Chemical Engineering Journal 462 (2023) 142089

Contents lists available at ScienceDirect

Chemical Engineering Journal

journal homepage: www.elsevier.com/locate/cej

A dynamic nonlinear optimization framework for learning data-driven reduced-order microkinetic models

Fernando Lejarza [a], Elsa Koninckx [c], Linda J. Broadbelt [c,*], Michael Baldea [a,b,**]

[a] The University of Texas at Austin, McKetta Department of Chemical Engineering, 200 E Dean Keeton St, Austin, TX 78712-1589, USA
[b] The University of Texas at Austin, Oden Institute for Computational Engineering and Sciences, 201 E 24th St, Austin, TX 78712-1589, USA
[c] Northwestern University, Department of Chemical and Biological Engineering, 2145 Sheridan Road, Evanston, IL 60208, USA

ARTICLE INFO

Keywords:
Microkinetic modeling
Governing equations discovery
Machine learning
Reaction mechanism

ABSTRACT

The use of kinetic models is key for analyzing, designing, controlling, and optimizing the manufacturing of high value chemicals. In particular, microkinetic modeling relies on large-scale networks of elementary reaction steps that map the conversion of feed materials to final products by considering a large number of pathways and intermediate species for a given reactor configuration. Intuitively, several of the (often automatically generated) reaction steps might be redundant or insignificant, and thus determining the true governing reaction network is critical to understanding and modeling the underlying chemistry. This work introduces a nonlinear dynamic optimization framework for discovering governing reaction networks from data, whereby both the model structure (the elementary reaction steps) and the model parameters (reaction rate constants, pre-exponential factors, and activation energies) are simultaneously learned from composition time series data. The proposed framework can also achieve dimensionality reduction to produce accurate reduced-order microkinetic models that are computationally parsimonious and thus better suited for applications involving simulation and optimization. Two numerical examples of different dimensions are presented to illustrate the key properties of our approach.

1. Introduction

Mathematical modeling of chemical reactions is an essential computational tool that enables optimal decision-making in chemical engineering and industrial chemistry, ranging from process design to production planning, scheduling and control. A large fraction of kinetic studies focus on *macrokinetics* [1], an approach that involves using macroscopic quantities such as inlet/initial composition, pressure, catalyst properties, and temperature to describe the evolution of a reacting system and to predict quantities of interest such as outlet/final composition. While macrokinetic models are valuable (e.g., in applications relating to reactor design and control), their accuracy and the physical insight they provide are often limited by several factors, such as assumptions relating to species in equilibrium, rate-determining steps, and intermediate species. Based on these presupposed structural properties, regression techniques are then used to fit experimental data and estimate the values of the relevant rate constants, reaction orders, and activation energies [2–5]. Considering the often complex nature of the true (and likely unknown) underlying reaction mechanism, macrokinetic models offer a simpler representation that is more computationally convenient for simulation and optimization purposes. Previous examples of macrokinetic modeling employ lumped species models and related "black-box" representations to describe the myriad of reactions involving complex natural feedstocks, such as for crude oil hydrocracking [6,7] and biomass pyrolysis [8,9].

On the other hand, approaches based on *microkinetic* (MK) modeling consider a large set of possible elementary reactions involving the initial components, generally with minimal assumptions regarding the equilibrium rates and the rate-determining steps, in order to understand the true underlying reaction mechanism and more accurately predict the distribution of product species [1]. Rate constants and related parameters are estimated from first principles calculations (e.g., using established relationships such as transition state theory), and in turn, insights regarding rate-determining reactions, concentrations or surface coverage of intermediates, reaction orders, and apparent activation

* Corresponding author.
** Corresponding author at: The University of Texas at Austin, McKetta Department of Chemical Engineering, 200 E Dean Keeton St, Austin, TX 78712-1589, USA.
E-mail addresses: lejarza@utexas.edu (F. Lejarza), ekoninckx@u.northwestern.edu (E. Koninckx), broadbelt@northwestern.edu (L.J. Broadbelt), mbaldea@che.utexas.edu (M. Baldea).

https://doi.org/10.1016/j.cej.2023.142089
Received 17 November 2022; Received in revised form 1 February 2023; Accepted 20 February 2023
Available online 28 February 2023

energies are obtained as a result of microkinetic analysis [5]. Thus, microkinetic analysis allows for identifying the governing reaction mechanism, which is defined as the parsimonious set of elementary reaction steps that fully define the evolution of species concentrations over an independent domain of interest (time or space, depending on the reactor configuration). Relative to the previously discussed macrokinetic expressions, such governing mechanistic knowledge is expected to provide improved extrapolation ability (e.g. at different operating conditions, where the assumptions invoked in macrokinetic modeling might not hold), and more accurate predictions [1], and can be highly informative in guiding the design and development of novel catalytic materials [5]. However, the increased resolution of MK models comes at the cost of higher dimensionality (that is, a greater number of variables and equations are needed to describe the system) and thus greater computational complexity for the aforementioned prediction and process optimization tasks. Furthermore, the equations may involve highly nonlinear expressions with respect to the species concentrations (e.g., when there is competition at the catalyst surface), as well as with respect to the model parameters and reaction conditions (e.g., to capture the exponential relationship between rate constants and temperature). The magnitudes of the rate constants of the different elementary reaction steps are also likely to be vastly different depending on the reaction conditions, resulting in complex multi-scale models, leading to stiffness and related numerical difficulties. Additionally, the presence of intermediate species that are typically not measurable (i.e., whose composition cannot be resolved experimentally) and are not observable (i.e., compositions for unmeasurable species cannot be estimated from available composition measurements) in the reaction networks associated with microkinetic models render the data-driven identification of governing mechanisms a difficult task [10,11]. These challenges call for new modeling and computational techniques for discovering and reducing the complexity of microkinetic models, without directly invoking restrictive or potentially erroneous assumptions (regarding e.g., equilibrium, rate-determining steps, abundant intermediates) as in macrokinetic modeling, that can significantly worsen model resolution and predictive performance.

Physics-informed machine learning and optimization methods provide a promising platform to develop new data-driven computational modeling techniques for microkinetic phenomena, by combining kinetic knowledge with data obtained from experimentation or simulation. To this end, this paper introduces a machine learning framework based on nonlinear dynamic optimization that addresses two critical goals:

- **Goal 1**: Given a large set of candidate elementary reaction steps and experimental or simulated concentration data, *find the governing microkinetic model* by simultaneously estimating the associated kinetic parameters (e.g., rate constants, activation energies, pre-exponential factors) and systematically eliminating insignificant/redundant elementary reaction steps.
- **Goal 2**: Given a governing microkinetic model and experimental/simulated concentration data for measurable species, *reduce model complexity* by finding a smaller *approximate* subset of elementary reactions and associated kinetic parameters that capture the dynamics underpinning the data to a desired level of accuracy.

The proposed approach begins by using automatic network generation and first principles arguments to build a large library of candidate reaction steps (including associated kinetic expressions), from which a subset of elementary reaction steps can be selected to best explain the available data. An efficient sequential elimination heuristic scheme is proposed to address the high-dimensional nonlinear optimization problem associated with the model discovery formulation. The proposed heuristic consists of evaluating the predictive performance of the candidate elementary reaction steps at explaining the measurable system behavior (as reflected by the data), together with a regularization function to sequentially eliminate elementary reaction steps and balance the predictive ability and complexity of the resulting model.

While some physical insight and model resolution (e.g., the ability to accurately predict the distribution of intermediate species) are typically lost when performing model reduction, the resulting lower-order models are valuable in practical settings for industrial applications such as reactor design, optimization and control. The underlying optimization problem is a weak dynamic nonlinear program (DNLP) that can cope with: (i) the nonlinear parameterizations that often arise in the rate expressions (e.g., we consider Arrhenius-type temperature dependence of the rate constants), (ii) the multi-scale nature of the resulting dynamical system (i.e., the set of rate equations used to predict concentration over time or space), (iii) species that are present in the candidate elementary reaction steps but whose concentrations are not measurable. The constrained nature of the underlying optimization problem naturally allows for including lower and upper bounds on the estimated parameters (e.g., rate constants, activation energies, and pre-exponential factors) to account for additional available first principles knowledge and to improve the convergence of the solution. More importantly, such first principles information can be leveraged to ensure that the discovered equations satisfy a certain model hierarchy when dealing with multi-scale systems (i.e., constraints can be based on atomistic concepts and transition state theory, kinetic models, diffusion and transport equations, as well as reactor design equations). Similarly, the parameters of the model (at the different length and time scales) can be bounded from below and from above to impose hierarchy with respect to fast and slow reaction dynamics and distribution of species at equilibrium.

The remainder of the paper is organized as follows. Section 2 provides an overview of the literature pertaining to microkinetic modeling and related applications of machine learning techniques, presenting existing contributions and highlighting their benefits and shortcomings. Section 3 describes the key fundamental principles employed in the derivation of microkinetic models. Section 4 outlines the governing equation discovery problem, and introduces previous efforts based on sparse regression to set the context for the present work. Section 5 describes the main contribution by introducing the nonlinear optimization problem and heuristic solution algorithm used for learning microkinetic models from data and for performing dimensionality reduction. Section 6 discusses extensive numerical results for two benchmark reaction network examples that are used to demonstrate the key properties and applicability of the proposed approach. Lastly, Section 7 provides concluding remarks and outlines important directions of future research.

2. Literature review

Microkinetic modeling can generally be thought of as a two-step process: (i) reaction network generation, whereby a large set of possible elementary reactions mapping the conversion of feed materials into final products is defined, and (ii) the solution of a system of reaction rate and reactor design equations by which (given initial conditions and a reactor configuration) the trajectories of species concentrations in time or space can be numerically estimated, if reasonable parameter values are available for the reactions defined in (i). The first step, network generation, refers to producing all possible elementary reaction steps of a candidate reaction mechanism, noting that some species may not be present in meaningful amounts, and some reactions may have insignificant rates. The resulting set of reactions can be prohibitively large (on the order of thousands of species and reactions). While initial efforts in modeling microkinetic phenomena relied on manual derivation of the aforementioned set of elementary reactions, several computational tools for automatic network generation have been developed more recently [12]. Automatic network generation schemes allow for a systematic computational exploration of all combinations of

species and reaction types to ensure that as many elementary reactions as possible are considered, without making any assumptions regarding a rate-determining step or pathway. For hydrocarbon processing, for example, the typical network generation framework begins by defining the number of carbon atoms in the starting material (e.g., C_3), and involves some user decisions regarding the complexity of the model, such as specifying the largest number of carbon atoms considered in the products C_n, and the degree of branching and/or cycle formation. A comprehensive review and discussion of automatic network generation software and algorithms can be found elsewhere [12].

The set of elementary reaction steps (and associated rate constant expressions) is then combined with reactor design equations (the material and energy balances for the reactor system), which reflect the experimental setup where data are collected. Reaction rate constants are typically defined using established expressions, such as the Arrhenius temperature dependence as in prior works e.g. [13,14], for which transition-state theory, the Evans–Polanyi relation [15], heats of reaction, etc. can be used to derive the associated kinetic parameters (e.g., pre-exponential factors and activation energies). In this case, domain knowledge plays a critical role in determining which expressions to employ, as they might vary significantly across fields (e.g., the reaction rate constant expressions that govern hydrocarbon oligomerization on solid-state catalysts are likely different from those used to model the kinetics of gene expression for a disease).

The candidate microkinetic model is then defined by the resulting system of differential and algebraic equations (DAEs), whose solution determines the evolution of the species concentrations for a given span of the independent variable(s), such as time or space depending on the reactor configuration under consideration. Along with the aforementioned automatic network generation software, numerous computational tools have been developed to solve the DAE systems representing the microkinetic models [16–18].

The application range of microkinetic models is extensive, including catalytic reactions [13,14,19], pyrolysis reactions [20–22], polymerization reactions [23,24], and biological and metabolic networks [25,26]. Modern automatic network generation frameworks can produce large-scale candidate MK models with thousands of molecular species and interconnected by thousands of elementary reaction steps. Simulation can be computationally challenging owing to the large size and stiffness of the corresponding DAE model. An even more difficult problem arises when attempting to use experimental or simulation data to perform parameter estimation (that is, finding the optimal values of parameters such as pre-exponential factors and activation energies), which requires solving a large-scale nonlinear, likely non-convex, optimization problem. Furthermore, leveraging these complex reaction network models directly for applications in reactor and/or process design, optimization and control is most likely computationally intractable, particularly for applications that require real-time and high-frequency decision-making (e.g., model predictive control [27]).

Hence, the complex nature of MK models calls for approaches to decrease the computational expense, such as lumping species [28], whereby complex chemistries are represented using smaller sets of reaction families based on, for example, the carbon number, branching degree and ion position, while still maintaining high overall product specificity for both gaseous and adsorbed species. MK model reduction [29] can also be performed by removing reactions and species that are deemed unimportant during the network generation phase based on user intuition or any available empirical evidence, and only preserving reaction intermediates that best direct the MK model to predict the observable species. A related model reduction approach leverages genetic algorithms to reduce the size of kinetic models [30] by means of pruning an initial set of reactions which, given the combinatorial nature of genetic algorithms, may not be scalable to larger networks of reactions.

The application of data science and machine learning methods to deriving a fundamental understanding of dynamical systems based on data has witnessed a rapid growth over the past decades. One relevant machine learning problem is *surrogate modeling*, often referred to as *system identification* when applied to dynamical systems and time series data, in which a model of arbitrary (but known and typically conveniently chosen) mathematical structure is learned via statistical methods and is subsequently employed to predict system outputs (e.g., concentration at a given time t_i) given system inputs (e.g., concentrations at previous times $t_j < t_i$) [31]. Model formats range from linear state space models, to decision tree ensembles, to recurrent neural network architectures; the training data sets used typically consist of time series information regarding composition of species that can be experimentally measured or simulated based on some a priori available model. The sole focus of surrogate modeling is to maximize predictive ability relative to a given error metric, while regularization techniques might be employed to prevent overfitting the training data (i.e., avoiding outcomes whereby the learned function aligns too closely to the training data resulting in poor generalization performance to "unseen" testing data). Nevertheless, surrogate models are often "black-box" in nature and do not typically provide physical insight into the underlying system dynamics. Hence, the use of surrogate models is often restricted to predictions within the domain/range of the training data, and employing them for extrapolation tasks should be avoided. Moreover, the often complex and highly non-convex model structure generally hinders their implementation in optimization and model-based control applications.

On the other hand, the task referred to as *parameter estimation*, which has been more commonly employed for building models of microkinetic phenomena (e.g., [28]), is based on the availability of a model structure that is rooted in physical and chemical arguments (rather than selected based on mathematical convenience as is the case with surrogate modeling), and training data are used to infer values for the model parameters (e.g., determining pre-exponential factors and activation energies for a known reaction network). The parameter estimation problem has been the main focus of applying machine learning to microkinetic modeling in the literature [32], aiming to improve the accuracy in estimating the enthalpy and entropy of adsorbed species [33], as well as to quantify model uncertainty and parameter sensitivity [34]. Similarly, machine learning can be used to correct for the mean-field approximation (by which the adsorbed species are homogeneously distributed on the catalyst surface and is inherent to most MK models), by using neural networks to learn instantaneous reaction rates from lattice Monte Carlo simulation samples for different surface coverage values [35]. Numerous optimization-based approaches have been proposed for kinetic parameter estimation, considering a variety of data types and sources, as well as employing different modeling paradigms [36–39]. Recently, a reduced-order kinetic (ROK) modeling framework for the catalytic oligomerization of propylene [13] was proposed in [40], where the structures of six different a priori established ROK models were postulated by lumping species based on their number of carbon atoms and using different simplifying assumptions regarding the underlying reaction kinetic parameters. The proposed ROK models were then evaluated for a reactor design optimization problem.

A related but fundamentally different problem in machine learning corresponds to the *discovery of governing equations* from data [41–44], which involves simultaneously finding **both** the model structure **and** the parameters that best explain the evolution of a system's state (i.e., variables of interest of a dynamical system, such as concentration or partial pressure) in relation to the available data.

Definition 1. Governing equations are a minimal-order mathematical representation of the behavior of the states of a system as a function of initial and boundary conditions, and potentially a set of independent variables (e.g., externally manipulated inputs used to control or excite the system). These expressions, typically in the form of ordinary or partial differential equations, are conventionally obtained from rigorous first principles analysis, such as the conservation laws

or knowledge-based phenomenological derivation, to reflect microscopic and macroscopic closures of theoretical (e.g., atomistic, kinetic) arguments.

In the context of microkinetic dynamics, and in contrast to the aforementioned parameter estimation methods, the governing equation discovery task consists of finding the minimal (i.e., smallest) set of elementary reaction steps and associated kinetic parameters that fully describe the evolution of all the (measured and unmeasured) chemical species in the system. Recent works focusing on discovering governing equations from data for a general dynamical system [42,43] are based on defining an initially large library of candidate nonlinear basis functions (this would correspond to a large set of possible elementary reaction steps for the current application to MK model discovery) with the expectation that only a small subset thereof is present in the governing equations; thus, the dynamics are said to be sparse with respect to the starting library of basis functions, expressions or reactions. Sparse identification of nonlinear dynamics (SINDy) [42] relies on sparse (nonlinear) regression, such as sequentially-thresholded least squares regression or regularized regression, to determine the few basis functions and associated coefficients that best predict the system state derivatives, which are estimated from available measurements. An extension, referred to as reactive SINDy [45], was tailored for learning reaction networks from constant-temperature, non-equilibrium concentration data by determining the isothermal rate constants of the active reactions. An inherent limitation is the assumption that the dynamics of the system can be represented by a *linear* combination of the basis functions or elementary reactions, without being able to cope with basis functions that are nonlinear in their parameters as is commonly the case in reaction rate expressions (e.g., Arrhenius-type). Furthermore, past works [42,45] rely on directly regressing the numerically estimated state derivatives with respect to the reaction rates evaluated for the available state measurements, which requires that *all* states be measurable, and cannot be intuitively applied to circumstances where some states are unmeasured, which is almost always the case when modeling systems of microkinetic reactions for which the concentration of intermediate states cannot typically be measured. More recently, [43] introduced a more flexible *weak* nonlinear mathematical programming approach that can cope with such nonlinearities and employs implicit discretization methods to handle (i) stiff system dynamics as the ones that are expected to arise from MK reaction networks, as well as (ii) system states that are unmeasured. The framework in [43] was subsequently implemented for discovering governing equations for non-isothermal chemical reactors including both batch [44] and continuous [46] systems, though only a simple first-order reaction was considered.

The present work leverages the results developed in [43,46] to introduce a nonlinear dynamic optimization approach for discovering governing microkinetic models starting from an initial (automatically generated) reaction network and data regarding species concentration in time/space. That is, *both* the structure *and* optimal coefficients are learned. Addressing Goals 1 and 2 introduced in the previous section, the present framework can be employed to (i) determine the governing microkinetic mechanism given a large initial library of candidate elementary reaction steps, and (ii) to reduce the order of an existing MK model and identify a more parsimonious *approximate* reaction network model of arbitrary accuracy that is better suited for optimization and control applications.

Definition 2. A reduced-order model (ROM) is a lower-dimensional approximation of the minimal order governing equations introduced in Definition 1. The number of differential variables and thus equations is lower than that of the governing model, which implicitly means that some species and/or reactions are, e.g., eliminated or lumped. The response of the reduced-order model is "close" (based on a metric of choice), but not identical to, that of the governing equations. Note that the ROM might not be unique based on the metric of choice, as well as on the structural properties of the problem.

The merit of such ROMs lies in superior numerical properties (e.g., lower condition number/less stiffness, faster solution times) relative to the governing equation model. Reduced-order models can be used to accelerate and robustify both online calculations (e.g., process control, process optimization) and off-line calculations (e.g., reactor design, process design) and therefore play an important role in the deployment of microkinetic models in practical settings. While ROMs are built to maintain good predictive performance for measured species, their approximate nature is expected to result in losing some of the mechanistic insight provided by the higher-order governing equations, as well as in poorer extrapolation capabilities for vastly different operating regimes. Therefore, the use and deployment of ROMs in practical applications should be restricted to the operating regimes that are reflected in the training data, unless the predictive (extrapolation) performance can be reliably validated for a wider range of operating conditions. These aspects of the performance of the ROM will be discussed in detail as we introduce the main components of the present framework (and resulting models) and present the results of our numerical experiments.

3. Fundamentals of microkinetic models and system dynamics

3.1. Kinetic parameters and rate equations

This section presents the general building blocks of microkinetic models, as well as expressions for rate constants and the reaction rate equations for each elementary reaction step. We consider a system of m elementary reaction steps indexed by $j \in \mathcal{M} = \{1, \dots, m\}$, consisting of n species indexed by $i \in \mathcal{N} = \{1, \dots, n\}$. The reaction rate constant of reaction j, denoted by k_j, is a function of the reaction temperature T and follows an Arrhenius dependence

$$k_j(T) = A_j \exp\left(-\frac{E_{aj}}{RT}\right) \tag{1}$$

where A_j and E_{aj} denote, respectively, the Arrhenius pre-exponential factor and the activation energy for reaction j, and R is the universal gas constant. The pre-exponential factor values can be computed using transition state theory as

$$A_j = \left(\frac{k_B T}{h}\right) \exp\left(\Delta S_j^{\neq}/R\right) \exp\left(1 - \Delta n_j^{\neq}\right) \left(c^0\right)^{\Delta n_j^{\neq}} \tag{2}$$

where k_B is Boltzmann's constant, h is Planck's constant, $\Delta S^{\neq}$ is the entropy change between the reactants and the transition state or activated complex, $\Delta n^{\neq}$ is the change in the number of moles between the reactants and the transition state, and c^0 is the standard reference state which is included for unit consistency [14]. First principles atomistic simulations (e.g., density functional theory (DFT)) can be used to derive $\Delta S^{\neq}$ (e.g., [47]) in order to estimate the value of the pre-exponential factor of a given reaction when there are no experimental data to regress the kinetic parameters of the model in (1). Such atomistic simulations are neither performed in nor strictly needed for the present work, but as it will be described subsequently, when such type of data are available they can be leveraged to improve convergence to the governing equations. Further, Evans–Polanyi relations [15], which assume that the difference in activation energy between two reactions of the same family is proportional to the difference of their enthalpy of reaction, can be used to calculate activation energies as follows

$$E_{aj} = \begin{cases} E_0 + \alpha \Delta H_{Rj} & \text{if } \Delta H_R \leq 0 \\ E_0 + (1-\alpha)\Delta H_{Rj} & \text{if } \Delta H_R > 0 \end{cases} \tag{3}$$

where E_0 is the intrinsic energy barrier for a given reaction family, $\alpha \in [0, 1]$ is a transfer coefficient that describes the position of the transition state along the reaction coordinate, and ΔH_{Rj} is the enthalpy of reaction j, which is defined (for each elementary reaction step) as the sum of the enthalpies of formation of the products and the reactants weighted by their stoichiometric coefficients.

The reaction rate law is then given by the product of the rate constant and the species concentrations raised to the power of the stoichiometric coefficients in the elementary reaction step. Generally, the rate of reaction j can be expressed using mass-action kinetics as

$$\gamma_j(T) = k_j(T) \prod_{i \in \tilde{\mathcal{N}}_j} C_i^{\phi_{ij}} \tag{4}$$

where the set $\tilde{\mathcal{N}}_j \subseteq \mathcal{N}$ is the set of species involved in a given reaction j, ϕ_{ij} denotes the stoichiometric coefficient of species i participating in reaction j, and C_i denotes the concentration of species i. From a systems-theoretical perspective, the description of the evolution of the concentration of the reacting species in time is given by a dynamical system, with the concentrations being the system states, that are denoted by $\mathbf{x}(t) \in \mathbb{R}^{|\mathcal{N}|}$ [42,43] (i.e., $\mathbf{x}(t) = [C_1\ C_2\ \dots\ C_{n-1}\ C_n]^T$).

3.2. Reactor design equations

In microkinetic modeling, given kinetic parameters such as A and E_a, a system of differential and algebraic equations corresponding to the reactor design equations and the reaction rates is solved simultaneously to determine composition profiles of all species. The design equations are the macroscopic material balances for the reactor system and can be generally expressed in terms of the number of moles of a species i as

$$F_i^0 - F_i + G_i = \frac{dN_i}{dt} \quad \forall i \in \mathcal{N} \tag{5}$$

where F_i^0 is the input molar flow rate, F_i is the output molar flow rate, G_i is the generation or consumption rate due to chemical reactions, N_i is the number of moles of species i, and the differential term corresponds to the accumulation or depletion of species i over time. Different versions of (5) can be derived based on the composition units of choice. Intuitively, different assumptions (e.g. perfect mixing, constant volume, uniform catalyst distribution, etc.) are imposed on (5) to derive the design equations of different reactor configurations such as batch, continuous-stirred tank (CSTR) and plug flow reactors [48]. The rate equations introduced in (1) and (4) are linked to the material balance via the generation term G_i for each species i. Furthermore, it should be noted that, when modeling non-isothermal conditions, energy balances must be considered in addition to the material balances in (5) [46]. For example, if the concentrations are assumed to be uniform within a well-mixed, constant volume batch reactor (i.e., $F_i^0 = F_i = 0$) the generation term for species i is

$$G_i = V r_i \quad \forall i \in \mathcal{N} \tag{6}$$

where V is the reactor volume and r_i is the rate of formation/consumption of species i per unit volume, which is given by

$$r_i = \sum_{j \in \mathcal{M}_i^+} \phi_{ij} \gamma_j - \sum_{j \in \mathcal{M}_i^-} \phi_{ij} \gamma_j \quad \forall i \in \mathcal{N} \tag{7}$$

where $\mathcal{M}_i^+$ and $\mathcal{M}_i^-$ are used to respectively denote the sets of reactions that generate and consume species i, and γ_j is defined by the rate law in (4). For the remainder of the paper, we consider the evolution of concentration in time for the sake of notation simplicity. Similar equations can be derived for variables representing other composition measures or for plug flow reactors (PFR) [49] where the independent variable of interest is space (i.e. the axial coordinate of the reactor). Solving the system of equations in (5) for a given set of initial conditions yields the values of the concentration of species over time for the batch reactor. If the MK model parameters (e.g., E_a, A, ΔH_R, ϕ, etc.) are unknown and cannot be reliably calculated based on first principles concepts, they must be estimated from data by solving the associated parameter estimation problem using regression techniques.

4. Fundamental concepts of discovering governing equations

Having introduced the main building blocks of MK models and the first principles arguments used to derive their associated parameters, this section presents the fundamental elements of a framework for discovering governing equations. We assume that there exists a set $\tilde{\mathcal{N}} \subseteq \mathcal{N}$ of $|\tilde{\mathcal{N}}| = \tilde{n} < n$ species whose concentrations are measured experimentally or are computed via simulation, and thus for which the concentration data as a function of time are available. Additional measurements such as temperature, volume, etc. may also be available, depending on the reactor configuration and experimental/simulation set up, for which material and energy balances must also be considered. In practical settings, where models might not be available a priori or where simulation might be computationally prohibitive, measurable species (that is, $i \in \tilde{\mathcal{N}}$) typically correspond to gaseous species whose concentrations as a function of time can be resolved experimentally using e.g. spectroscopy and chromatography techniques. As indicated previously in Section 1, the distribution of adsorbed surface intermediates cannot typically be determined directly using experimental methods (that is, $i \in \{\mathcal{N} - \tilde{\mathcal{N}}\}$ and these species are said to be *unmeasurable*). We denote the available N training data points as $\tilde{C}_i$, which are available at sampling times $\{\tilde{t}_0, \dots, \tilde{t}_N\}$ that need not be uniformly spaced, and can be organized in a data matrix of the form of

$$\tilde{\mathbf{C}} = \begin{bmatrix} \tilde{C}_1(t_1) & \tilde{C}_2(t_1) & \dots & \tilde{C}_{\tilde{n}}(t_1) \\ \tilde{C}_1(t_2) & \tilde{C}_2(t_2) & \dots & \tilde{C}_{\tilde{n}}(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{C}_1(t_N) & \tilde{C}_2(t_N) & \dots & \tilde{C}_{\tilde{n}}(t_N) \end{bmatrix} \in \mathbb{R}^{N \times |\tilde{\mathcal{N}}|} \tag{8}$$

4.1. Regression-based discovery of governing equations

Early works on discovering governing equations from data [42], and applications thereof to reaction networks [45], are based on sparse regression techniques that involve solving an optimization problem of the form of

$$\begin{aligned} \mathcal{K}^* \in \arg\min_{\mathcal{K}} \ & \frac{1}{2N} \|\dot{\tilde{\mathbf{C}}} - \Theta(\tilde{\mathbf{C}})\mathcal{K}\|_F^2 + \lambda\rho\ell_1(\mathcal{K}) + \frac{\lambda(1-\rho)}{2}\ell_2(\mathcal{K})^2 \\ \text{s.t.} \ & \mathcal{K} \geq 0 \end{aligned} \tag{9}$$

where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix, $\tilde{\mathbf{C}}$ is the data matrix for the measured species concentration given in (8), and $\dot{\tilde{\mathbf{C}}}$ denotes the estimated derivative for the data matrix, which is typically obtained using numerical techniques such as finite differences [45]. The set of possible elementary reaction steps generated as discussed earlier is denoted with $\Theta : \mathbb{R}^{N \times \tilde{\mathcal{N}}} \to \mathbb{R}^{N \times \tilde{\mathcal{N}} \times \mathcal{M}}$, whose entries correspond to the product terms in the right-hand side of (4). That is

$$\Theta_{kj}(\tilde{\mathbf{C}}) = \prod_{i \in \tilde{\mathcal{N}}_j} (\tilde{C}_i(t_k))^{\phi_{ij}} \quad \forall t \in \{0, \dots, N\}, j \in \mathcal{M} \tag{10}$$

which is a (nonlinear) function of the species concentrations, noting that the term $\Theta(\tilde{\mathbf{C}})\mathcal{K}$ corresponds to a linear combination of the reaction rates to reflect the species balances similarly to those in (7). That is, $\Theta_{kj}(\tilde{\mathbf{C}})\mathcal{K}_j$ corresponds to the instantaneous reaction rate of reaction j at time t_k, and the ith column $(\Theta(\tilde{\mathbf{C}})\mathcal{K})_i$ corresponds to the material balance of species i. The goal is to find the optimal values of the reaction rate constants $\mathcal{K}^*$ (i.e., the coefficients associated with each nonlinear rate function), where the rate constant vector is $\mathcal{K} = [k_1, k_2, \dots, k_m]^T \in \mathbb{R}_{\geq 0}^{|\mathcal{M}|}$, that minimizes the objective function in (9). In brief, the regression problem in (9) attempts to find the best values of the reaction rate constants for each elementary reaction step in the basis, by minimizing the mean squared error between the available data and the values predicted by the model, plus a function that penalizes the magnitude of the rate constant estimates.

In (9), $\lambda \in \mathbb{R}_{\geq 0}$ is a hyperparameter (i.e., a user decision related to the implemented discovery algorithm) corresponding to the regularization coefficient that dictates the degree of sparsity (i.e., the extent to which reactions are eliminated from the starting candidate set of reactions by setting $k_j = 0$) in the discovered reaction network, and $\rho \in [0, 1]$ is a hyperparameter that linearly balances between ℓ_1 and ℓ_2 norm regularization functions. Performing regularization is

critical in the discovery of governing reaction networks as, without regularization (note that the case when $\lambda = 0$ corresponds to ordinary least-squares regression), all elementary reaction steps in the material balance of the measured species will be assigned non-zero rate constant values. Increasing values of λ imposes a greater penalty on the values of the estimated rate constants, and thus can shrink the estimated rate constant values to zero. In particular, the ℓ_1 norm regularization penalty (also known as LASSO regression when $\rho = 1$), which is given by

$$\ell_1(\mathcal{K}) = \|\mathcal{K}\|_1 = \sum_{j\in\mathcal{N}} |k_j| \tag{11}$$

will typically perform better at sparsifying the library of elementary reaction steps (that is, at eliminating as many elementary reaction steps as possible by forcing their reaction rate constants to $k_j = 0$) relative to the ℓ_2 norm regularization penalty (also known as Ridge regression when $\rho = 0$), which is given by

$$\ell_2(\mathcal{K}) = \|\mathcal{K}\|_2 = \sqrt{\sum_{j\in\mathcal{N}} k_j^2} \tag{12}$$

Both regularization approaches increase bias and reduce variance in the model to avoid overfitting the training data, and typically result in models that have improved generalization performance beyond the training data. In the context of sparse regression, LASSO regression is better suited for the discovery of governing equations as it naturally performs "feature selection" (selecting the appropriate basis functions or, in this context, candidate reaction steps) relative to Ridge regression [50]. The value of hyperparameter λ is typically tuned via cross-validation to ensure an adequate balance between model complexity and predictive performance, a strategy that will be illustrated in the numerical experiments in Section 6. The combination of the two aforementioned norms corresponds to elastic net regression, for which the additional hyperparameter ρ must be determined during cross validation. For the remainder of the paper we consider a generic regularization function $\ell(\cdot)$ that is a convex combination of the ℓ_1 and ℓ_2 norms, that is

$$\ell(\cdot) = \rho\ell_1(\cdot) + \frac{1-\rho}{2}\ell_2(\cdot)^2 \tag{13}$$

Discovery frameworks that involve optimization problems of the form of (9) directly evaluate the various basis functions/candidate elementary reaction steps on the measurement data and directly regress (and often numerically estimate) the derivative $\dot{\tilde{\mathbf{C}}}$. While numerous extensions and applications of these formulations have been reported in the literature, they suffer from several inherent limitations:

- The rates of change (i.e., the derivatives $\dot{\tilde{\mathbf{C}}}$) of the concentration of the species are directly estimated from the measured data. The results of such numerical differentiation calculations are highly sensitive to measurement noise and are impacted by sparsely sampled data, with a potential for considerable approximation error.
- Similarly, stiff cases where the values of the rates of change of the state variables (i.e., compositions) span multiple orders of magnitude (as can often be the case for microkinetic systems) can result in ill-conditioned instances of the optimization problem (9) leading to poor discovery performance.
- The discovered governing equations are assumed to be linear with respect to the unknown parameters $\mathcal{K}$ (i.e. as given by $\Theta(\mathbf{C})\mathcal{K}$), and therefore, these ideas are not applicable to discovering microkinetic models with nonlinear parametrizations arising from non-isothermal data. Equivalently, it is not possible to directly discern the pre-exponential factors, activation energies, and heats of reaction in the rate constant relationships introduced in Section 3.1.
- Owing to the implicit assumption that all state variables are directly measurable or local functions of the measurable data (i.e., they are observable), approaches based on (9) cannot cope with unmeasurable/unobservable variables, such as the surface coverage of adsorbed intermediate species.

In light of the above, discovery methods of the type (9) are not anticipated to perform well for microkinetic systems. Several modifications to the underlying optimization problem (9) have been proposed. An "implicit" approach (which by definition is based on considering the state derivatives in the set of candidate basis functions of the form of $\Theta(\tilde{\mathbf{C}}, \dot{\tilde{\mathbf{C}}})$) was introduced to infer governing equations with rational functions [51]. In particular, *weak formulations* [52–55] avoid the estimation of the derivative $\dot{\tilde{\mathbf{C}}}$ by directly integrating the candidate governing dynamics, which results in improved stability under noisy measurements. The general structure of the optimization problem is

$$\begin{aligned} \mathcal{K}^* \in \arg\min_{\mathcal{K}} \ & \frac{1}{2N}\sum_{i\in\tilde{\mathcal{N}}} \|\tilde{C}_i - C_i(t)\|_F^2 + \lambda\ell(\mathcal{K}) \\ \text{s.t.} \ \ & C_i(t) = \int \Theta(C(t))\mathcal{K}dt \ \ \forall i \in \{1,\dots,n\} \\ & \mathcal{K} \geq 0 \end{aligned} \tag{14}$$

Different methods have been reported for computing the integrated candidate governing equations, such as piecewise constant quadrature [52], Runge–Kutta methods [55], and orthogonal collocation on finite elements [43]. Note that formulation (14) can cope with unmeasured species by only considering data corresponding to measured species $i \in \tilde{\mathcal{N}} \subseteq \mathcal{N}$ in the model performance metric. To the best of our knowledge, however, these methods have not yet been demonstrated for the data-driven discovery of governing reaction networks.

5. Dynamic nonlinear optimization framework for microkinetic model discovery

In this section, we introduce our main methodological contribution addressing the limitations discussed in the previous section. More specifically to the problem at hand, in this section we distinguish between isothermal and non-isothermal microkinetic model discovery, and provide details on each of the underlying optimization problems. In the isothermal case, when a set of experiments is performed at the same temperature, the discovery problem can be formulated as a regularized dynamic nonlinear optimization problem to determine a (sparse) vector of reaction rate constants k that best model the behavior of the system. The optimization problem is given by

$$\begin{aligned} \min_{C,k} \ & \epsilon + \lambda_k\ell(k) \\ \text{s.t.} \ \ & \epsilon = \frac{1}{2N}\sum_{i\in\tilde{\mathcal{N}}} w_i \int_{t_0}^{t_f} \|\tilde{C}_i(t) - C_i(t)\|_2^2 dt \\ & \text{Rate equations (4)} \\ & \text{Reactor design equations (5)} \\ & k_j \in [k_j^L, k_j^U] \ \ \forall j \in \{1,\dots,m\} \end{aligned} \tag{15}$$

where $\lambda_k \in \mathbb{R}_{\geq 0}$ is the regularization coefficient for the reaction rate constants and $\ell(k)$ is the regularization function as in (13). Similarly to (9), the objective function is a linear combination of the error metric denoted by ϵ, which corresponds to the mean squared error between measurements and model predictions, and the regularization penalty that favors models that are of lower complexity. The weight factor $w_i \in \mathbb{R}_{>0}$ is set to $w_i = 1/\max \tilde{\mathbf{C}}_i$ and is used to scale the values of the measured species concentrations such that they are all given similar importance in the optimization problem, thus accounting for species with low concentration values in the training data set. Lastly, $k_j^L, k_j^U \in \mathbb{R}_{>0}$ correspond to lower and upper bounds, respectively, imposed on the reaction rate constants being estimated. Introducing these bounds in the optimization problem (15) has the computational

advantage of reducing the feasible space and ensures that the recovered rate constants are within bounds that are physically meaningful (based on first principles or previous knowledge of the reaction family).

While the reaction rate constants k_j can be inferred directly by solving (15) for data corresponding to experiments performed at the same temperature, resolving both kinetic parameters (activation energies and pre-exponential factors) of the Arrhenius relationship (1) requires reaction data for at least two different temperatures. We denote by $\mathcal{T}$ the set of temperatures at which experimental data are available. Thus in the case of non-isothermal MK network discovery, the resulting optimization problem has a similar structure to the one in (15), with appropriate modifications to account for the data at different temperatures.

$$\begin{aligned}
&\min_{C_T,A,E_a} \quad \epsilon + \lambda_A \ell(A)\\
&\text{s.t.} \quad \epsilon = \frac{1}{2N}\sum_{T\in\mathcal{T}}\sum_{i\in\mathcal{N}} w_{i,T}\int_{t_i}^{t_f} \|\tilde{C}_{i,T}(t) - C_{i,T}(t)\|_2^2 dt\\
&\qquad \text{Arrhenius equation (1)}\\
&\qquad \text{Rate equations (4)}\\
&\qquad \text{Reactor design equations (5)}\\
&\qquad A_j \in [A_j^L, A_j^U],\ E_{aj} \in [E_{aj}^L, E_{aj}^U]\ \ \forall j \in \{1,\dots,m\}
\end{aligned} \tag{16}$$

Note that the concentration profiles predicted by the model (16) are a function of temperature (and thus denoted by C_T), and that A and E_a are estimated instead of k in (15). Intuitively, these modifications result in a higher dimensional optimization problem (the number of state variables and constraints scales with $|\mathcal{T}|$ in addition to including twice as many kinetic parameters) than (15). Furthermore, the nonlinearity and non-convexity with respect to the decision variables resulting from the Arrhenius Eq. (1) render (16) a much more challenging optimization problem relative to (15). Further, while we have that

$$\lim_{E_{aj}\to+\infty} k_j(T) = 0 \tag{17}$$

by which, for a fixed value of the pre-exponential factor A_j, reactions can be eliminated by setting the activation energy to infinity, we have that $E_{aj} \le E_{aj}^U$ and from first principles it follows that $E_{aj}^U \ll +\infty$. Furthermore, eliminating elementary reaction steps by means of setting $E_{aj} = +\infty$ poses numerical difficulties and renders the optimization problem (16) unbounded. Therefore, in the case of (16), regularization with respect to the pre-exponential factor (that is, the extent of regularization dictated by λ_A) will be used to impose sparsity on the discovered reaction network, since the value of the activation energy (for $E_{aj} \le E_{aj}^U \ll +\infty$) does not have a structural impact on the microkinetic model discovered (i.e., does not influence whether an elementary reaction step is included in the model or not). Tuning λ_A will be further discussed in the numerical experiments section.

Remark 1. The fact that the model in (16) is learned based on the macroscopic (i.e., measurable) data, as well as a large set of possible elementary reaction steps, constitutes a **structural deficiency** of the proposed approach. In general, it cannot be expected, as discussed in previous sections, that the microscopic kinetics will be fully resolved based on macroscopic measurements when the system is unobservable [10,11]. This deficiency can be partially addressed by: (i) measuring the composition of additional species and in particular that of key surface intermediates (or at least inferring the dominant surface species), and (ii) adding bias to the model towards certain reaction intermediates or reaction rate constants based on first principles calculations (e.g., obtained from DFT simulations) which can be implemented by appropriately modifying the regularization penalty, the parameter bounds, and initial values.

Remark 2. Parameter bounds can be used to impose hierarchy in the resulting model (e.g., by enforcing that rate constants follow a certain relative value based on their reaction families). Similarly, terminal constraints of the form of $C_i(t_f) \in [\psi_i^L, \psi_i^U]$ with $\psi_i^L, \psi_i^U \in \mathbb{R}_{\ge 0}$ can be included in (15) and (16) to ensure that a certain distribution of species is observed at equilibrium. Moreover, hierarchy in the multi-scale sense can be enforced by incorporating additional constraints to ensure that the known physics and chemistries at different length and time scales are obeyed (e.g., by including mass and heat transport equations in combination with the previously described kinetic and reactor design equations).

5.1. Model discretization

An important challenge posed by optimization problems of the type of (15) and (16) is that the system dynamics (i.e., the coupled reaction rate and reactor design equations) are formulated as ordinary differential equations with a continuous independent variable. These infinite-dimensional optimization problems must then be discretized and reformulated on a finite dimensional domain. The ODEs (5) are converted to a system of nonlinear algebraic equations and are then solved numerically. Recent works [43,46] showed that orthogonal collocation on finite elements is an effective way to discretize continuous time ODEs with stiff nonlinear dynamics. The discretization scheme is A-stable, high-order, and can handle non-smooth events at the element boundaries (e.g., non-smooth or discontinuous profiles of the reactor feed) [56].

Recalling that $\mathbf{x}$ denotes the system states (e.g. the species concentrations), the dynamics (e.g., the design equation in (5)) can be written in the generic form

$$\frac{d\mathbf{x}(t)}{dt} = f(\mathbf{x}(t)) \tag{18}$$

where $f(\cdot) : \mathbb{R}^{n_x} \to \mathbb{R}^{n_x}$ and n_x is the number of states (e.g., the number of species whose concentrations are modeled as a function of time). The collocation equations corresponding to system (18) form a set of algebraic constraints given by

$$\begin{aligned}
&\left.\frac{d\mathbf{x}(t)}{dt}\right|_{t_{ij}} = \frac{1}{h_i}\sum_{k=0}^{K}\mathbf{x}_{ik}\frac{d\mathcal{L}_k(\tau_j)}{d\tau}, && j\in\{1,\dots,K\},\ i\in\{1,\dots,N\}\\
&\left.\frac{d\mathbf{x}(t)}{dt}\right|_{t_{ij}} = f(\mathbf{x}(t_{ij})), && j\in\{1,\dots,K\},\ i\in\{1,\dots,N\}\\
&\mathbf{x}(t_{i+1,0}) = \sum_{k=0}^{K}\mathcal{L}_k(1)\mathbf{x}(t_{ik}), && i\in\{1,\dots,N-1\}
\end{aligned} \tag{19}$$

where K is the number of collocation points on each of the N finite elements, h_i is the length of finite element i, $t_{i-1} = t_{ij} - \tau_j h_i$, and the state variable $\mathbf{x}(t)$ is interpolated using Lagrange polynomials $\mathcal{L}_k$ as follows

$$\begin{aligned}
&\mathbf{x}(t) = \sum_{k=0}^{K}\mathcal{L}_k(\tau)\mathbf{x}_{ik}, \qquad t\in[t_{i-1},t_i],\ \tau\in[0,1]\\
&\mathcal{L}_k(\tau) = \prod_{j=0,\neq k}^{K}\frac{\tau-\tau_j}{\tau_k-\tau_j}
\end{aligned} \tag{20}$$

The optimal choice of the interpolation points τ_j is derived in detail in [56, Theorem 10.1]. While (as with any discretization scheme) the approximation error of (19) is evidently nonzero, in this case the truncation error is $\mathcal{O}(h^{2K-1})$ for Lagrange–Radau collocation points [56].

5.2. Training data

Problems (15) and (16) amount to a supervised learning process corresponding to constrained nonlinear regression. The training data used are concentration values over time or space for species that are measurable experimentally or for which data can be generated via simulation. While it is not possible to determine the precise data requirements, as is often the case when machine learning techniques are

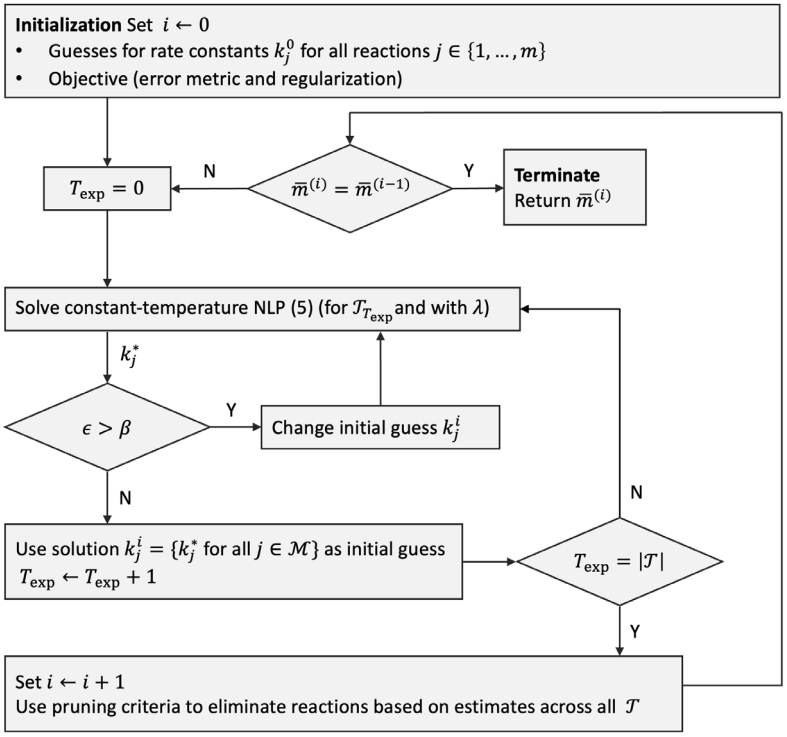

Fig. 1. Sequential elimination heuristic algorithm for discovering MK models from data sets collected at different temperatures.

used, it is important that the data sets used for training reflect both the fastest and the slowest dynamics of the system, as well as the different steady states that the system may have. Such a requirement implies that experiments and simulations must be conducted for sufficiently long time spans to allow the system to reach steady state. We note that an earlier version of the proposed framework was demonstrated for systems with periodic steady states [43] including oscillatory reactive systems [46]. The proposed framework allows for discovering dynamics that can capture steady state multiplicity or limit cycle behavior of the system so long as these conditions are reflected in the training data (e.g. the catalytic oxidation of CO on Pt catalysts [57,58]) and the elementary reaction steps in the initial basis are able to reproduce the observed dynamics.

Remark 3. Similar modifications to the ones made to (15) in (16) to account for multiple temperatures can be made to account for data that reflect different initial conditions of the gas phase. Given the nonlinear nature of the system, different initial conditions can result in distinct dynamic behavior and convergence to different steady states. Leveraging a greater number and variety of data sets (i.e., by varying temperature or initial conditions) is expected to improve the learning outcome and result in models that more accurately represent different kinetic regimes, and hence are more general and have greater practical utility.

As for the sampling frequency of the training data, which describes the number of samples collected per unit time or space, there are also no explicit requirements. A lower bound on the required sampling frequency can be obtained using Nyquist's theorem, which establishes that the minimum sampling rate needed to reconstruct a signal must be at least twice the frequency of the highest frequency component. A higher sampling frequency might be desirable as it provides greater resolution on the observed dynamics and can improve robustness to measurement noise. An important consideration regarding the proposed approach is that the training data must be available at the collocation points defined previously in Section 5.1. While it might be difficult to enforce this condition when collecting data experimentally, polynomial interpolation techniques (as described in [43]) can be used to resample the training data to align with the discretization scheme used. In such cases where interpolation is required, a sampling frequency greater than the lower bound obtained with the Nyquist theorem is particularly desirable to minimize the approximation error from resampling the data.

5.3. Solution strategy for the isothermal case

In principle, the governing MK model and a reduced representation thereof can be obtained by directly solving the optimization problem when suitable values of the regularization penalties are selected. Solving (16) to optimality and tuning associated regularization coefficients can be computationally demanding. Therefore, it is helpful to define a systematic approach for reducing the dimensions of the problem prior to attempting to directly solve (16). Per the previous discussion, and noting that the single—temperature optimization formulation in (15) has fewer variables and a more favorable structure than the problem in (16), we propose a heuristic strategy for reducing the size of (16) which consists of pruning/eliminating elementary reaction steps based on the repeated solution of (15).

A flowchart of the proposed sequential heuristic algorithm is shown in Fig. 1. Starting with all parameters and data that define optimization problem (15) and initial guesses k_j^0 for all the reaction rate constants (e.g. guided using first principles knowledge as discussed in Section 3), (15) is solved using the measured species concentration profiles over time for different temperature settings. Since elementary reaction rate constants increase with temperature, we consider the set $\mathcal{T}$ ordered as a function of increasing temperature, and (15) is initialized with a low value (e.g., close to k^L) for the rate constants and subsequently solved. Considering the likely large-scale and non-convex nature of the problem, it is possible that the nonlinear optimization solver terminates at an unsatisfactory solution (that is, having an error value ϵ greater than a predefined error tolerance β), whether this solution is a true (local) minimum, or is a feasible solution for which optimality cannot be established but the optimization solver terminates based on internal

settings/criteria (e.g., minimal change in the objective function but large gradient). It also permits that the optimization problem may be infeasible or unbounded. In either case, if the error ϵ is greater than the tolerance β, it may be possible to achieve convergence to a better solution by changing the initialization strategy used to solve (15). One way to do so is to progressively increment the initial rate constant values, which can be justified by the fact that the heuristic is started with low-value initial guesses and data are considered in order of increasing temperature. Other heuristic initialization strategies can be employed, such as lowering the initial values of rate constants corresponding to reactions that are speculated to have low importance, and increasing the initial values of rate constants corresponding to reactions that are speculated to have high importance. Repeated initialization is performed until convergence to an acceptable local minimum is achieved. Otherwise, structural modifications to the MK model currently under consideration must be made (e.g., including additional elementary reactions steps, widening the bounds on the reaction rate constants). Furthermore, the estimated rate constant values are successively used as the initial guesses of subsequent (higher temperature) instances of optimization problem (15). In Section 6, we present extensive numerical experiments in which we study the variability of the discovered model as a function of different initial rate constant values sampled at random.

Once estimates for k_j are available for all $T \in \mathcal{T}$, reactions with the lowest average rate constant value are eliminated from the basis, similarly to sequentially thresholded least squares regression [42] and the pruning criteria employed in the reactive SINDy framework [45]. Having determined a smaller set of elementary reactions that maintain an acceptable predictive performance, the procedure is repeated starting with this reduced basis of candidate elementary reactions. The algorithm terminates once the number of reactions in the basis does not change from one iteration to the next. As will be discussed in the subsequent section, an additional termination criterion employed corresponds to checking for significant increases in the error metric(s) (e.g., mean squared error, Akaike information criterion, Bayesian information criterion), following a similar idea to the one introduced in [59]. For example, if the error increases substantially after eliminating a given reaction, it is likely that a reaction present in the set of true governing equations was eliminated, in which case the pruning step would be reversed, and if no other reactions can be eliminated then the algorithm terminates.

Once the algorithm illustrated in Fig. 1 terminates, yielding a set of reactions of size $\tilde{m}$ (with $\tilde{m} \leq m$), solving optimization problem (16) is expected to be more tractable based on the extent of pruning. If no further reduction of the basis of reactions is desired, then (16) can be solved without regularization (i.e., by setting $\lambda_A = 0$ since the pre-exponential factor dictates which reactions are active or inactive). Otherwise, further model reduction can be performed by setting $\lambda_A > 0$, but a range of values of the regularization coefficient should be evaluated to ensure that the final discovered model has the appropriate balance of predictive ability and complexity. Intuitively, the structural deficiency explained in Remark 1 accentuates as more regularization is performed (through the sequential elimination or by increasing λ_A). These concepts are demonstrated in the subsequent section via two case studies.

6. Numerical experiments

In this section we describe comprehensive numerical experiments to illustrate the implementation of our framework and benchmark its performance. We consider two different but related outcomes. First, we show that the framework can be used for discovering governing microkinetic models (**Goal 1**) from an initial set of elementary reaction steps corresponding to different pathways (and associated intermediates) from the feed to the outlet species. Second, we focus on dimensionality reduction (**Goal 2**) to demonstrate that the proposed approach can identify a more parsimonious microkinetic model without losing predictive ability within the domain of the training data, as well as having good extrapolation performance (e.g., at lower and higher temperatures than those reflected in the training data). We consider dynamics of the form of (7) for a constant-volume batch reactor, noting that the framework is expected to cope with different reactor configurations by using the relevant reactor design equations as discussed earlier.

All computations were performed on a PC running Windows 7 64-bit, with a 3.6 GHz Intel Core i7-7700 processor and 32 GB RAM. The numerical analysis was performed in Python 3.8.3. The dynamic optimization problems (15) and (16) were implemented in Pyomo [60] and were discretized using the `pyomo.DAE` package [61], using orthogonal collocation on finite elements with the default Lagrange–Radau polynomials. The resulting discrete optimization problems were solved using CONOPT [62], a generalized reduced-gradient nonlinear solver using default settings and the absolute optimality criterion set to `optca=` 0.01. The training data for the ensuing numerical experiments were generated via simulation by integrating the true governing microkinetic model at different temperatures. Only the concentrations of the gas-phase species as a function of time are used for training.

6.1. Experiment 1: Prototype reaction network discovery

We begin by considering a prototype reaction network with overall kinetics given by

$$\mathrm{A} + \mathrm{B} \longrightarrow \mathrm{C} \tag{21}$$

which follows a mechanism governed by the elementary reaction steps and associated kinetic parameters given in Table 1. This prototype example is used to establish the foundations of the proposed framework and to illustrate some of the main concepts discussed in the previous sections.

In the reaction mechanism depicted in Table 1, gaseous species A and B must both first adsorb on the free sites on the catalyst surface, which are denoted with *, to yield adsorbed species A$*$ and B$*$. These adsorption reactions are first-order and monomolecular with respect to the gas-phase species. The adsorbed species react irreversibly at the surface to form species C and liberate two free sites, without forming an adsorbed intermediate (e.g., C*). This prototype example thus consists of five elementary reactions and six species, for which activation energies and pre-exponential factors are shown in Table 1. To illustrate the performance of our framework we assume that several other elementary reactions may be occurring in addition to the true governing reaction steps shown in Table 1. These additional reactions are given by:

$$\begin{aligned} &\mathrm{A} + 2\,* \rightleftharpoons \mathrm{A}** \\ &\mathrm{B} + 2\,* \rightleftharpoons \mathrm{B}** \\ &\mathrm{A}** + \mathrm{B}** \rightleftharpoons \mathrm{C}** + 2\,* \\ &\mathrm{C}** \rightleftharpoons \mathrm{C} + 2\,* \end{aligned} \tag{22}$$

Including these additional spurious reactions results in a system of 13 elementary reactions and 9 species (three are gaseous, five are adsorbed intermediates, and one corresponds to the available free sites). It should be noted that the additional spurious adsorption steps in (22) are second-order reactions on a dual site as opposed to the monomolecular adsorption reactions shown in Table 1. A complete list of all the bounds and initial values used for the true and additional reaction parameters (i.e., rate constants, pre-exponential factors, and activation energies) can be found in the Supplementary Information. For this example, we consider experiments carried out at four different temperatures $\mathcal{T} = \{420\ \mathrm{K}, 430\ \mathrm{K}, 440\ \mathrm{K}, 450\ \mathrm{K}\}$. The resulting dynamic model (as per the reaction rate and design equations introduced in Section 3) for this prototype reaction network was discretized using 75 finite elements and 5 collocation points per finite element. For the microkinetic model

Table 1
Experiment 1: Prototype reaction network example and associated true parameter values for forward and reverse reactions.

Reaction	Forward		Reverse	
	E_a^f (kcal mol^{-1})	A^f	E_a^r (kcal mol^{-1})	A^r
A + ∗ ⇌ A ∗	0	10^{-3} †	25.1	10^{13} ‡
B + ∗ ⇌ B∗	0	10^{-3} †	23.9	10^{13} ‡
A∗ + B∗ → C + 2∗	13.6	2×10^{9} ‡	–	–

† Units of molecule Pa^{-1} site^{-1} s^{-1} ‡ Units of molecule site^{-1} s^{-1}.

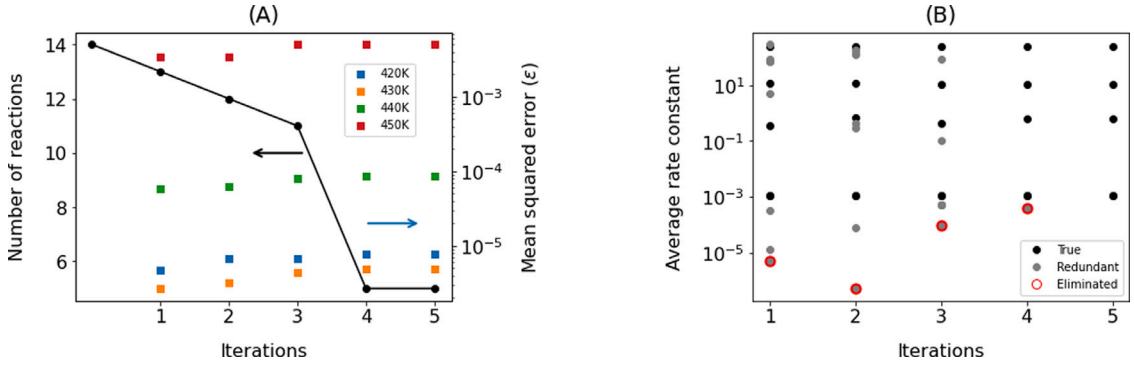

Fig. 2. Experiment 1: Performance of the proposed heuristic algorithm in discovering prototype reaction network example as a function of iterations performed until convergence is reached. (A) Model complexity (left), as measured by the number of elementary reaction steps included in the model, and predictive ability (right), using mean squared error. (B) Average estimated rate constant values highlighting true reactions, redundant reactions, and the elementary reaction steps eliminated at a given iteration.

defined for the reactions in Table 1 and in (22), the resulting nonlinear program had 6,762 rows, 6,775 columns and 60,743 non-zeros for the constant-temperature problem (15), and 27,040 rows, 27,066 columns and 305,983 non-zeros for the variable-temperature problem (16).

Fig. 2 shows the progress of the sequential elimination heuristic algorithm (shown in Fig. 1) in eliminating spurious reaction steps from the initial network. Fig. 2(A) shows the number of reactions (as a measure of model complexity) and the mean squared error (MSE) (as a measure of the predictive ability of the model) as a function of the algorithm iterations. It is evident that the number of elementary reaction steps included in the model decreases consistently as the algorithm proceeds. On the other hand, the MSE for each of the temperature experiments remains relatively constant, which is indicative of the fact that spurious model components are being pruned. It should also be noted that eliminating reaction steps beyond the iterations in Fig. 2 results, in this case, in model structures that are ostensibly mechanistically incorrect due to violations of mass conservation (e.g., the only elementary reaction responsible for the generation of species C is eliminated). These structural modeling errors lead to significant increases in the objective function value and even termination of the optimization solver due to failing to identify a feasbile solution. Either of these outcomes constitutes an important additional stopping criterion for the discovery algorithm, indicating that the model deteriorated with the latest pruning step. Similarly, Fig. 2(B) shows the average value (computed for all temperature experiments) of the rate constant values estimated at each iteration, for both the spurious and the true governing elementary reaction steps involved in the model. Rate constant values highlighted in red correspond to the reactions eliminated at a given iteration, which have the lowest estimate values on average. It should also be noted that for systems featuring a broad range of reaction rate constant values (occurring commonly in microkinetic models with many elementary reaction steps that occur over different time scales), reactions having small rate constant values might be eliminated when the pruning criteria are based on magnitude, calling for evaluating and monitoring the predictive performance of the discovered model using the MSE for a range of different experimental conditions as shown in Fig. 2(A). The predictive performance of the final model will be optimal relative to the error metric using data for all temperatures available using (16), so that the model discovered is not biased towards low or high temperatures within the training regime.

Lastly, once the sequential elimination heuristic has terminated, the variable-temperature optimization problem (15) is solved to estimate the kinetic parameters A and E_a for the resulting lower-dimensional network. In this case, since no further reactions are eliminated from the basis obtained via the heuristic algorithm, no regularization was employed in solving (15).

As a final validation step, the resulting kinetic model is simulated by integrating the discovered equations, and the resulting trajectories are compared to the available testing data, preferably "unseen" and not used at any stage during model training. These results are displayed in Fig. 3 for the four temperatures considered, where the top subplots correspond to the evolution of the partial pressures of the gaseous species and the bottom plots correspond to the evolution of the site coverage of the surface species. Clearly, and not surprisingly, the model accurately predicts the concentration profiles of the gaseous species, as these were the data used for discovery. Moreover, since the correct reaction steps were discovered, the resulting model also accurately predicts the evolution of the (unmeasurable) surface species. The resulting model parameters are presented in Table 2; their values are close to those used to generate the training data shown in Table 1.

6.2. Experiment 2: Model reduction for oligomerization microkinetic model

We now consider a more complex and industrially relevant example corresponding to the oligomerization of propene to C_6 hydrocarbons. The $C_3 \longrightarrow C_6$ dimerization model was generated using the NetGen automatic network generation framework [16,17]. A MK model comprising 15 species (7 catalytic surface species and 8 gaseous species) and 32 reaction steps with Arrhenius-type rate constants was used to generate the data representing concentration profiles, which will later be used by the proposed machine learning framework to discover the reaction network. The simulated reactor had a volume of 10 cm^3 and contained catalyst with 9×10^{17} active catalytic sites. The Sankey diagram in Fig. 4 shows the relative magnitude of the fluxes (molecule per site per second) of the species predicted by the MK model. Reactions branching from the methyl shift step (boxed in Fig. 4) are considered as redundant steps in the model (i.e., their rates were set to zero when

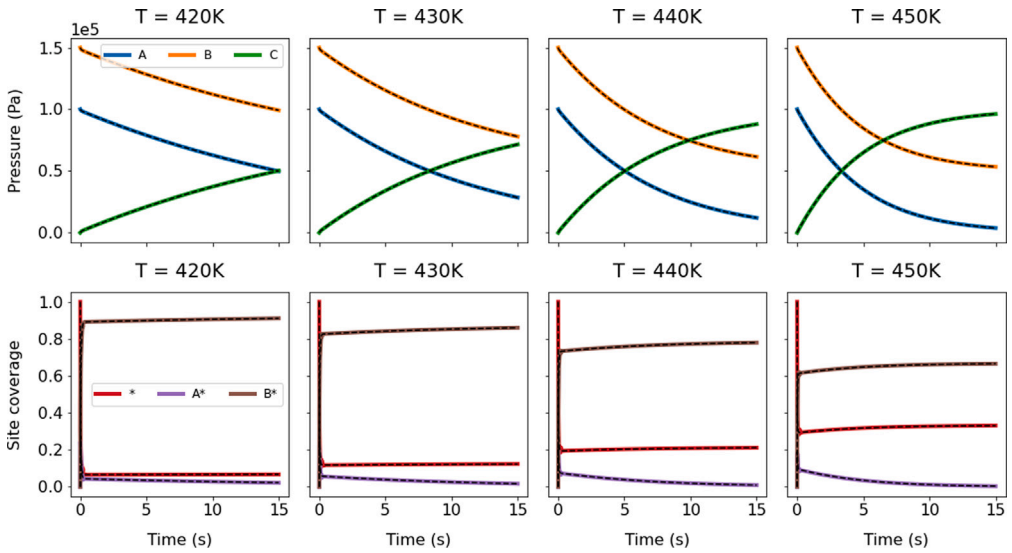

Fig. 3. Experiment 1: Comparison of data generated by simulating the true microkinetic model and the governing equations discovered by the proposed framework for different temperature settings. Top plots correspond to gaseous species, and bottom plots correspond to surface species. Dashed lines correspond to data obtained by simulating the true model.

Fig. 4. Experiment 2: Sankey diagram showing the flux (molecule per site per second) of species in the oligomerization of propene up to C_6 compounds. Reactions and species boxed in gray segue from the methyl shift step and are not modeled when generating training data, but are included in the initial library of reaction steps used for model discovery.

Table 2
Experiment 1: Parameters estimated for the prototype reaction network example. Note that eliminated (redundant) reactions are not shown as they all were assigned $A = E_a = 0$.

Reaction	Forward		Reverse	
	E_a^f (kcal mol^{-1})	A^f	E_a^r (kcal mol^{-1})	A^r
A + ∗ ⇌ A∗	6.04×10^{-2}	1.07×10^{-3} †	26.3	3.94×10^{13} ‡
B + ∗ ⇌ B∗	6.55×10^{-2}	1.08×10^{-3} †	23.9	1.01×10^{13} ‡
A∗ + B∗ → C + 2 ∗	13.6	1.93×10^{9} ‡	–	–

† Units of molecule Pa^{-1} site^{-1} s^{-1} ‡ Units of molecule site^{-1} s^{-1}.

generating the training data but were nevertheless included in the initial library of reaction steps used for model discovery). It is important to note that only composition data pertaining to the gaseous species (left- and right-most species in Fig. 4) are used for model discovery and reduction, while all surface species are assumed to be unmeasurable, and hence, no associated data are used in the construction of the error ϵ in (15) and (16). We employ data corresponding to three temperatures, that is $\mathcal{T} = \{420\text{ K}, 430\text{ K}, 440\text{ K}\}$. The dynamics of the MK model were discretized using 80 finite elements and 3 collocation points per finite element. For the MK model constructed with the initial (large) library of

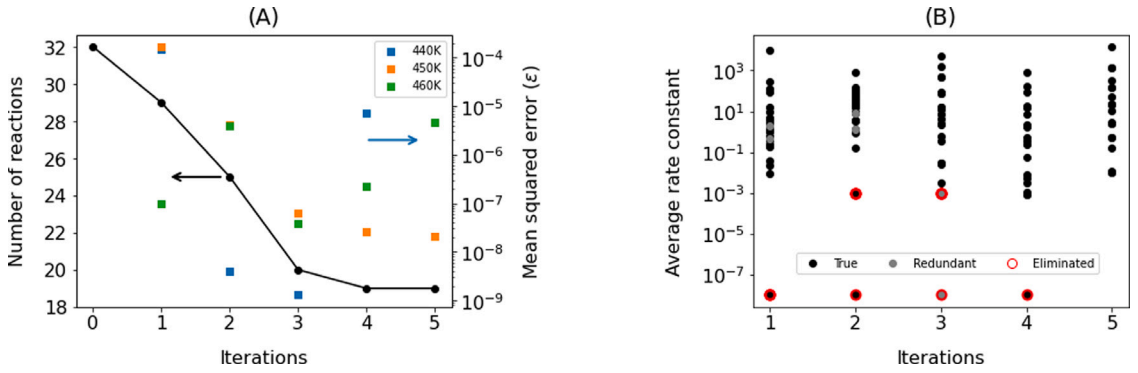

Fig. 5. Experiment 2: Performance of proposed heuristic algorithm in discovering C_3 to C_6 oligomerization microkinetic model as a function of iterations performed to convergence. (A) Model complexity (left), as measured by the number of reactions in the model, and predictive ability (right), using mean squared error. (B) Average rate constant values highlighting true reaction steps, spurious reaction steps, and reaction steps that get eliminated at a given iteration.

elementary reactions, the resulting nonlinear program with 7,218 rows, 7,250 columns and 63,129 non-zeros for the constant-temperature problem (15), and 21,649 rows, 21,713 columns and 252,718 non-zeros for the variable-temperature problem (16).

We begin by exploring the performance of the sequential elimination heuristic as the algorithm proceeds. Fig. 5 shows the number of elementary reaction steps in the basis, the mean squared error, and the value of the estimated rate constants for each iteration of the heuristic. Unlike the results for the prototype network in Fig. 2(A), the mean squared error shown in Fig. 5(A) roughly decreases as the iterations progress and as more elementary reaction steps are eliminated from the basis. It is likely that, given the problem complexity, it becomes easier to find better local optimal solutions as the problem size decreases with each pruning step. From Fig. 5(B), we observe that both redundant and governing reaction steps are eliminated from the basis as the iterations progress, in order to find a more parsimonious representation of the microkinetic model. In this case, we also observe that for some reactions the optimal solution obtained does not result in a change in the rate constant value from the initial value provided (e.g., iterations two and three, where reactions with rate constant value of 10^{-3} are also eliminated). This is typically the case for reactions that have no impact on the material balances of the measured species, which can take arbitrary values, and thus provide no additional predictive ability to the model. These elementary reaction steps can be pruned from the current library. Fig. 5(A) indicates that the heuristic algorithm terminates in five iterations, resulting in a network of 19 reactions.

Since the optimal solution of the NLP in (15) is expected to be sensitive to the initial value selection, we explore the variability of the resulting rate constant estimates when randomly perturbing the initial values of the decision variables. Figure S1 in the Supplementary Information shows the distribution of rate constant estimates for several random initializations, and indicates that in general the estimates are closely grouped around the median. More importantly, Figure S2 in the Supplementary Information shows that the estimates of the equilibrium constants for reversible reactions have a low spread for different initial rate constant values used to solve the NLP in (15). These results indicate that while different initial values may result in variability of the individual rate constant estimates, they do not significantly alter the equilibrium of species, which confirms model consistency with respect to initialization. Further, it should be noted that the constant temperature NLP (15) required a median solution time of 19.8 s (with interquartile range of 9.6 s), and hence, such sensitivity analysis with respect to initial conditions can be performed efficiently and should be conducted when applying this framework.

Upon termination of the heuristic algorithm, it may be possible to continue reducing the number of reactions in the network by solving the regularized optimization problem (16). Fig. 6 shows the number of reactions (i.e., model complexity), and mean squared error as a

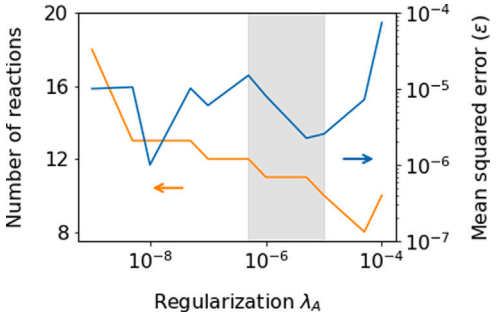

Fig. 6. Experiment 2: Model-order reduction via regularized optimization. Results correspond to model complexity (left), as measured by the number of reactions in the model, and predictive ability (right), using the mean squared error, as a function of the values of λ_A.

function of the choice of values of regularization hyperparameter λ_A. Since the values of the activation energy (for $E_{aj}^U < \infty$) have no direct impact on the complexity of the network, no regularization was performed with respect to E_a (i.e., $\lambda_{E_a} = 0$), and setting $A_j = 0$ is the only way of eliminating a reaction from the basis. Increasing values of λ_A result in decreasing model complexity with no major reduction in the predictive ability of the current reaction network. In this case, we were able to accomplish reduction to 11 total reactions (from 19 after termination of the sequential elimination heuristic algorithm) for values of regularization penalty $\lambda_A \sim 1 \times 10^{-6}$ within the shaded gray region in Fig. 6. While the mean squared error for gas-phase species remains relatively stable for increasing regularization in Fig. 6, the discovered reduced-order model becomes progressively less accurate in describing the distribution of intermediate adsorbed species.

Fig. 7 shows a comparison of the data used for model discovery and reduction (corresponding only to the gaseous species) and the true trajectories of the surface intermediates to the model predictions (i.e., the discovered microkinetic model). For brevity, only results corresponding to 450 K data are displayed. Similarly to the previous results shown in Fig. 3, the trajectories for the partial pressures of the gaseous species predicted by the discovered model accurately match those generated using the true model. For the present oligomerization example, since several intermediate reaction steps were eliminated in the reduced-order model, not all the concentration trajectories predicted by the discovered model for the surface species exactly match those of the true model that was used to generate the training data. Nonetheless, the trajectories of the fraction of available active sites and adsorbed propene (that is, species `msp1` and `msp3`, respectively, in Fig. 7) are accurately predicted by the discovered model, thus providing good insight into the overall surface coverage of the catalyst over time. These results illustrate the structural deficiencies described in Remark 1 that

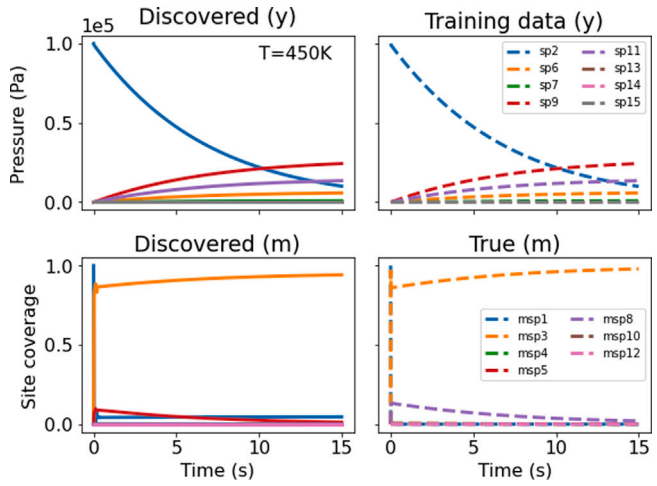

Fig. 7. Experiment 2: Comparison of data generated by simulating the true governing microkinetic model and the reduced-order model identified by the proposed framework. Only data for 450 K are shown. The top plots correspond to the gaseous species, and bottom plots correspond to the surface species. The plots on the left correspond to the model discovered by the proposed framework, and the plots on the right correspond to the data generated using the true model.

arise from: (1) the unobservable nature of the microkinetic reaction network model (that is, even if all of the reactions in the governing equations were kept in the discovered model and all species were measurable), as well as from (2) the approximate nature of the reduced-order model, as the model is structurally different from the governing mechanism. Most importantly, however, it is clear that the discovered model (and the underlying discretization approach) accurately capture the multi-scale nature of the true microkinetic model. Fig. 7 shows that the discovered ROM accurately represents the fast initial adsorption dynamics, and the subsequent slow oligomerization steps.

Having demonstrated that that the proposed framework yields a reduced-order approximate microkinetic model that accurately predicts the dynamics reflected in the training data (i.e., concentration of gas phase species) and with lower fidelity predicts the site coverage on the surface, we next assess extrapolation performance, which is important for the broader application of the discovered model in process optimization and control. Fig. 8 shows the parity plots comparing testing data for temperatures beyond the domain considered during training, and reduced-order model predictions for different temperatures for all gaseous species involved. For visualization purposes, the curves were shifted horizontally to accommodate data for all temperatures and results are displayed in arbitrary units on the axes. Although only training data corresponding to $\{440\ \text{K}, 450\ \text{K}, 460\ \text{K}\}$ were used, we note that the discovered reduced-order reaction network predicts the behavior of the true system at lower (i.e., $<$ 440 K) and higher (i.e., $>$ 460 K) temperatures than those reflected in the training data. This favorable extrapolation performance indicates that the discovered model can be reliably used in e.g. reactor design optimization considering a broader range of operating temperature which, because of the larger feasible space, is expected to result in improved overall decision-making performance.

Fig. 9 shows the mean squared error for the prediction of the concentration of each measured species as a function of temperature, ranging from 350 K to 550 K. Since the prediction error for each species concentration is displayed individually, it should be noted that the MSE displayed in Fig. 9 does not employ the normalization weight w_i introduced in the objective of (15) and (16), which is the metric shown in previous plots (Figs. 2, 5, 6) as indicated by ϵ in the labels of the axes. The discovered model has good performance within the domain of the training data (440 K to 460 K), where the error is typically the lowest near 450 K. While the mean squared error increases for lower and higher temperatures than those covered by the training data, the discovered reaction network model maintains reasonable predictive performance for the entire temperature range considered. Noting that the gaseous species have partial pressures that are of $\sim \mathcal{O}(10^3) - \mathcal{O}(10^5)$ Pa, their trajectories predicted by the discovered MK model are of reasonable accuracy.

Similarly, Fig. 10 shows the coefficient of determination (R^2) and the slope of the parity plots in Fig. 8 for the measured and predicted species trajectories, for which a value of one for both metrics indicates near perfect predictive performance. From these results it becomes clear that the learned reduced-order model maintains good extrapolation performance in a $\pm$50 K interval beyond the temperature range of the training data set, confirming the usefulness for practical applications such as optimal reactor design and control. The prediction quality of the ROM degrades considerably outside of this temperature regime. The physical intuition for this performance degradation is as follows: at least in part, model reduction is achieved by eliminating elementary reaction steps such that the remaining reactions become largely irreversible. The resulting ROM is in a sense similar to a macrokinetic model derived based on certain equilibrium or dominant species assumptions, except the ROM leverages data to arrive to such simplifications. As such, it is not surprising that at considerably higher and lower operating temperatures the accuracy of the concentration values predicted by the ROM deteriorates. In Figure S3 of the Supplementary Information we show that a model of higher complexity (of 19 elementary reactions) has in general better extrapolation performance than the current ROM of 11 elementary reactions. These results illustrate the trade-off between complexity and predictive performance which can be controlled by tuning λ_A based on the desired use case.

To gain further insight on extrapolation performance of the models resulting from the proposed framework, Fig. 11 shows the absolute error of the discovered model predictions relative to data generated by simulating the true reaction network, as a function of both time and the operating temperature. Although model performance is seen to deteriorate over longer time horizons for lower and higher temperatures, the absolute error is in general small compared to the magnitude of concentration values of the training data. In practice, nonetheless, it might be desirable to obtain experimental and/or simulation training data that are as close as possible to the operating conditions of the process in order to avoid sub-optimal or infeasible decision-making when using the model (resulting in e.g. considerable economic losses or safety risks).

Lastly, Fig. 12 shows the Sankey diagram of the discovered ROM of the oligomerization reaction network. While it is clear that the resulting reaction network is of lower complexity (even less complex than it is apparent, since reverse reaction steps are present but not shown in Fig. 4, but were eliminated in Fig. 12), the distribution of end products of the true (high dimensional) reaction network used to generate the training data are in agreement with that predicted by simulating the reduced-order discovered model. Even though these constitute the distribution of species at equilibrium, we have shown based on the results in previous figures that the discovered microkinetic model accurately predicts the temporal dynamics of the gaseous and (to a lesser extent) the surface species. By comparing Figs. 4 and 12, it is noticeable that the reduction in the number of elementary reactions is the result of representing the generation of each gas species with a single pathway. For example, in Fig. 4 species `sp6` is generated via two different pathways, while the reduced-order network in Fig. 12 shows that `sp6` is produced via a unique pathway from propene (i.e., `sp2`). The model reduction in this case occurs with respect to the reaction rates and not with respect to the species (typically denoted as "lumping" as e.g. in [49]), offering reduced-order modeling capabilities but providing resolution at the individual species level. Table 3 shows a complete list of all kinetic parameters used to generate the training data, as well as the kinetic parameters estimated in the final reduced-order reaction network.

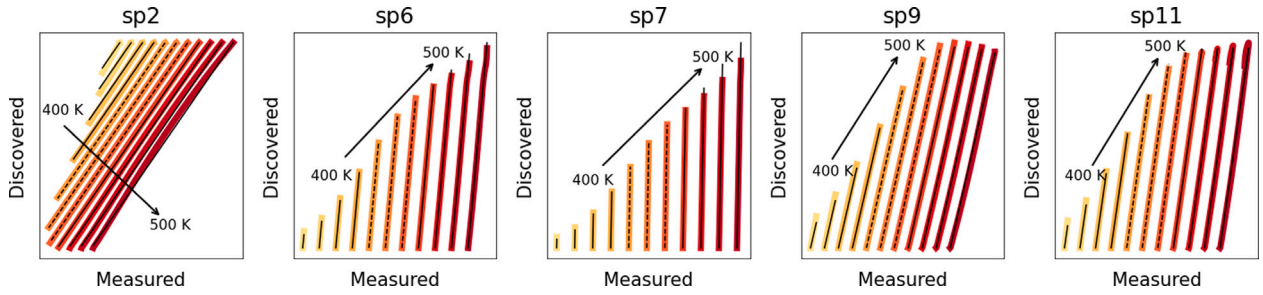

Fig. 8. Experiment 2: Parity plots comparing training data against predictions by the reduced-order reaction network. Black lines correspond to the (shifted) identity line of training (measurement) data, and colored lines correspond to the measurement versus model prediction data (discovered). Dashed black lines indicate the temperature instances used for training $\{440, 450, 460\}$ K.

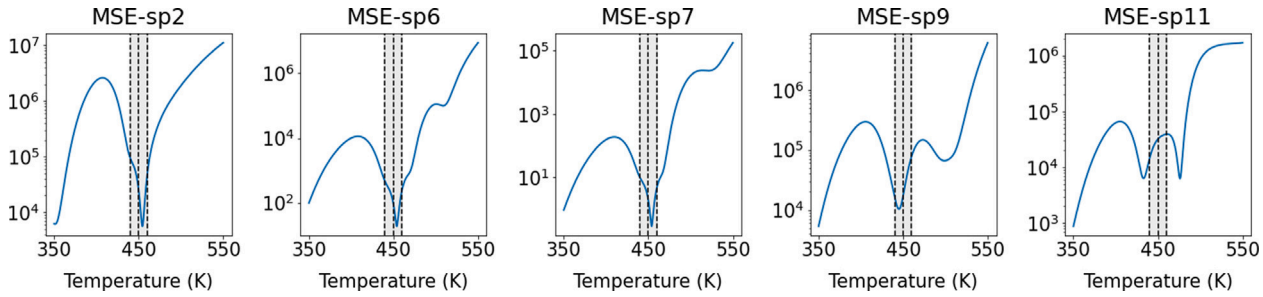

Fig. 9. Experiment 2: Mean squared prediction error corresponding to a sample of measured (gas phase) species as a function of temperature (y-axes have units of Pa^2). The vertical shaded gray region corresponds to the temperature range of the training data, and vertical dashed black lines indicate the three specific temperatures used for generating training data.

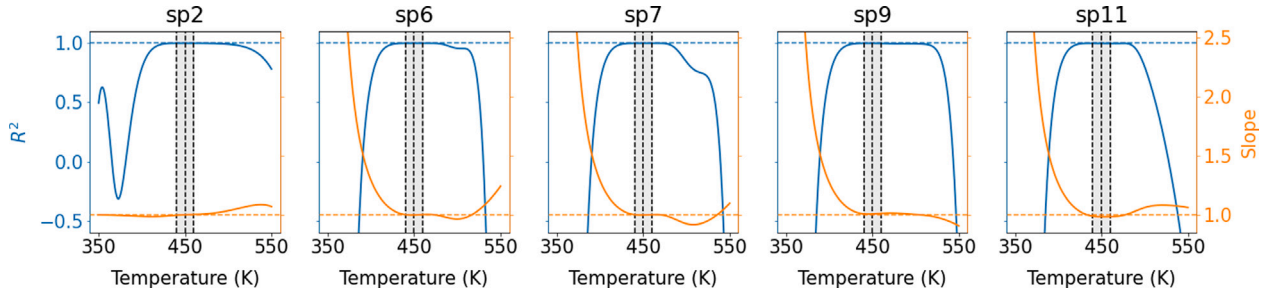

Fig. 10. Experiment 2: R^2 and slope of parity plots corresponding to different measured (gas phase) species as a function of reaction data simulated at different temperatures. The shaded gray region corresponds to the range of the training data, and vertical dashed black lines indicate the three specific temperatures used for generating training data.

7. Conclusions and directions of future research

In this work we have introduced a framework for addressing two important tasks in microkinetic modeling: (1) data-driven discovery of the governing microkinetic model from a large-set of candidate elementary reaction steps, and (2) model reduction by compressing the number of reaction pathways and species that map the input/inlet materials to the final/outlet products. The proposed framework offers several advantages relative to approaches previously reported in the literature, in that it can cope with unmeasurable species (which arise commonly as adsorbed intermediates in microkinetic models), it can handle reaction rate constants that are nonlinear functions of unknown parameters (as in the temperature-dependent case), and is well-suited to handle the often multi-scale and stiff nature of microkinetic models. An efficient sequential elimination heuristic was proposed to account for the large-scale nature of the associated dynamic nonlinear optimization problem. The main properties and performance of the proposed discovery framework were illustrated by means of two case studies: one corresponding to a prototype reaction network and another corresponding to the catalytic oligomerization of propene. While some physical interpretability is lost when reducing model order, the resulting models are considerably more parsimonious and thus valuable in process design optimization and control applications which do not require complete mechanistic resolution.

An important direction of future research is improving the scalability of the discovery framework by exploiting the sparse nature of the discretization via collocation on finite elements. Furthermore, it should be noted that the sequential elimination heuristic can be readily parallelized for improved computational efficiency upon appropriate initialization of the optimization problems associated with each of the temperature experiments. In this sense, Bayesian optimization techniques could be used to determine the optimal temperature experiments to perform to maximize the likelihood of discovery and simultaneously minimize the costs of experimentation. Lastly, an interesting

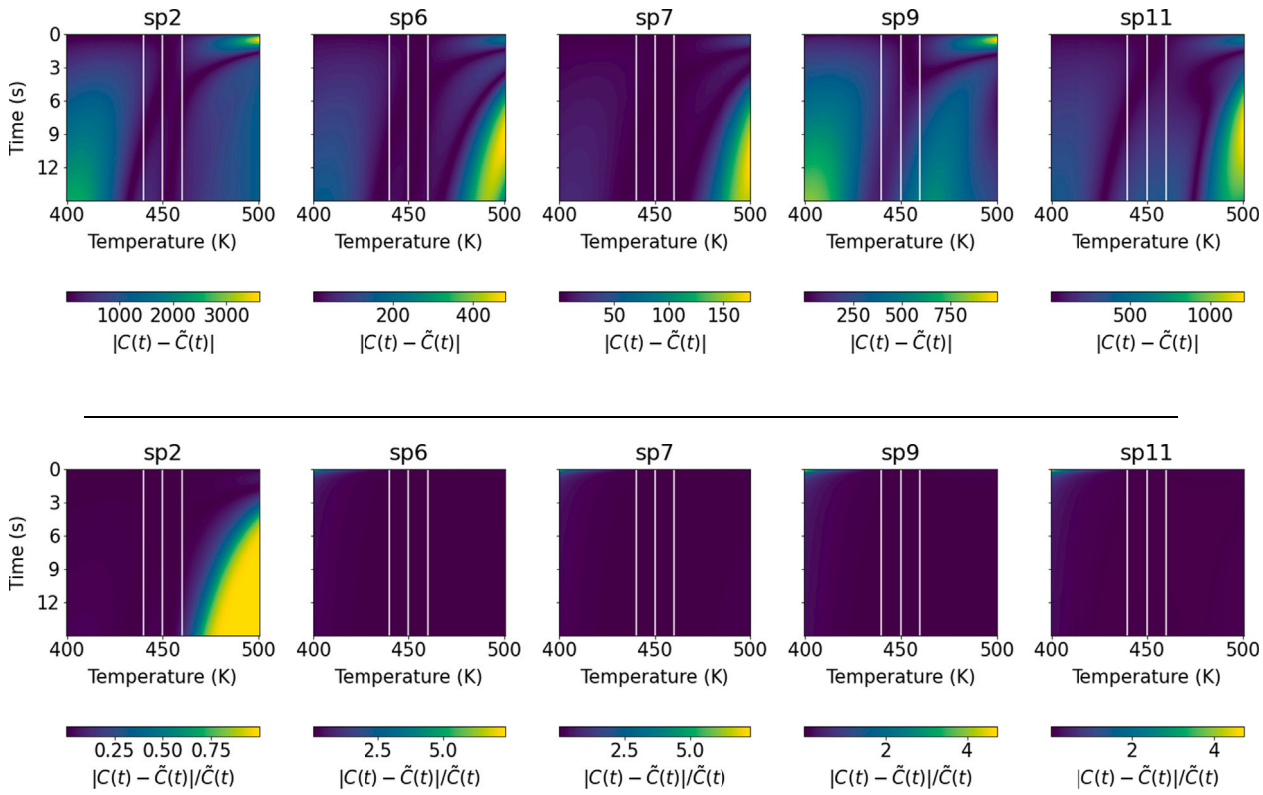

Fig. 11. Experiment 2: Top and bottom plots show, respectively, the absolute error and absolute percentage error for the different measured (gas-phase) species as a function of temperature and time of the reaction.

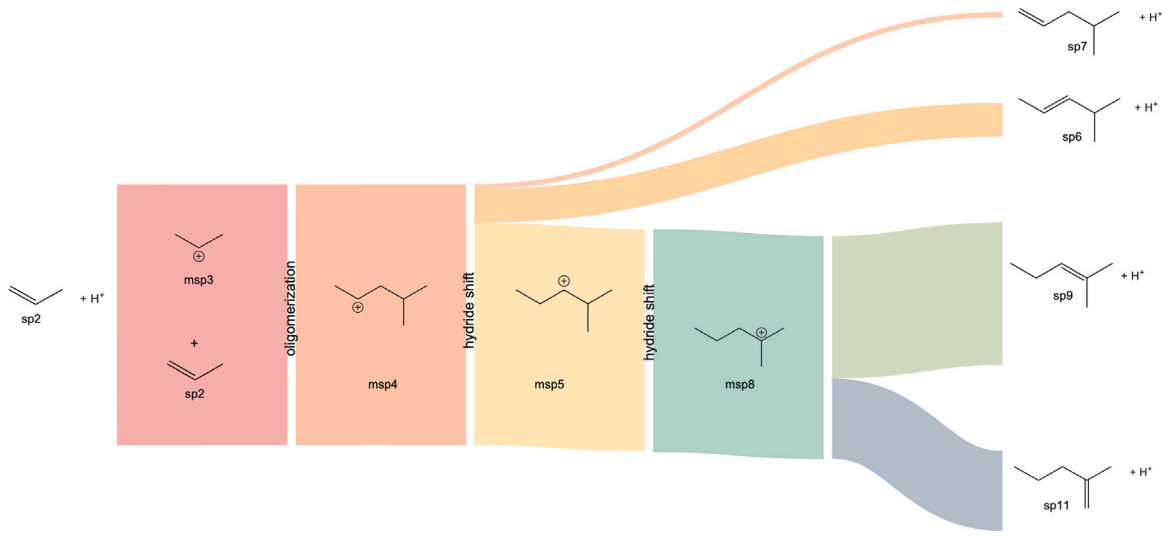

Fig. 12. Experiment 2: Sankey diagram showing the flux (molecule per site per second) of species in the oligomerization of propene up to C_6 compounds, for the reaction network discovered by the proposed machine learning framework.

next research step would consider the implementation of the proposed framework to different application domains of microkinetic modeling such as large scale biomass pyrolysis, gene regulation in metabolic networks, and systems involving novel catalytic materials. A candidate example to investigate in the future is the heterogeneous oxidation of carbon monoxide on platinum catalysts, where the proposed approach could be leveraged to discover reduced-order models that can reliably capture and help explain the oscillatory nature of the underlying microkinetic system, as well as its steady state multiplicity.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Table 3
Experiment 2: List of reaction rates and model parameters for the full-order network (the one used for generating training data) and that discovered via the proposed framework. Each reaction is considered reversible, with a forward and reverse reaction. Reaction rate parameters highlighted in red correspond to the eliminated reaction steps (i.e., $A = 0$) for the discovered reaction network. Pre-exponential factors (A) indicated with ‡ have units of Pa^{-1} s^{-1}; otherwise units are s^{-1}. † reaction rates correspond to steps involving methyl shift reactions, which are not used for generating data but are included as redundant steps in the reaction library used in the discovery process.

Model rates		Parameters of full-order model		Parameters of the discovered model	
Rate expression	*Reaction type*	A (s^{-1})	E_a (kcal mol^{-1})	A (s^{-1})	E_a (kcal mol^{-1})
k[0] × y[m] × y[sp2];	protonation	7.20E+04‡	6.81	2.01E−01‡	0.00E+00
k[1] × y[msp3];	deprotonation	6.40E+14	34.11	4.65E−01	0.00E+00
k[2] × y[sp2] ×y[msp3];	oligomerization	3.80E+04‡	17.41	9.85E+05‡	2.03E+01
k[3] × y[msp4];	betascission	4.20E+13	23.31	0.00E+00	0.00E+00
k[4] × y[msp4];	onetwohshift	8.10E+12	10.00	3.74E+04	3.99E−01
k[5] × y[msp4];	deprotonation	6.40E+14	26.56	0.00E+00	3.50E+01
k[6] × y[sp6] × y[m];	protonation	7.20E+04‡	10.05	0.00E+00	2.70E+01
k[7] × y[msp4];	deprotonation	6.40E+14	27.97	8.90E+04	6.25E+00
k[8] × y[sp7] × y[m];	protonation	7.20E+04‡	9.44	0.00E+00	3.50E+01
k[9] × y[msp5];	onetwohshift	8.10E+12	10.00	6.06E+04	5.89E+00
k[10] × y[msp5];	onetwohshift	8.10E+12	8.15	2.12E+06	9.55E+00
k[11] × y[msp5];	deprotonation	6.40E+14	26.56	1.03E+06	1.12E+01
k[12] × y[m] × y[sp6];	protonation	7.20E+04‡	10.05	0.00E+00	3.50E+01
k[13] × y[msp5];	deprotonation	6.40E+14	25.37	0.00E+00	2.67E+01
k[14] × y[sp9] × y[m];	protonation	7.20E+04‡	10.56	0.00E+00	0.00E+00
k[15] × y[msp5] †;	onetwomethylshift	2.50E+12	10.00	0.00E+00	1.34E+01
k[16] × y[msp8];	onetwohshift	8.10E+12	11.85	1.09E+04	8.91E+00
k[17] × y[msp8];	deprotonation	6.40E+14	27.96	7.20E+03	0.00E+00
k[18] × y[m] × y[sp9];	protonation	7.20E+04‡	9.44	5.70E+04‡	6.73E+00
k[19] × y[msp8];	deprotonation	6.40E+14	29.22	1.14E+04	0.00E+00
k[20] × y[sp11] × y[m];	protonation	7.20E+04‡	8.90	2.03E+05‡	5.95E+00
k[21] × y[msp10] †;	onetwohshift	8.10E+12	8.15	0.00E+00	0.00E+00
k[22] × y[msp10] †;	deprotonation	6.40E+14	25.98	0.00E+00	0.00E+00
k[23] × y[m] × y[sp13] †;	protonation	7.20E+04‡	10.30	0.00E+00	0.00E+00
k[24] × y[msp10] †;	deprotonation	6.40E+14	28.27	0.00E+00	0.00E+00
k[25] × y[m] × y[sp14] †;	protonation	7.20E+04‡	9.31	0.00E+00	0.00E+00
k[26] × y[msp10] †;	onetwomethylshift	2.50E+12	10.00	0.00E+00	6.84E+00
k[27] × y[msp12] †;	onetwohshift	8.10E+12	11.86	0.00E+00	1.00E+01
k[28] × y[msp12] †;	deprotonation	6.40E+14	28.57	0.00E+00	0.00E+00
k[29] × y[sp13] × y[m] †;	protonation	7.20E+04‡	9.18	0.00E+00	0.00E+00
k[30] × y[msp12] †;	deprotonation	6.40E+14	29.78	0.00E+00	0.00E+00
k[31] × y[sp15] × y[m] †;	protonation	7.20E+04‡	8.66	0.00E+00	0.00E+00

Acknowledgments

Partial financial support for this work is gratefully acknowledged from The National Science Foundation (NSF), USA, through the NSF CAREER Award No. 1454433 (recipient: M.B.), NSF Graduate Research Fellowships Program (GRFP), USA grant number DGE-1842165 (recipient: E.K.), and NSF Cooperative Agreement, USA No. EEC-164772 (recipient: L.J.B.). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. The authors also acknowledge the support of the Belgium American Education Foundation (BAEF) fellowship to E.K., and the Donald D. Harrington Fellowship, USA to F.L.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.cej.2023.142089.

References

[1] H. Lynggaard, A. Andreasen, C. Stegelmann, P. Stoltze, Analysis of simple kinetic models in heterogeneous catalysis, Prog. Surf. Sci. 77 (3–4) (2004) 71–137, http://dx.doi.org/10.1016/j.progsurf.2004.09.001.

[2] F.G. Botes, B. van Dyk, C. McGregor, The development of a macro kinetic model for a commercial Co/Pt/Al_2O_3 Fischer-Tropsch catalyst, Ind. Eng. Chem. Res. 48 (23) (2009) 10439–10447, http://dx.doi.org/10.1021/ie900119z.

[3] S. Mousavi, A. Zamaniyan, M. Irani, M. Rashidzadeh, Generalized kinetic model for iron and cobalt based Fischer–Tropsch synthesis catalysts: Review and model evaluation, Appl. Catal. A: General 506 (2015) 57–66, http://dx.doi.org/10.1016/j.apcata.2015.08.020.

[4] C. Albano, J. Papa, M. Ichazo, J. González, C. Ustariz, Application of different macrokinetic models to the isothermal crystallization of PP/talc blends, Compos. Struct. 62 (3) (2003) 291–302, http://dx.doi.org/10.1016/j.compstruct.2003.09.028.

[5] A.H. Motagamwala, J.A. Dumesic, Microkinetic modeling: a tool for rational catalyst design, Chem. Rev. 121 (2) (2020) 1049–1076, http://dx.doi.org/10.1021/acs.chemrev.0c00394.

[6] J. Ancheyta, S. Sánchez, M.A. Rodríguez, Kinetic modeling of hydrocracking of heavy oil fractions: A review, Catal. Today 109 (1) (2005) 76–92, http://dx.doi.org/10.1016/j.cattod.2005.08.015.

[7] S.A. Qader, G.R. Hill, Hydrocracking of gas oil, Ind. Eng. Chem. Process Des. Dev. 8 (1) (1969) 98–105, http://dx.doi.org/10.1021/i260029a017.

[8] C.D. Blasi, Modeling chemical and physical processes of wood and biomass pyrolysis, Prog. Energy Combust. Sci. 34 (1) (2008) 47–90, http://dx.doi.org/10.1016/j.pecs.2006.12.001.

[9] J.E. White, W.J. Catallo, B.L. Legendre, Biomass pyrolysis kinetics: A comparative critical review with relevant agricultural residue case studies, J. Analyt. Appl. Pyrolysis 91 (1) (2011) 1–33, http://dx.doi.org/10.1016/j.jaap.2011.01.004, URL https://www.sciencedirect.com/science/article/pii/S0165237011000088.

[10] A.F. Villaverde, A. Barreiro, A. Papachristodoulou, Structural identifiability of dynamic systems biology models, PLoS Comput. Biol. 12 (10) (2016) e1005153, http://dx.doi.org/10.1371/journal.pcbi.1005153.

[11] O.-T. Chis, J.R. Banga, E. Balsa-Canto, Structural identifiability of systems biology models: a critical comparison of methods, PLoS One 6 (11) (2011) e27755, http://dx.doi.org/10.1371/journal.pone.0027755.

[12] S. Vernuccio, L.J. Broadbelt, Discerning complex reaction networks using automated generator, AIChE J. 65 (8) (2019) e16663, http://dx.doi.org/10.1002/aic.16663.

[13] S. Vernuccio, E.E. Bickel, R. Gounder, L.J. Broadbelt, Microkinetic model of propylene oligomerization on Brønsted acidic zeolites at low conversion, ACS Catal. 9 (10) (2019) 8996–9008, http://dx.doi.org/10.1021/acscatal.9b02066.

[14] E. Koninckx, R. Gounder, J.W. Thybaut, L.J. Broadbelt, Kinetic modeling of ethene oligomerization on bifunctional nickel and acid β zeolites, Ind. Eng. Chem. Res. 61 (11) (2022) 3860–3876, http://dx.doi.org/10.1021/acs.iecr.1c04105.

[15] M. Evans, M. Polanyi, Inertia and driving force of chemical reactions, Trans. Faraday Soc. 34 (1938) 11–24, http://dx.doi.org/10.1039/TF9383400011.

[16] L.J. Broadbelt, S.M. Stark, M.T. Klein, Computer generated pyrolysis modeling: On-the-fly generation of species, reactions, and rates, Ind. Eng. Chem. Res. 33 (2019) 790–799, http://dx.doi.org/10.1021/ie00028a003.

[17] L.J. Broadbelt, S.M. Stark, M.T. Klein, Computer generated reaction networks: on-the-fly calculation of species properties using computational quantum chemistry,

Chem. Eng. Sci. 49 (24) (1994) 4991–5010, http://dx.doi.org/10.1016/0009-2509(94)00326-2.
[18] J. Thybaut, G. Marin, Single-event MicroKinetics: Catalyst design for complex reaction networks, J. Catalysis 308 (2013) 352–362, http://dx.doi.org/10.1016/j.jcat.2013.08.013.
[19] G. Marsden, P. Kostetskyy, R.-S. Sekiya, A. Hoffman, R.G. Songhyun Lee, D. Hibbitts, L.J. Broadbelt, Quantifying effects of active site proximity on rates of methanol dehydration to dimethyl ether over chabazite zeolites through microkinetic modeling, ACS Materials Au 2 (2) (2022) 163–175, http://dx.doi.org/10.1021/acsmaterialsau.1c00057.
[20] B.R. Hough, D.T. Schwartz, J. Pfaendtner, Detailed kinetic modeling of lignin pyrolysis for process optimization, Ind. Eng. Chem. Res. 55 (34) (2016) 9147–9153, http://dx.doi.org/10.1021/acs.iecr.6b02092.
[21] L.D. Dellon, C.-Y. Sung, D.J. Robichaud, L.J. Broadbelt, 110th anniversary: Microkinetic modeling of the vapor phase upgrading of biomass-derived oxygenates, Ind. Eng. Chem. Res. 58 (33) (2019) 15173–15189, http://dx.doi.org/10.1016/j.copbio.2019.02.005.
[22] A.J. Yanez, P. Natarajan, W. Li, R. Mabon, L.J. Broadbelt, Coupled structural and kinetic model of lignin fast pyrolysis, Energy Fuels 32 (2018) 1822–1830, http://dx.doi.org/10.1021/acs.energyfuels.7b03311.
[23] R.D. Priestley, C.J. Ellison, L.J. Broadbelt, J.M. Torkelson, Structural relaxation of polymer glasses at surfaces, interfaces, and in between, Science 309 (5733) (2005) 456–459, http://dx.doi.org/10.1126/science.1112217.
[24] L.H. Oakley, F. Casadio, K.R. Shull, L.J. Broadbelt, Modeling the evolution of crosslinked and extractable material in an oil-based paint model system, Angew. Chem. 57 (2018) 413–7417, http://dx.doi.org/10.1002/anie.201801332.
[25] Z. Ni, A.E. Stine, K.E. Tyo, L. J.Broadbelt, Curating a comprehensive set of enzymatic reaction rules for efficient novel biosynthetic pathway design, Metab. Eng. 65 (2021) 79–87, http://dx.doi.org/10.1016/j.ymben.2021.02.006.
[26] J. Strutz, J. Martin, J. Greene, L.J. Broadbelt, K. Tyo, Metabolic kinetic modeling provides insight into complex biological questions, but hurdles remain, Curr. Opin. Biotechnol. 59 (2019) 24–30, http://dx.doi.org/10.1016/j.copbio.2019.02.005.
[27] J.B. Rawlings, D.Q. Mayne, Model predictive control: Theory and design, 2, Nob Hill Pub. Madison, Wisconsin, 2020, URL https://sites.engineering.ucsb.edu/~jbraw/mpc/.
[28] E. Koninckx, J.G. Colin, L.J. Broadbelt, S. Vernuccio, Catalytic conversion of alkenes on acidic zeolites: Automated generation of reaction mechanisms and lumping technique, ACS Engineering Au 2 (3) (2022) 257–271, http://dx.doi.org/10.1021/ef980259r.
[29] P.V. Joshi, H. Freund, M.T. Klein, Directed kinetic model building: Seeding as a model reduction tool, Energy Fuels 13 (4) (1999) 877–880, http://dx.doi.org/10.1021/ef980259r.
[30] K. Edwards, T. Edgar, V. Manousiouthakis, Kinetic model reduction using genetic algorithms, Comput. Chem. Eng. 22 (1) (1998) 239–246, http://dx.doi.org/10.1016/S0098-1354(96)00362-6.
[31] A. Bhosekar, M. Ierapetritou, Advances in surrogate based modeling, feasibility analysis, and optimization: A review, Comput. Chem. Eng. 108 (2018) 250–267, http://dx.doi.org/10.1016/j.compchemeng.2017.09.017.
[32] S. Rangarajan, H. Tian, Improving the predictive power of microkinetic models via machine learning, Curr. Opin. Chem. Eng. 38 (2022) 100858, http://dx.doi.org/10.1016/j.coche.2022.100858.
[33] H. Tian, S. Rangarajan, Predicting adsorption energies using multifidelity data, J. Chem. Theory Comput. 15 (10) (2019) 5588–5600, http://dx.doi.org/10.1021/acs.jctc.9b00336.
[34] H. Tian, S. Rangarajan, Computing a global degree of rate control for catalytic systems, ACS Catal. 10 (22) (2020) 13535–13542, http://dx.doi.org/10.1021/acscatal.0c03150.
[35] H. Tian, S. Rangarajan, Machine-learned corrections to mean-field microkinetic models at the fast diffusion limit, J. Phys. Chem. C 125 (37) (2021) 20275–20285, http://dx.doi.org/10.1021/acs.jpcc.1c04495.
[36] L. Biegler, J. Damiano, G. Blau, Nonlinear parameter estimation: a case study comparison, AIChE J. 32 (1) (1986) 29–45, http://dx.doi.org/10.1002/aic.690320105.
[37] W. Chen, L.T. Biegler, S.G. Muñoz, An approach for simultaneous estimation of reaction kinetics and curve resolution from process and spectral data, J. Chemom. 30 (9) (2016) 506–522, http://dx.doi.org/10.1002/cem.2808.
[38] W. Chen, L.T. Biegler, S.G. Muñoz, Kinetic parameter estimation based on spectroscopic data with unknown absorbing species, AIChE J. 64 (10) (2018) 3595–3613, http://dx.doi.org/10.1002/aic.16334.
[39] T. Krumpolc, D. Trahan, D. Hickman, L. Biegler, Kinetic parameter estimation with nonlinear mixed-effects models, Chem. Eng. J. 444 (2022) 136319, http://dx.doi.org/10.1016/j.cej.2022.136319.
[40] K. Ghosh, S. Vernuccio, A.W. Dowling, Nonlinear reactor design optimization with embedded microkinetic model information, Front. Chem. Eng. 4 (2022) 898685, http://dx.doi.org/10.3389/fceng.2022.898685.
[41] M. Schmidt, H. Lipson, Distilling free-form natural laws from experimental data, Science 324 (5923) (2009) 81–85, http://dx.doi.org/10.1126/science.1165893.
[42] S.L. Brunton, J.L. Proctor, J.N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, Proc. Natl. Acad. Sci. 113 (15) (2016) 3932–3937, http://dx.doi.org/10.1073/pnas.1517384113.
[43] F. Lejarza, M. Baldea, Data-driven discovery of the governing equations of dynamical systems via moving horizon optimization, Sci. Rep. 12 (1) (2022) 1–15, http://dx.doi.org/10.1038/s41598-022-13644-w.
[44] F. Lejarza, M. Baldea, Sparse mathematical programming for fundamental learning of governing equation, in: M. Soroush, R.D. Braatz (Eds.), Artificial Intelligence in Manufacturing, Elsevier, 2023, *(To appear)* URL https://www.elsevier.com/books/artificial-intelligence-in-manufacturing/soroush/978-0-323-99134-6.
[45] M. Hoffmann, C. Fröhner, F. Noé, Reactive SINDy: Discovering governing reactions from concentration data, J. Chem. Phys. 150 (2) (2019) 025101, http://dx.doi.org/10.1063/1.5066099.
[46] F. Lejarza, M. Baldea, Discovering governing equations via moving horizon learning: The case of reacting systems, AIChE J. 68 (6) (2022) e17567, http://dx.doi.org/10.1002/aic.17567.
[47] C.M. Nguyen, B.A. De Moor, M.-F. Reyniers, G.B. Marin, Isobutene protonation in H-FAU, H-MOR, H-ZSM-5, and H-ZSM-22, J. Phys. Chem. C 116 (34) (2012) 18236–18249, http://dx.doi.org/10.1021/jp304081k.
[48] H.S. Fogler, Elements of Chemical Reaction Engineering, Pearson Educacion, 1999, URL http://websites.umich.edu/~elements/5e/.
[49] E. Koninckx, R. Gounder, J.W. Thybaut, L.J. Broadbelt, Kinetic modeling of ethene oligomerization on bifunctional nickel and acid β zeolites, Ind. Eng. Chem. Res. 61 (11) (2022) 3860–3876, http://dx.doi.org/10.1021/acs.iecr.1c04105.
[50] T. Hastie, R. Tibshirani, J.H. Friedman, J.H. Friedman, The elements of statistical learning: data mining, inference, and prediction, vol. 2, Springer, 2009, URL https://hastie.su.domains/ElemStatLearn/.
[51] N.M. Mangan, S.L. Brunton, J.L. Proctor, J.N. Kutz, Inferring biological networks by sparse identification of nonlinear dynamics, IEEE Trans. Mol. Biol. Multi-Scale Commun. 2 (1) (2016) 52–63, http://dx.doi.org/10.1109/TMBMC.2016.2633265.
[52] H. Schaeffer, S.G. McCalla, Sparse model selection via integral terms, Phys. Rev. E 96 (2) (2017) 023302, http://dx.doi.org/10.1103/PhysRevE.96.023302.
[53] P.A. Reinbold, D.R. Gurevich, R.O. Grigoriev, Using noisy or incomplete data to discover models of spatiotemporal dynamics, Phys. Rev. E 101 (1) (2020) 010203, http://dx.doi.org/10.1103/PhysRevE.101.010203.
[54] D.A. Messenger, D.M. Bortz, Weak SINDy: Galerkin-based data-driven model selection, Multiscale Model. Simul. 19 (3) (2021) 1474–1497, http://dx.doi.org/10.1137/20M1343166.
[55] P. Goyal, P. Benner, Discovery of nonlinear dynamical systems using a Runge–Kutta inspired dictionary-based sparse regression approach, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 478 (2262) (2022) 20210883, http://dx.doi.org/10.1098/rspa.2021.0883.
[56] L.T. Biegler, Nonlinear Programming: Concepts, Algorithms, and Applications To Chemical Processes, SIAM, 2010, http://dx.doi.org/10.1137/1.9780898719383, URL https://epubs.siam.org/doi/book/10.1137/1.9780898719383.
[57] G.v. Yablonskii, V. Bykov, V. Elokhin, A. Gorban, Kinetic Models of Catalytic Reactions, Elsevier, 1991, URL https://www.elsevier.com/books/kinetic-models-of-catalytic-reactions/yablonskii/978-0-444-88802-0.
[58] G.B. Marin, G.S. Yablonsky, D. Constales, Kinetics of Chemical Reactions: Decoding Complexity, John Wiley & Sons, 2019, URL https://www.wiley.com/en-us/Kinetics+of+Chemical+Reactions:+Decoding+Complexity,+2nd+Edition-p-9783527342952.
[59] N.M. Mangan, J.N. Kutz, S.L. Brunton, J.L. Proctor, Model selection for dynamical systems via sparse regression and information criteria, Proc. R. Soc. A: Math. Phys. Eng. Sci. 473 (2204) (2017) 20170009, http://dx.doi.org/10.1098/rspa.2017.0009.
[60] W.E. Hart, C.D. Laird, J.-P. Watson, D.L. Woodruff, G.A. Hackebeil, B.L. Nicholson, J.D. Siirola, et al., Pyomo-optimization modeling in python, vol. 67, Springer, 2017, http://dx.doi.org/10.1007/978-3-319-58821-6.
[61] B. Nicholson, J.D. Siirola, J.-P. Watson, V.M. Zavala, L.T. Biegler, Pyomo.dae: a modeling and automatic discretization framework for optimization with differential and algebraic equations, Math. Program. Comput. 10 (2) (2018) 187–223, http://dx.doi.org/10.1007/s12532-017-0127-0.
[62] A.S. Drud, CONOPT—a large-scale GRG code, ORSA J. Comput. 6 (2) (1994) 207–216, http://dx.doi.org/10.1287/ijoc.6.2.207.