	-	-	[5, 5, 5]	[3,-]
<i>da</i>	-	$\frac{10}{4}$	$\frac{40}{60}$	$\frac{200}{40}$	$\frac{100}{80}$	-	-	-	-	-	-	[12, 12, 12]	[3,4]
<i>dc</i>	-	$\frac{20}{8}$	$\frac{16}{4}$	$\frac{12}{8}$	-	$\frac{1200}{400}$	$\frac{2000}{1600}$	-	-	-	-	[16, 16, 16]	[3,5]
<i>dcs</i>	-	$\frac{20}{8}$	$\frac{24}{12}$	$\frac{16}{40}$	-	-	-	$\frac{640}{960}$	$\frac{160}{80}$	$\frac{560}{160}$	$\frac{640}{240}$	[20, 20, 20]	[3,7]

The cost function to be minimized for learning the solution of the ODEs using KINNs was defined in terms of mean-square error, MSE. The JAX library was utilized for the forward AD, allowing the obtainment of the gradients with respect to the SAs’ parameters, which were optimized with the *Adam* algorithm [87]. Since ICs are structurally enforced by C_{ic}^D , the ODE solutions are found by minimizing j_d , i.e. finding ω_s such that the state derivatives from the physical model given SA state estimates, \mathbf{x} , match those estimated by forward AD of the SAs with respect to their single value time input. All code necessary for reproducing the results of this work are provided in the Supplementary Material and via GitHub (https://github.com/gusmaogabriels/kinn/tree/paper_reg).

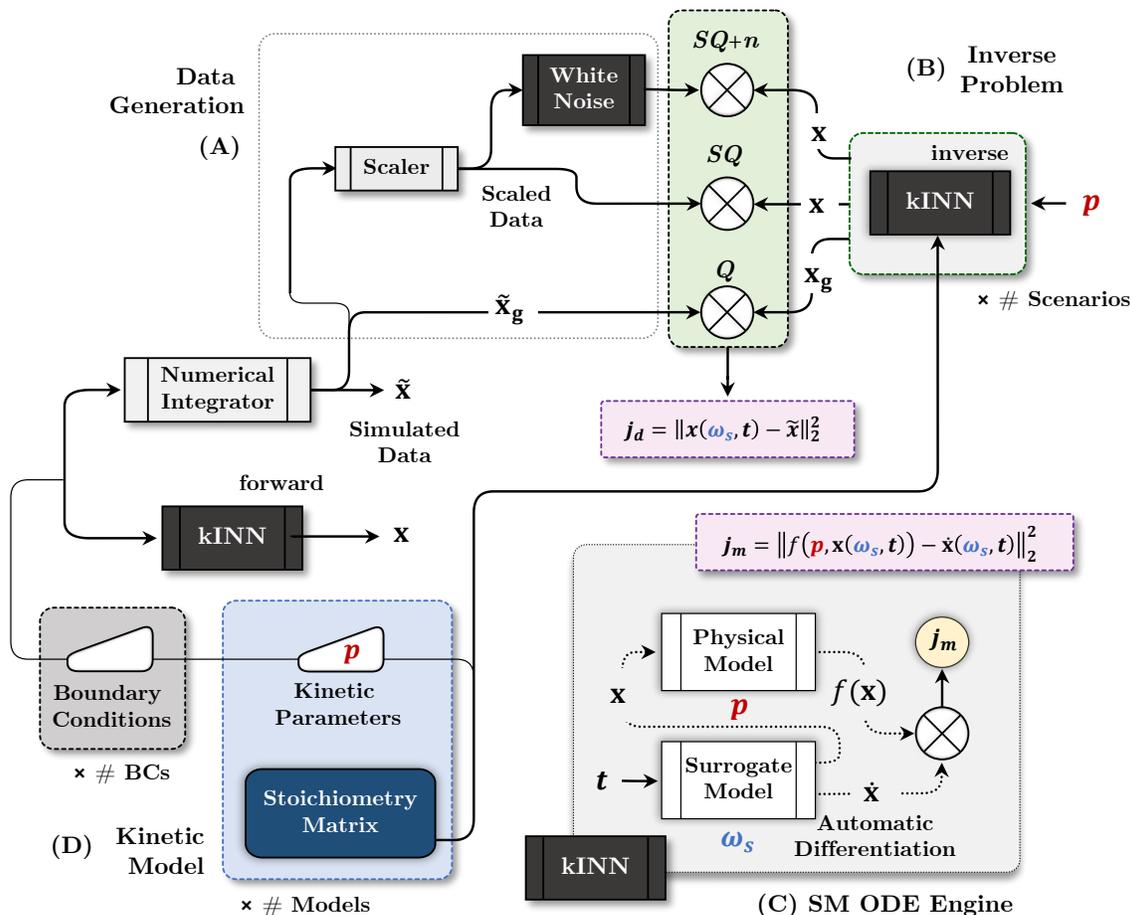


Figure 2: Data generation (A) from stiff-nonstiff algorithm LSODA for three case studies: quantitative data for gas phase only (Q), quantitative gas phase and scaled semiquantitative coverage fraction without (SQ) and with added noise ($SQ+n$). Forward (D, KINN block) and inverse (B) KINNs scheme. KINN block (C) and kinetic model structure and inputs (D).

3 Results and Discussions

3.1 Forward KINNs

Three-hidden-layer NNs with *tanh*, *swish* and *tanh* activation functions respectively were utilized as basis for the solution of forward KINNs. For the overall reversible reaction $A \rightleftharpoons B + C$ represented by each reaction network type, several anecdotal examples for the forward problem were solved given a set of diverse *Dirichlet* ICs in terms of x_A , x_B and x_C at $t = 0$, and sets of rate constants chosen to portray different degrees of stiffness, i.e. faster rates for surface reactions, as in Table 2. For all reaction types, hyperbolic tangent *Dirichlet*-type boundary-condition operators, C_{ic}^D , were applied to SAs, as in section 2.2.1. Ancillary single-hidden-layer FFNNs with three (*a*- and *da*-type) or six (*dc* and *dcs*-type) neurons and *swish* activation functions were trained in parallel for each forward problem to allow for a time-dependent characteristic time or gain for the IC operator, which provides a learning gain-layer in the case of varying stiffness, as defined in section 2.2.2. The forward KINNs are defined in terms of their associated reaction network (stoichiometry matrix), ICs and kinetic parameters. Normalization constraints are not included in the forward KINNs since they bring additional non-linearity that is unnecessary given that the problems are well-defined by their ODEs and ICs.

Table 3 allows the comparison of the numerical results with the SA derived for the forward KINNs. The set of ICs was defined so as to represent the two scenarios: one in which there are similar concentrations of both reactants, (IC 1, $\{x_A, x_B, x_C, x_*\}_{t_0} = \{0.6, 0.4, 0.0, 1.0\}$), and another to represent the case where considerable amount of product

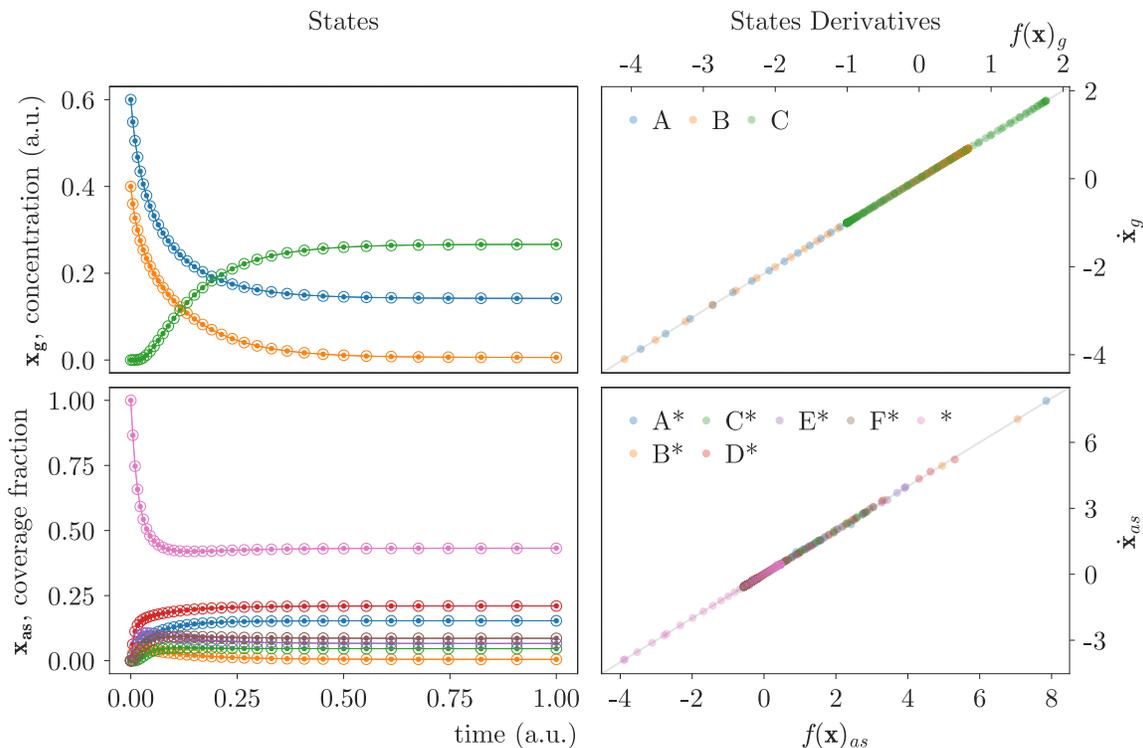


Figure 3: Reaction network type *dcs* forward solution with IC of $\{x_A, x_B\}_{t_0} = \{0.6, 0.4\}$ and clean surface (IC 1). States solution (top) (SA, open circles; numerical solution, closed markers) under minimization of (3.1) (bottom) for observable variables (unbound species, top) and latent variables (coverages, bottom).

is present, (IC 2, $\{x_A, x_B, x_C, x_*\}_{t_0} = \{0.2, 0.3, 0.5, 1.0\}$). The two conditions were chosen to represent chemical transformations that predominantly run in the forward or reverse reaction, and thus inform further attempts to solve the related inverse problems. The performance metrics are defined as a function of the collocation points about which the SA is trained, which is exemplified in Fig. 3 by the *dcs*-reaction network type solution for IC 1. For all scenarios, about 100 collocation points were sampled in a logarithm space in the time domain, such that there is higher probability that points be placed in regions of higher derivatives near the IC. To refine the solutions, SAs were trained in three stages with varying training step size from 10^{-4} to 10^{-6} , over a total of 10^3 epochs, and 10^2 iterations per epoch. The convergence criteria was that either $j_t < 10^{-12}$ or that the maximum number of epochs was reached. Evidently, absolute error metrics such as MAE and MSE cannot be compared between different MKMs and ICs, since different ODEs may exhibit different stiffness and therefore larger or smaller derivative values in different scales as compared to actual state values. Irrespective of MKM and ICs, the coefficient of determination (r^2) for all examples is unity (within round-off error). Such results support the hypothesis that NNs can be utilized as general bases for the solution of forward KINNs, which is the underpinning assumption for their utilization in the solution of inverse KINNs.

In Fig. 3 the final solution for the type-*dcs* MKM under the reactant-only IC (IC 1) is shown. The predicted states (open circles) line up over the numerical solution (solid circles) for both the observable concentrations \mathbf{x}_g (top) and coverage fractions \mathbf{x}_{as} (bottom). The minimization of the underlying forward KINNs cost function j_m can be inferred by the parity plot between physical (kinetic) model derivatives from estimated states, and their derivative estimates from forward AD. For the remaining cases, the standardized derivative parity plots for the observed and latent states can be compared with the coefficient of determination reported in Table 3 for which coefficients of determinations are computed as the average squared correlation coefficient over different states.

3.2 Inverse KINNs

The forward numerical solution data for the different reaction network types were utilized as the ground truth data for the inverse KINN studies. We define three scenarios or categories for the inverse KINNs involving heterogeneous models, i.e. reaction networks other than type-*g* only, as portrayed by scheme (A) in Fig. 2. Unlike the forward KINNs,

Table 3: Forward KINNs Results Summary - Performance Metrics (vs Ground Truth)

Type	IC	Observed States Derivatives			Latent States Derivatives		
		r ²	MAE	MSE	r ²	MAE	MSE
<i>g</i>	1	1.00	5.50×10^{-4}	3.92×10^{-7}	-	-	-
	2	1.00	1.79×10^{-5}	4.27×10^{-10}	-	-	-
<i>da</i>	1	1.00	2.51×10^{-3}	1.12×10^{-5}	1.00	2.97×10^{-3}	1.84×10^{-5}
	2	1.00	8.93×10^{-3}	2.49×10^{-4}	1.00	1.73×10^{-2}	1.32×10^{-3}
<i>dc</i>	1	1.00	3.30×10^{-3}	2.09×10^{-5}	0.999	1.26×10^{-2}	4.70×10^{-4}
	2	1.00	2.78×10^{-3}	1.68×10^{-5}	1.00	7.38×10^{-3}	1.42×10^{-4}
<i>dcs</i>	1	1.00	2.18×10^{-3}	9.87×10^{-6}	1.00	7.13×10^{-3}	4.00×10^{-4}
	2	1.00	6.89×10^{-4}	8.45×10^{-7}	1.00	3.33×10^{-3}	3.58×10^{-5}

whose solutions for the two ICs are independent of each other, the inverse KINNs are interdependent since they share, and thus act on, the same kinetic model and its kinetic parameters encoded in \mathbf{p} . The solution of inverse KINNs can involve multiple experimental (observed) data to which independent SAs are trained under a single underlying physical model. In this section, all inverse KINN estimated model parameters were obtained by the simultaneous interpolation of IC 1 and 2 data from the numerical solution of the MKM ODEs.

We start by analyzing the *SQ* and *SQ+n* scenarios, since these scenarios contain the most information about the system. The normalization constraint operator, C_N , plays a major role in the learning of scaling factors for the *SQ* scenarios, since it constrains the scaling factors to describe normalized (fractional) coverages, and indirectly imposes a DAE constraint to ODE solution. Importantly, the regularization parameter, α , (4.1) must be determined for each specific reaction network type and associated data (as in any method where hyperparameters need be optimized). The optimal α range can be inferred by simultaneously minimizing the residual terms with respect to data, j_d , and to the physical model, j_m . Since this involves a trade-off between the two parts of the objective function, there will be a Pareto frontier of optimal models as a function of α . We utilized sensitivity analyses on the residual terms for the anecdotal *dcs*-type reaction network scenarios to estimate their Pareto frontiers. We successively increased α from 10^{-6} to 10^6 and further decreased it back to 10^{-6} to assess the non-dominated points (i.e. Pareto-optimal set of points with optimal trade-off between the two objectives) in the estimated Pareto sets for the *SQ* and *SQ+n* scenarios. For each value of α , the same SA learning strategy in terms of convergence criteria, number of epochs, iteration per epoch and solution refinement as defined in section 3.1 was adopted.

Fig. 4(B) shows the learned Pareto sets as α is increased (tightened) with open markers from left to right. The SA begins by learning the underlying ODE structure (low α , low j_m , high j_d). As α increases, j_d monotonically decreases with a corresponding increase in j_m , resulting in a forward concave set with an inflection point for j_m around $\log(j_m) \approx -4$ (open markers) for both the *SQ* (black) and *SQ+n* (red) scenarios. In Fig. 4(A), the steepest improvement in the correlation between the ground-truth and learned model parameter is located in the same region of the concave tightening-Pareto inflection point α . When relaxing α with the same SAs trained in the tightening direction, the Pareto set becomes convex, Fig. 4(A), as expected for a multiobjective minimization problem between two cost functions. This tightening-relaxation strategy and the observed Pareto-set behavior allow the proper estimation of α values in the Pareto set that provide an equilibrated trade-off between satisfying the MKM ODE and interpolating the observed data. Specifically, in the case of *SQ+n*, further tightening α beyond the inflection point leads to little improvement (plateau in the red curve, Fig. 4(B) in the inverse problem solution, instead further fitting the SA to data noise (overfitting)). This can be inferred from the trend in the correlation between learned and ground truth model parameters in Fig. 4(A), which reaches a stable value for the same range of α where the plateau is observed in Fig. 4(B) over successive tightening learning steps. Additionally, if α is further increased, a second inflection point is expected, where there is a detachment of the SA from the underlying MKM ODE and a steep reduction in j_d due to overfitting. In the case of *SQ*, the basis slack region in (B) portrays the inability of the specific choice of SA architecture (i.e. number of layers, activation function) to satisfy both the MKM ODE (physical model) and interpolate the ground-truth data. Such an argument is supported by the universal approximation theorem [77], which states that, with a sufficient number of neurons, states and their associate derivatives can be approximated to arbitrary accuracy. Nevertheless, it ensues from the universal approximation theorem that there should exist pathological scenarios for sufficiently large or deep NNs where exact interpolation of datapoints is obtained, while excess degrees of freedom allows steep kinks to be created about the interpolated points. In such a scenario, derivatives can span a wider range of possible kinetic parameters that would still satisfy the underlying physics constraints. We opted for constructing NNs that are sufficiently large to adhere to the observed data, but not too large to enable multiple degenerate physical solutions. Fig. 4 also illustrates, in (C),

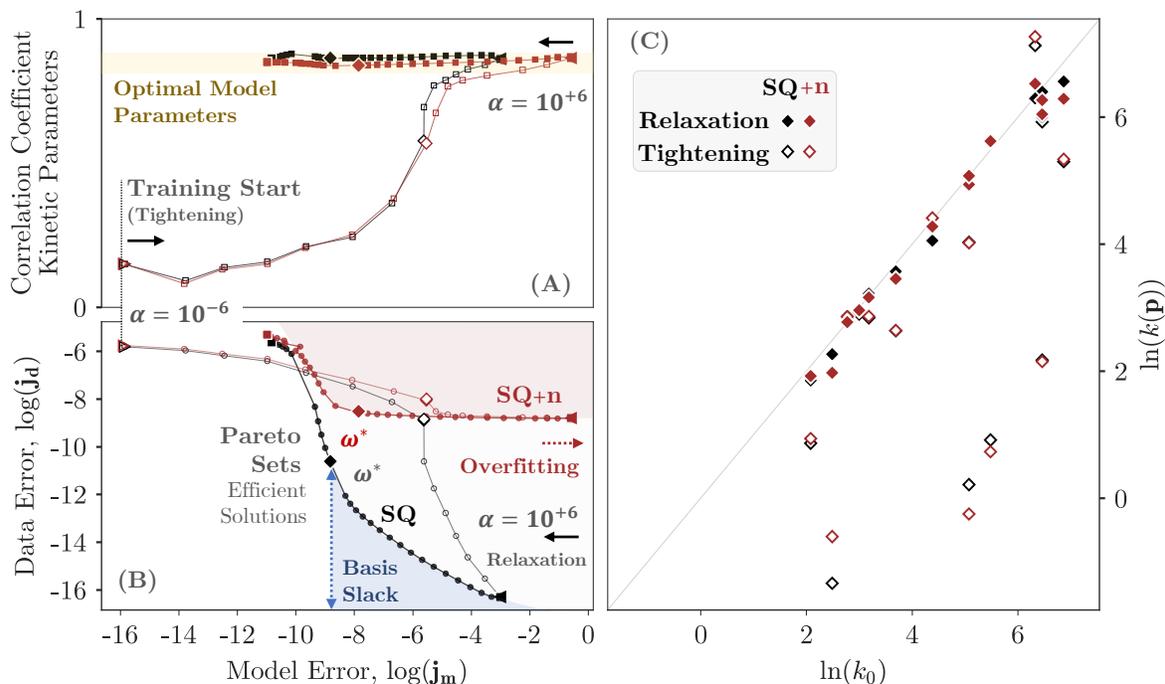


Figure 4: Inverse KINN - Regularization Sensitivity Analysis: Pareto Set Estimate. Regression correlation coefficient between fitted and actual kinetic parameters (A), where training (tightening, open markers) starts at the lower left corner at low data regularization (α), increasing from left to right as a function of model error \mathbf{j}_m . The relaxation series succeeds from right to left (closed markers), where model parameters exhibit minor improvement; however data (\mathbf{j}_d) and model- (\mathbf{j}_m) error evolution as a function of α define the Pareto frontier in (B) from which non-dominated (efficient) solutions can be obtained. Optimal parameters for scenarios SQ and $SQ+n$ (black) and $SQ+n$ (red) are shown in (C) at the open and closed inflection points in (B). The basis slack region in (B) conveys the inherent constraint imposed by the choice of simple NN architecture to the derivatives (model) space, which translates into a compromise between interpolation and satisfying the physical model as α is relaxed.

that the learned physical model parameters at the relaxation inflection point are highly correlated with the ground-truth parameters irrespective of noise.

In all models in this work, the tightening steps started with $\alpha = 10^{-6}$, with 10-fold α increment between each training cycle, and the relaxation steps reduce α by the finer increments, with the model at the inflection point in relaxation taken as the optimal model. Table S1 in the Supplementary Material compiles the performance metrics for the observable variables states and their derivatives for the four different reaction network types, two ICs and three inverse scenarios. Accurate matches between observed states within noise limits and highly correlated derivatives are observed for all cases. In terms of latent variables, the performance metrics in Table 4 show the agreement between simulated and fitted latent states and their derivatives, irrespective of reaction network type, for scenarios SQ and $SQ+n$. As for the Q scenarios, even with lack of direct knowledge on surface coverages, the surface dynamics information can be indirectly retrieved for model da . Hence, the latent (bound) state profile can be inferred from unbound data only. With the increase in depth, the correlation between predicted states starts to fade. This occurs because adsorbed species (type c and s) coverages are not directly measured in the Q scenario, thus there exist multiple solutions for the combination of their evolution over time and the kinetic parameters associated with their reactions. The temporal profiles for the final solutions for observed and latent states as well as their derivative (SA vs. physics model) are reported in the Supplementary Material, Figs. S2.1 to S2.6, S3.1 to S3.6 and S4.1 to S4.6

To showcase the robustness of SA-based inverse KINNs, we consider the final refined solution profiles for the $SQ+n$ scenario for the dcs -type reaction network with IC 1 in Fig. 5. In the $SQ+n$ scenarios, not only are kinetic parameters learned, but also the scaling factor that connects intermediate species abundance as a signal or intensity to fractional coverages. For the optimized regularization, which is dependent on the α value, the underlying SA interpolation suppresses the effect of the added homoscedastic noise. Larger α values would lead to predicted points (interpolation, open markers) moving further towards the measured data (closed circle markers), resulting in SA

Table 4: Inverse KINNs Results Summary for reaction types da , dc and dcs - latent states (vs. ground truth, right) and derivatives (physical model, left) error metrics.

Type	IC	Mode	Latent States Derivatives			Latent States (ground truth)		
			r^2	MAE	MSE	r^2	MAE	MSE
da	1	Q	1.00	9.59×10^{-3}	2.99×10^{-4}	0.972	3.17×10^{-2}	1.40×10^{-3}
		SQ	0.944	0.131	0.325	0.999	1.09×10^{-3}	2.30×10^{-6}
		$SQ+n$	0.941	7.62×10^{-2}	1.90×10^{-2}	0.863	1.56×10^{-2}	4.79×10^{-4}
	2	Q	1.00	7.59×10^{-3}	1.69×10^{-4}	0.998	3.48×10^{-2}	1.69×10^{-3}
		SQ	0.998	7.82×10^{-2}	4.47×10^{-2}	1.00	1.26×10^{-3}	2.23×10^{-6}
		$SQ+n$	1.00	4.47×10^{-2}	3.37×10^{-3}	0.948	1.79×10^{-2}	4.15×10^{-4}
dc	1	Q	1.00	1.91×10^{-2}	5.92×10^{-4}	0.965	0.118	2.32×10^{-2}
		SQ	0.999	2.63×10^{-2}	3.50×10^{-3}	1.00	6.39×10^{-4}	6.21×10^{-7}
		$SQ+n$	0.983	5.96×10^{-2}	5.93×10^{-3}	0.895	2.67×10^{-2}	1.29×10^{-3}
	2	Q	0.996	1.04×10^{-2}	1.54×10^{-4}	0.940	0.151	3.68×10^{-2}
		SQ	1.00	1.13×10^{-2}	4.60×10^{-4}	1.00	6.64×10^{-4}	6.96×10^{-7}
		$SQ+n$	0.988	4.08×10^{-2}	3.70×10^{-3}	0.988	3.24×10^{-2}	2.05×10^{-3}
dcs	1	Q	0.975	1.87×10^{-2}	8.11×10^{-4}	0.374	0.146	3.84×10^{-2}
		SQ	1.00	5.53×10^{-3}	8.53×10^{-5}	1.00	2.66×10^{-3}	1.28×10^{-5}
		$SQ+n$	1.00	1.28×10^{-2}	3.24×10^{-4}	0.996	1.47×10^{-2}	3.89×10^{-4}
	2	Q	0.736	1.81×10^{-2}	9.85×10^{-4}	0.236	0.181	5.76×10^{-2}
		SQ	1.00	9.30×10^{-3}	2.51×10^{-4}	1.00	2.68×10^{-3}	1.30×10^{-5}
		$SQ+n$	0.998	2.11×10^{-2}	1.44×10^{-3}	0.989	1.06×10^{-2}	1.45×10^{-4}

overfitting with respect to data (i.e. the resulting curve would not satisfy the differential equations of the kinetic model). In contrast to the forward KINNs, ICs are not explicitly provided, so all observed data act as a composite set of ICs whose weight is conveyed by j_d . As depicted by Eq. (4.1), the efficient solutions are given by Pareto-optimal (i.e. non-dominated) points that minimize the multiobjective optimization function j_t , involving j_d and j_m . The latter, which represents the error in satisfying the MKM ODE, has similar structure as the one for forward KINNs except that \mathbf{p} becomes a variable for which to solve. In inverse KINNs, as the SA is trained to minimize Eq. (4.1), the kinetic parameters are learned from the underlying physical model (MKM). For all studied cases, kinetic parameters, \mathbf{p} , were randomly uniformly initialized at scale of 10^{-2} , whereas ground-truth values would range approximately from $0 \leq \ln(k_0) \leq 8$, as can be inferred from the natural logarithm of the values presented in Table 2. The final refined solution for the $SQ+n$ scenario of dcs -type reaction network, with IC 1 and 2 solved (interpolated) in parallel is shown in Fig. 6. These results are illustrative of cases in which knowledge on surface coverages over time, despite the presence of noise, can provide a description of the system dynamics from which kinetic constants can be retrieved. This highlights the importance of operando spectroscopic techniques, especially for complex reaction networks involving reactions between adsorbed intermediates.

A summary of final refined regression results for all inverse KINNs scenarios is given by Table 5. All scenarios comprise the simultaneous training of SA weights and the natural logarithm of the true kinetic model parameters for ICs 1 and 2. For the purely homogeneous g reaction networks, kinetic parameters are easily retrieved with MAE below 10^{-3} for the Q scenario (in the absence of noise). The inverse KINNs for reaction networks of types g , d and da may also be solved, since they involve reactions between unbound species and corresponding bound molecules, but not reactions between intermediate species, as can be seen from the results for reaction network types g and da . As for reaction networks that include c and g -type reactions, the results support the need for surface composition information over time, as depicted by the detached correlation coefficients and high MAE for type c reactions. In the availability of surface composition information (i.e. signal, intensity), forward and reverse kinetic parameters can be regressed even for the highest-depth reactions, those between surface intermediates (s -type). Notably, the framework presented allows for spectroscopic information about any subset of species to be included, making it possible to quantitatively combine transient operando spectroscopic data with transient kinetic data.

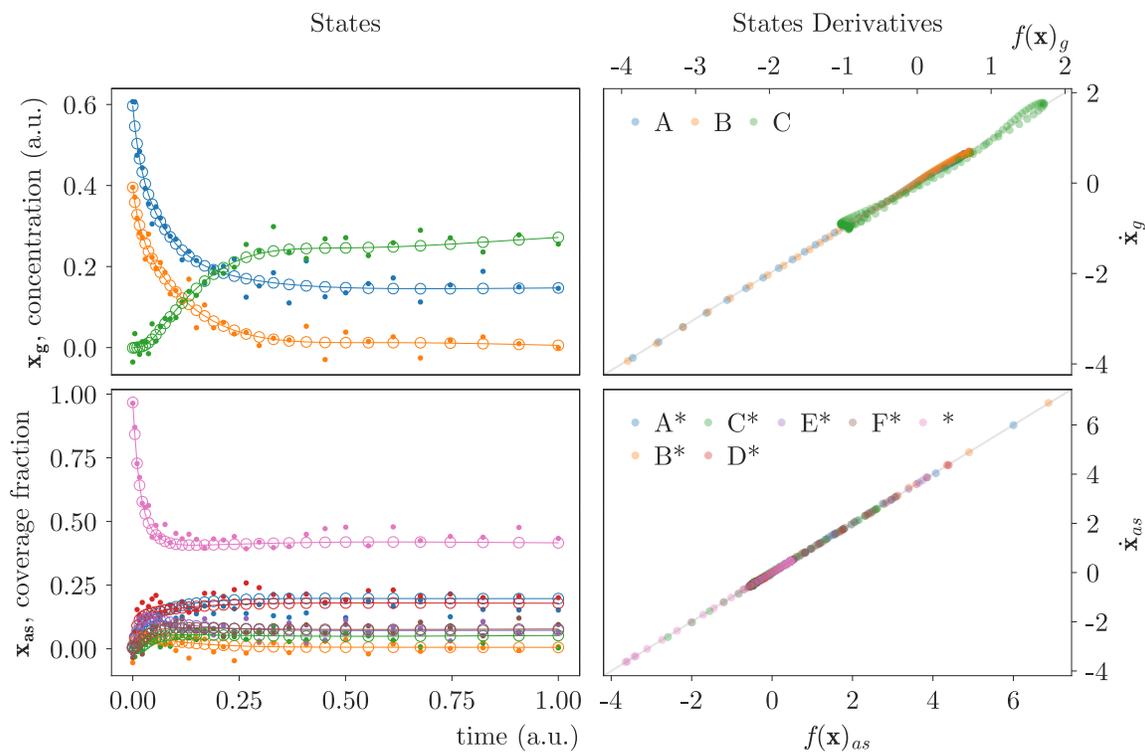


Figure 5: Reaction network type *dcs* inverse solution. States solution (left) (KINN approximation, open circles; *SQ*+ndata, closed markers) under minimization of (4.1) (right) for observable variables (unbound species, top) and latent variables (coverages, bottom).

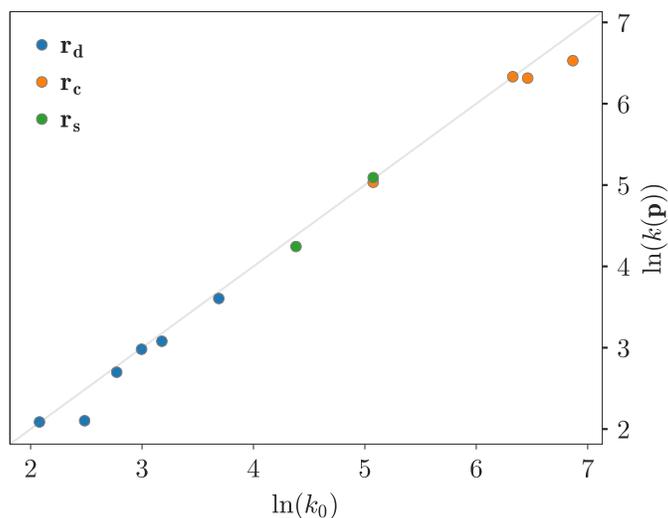


Figure 6: Reaction network type *dcs*, *SQ*+n, inverse solution. Natural logarithms of ground truth parameter $\ln(k_0)$ vs. regressed parameters, $\ln(k(\mathbf{p})) = \mathbf{p}$.

Table 5: Inverse KINNs regression performance metrics summary for type g , da , dc and dcs reactions types and scenarios Q , SQ and $SQ+n$.

Type	Mode	MAE - $\log(k_0)$					total ρ
		g	d	a	c	s	
g	Q	9.90×10^{-4}	-	-	-	-	1.00
	$Q+n$	2.29×10^{-4}	-	-	-	-	1.00
da	Q	-	0.161	9.76×10^{-2}	-	-	0.991
	SQ	-	5.95×10^{-2}	1.16×10^{-2}	-	-	0.999
	$SQ+n$	-	0.150	8.11×10^{-2}	-	-	0.999
dc	Q	-	0.520	-	1.80×10^1	-	-0.523
	SQ	-	1.70×10^{-2}	-	7.32×10^{-2}	-	1.00
	$SQ+n$	-	0.236	-	0.420	-	0.996
dcs	Q	-	5.06	-	1.02×10^1	1.90	0.205
	SQ	-	4.10×10^{-2}	-	0.116	6.61×10^{-2}	0.999
	$SQ+n$	-	0.110	-	0.114	9.62×10^{-2}	0.997

3.3 Discussion of Numerical Methods Aspects

As with any numerical technique, there are numerous factors to consider when setting up and solving problems with KINNs. This section contains a brief discussion of each of these factors, the choices used in this work, and the possibility of other options.

Architecture Three-layer NNs have proven satisfactorily flexible to accommodate the different dynamics between the latent and observable variable domains in the inverse and forward problems in this study. Deeper NNs appear to be particularly important when trigonometric ICs, which encompass DAE constraints, are enforced onto the output of the latent variables in inverse KINNs. Larger or smaller optimized NN architectures and associated activation functions could have produced similar or better results; however, NN architecture optimization is beyond the scope of this work. For non-normalized non-stiff problems, a single *swish* hidden layer suffices to solve the MKM ODEs. For solutions that exhibit stiff behavior in certain regions and flat response in others, the characteristic time can be modelled as a function of the independent variable (e.g., time), which was implemented for all forward KINNs in this work in accordance with the mathematical structure proposed section [2.2.2](#)

Training Given the high number of weights to be optimized from the SA and underlying NNs, and the possibility of non-convexity especially in cases with large number of reaction intermediates (species of type s), we opt for the current standard algorithm for deep NN training, *Adam*, which combines a stochastic gradient descent optimizer with adaptive momentum. Although reliant on a noisy average estimation, SGD has been demonstrated to provide fast convergence and associated parameters, even compared to more elaborate optimizers [\[88\]](#). The RMSProp built onto SGD in *Adam* is known to result in improved performance in on-line settings. Considering these properties and the literature-reported success of *Adam* in NN settings [\[87\]](#), it was chosen as an optimizer in both forward and inverse settings.

Sampling & Regularization As regards the sampling procedure, other strategies could have been utilized instead of sampling evenly spaced, fixed points in the logarithm space. Cross-validation routines in inverse KINN schemes could prevent overfitting for α values beyond the tightening inflection point in Fig. [4](#). Furthermore, pre-training the SA with low α is paramount, especially in the case of noisy data, since it allows the SA to first learn the MKM ODE, which is further presented the observed data to which to interpolate. Such a procedure is also important when there are too few observed data points and large NNs, to prevent the physical model parameters from rapidly squashing at the beginning of training cycle and to avoid the issue of vanishing gradients. Proper scaling of the physical model parameters is also beneficial to accelerate the learning procedure in our current inverse KINNs setup, since it provides the learning (optimization) algorithm *Adam* with parameters of similar order of magnitude.

Inverse Problem In this work, we combine the optimization of the SA and kinetic model parameters into the NN training algorithm. We suggest that the solution for this combined approach may serve as initial guess to accelerate the solution of methods with mathematically rigorous formulations from the standpoint of optimization, such as the adjoint state method [\[89\]](#) that enforces Karush-Kuhn-Tucker, KKT [\[90, 91\]](#) optimality conditions. Such methods may vastly benefit from pre-optimized initial guesses, since their deterministic behavior, especially for the adjoint formulation, often leads to local convergence when deterministic gradient-based optimization algorithms are utilized.

MKMs & KINNs Analysis of the inverse solutions for the different reaction network types examples of varying complexity or depth (section 2.1) highlights the need for operando techniques to allow for information related to species of type c and s to be retrieved. The SQ and $SQ+n$ pose similar scenarios with respect to operando techniques. Coverage fraction time series obtained from the forward numerical solution of the initial value problems are rescaled with their multiplication by a random positive scalar, emulating resolved spectroscopic signals. Inverse KINNs therefore entail learning the kinetic parameters associated with the MKM as well as the scaling factor that maps signal intensities to coverages of intermediate species, which is only possible due to the normalization operator that enforces DAE constraints to the ODE solutions. Accurate parameter estimation for the complex dcs mechanism is supporting evidence that inverse KINNs, seen as either physics-regularized interpolation of data or data-regularized SA-based solution of ODEs, are a robust method for the estimation of kinetic model parameters.

4 Conclusion

In this work, we use artificial neural networks (NNs) as basis functions for microkinetic models (MKMs), proposing strategies to create surrogate approximators (SAs) for the solution of the kinetic ODEs. We present a general classification and notation for catalytic reaction networks based on reaction and species types and the phases of species involved. This notation provides an immediate understanding of the extent or depth of dissociation between the observable or measurable states and the underlying intermediates, and hence the complexity of attempting mechanism elucidation. The suitability of NNs as basis functions for the solution of ordinary differential equations is demonstrated by their ability to solve kinetic forward problems. Structural ICs are imposed on the underlying NNs to enforce ODE ICs as in initial value problems. Furthermore, we also propose the inclusion of ancillary NNs to learn the characteristic timescale of the ODEs, which is important in the case of stiffness about the IC. Normalization of the coverage fractions of intermediates is enforced by separating bulk and surface species into two independent neural networks, which structurally enforces a DAE-type of constraint to the ODE solution.

Using a lumped approach for the simultaneous training of NNs weights and kinetic model parameters in a multiobjective optimization framework, we demonstrate the ability to utilize NNs as basis functions and NN training algorithms to retrieve kinetic parameters from data generated for anecdotal MKM ODEs. The optimization problem is framed in terms of a regularization parameter that can be used to guide the training cycles by (1) low regularization, allowing the SAs to learn or adapt to the physical model ODE, slowly transitioning to (2) high regularization to force the SA to interpolate observed data while parameters of the physical model are simultaneously learned. The regularization parameters convey the relative variance between errors in the physical model (derivatives) space and those associated with states (observed data). This framework can be utilized to estimate Pareto sets and establish associated stopping criteria for training cycles.

Our SA-based approach for the solution of inverse problem was tested against different types of reaction networks. We show that for homogeneous reactions, when the concentration of species over time are known, or for reactions involving non-dissociated adsorbed chemical species, the approach can be utilized to readily retrieve parameters associated with the underlying kinetic model. For more complex scenarios, where there is incomplete knowledge about concentration of intermediate species on the surface of a catalyst, we assessed the possibility of using resolved operando spectroscopic signals for the adsorbed species to extend regression capabilities. Scaling or calibration factors are included as variables in the training process, and we show that our approach can retrieve such factors as well as the kinetic model parameters for these complex systems. Although our analyses are limited to cases where data can be represented as derived from closed, concentrated-parameter models (i.e. models where parameters are spatially uniform), we suggest that SA-based formulation of inverse kinetic ODEs can be applied to transient techniques for the acquisition of kinetic parameters from experimental data, such as temporal analysis of products (TAP) or step-response experiments.

Acknowledgements

The authors acknowledge the U.S. Department of Energy for financial support through contract DE-FE0031719. We are grateful to Dr. John Kitchin and acknowledge his early formative work on the utilization of neural networks for the solution of simple coupled forward kinetics ODEs available via his blog kitchingroup.cheme.cmu.edu/blog. We are also grateful to Dr. Ashi Savara for comments on the pre-print of this manuscript.

Conflict of Interest

The authors declare the nonexistence of conflicts of interest.

References

- [1] Yong-ha Kim, Lydia K. Park, Sotira Yiacoumi, and Costas Tsouris. Modular Chemical Process Intensification: A Review. *Annual Review of Chemical and Biomolecular Engineering*, 8(1):359–380, 6 2017.
- [2] Andrew J. Medford, M. Ross Kunz, Sarah M. Ewing, Tammie Borders, and Rebecca Fushimi. Extracting Knowledge from Data through Catalysis Informatics. *ACS Catalysis*, 8(8):7403–7429, 8 2018.
- [3] Michel Boudart and G. Djega-Mariadassou. *Kinetics of Heterogeneous Catalytic Reactions*. Princeton University Press, Princeton, 1 1984.
- [4] J. A. Dumesic. The Microkinetics of heterogeneous catalysis. *ACS professional reference book*, 1993.
- [5] G Jones, T Bligaard, F Abild-Pedersen, and J K Nørskov. Using scaling relations to understand trends in the catalytic activity of transition metals. *Journal of Physics: Condensed Matter*, 20(6):064239, 2 2008.
- [6] Zachary W. Ulissi, Andrew J. Medford, Thomas Bligaard, and Jens K. Nørskov. To address surface reaction network complexity using scaling relations machine learning and DFT calculations. *Nature Communications*, 8(1):14621, 4 2017.
- [7] Carsten Stegelmann, Anders Andreasen, and Charles T. Campbell. Degree of Rate Control: How Much the Energies of Intermediates and Transition States Control Rates. *Journal of the American Chemical Society*, 131(23):8077–8082, 6 2009.
- [8] Kevin McBride and Kai Sundmacher. Overview of Surrogate Modeling in Chemical Process Engineering, 3 2019.
- [9] Jens K. Nørskov, Frank Abild-Pedersen, Felix Studt, and Thomas Bligaard. Density functional theory in surface chemistry and catalysis, 1 2011.
- [10] Andrew J Medford, Chuan Shi, Max J Hoffmann, Adam C Lausche, Sean R Fitzgibbon, Thomas Bligaard, and Jens K Nørskov. CatMAP: A Software Package for Descriptor-Based Microkinetic Mapping of Catalytic Trends. *Catalysis Letters*, 145(3):794–807, 3 2015.
- [11] James A. Dumesic, George W. Huber, and Michel Boudart. Principles of Heterogeneous Catalysis. In *Handbook of Heterogeneous Catalysis*, pages 29–55. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 3 2008.
- [12] R. B. Getman, W. F. Schneider, A. D. Smeltz, W. N. Delgass, and F. H. Ribeiro. Oxygen-Coverage Effects on Molecular Dissociations at a Pt Metal Surface. *Physical Review Letters*, 102(7):076101, 2 2009.
- [13] Adam C. Lausche, Andrew J. Medford, Tuhin Suvra Khan, Yue Xu, Thomas Bligaard, Frank Abild-Pedersen, Jens K. Nørskov, and Felix Studt. On the effect of coverage-dependent adsorbate–adsorbate interactions for CO methanation on transition metal surfaces. *Journal of Catalysis*, 307:275–282, 11 2013.
- [14] A. B. Mhadeshwar, J. R. Kitchin, M. A. Barteau, and D. G. Vlachos. The role of adsorbate-adsorbate interactions in the rate controlling step and the most abundant reaction intermediate of NH₃ decomposition on Ru. *Catalysis Letters*, 96(1-2):13–22, 7 2004.
- [15] Abhijit Chatterjee and Dionisios G. Vlachos. An overview of spatial microscopic and accelerated kinetic Monte Carlo methods. *Journal of Computer-Aided Materials Design*, 14(2):253–308, 3 2007.
- [16] M. Pineda and M. Stamatakis. Beyond mean-field approximations for accurate and computationally efficient models of on-lattice chemical kinetics. *Journal of Chemical Physics*, 147(2):024105, 7 2017.
- [17] Max J. Hoffmann and Thomas Bligaard. A Lattice Kinetic Monte Carlo Solver for First-Principles Microkinetic Trend Studies. *Journal of Chemical Theory and Computation*, 14(3):1583–1593, 3 2018.
- [18] Mie Andersen, Craig P. Plaisance, and Karsten Reuter. Assessment of mean-field microkinetic models for CO methanation on stepped metal surfaces using accelerated kinetic Monte Carlo. *Journal of Chemical Physics*, 147(15):152705, 10 2017.
- [19] Maarten K. Sabbe, Marie-Françoise Reyniers, and Karsten Reuter. First-principles kinetic modeling in heterogeneous catalysis: an industrial perspective on best-practice, gaps and needs. *Catalysis Science & Technology*, 2(10):2010, 10 2012.
- [20] Sebastian Matera, William F. Schneider, Andreas Heyden, and Aditya Savara. Progress in Accurate Chemical Kinetic Modeling, Simulations, and Parameter Estimation for Heterogeneous Catalysis. *ACS Catalysis*, 9(8):6624–6647, 8 2019.
- [21] Karsten Reuter. Ab Initio Thermodynamics and First-Principles Microkinetics for Surface Catalysis. *Catalysis Letters*, 146(3):541–563, 3 2016.

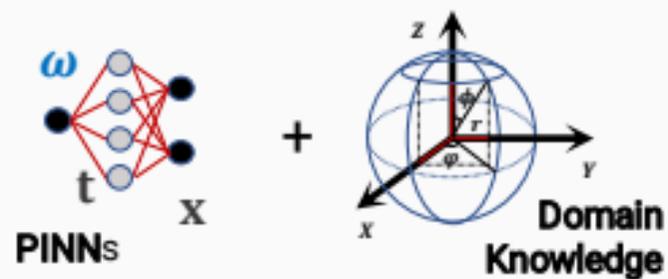
- [22] Yuekang Jin, Guanghui Sun, Zhengming Wang, Haibin Pan, Lingshun Xu, Hong Xu, and Weixin Huang. Elementary surface reactions on Co(0001) under Fischer-Tropsch synthesis conditions. *Journal of Physical Chemistry C*, 121(39):21535–21540, 2017.
- [23] Pei Pei Chen, Jin Xun Liu, and Wei Xue Li. Carbon Monoxide Activation on Cobalt Carbide for Fischer-Tropsch Synthesis from First-Principles Theory. *ACS Catalysis*, pages 8093–8103, 2019.
- [24] Lars C. Grabow, Britt Hvolbæk, and Jens K. Nørskov. Understanding trends in catalytic activity: The effect of adsorbate-adsorbate interactions for Co oxidation over transition metals. *Topics in Catalysis*, 53(5-6):298–310, 2010.
- [25] Xiao Li and Lars C. Grabow. Evaluating the benefits of kinetic Monte Carlo and microkinetic modeling for catalyst design studies in the presence of lateral interactions. *Catalysis Today*, 3 2021.
- [26] Javier Pérez-Ramírez, Rob J. Berger, Guido Mul, Freek Kapteijn, and Jacob A. Moulijn. Six-flow reactor technology a review on fast catalyst screening and kinetic studies. *Catalysis Today*, 60(1):93–109, 7 2000.
- [27] Junmei Wei and Enrique Iglesia. Isotopic and kinetic assessment of the mechanism of reactions of CH₄ with CO₂ or H₂O to form synthesis gas and carbon on nickel catalysts. *Journal of Catalysis*, 224(2):370–383, 6 2004.
- [28] J. C. De Deken, E. F. Devos, and G. F. Froment. STEAM REFORMING OF NATURAL GAS: INTRINSIC KINETICS, DIFFUSIONAL INFLUENCES, AND REACTOR DESIGN. In *ACS Symposium Series*, pages 181–197. ACS, 1982.
- [29] Jianguo Xu and Gilbert F. Froment. Methane steam reforming, methanation and water-gas shift: I. Intrinsic kinetics. *AIChE Journal*, 35(1):88–96, 1 1989.
- [30] Toru Numaguchi and Katsutoshi Kikuchi. INTRINSIC KINETICS AND DESIGN SIMULATION IN A COMPLEX REACTION NETWORK; STEAM-METHANE REFORMING. In *Tenth International Symposium on Chemical Reaction Engineering*, pages 2295–2301. Elsevier, 1 1988.
- [31] Srinivas Rangarajan, Aditya Bhan, and Prodromos Daoutidis. Language-oriented rule-based reaction network generation and analysis: Description of RING. *Computers & Chemical Engineering*, 45:114–123, 10 2012.
- [32] Masayoshi Kobayashi and Haruo Kobayashi. Application of transient response method to the study of heterogeneous catalysis. II. Mechanism of catalytic oxidation of carbon monoxide on manganese dioxide. *Journal of Catalysis*, 27(1):108–113, 1972.
- [33] Masayoshi Kobayashi and Haruo Kobayashi. Application of transient response method to the study of heterogeneous catalysis. III. Simulation of carbon monoxide oxidation under an unsteady state. *Journal of Catalysis*, 27(1):114–119, 1972.
- [34] Masayoshi Kobayashi and Haruo Kobayashi. Application of transient response method to the study of heterogeneous catalysis. I. Nature of catalytically active oxygen on manganese dioxide for the oxidation of carbon monoxide at low temperatures. *Journal of Catalysis*, 27(1):100–107, 1972.
- [35] P. Biloen. Transient kinetic methods. *Journal of Molecular Catalysis*, 21(1-3):17–24, 10 1983.
- [36] Carroll O. Bennett. The Transient Method and Elementary Steps in Heterogeneous Catalysis. *Catalysis Reviews*, 13(1):121–148, 1 1976.
- [37] J. T. Gleaves, J. R. Ebner, and T. C. Kuechler. Temporal Analysis of Products (TAP) — A Unique Catalyst Evaluation System with Submillisecond Time Resolution. *Catalysis Reviews*, 30(1):49–116, 2 1988.
- [38] John T. Gleaves, Gregory Yablonsky, Xiaolin Zheng, Rebecca Fushimi, and Patrick L. Mills. Temporal analysis of products (TAP)-Recent advances in technology for kinetic analysis of multi-component catalysts. *Journal of Molecular Catalysis A: Chemical*, 315(2):108–134, 1 2010.
- [39] K. Morgan, N. Maguire, R. Fushimi, J. T. Gleaves, A. Goguet, M. P. Harold, E. V. Kondratenko, U. Menon, Y. Schuurman, and G. S. Yablonsky. Forty years of temporal analysis of products, 2017.
- [40] Robert J. Madix. The application of flash desorption spectroscopy to chemical reactions on surfaces: Temperature programmed reaction spectroscopy. <http://dx.doi.org/10.1080/10408437808243436>, 7(2):143–152, 2006.
- [41] Christian Reece, Mathilde Luneau, and Robert J. Madix. Dissecting the Performance of Nanoporous Gold Catalysts for Oxygen-Assisted Coupling of Methanol with Fundamental Mechanistic and Kinetic Information. *ACS Catalysis*, 9(5):4477–4487, 5 2019.
- [42] Christian Reece and Robert J. Madix. Moving from Fundamental Knowledge of Kinetics and Mechanisms on Surfaces to Prediction of Catalyst Performance in Reactors. *ACS Catalysis*, pages 3048–3066, 2021.
- [43] Aditya Savara. Simulation and fitting of complex reaction network TPR: The key is the objective function. *Surface Science*, 653:169–180, 11 2016.

- [44] J. Happel. Transient tracing. *Chemical Engineering Science*, 33(11):1567, 1978.
- [45] Cristian Ledesma, Jia Yang, De Chen, and Anders Holmen. Recent approaches in mechanistic and kinetic studies of catalytic reactions using SSITKA technique, 12 2014.
- [46] Schohn L. Shannon and James G. Goodwin. Characterization of Catalytic Surfaces by Isotopic-Transient Kinetics during Steady-State Reaction. *Chemical Reviews*, 95(3):677–695, 1995.
- [47] Rob J. Berger, Freek Kapteijn, Jacob A. Moulijn, Guy B. Marin, Juray De Wilde, Maria Olea, De Chen, Anders Holmen, Luca Lietti, Enrico Tronconi, and Yves Schuurman. Dynamic methods for catalytic kinetics, 6 2008.
- [48] Preeti Aghalayam, Young K. Park, and Dionisios G. Vlachos. Construction and optimization of complex surface-reaction mechanisms. *AIChE Journal*, 46(10):2017–2029, 2000.
- [49] P. Aghalayam, Y. K. Park, and D. G. Vlachos. A detailed surface reaction mechanism for CO oxidation on Pt. *Proceedings of the Combustion Institute*, 28(1):1331–1339, 1 2000.
- [50] Patricia Rubert-Nason, Manos Mavrikakis, Christos T. Maravelias, Lars C. Grabow, and Lorenz T. Biegler. Advanced solution methods for microkinetic models of catalytic reactions: A methanol synthesis case study. *AIChE Journal*, 60(4):1336–1346, 4 2014.
- [51] J. M. Caruthers, J. A. Lauterbach, K. T. Thomson, V. Venkatasubramanian, C. M. Snively, A. Bhan, S. Katare, and G. Oskarsdottir. Catalyst design: Knowledge extraction from high-throughput experimentation. *Journal of Catalysis*, 216(1-2):98–109, 5 2003.
- [52] Jonas Sjöblom and Derek Creaser. New approach for microkinetic mean-field modelling using latent variables. *Computers & Chemical Engineering*, 31(4):307–317, 2 2007.
- [53] A. B. Mhadeshwar, H. Wang, and D. G. Vlachos. Thermodynamic Consistency in Microkinetic Development of Surface Reaction Mechanisms. *Journal of Physical Chemistry B*, 107(46):12721–12733, 11 2003.
- [54] Adam Yonge, M. Ross Kunz, Rakesh Batchu, Zongtang Fang, Tobin Issac, Rebecca Fushimi, and Andrew J. Medford. TAPsolver: A Python package for the simulation and analysis of TAP reactor experiments. *Chemical Engineering Journal*, page 129377, 4 2021.
- [55] Henrik Topsøe. Developments in operando studies and in situ characterization of heterogeneous catalysts. *Journal of Catalysis*, 216(1-2):155–164, 5 2003.
- [56] Thomas Bligaard, R. Morris Bullock, Charles T. Campbell, Jingguang G. Chen, Bruce C. Gates, Raymond J. Gorte, Christopher W. Jones, William D. Jones, John R. Kitchin, and Susannah L. Scott. Toward Benchmarking in Catalysis Science: Best Practices, Challenges, and Opportunities. *ACS Catalysis*, 6(4):2590–2602, 4 2016.
- [57] Takashi Toyao, Zen Maeno, Satoru Takakusagi, Takashi Kamachi, Ichigaku Takigawa, and Ken-ichi Ichi Shimizu. Machine Learning for Catalysis Informatics: Recent Applications and Prospects. *ACS Catalysis*, 10(3):2260–2297, 2 2020.
- [58] Saurabh Bhandari, Srinivas Rangarajan, and Manos Mavrikakis. Combining Computational Modeling with Reaction Kinetics Experiments for Elucidating the In Situ Nature of the Active Site in Catalysis. *Accounts of Chemical Research*, 53(9):1893–1904, 9 2020.
- [59] Saurabh Bhandari, Srinivas Rangarajan, Christos T. Maravelias, James A. Dumesic, and Manos Mavrikakis. Reaction Mechanism of Vapor-Phase Formic Acid Decomposition over Platinum Catalysts: DFT, Reaction Kinetics Experiments, and Microkinetic Modeling. *ACS Catalysis*, 10(7):4112–4126, 4 2020.
- [60] Srinivas Rangarajan, Christos T. Maravelias, and Manos Mavrikakis. Sequential-Optimization-Based Framework for Robust Modeling and Design of Heterogeneous Catalytic Systems. *Journal of Physical Chemistry C*, 121(46):25847–25863, 11 2017.
- [61] Jeffrey A. Herron, Manos Mavrikakis, and Christos T. Maravelias. *Optimization Methods for Catalyst Design*, volume 38. Elsevier Masson SAS, 2016.
- [62] Alexander I.J. Forrester and Andy J. Keane. Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45(1-3):50–79, 1 2009.
- [63] Sushant Suhas Garud, I.A. Karimi, and Markus Kraft. Smart Sampling Algorithm for Surrogate Model Development. *Computers & Chemical Engineering*, 96:103–114, 1 2017.
- [64] ROBERT HECHT-NIELSEN. Theory of the Backpropagation Neural Network. In *Neural Networks for Perception*, pages 65–93. Elsevier, 1 1992.
- [65] Andreas Griewank and Andrea Walther. *Evaluating Derivatives Principles and Techniques of Algorithmic Differentiation Second Edition*.

- [66] I.E. Lagaris, Aristidis Likas, and D.I. Fotiadis. Artificial neural networks for solving ordinary and partial differential equations. *IEEE Transactions on Neural Networks*, 9(5):987–1000, 5 1998.
- [67] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.
- [68] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics Informed Deep Learning (Part II): Data-driven Discovery of Nonlinear Partial Differential Equations. *arXiv*, (Part II):1–19, 2017.
- [69] Yiping Lu, Aoxiao Zhong, Quanzheng Li, and Bin Dong. Beyond Finite Layer Neural Networks: Bridging Deep Architectures and Numerical Differential Equations. *35th International Conference on Machine Learning, ICML 2018*, 7:5181–5190, 10 2017.
- [70] Eldad Haber and Lars Ruthotto. Stable Architectures for Deep Neural Networks. *Inverse Problems*, 34(1), 5 2017.
- [71] Lars Ruthotto and Eldad Haber. Deep Neural Networks Motivated by Partial Differential Equations. *Journal of Mathematical Imaging and Vision*, 62(3):352–364, 4 2018.
- [72] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural Ordinary Differential Equations. *NIPS*, 109(NeurIPS):31–60, 6 2018.
- [73] Benny Avelin and Kaj Nyström. Neural ODEs as the deep limit of ResNets with constant weights. *Analysis and Applications*, 2020.
- [74] Per Stoltze. Microkinetic simulation of catalytic reactions. *Progress in Surface Science*, 65(3-4):65–150, 10 2000.
- [75] Gabriel S. Gusmão and Phillip Christopher. A general and robust approach for defining and solving microkinetic catalytic systems. *AIChE Journal*, 61(1):188–199, 1 2015.
- [76] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2(4):303–314, 12 1989.
- [77] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Networks*, 3(5):551–560, 1 1990.
- [78] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1991.
- [79] Moshe Leshno, Vladimir Ya Lin, Allan Pinkus, and Shimon Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6(6):861–867, 1 1993.
- [80] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Networks*, 3(5):551–560, 1 1990.
- [81] A.J. Meade and A.A. Fernandez. The numerical solution of linear ordinary differential equations by feedforward neural networks. *Mathematical and Computer Modelling*, 19(12):1–25, 6 1994.
- [82] A.J. Meade and A.A. Fernandez. Solution of nonlinear ordinary differential equations by feedforward neural networks. *Mathematical and Computer Modelling*, 20(9):19–44, 11 1994.
- [83] Roy Frostig, Matthew Johnson, and Chris Leary. Compiling machine learning programs via high-level tracing. In *SysML*, 3 2018.
- [84] James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, and Skye Wanderman-Milne. {JAX}: composable transformations of {P}ython+{N}um{P}y programs. <http://github.com/google/jax>, 2018.
- [85] Alan C. Hindmarsh. LSODE and LSODI, two new initial value ordinary differential equation solvers. *ACM SIGNUM Newsletter*, 15(4):10–11, 12 1980.
- [86] Linda Petzold. Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations. *SIAM Journal on Scientific and Statistical Computing*, 4(1):136–148, 3 1983.
- [87] Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. International Conference on Learning Representations, ICLR, 12 2015.
- [88] Yann Lecun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 5 2015.
- [89] Ibrahim Ayed, Emmanuel De Bézenac, Arthur Pajot, Julien Brajard, and Patrick Gallinari. Learning Dynamical Systems from Partial Observations, 2019.
- [90] W Karush. *Minima of Functions of Several Variables with Inequalities as Side Conditions*. University of Chicago, Department of Mathematics, 1939.

- [91] H W Kuhn and A W Tucker. Nonlinear Programming. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 481–492, Berkeley, Calif., 1951. University of California Press.

KINNs



$$\min_{\omega, p} \left\| \dot{x}(\omega, t) - M \times p \circ \phi(x(\omega, t)) \right\|_2^2$$

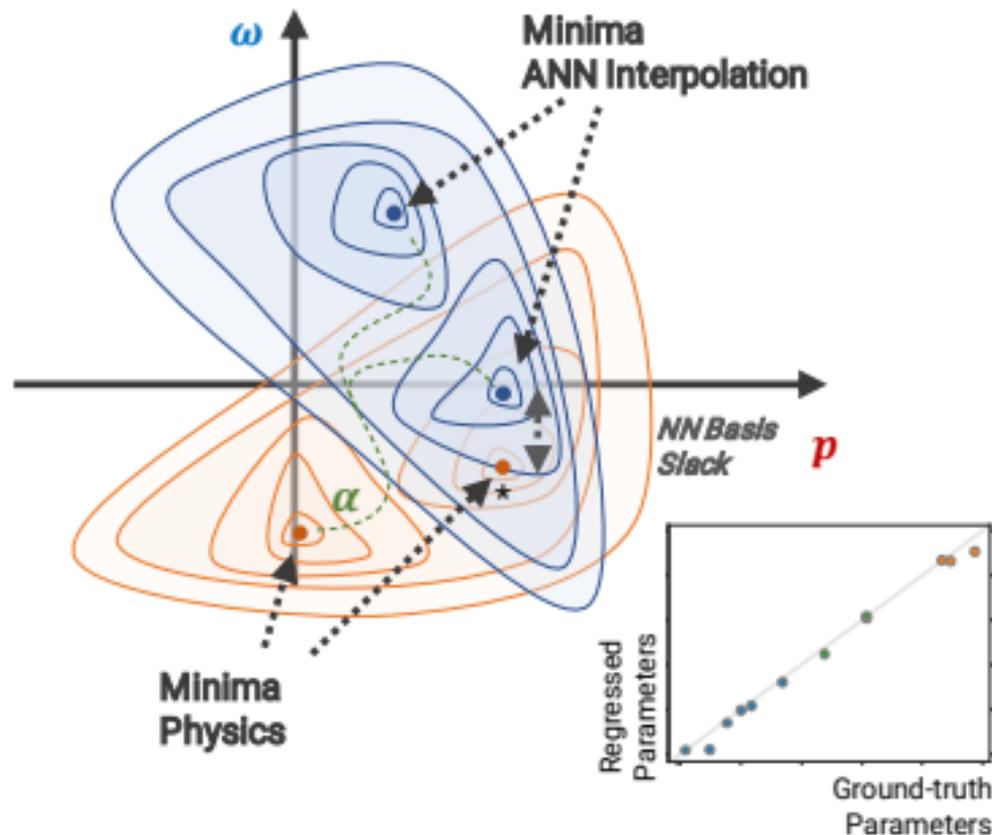
KINETICS (PHYSICS)

$$+ \alpha \left\| x(\omega, t) - \tilde{x}(t) \right\|_2^2$$

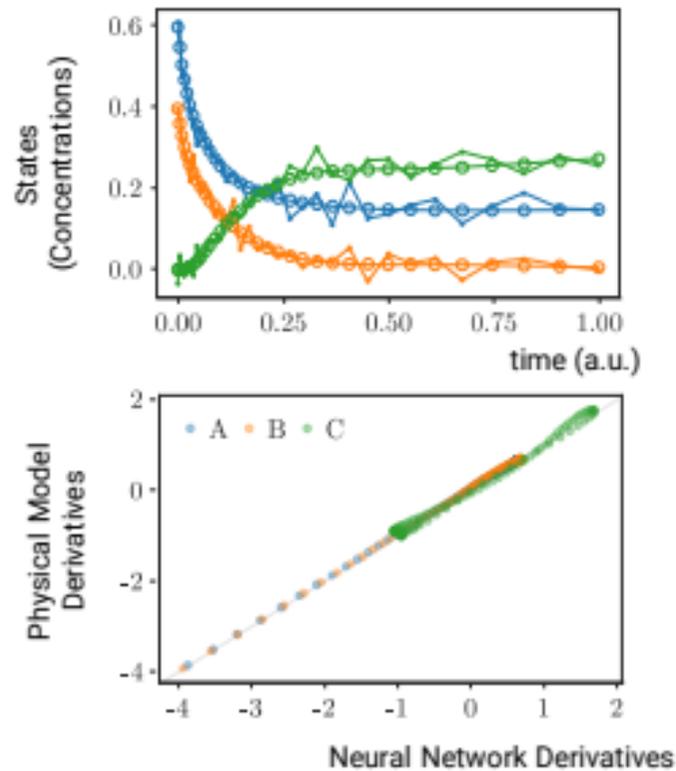
Cost Function

INTERPOLATION

Physics-regularized Interpolation or Data-regularized ODE solution



INTERPOLATION



KINETICS (PHYSICS)