When machine learning meets multiscale modeling in chemical reactions

Article *in* The Journal of Chemical Physics · September 2020

DOI: 10.1063/5.0015779

CITATIONS
17

READS
141

4 authors:

Wuyue Yang
Yanqi Lake Beijing Institute of Mathematical Sciences and Applications
17 PUBLICATIONS **537** CITATIONS
SEE PROFILE

Peng Liangrong
Minjiang University
18 PUBLICATIONS **469** CITATIONS
SEE PROFILE

Yi Zhu
Tsinghua University
67 PUBLICATIONS **1,136** CITATIONS
SEE PROFILE

Liu Hong
Sun Yat-Sen University
93 PUBLICATIONS **1,308** CITATIONS
SEE PROFILE

The Journal of Chemical Physics

When machine learning meets multiscale modeling in chemical reactions

Cite as: J. Chem. Phys. **153**, 094117 (2020); https://doi.org/10.1063/5.0015779
Submitted: 30 May 2020 . Accepted: 10 August 2020 . Published Online: 04 September 2020

Wuyue Yang, Liangrong Peng, Yi Zhu, and Liu Hong

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

When machine learning meets multiscale modeling in chemical reactions

Cite as: J. Chem. Phys. 153, 094117 (2020); doi: 10.1063/5.0015779
Submitted: 30 May 2020 • Accepted: 10 August 2020 •
Published Online: 4 September 2020

Wuyue Yang,[2] Liangrong Peng,[3] Yi Zhu,[2] and Liu Hong[1,a)]

AFFILIATIONS

[1]School of Mathematics, Sun Yat-sen University, Guangzhou 510275, People's Republic of China
[2]Yau Mathematical Sciences Center, Tsinghua University, Beijing 100084, People's Republic of China
[3]College of Mathematics and Data Science, Minjiang University, Fuzhou 350108, People's Republic of China

Note: This paper is part of the JCP Special Topic on Machine Learning Meets Chemical Physics.
[a)]**Author to whom correspondence should be addressed:** hongliu@sysu.edu.cn

ABSTRACT

Due to the intrinsic complexity and nonlinearity of chemical reactions, direct applications of traditional machine learning algorithms may face many difficulties. In this study, through two concrete examples with biological background, we illustrate how the key ideas of multiscale modeling can help to greatly reduce the computational cost of machine learning, as well as how machine learning algorithms perform model reduction automatically in a time-scale separated system. Our study highlights the necessity and effectiveness of an integration of machine learning algorithms and multiscale modeling during the study of chemical reactions.

I. INTRODUCTION

Accompanied with matter synthesis and decomposition, energy storage and release, and biofunction activation and deactivation, chemical reactions play a fundamental role in multiple disciplines,[1] including biology, chemical engineering, materials science, and so on. They help to model complicated phenomena in nature by an explicit reaction network, to allow the interpretation of observed data through quantitative mathematical equations, and to translate varied experimental conditions into tunable reaction rates and reaction orders. Due to their high complexity and nonlinearity, previous studies on chemical reactions heavily rely on sophisticated mathematical analysis and first-principle calculations, like quantum chemistry.[2]

The first mission of studies on chemical reactions is to obtain the proper mathematical model, which can interpret the observed phenomena and data. Even though there are some empirical laws and pre-knowledge on the reaction networks that may help to build the model, the parameters like reaction rates are usually deeply buried inside the massive data. Recent rapid developments of various machine learning algorithms, especially deep neural networks, make it possible to infer the reaction networks and parameters efficiently. For example, Mangan *et al.*[3] proposed an implicit sparse identification of nonlinear dynamics to infer hidden biochemical reaction networks, with emphasis on the rational nonlinear forms of the governing dynamics. Hu *et al.*[4] constructed a so-called ODENet (short for Ordinary Differential Equations Network), which was used for explicitly modeling the Lotka–Volterra type dynamics and actin growth in the presence of medium-level noise. From a stochastic perspective, the chemical reaction system was modeled as a continuous-time Markov chain, whose propensity function was reconstructed as a combination of the pre-designed basis functions based on maximization of the log-likelihood function.[5]

In application, Costello and Martin[6] showed that a supervised learning method can predict the metabolic pathway dynamics from proteomics data, which may be used to design various bioengineered systems. Yang *et al.*[7] revealed that the aggregation rates of amyloid proteins could be reliably estimated based on the feed-forward fully connected neural network and feature selection. In organic chemistry, given some reactants and external conditions, all possible reactions were ranked by a machine learning approach, including a reactive site classifier and a ranking model, with the top-ranked mechanism corresponding to the major products.[8] The Gaussian

process regression was utilized to construct the potential energy surface of the HO_x system,[9] which could reduce the computational cost and meanwhile guarantee the convergence with few training points. For more applications of machine learning to chemical reactions, including the supervised and unsupervised learning, see, e.g., Refs. 10–14.

The successful attempts of machine-learning-based modeling pave a new way for understanding the complicated dynamics of chemical reactions. However, most chemical reactions involve plenty of reactants, multiple potential reaction routines, diverse reaction rates, and so on. Without considering the intrinsic multi-component and multiscale nature of the system, direct applications of machine learning algorithms may face inevitable difficulties (see examples below for details). Motivated by the requirements on a real complex system, especially a simultaneous maintenance of the efficiency of macroscopic models and the accuracy of microscopic models, the view of multiscale modeling is introduced. It focuses on a proper separation of the system into several scales with minimum overlap, a correct characterization of relations among different levels of physical models, as well as a systematical procedure of coarse-graining.[15] Multiscale modeling offers a unified way to examine the chemical reactions by looking into the reactions occurring at different time scales and the relations among them. Therefore, it is expected that a proper integration of machine learning algorithms with ideas and methodologies from multiscale modeling and analysis will shed light into this field. Moreover, this is also the major motivation of our current study.

To be concrete, we will justify our arguments from two aspects: (1) By using the explicit correspondence between mesoscopic chemical master equations and macroscopic mass-action equations in the Kurtz's limit, the task of learning detailed probability distribution function (PDF) is converted into learning low-order moments. Obviously, the latter is much simpler. In this case, the computational cost of direct machine learning is greatly reduced by incorporating the idea of multiscale modeling. (2) When fast and slow reactions appear simultaneously in the same system, meaning there is a time-scale separation among the chemical reactions, the ODENet, a kind of machine learning algorithms with sparse identification, shows an astonishing ability of deriving simplified models under quasi-steady-state approximation (QSSA) automatically. Therefore, machine learning could help to model multiscale chemical reactions too. These two examples clearly demonstrate that a proper integration of machine learning and multiscale modeling will greatly facilitate our study of chemical reactions.

During the preparation of this paper, we notice two interesting papers published by E and his co-workers. The first one[16] presented a multiscale modeling framework for systems without scale-separation, which used the moment-closure of Boltzmann equations as an illustration. It is pertinent to the first example of our current study. Though ours is an ODE system and theirs is PDE (Partial Differential Equation)-based, the main ideas of the two are quite close to each other. In the second paper,[17] E *et al.* discussed the integration of machine learning with physics-based modeling, with emphasis on imposing physical constraints, like symmetry and frame-indifference, and obtaining optimal datasets. Although these two aspects have not been carefully examined in our study, we do believe they are essential for modeling based on machine learning.

The whole paper is organized as follows: The basic architecture of the ODENet, a special kind of machine learning algorithms, which is designed to derive the explicit form of ODEs from the pre-given time-series data, is introduced in Sec. II, along with the basic ideas and techniques for multiscale modeling and analysis for chemical reactions, including the Kurtz's limit from chemical master equations to mass-action equations, and the quasi-steady-state approximation. In Sec. III, we illustrate our key ideas through two examples—the development and differentiation of cells, as well as the self-regulatory gene transcription and translation. The power of an integration of machine learning and multiscale modeling can be clearly learned. The last section contains some discussions.

II. METHODS

A. Basic architecture of ODENet

The ordinary differential equations network was proposed[4,18] as a continuous version of the famous ResNet[19] for dealing with time-series data modeled by ordinary differential equations (ODEs). Mathematically, the consecutively repeating building blocks – each layer of a residual network can be expressed as $\mathbf{y}_{k+1} = \mathbf{y}_k + f(\mathbf{y}_k; \theta_k)$, where $\mathbf{y}_k$ is the output of k'th hidden layer, $\mathbf{y}_{k+1}$ is the output of (k + 1)'th hidden layer, and $f(\mathbf{y}_k; \theta_k)$ represents the function of a network layer parameterized by θ_k. After a simple algebraic transformation, we can get $\frac{\mathbf{y}_{k+1}-\mathbf{y}_k}{h} = \frac{f(\mathbf{y}_k;\theta_k)}{h}$ with the time interval h, which is the Euler's discretization scheme of ODEs,

$$\frac{d\mathbf{y}}{dt} = \frac{f(\mathbf{y};\theta)}{h}. \tag{1}$$

As a consequence, the forward propagation process of a residue network is actually equivalent to the numerical solution of a group of corresponding ordinary differential equations. Alternatively, it also means, if we use an ODE solver to solve the ODEs directly, the process of forward propagation in a residue network is accomplished too. This significant finding lays down the theoretical foundation of ODENet. The application of ODE solvers could easily cope with input data with unequal time intervals, fight against medium-level noises, control the numerical errors, and dynamically adjust its convergence rate.

To enhance the ability of learning the explicit governing ODEs from the pre-given time-series data, in a previous study, we combined the ODENet with symbolic regression and sparse identification.[4] Symbolic regression means the explicit form of $f(\mathbf{y};\theta)$, which is characterized through coefficients θ by expanding $f(\mathbf{y})$ on a complete set of orthogonal basis functions $\Gamma(\mathbf{y})$, i.e. $f(\mathbf{y};\theta) = \theta\Gamma(\mathbf{y})$. Consequently, the learning of ODEs helps to determine the unknown coefficients θ from the data. In practice, polynomials are the most often used basis functions. Sparse identification means, in the loss function L, an additional regulation term $||\theta||_1$ is added in order to remove redundant free parameters θ as many as possible. Therefore, the loss function consists of two parts,

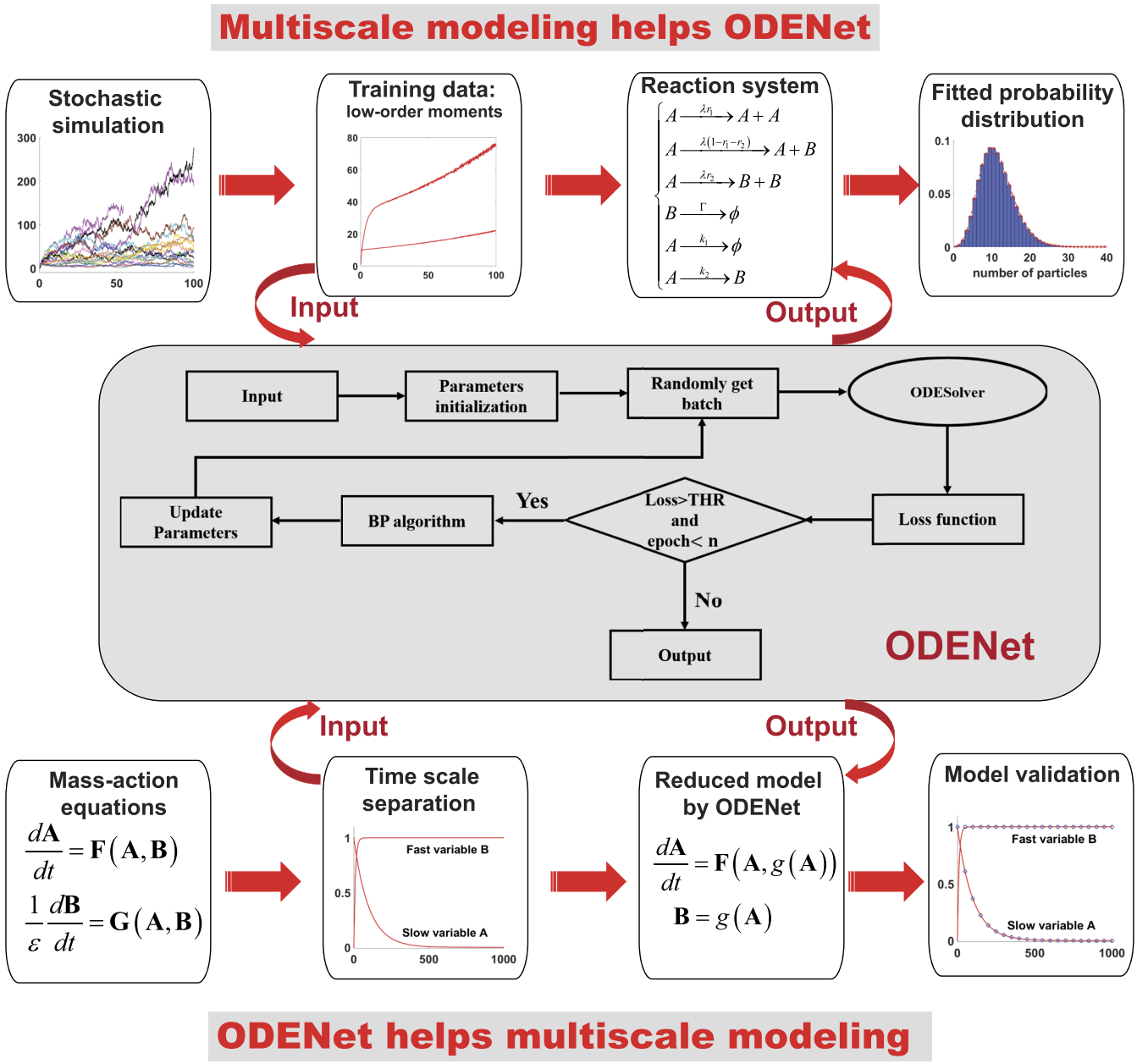

FIG. 1. An integration of ODENet with multi-scale modeling in the study of chemical reactions. The upper panel illustrates the ODENet-based learning procedure of the reaction mechanism under the help of multiscale modeling, while the lower panel gives the automatic procedure for model reduction aided by ODENet. The flowchart of ODENet is shown in the middle.

$$L = \|\mathbf{y} - \widehat{\mathbf{y}}\|_1 + \varepsilon\|\theta\|_1. \tag{2}$$

The first part controls the difference between the training data $\mathbf{y}$ and the predicted data $\widehat{\mathbf{y}}$ by ODENet, while the second part aims at a minimal model according to the Occam's razor. Here, ε is a hyper-parameter. To obtain the optimal parameters θ, the classical Back Propagation (BP) algorithm[20] is adopted to make an update, which will be repeated for many iterations until the loss function converges or is less than the threshold. Please see Fig. 1 for the flowchart of ODENet or refer to Ref. 4 for further details.

B. Multiscale modeling of chemical reactions

Without loss of generality, we consider a chemical system with N species and M reactions,[21]

$$\nu_{1j}S_1 + \nu_{2j}S_2 + \cdots + \nu_{Nj}S_N \xrightarrow{k_j} \nu'_{1j}S_1 + \nu'_{2j}S_2 + \cdots + \nu'_{Nj}S_N, \quad (3)$$

where $j = 1, 2, \ldots, M$, $k_j > 0$ denotes the rate constant of the reaction j. The nonnegative integers $\{\nu_{ij}\}$ and $\{\nu'_{ij}\}$ denote the stoichiometric coefficients of the reactants and products, respectively. The stoichiometric matrix is defined as $U = [(u_{ij})]_{N\times M}$ with elements being $u_{ij} = \nu'_{ij} - \nu_{ij}$.

1. Chemical master equations

We focus on the molecular number of species $(S_1, S_2, \ldots, S_N)$ represented by a stochastic variable $\mathbf{n} = (n_1, n_2, \ldots, n_N)^T$ in a reaction vessel of volume V. When the magnitude of $(n_1, n_2, \ldots, n_N)^T$ is relatively small compared with the Avogadro's constant, the randomness comes into play due to the intrinsic stochasticity of molecular collisions. From the perspective of ensemble average, we can denote the probability of the system in state $\mathbf{n}$ by $p(\mathbf{n}, t)$, where the time dependence is usually omitted as $p(\mathbf{n})$.

With respect to the reactions in (3), the probability distribution obeys the following chemical master equations (CMEs), in a compact form as

$$\frac{d}{dt}p(\mathbf{n}) = \sum_{j=1}^{M}[p(\mathbf{n}-\mathbf{u}_j)\Phi_j(\mathbf{n}-\mathbf{u}_j) - p(\mathbf{n})\Phi_j(\mathbf{n})], \quad (4)$$

accompanied by the initial condition $p(\mathbf{n})|_{t=0} = p_0(\mathbf{n})$. Here, $\mathbf{u}_j$ is the j'th column of stoichiometric matrix $U = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_M)$, and $\Phi_j(\mathbf{n})$ is the mesoscopic propensity function characterizing the probability $\Phi_j(\mathbf{n})dt$, for which the j'th reaction occurs once within the time interval $[t, t + dt)$.

In general, the state-dependent mesoscopic propensity function $\Phi_j(\mathbf{n})$ of CMEs is assumed to follow the laws of mass-action,

$$\Phi_j(\mathbf{n}) = k_j V \prod_{l=1}^{N}\left(V^{-\nu_{lj}} C_{n_l}^{\nu_{lj}}\right), \quad (5)$$

which is the product of the molecular number in a polynomial form and the rate coefficient k_j for $n_l \geq \nu_{lj}$, $\forall l = 1, 2, \cdots, N$. When there are not enough particles to form a reactant, saying S_l, such that $n_l < \nu_{lj}$, the propensity reduces to zero, $\Phi_j(\mathbf{n}) = 0$. Here, the combination number $C_{n_l}^{\nu_{lj}} = n_l!/[(\nu_{lj}!)(n_l - \nu_{lj})!]$.

2. Stochastic simulations

In most cases, the chemical master equations in (4) are a huge group of ordinary differential equations, which are quite computational consuming. Alternative efficient sampling algorithms are needed. The Gillespie algorithm (GA),[22] which is able to generate typical time evolutionary trajectories of species according to the reaction mechanisms and reaction rates in a stochastic way, may be the most famous one.

Gillespie implemented two stochastic simulation algorithms. The one is the direct method (DM) and the other is the first-reaction method (FRM). These two methods are theoretically equivalent, so we here only implement the first-reaction method. The FRM generates putative time for every reaction and chooses a time at which the corresponding reaction would occur while no other reaction occurred before that.

By independently running the Gillespie algorithm once again, statistics on the corresponding stochastic trajectories will converge to the correct probability distribution given by the chemical master equations. They constitute the training dataset to feed into machine learning algorithms.

3. Moment-closure equations in Kurtz's limit

Although CMEs provide a relatively accurate way to model general chemical reaction systems, it leads to a heavy burden in both modeling and experiments, since the dimensionality of the transition matrix is usually extremely high. Moreover, the time-consuming numerical simulations of CMEs become a common bottleneck when the number of species or reactions is large. In order to make a simplification, we turn to look at the mean density of species,

$$c_i = \sum_{\mathbf{n}} V^{-1} n_i p(\mathbf{n}). \quad (6)$$

To deduce the macroscopic kinetics of the concentration c_i, we multiply (4) by the number density $V^{-1}n_i$ and take summation over all admissible state $\{\mathbf{n}\}$ on both sides, which yields

$$\begin{aligned}\frac{d}{dt}\sum_{\mathbf{n}} V^{-1}n_i p(\mathbf{n}) &= \sum_{j=1}^{M}\sum_{\mathbf{n}} V^{-1}n_i[p(\mathbf{n}-\mathbf{u}_j)\Phi_j(\mathbf{n}-\mathbf{u}_j) - p(\mathbf{n})\Phi_j(\mathbf{n})] \\ &= \sum_{j=1}^{M} u_{ij}\left[\sum_{\mathbf{n}} V^{-1}p(\mathbf{n})\Phi_j(\mathbf{n})\right], \end{aligned} \quad (7)$$

where, in the last step, we have used the variable substitution $\mathbf{n} - \mathbf{u}_j = \mathbf{n}'$ and have neglected the boundary terms. Direct calculation shows that the volume density of the mesoscopic propensity function deduces $V^{-1}\Phi_j(\mathbf{n}) = \phi_j(V^{-1}\mathbf{n}) + \mathcal{O}(V^{-1})$, with $\phi_j(\mathbf{c}) = k_j \prod_{l=1}^{N} c_l^{\nu_{lj}}/\nu_{lj}!$ being the usual macroscopic propensity function.

Taking the limit of $V \to +\infty$, $\mathbf{n} \to +\infty$ while keeping $V^{-1}\mathbf{n}$ finite, we have the following mass-action equations (MAEs)

$$\frac{d}{dt}c_i(t) = \sum_{j=1}^{M}(\nu'_{ij} - \nu_{ij})\phi_j(\mathbf{c}), \quad (8)$$

on a nonnegative continuous state space $\{\mathbf{c}|\mathbf{c} \in \mathbb{R}^N_{\geq 0}\}$. The MAEs in (8) are the macroscopic description derived from the mesoscopic CMEs of the reaction system (3). A rigorous mathematical justification of the above limit process was first done by Kurtz in the 1970s.[23] A similar procedure can be carried out for high-order moments of PDF, like the second-order variance studied in the first example in Sec. III.

Remark II.1. According to the results proved by Kurtz,[23] in the limit of $V \to +\infty$, for any finite time, the solution of CMEs in (4) will converge in probability to the solution of the corresponding MAEs in (8), provided the initial conditions $\lim_{V\to+\infty} V^{-1}\mathbf{n}(t=0) = \mathbf{c}(t=0)$, which is a straightforward consequence of the Central Limit theorem. Our derivation above from the CMEs in (4) to MAEs in (8) for the reaction system (3) serves as a formal illustration of the Kurtz's theorem.

C. Model reduction by QSSA

Consider a very general chemical reaction system with time scale separation, which is written in an abstract matrix form,

$$\begin{cases} \dfrac{d\mathbf{A}}{dt} = \mathbf{F}(\mathbf{A},\mathbf{B}), \\ \dfrac{1}{\varepsilon}\dfrac{d\mathbf{B}}{dt} = \mathbf{G}(\mathbf{A},\mathbf{B}), \end{cases} \tag{9}$$

where **A** and **B** respectively stand for slow and fast variables after some kind of proper non-dimensionalization. $\varepsilon \ll 1$ is a small parameter characterizing the gap between fast and slow time scales in the dynamics.

With respect to the above dynamics, QSSA states that in the slow time scale dominated by the changes in **A**, **B** can be regarded as remaining at a dynamically equilibrium state (quasi-steady-state) due to their fast reactive nature, meaning approximately we have $\mathbf{G}(\mathbf{A}, \mathbf{B}) = 0$. If **B** can be uniquely solved from this algebraic relation, i.e. $\mathbf{B} = \mathbf{g}(\mathbf{A})$, the original time-scale separated dynamics could be simplified as

$$\frac{d\mathbf{A}}{dt} = \mathbf{F}(\mathbf{A},\mathbf{g}(\mathbf{A})), \quad \mathbf{B} = \mathbf{g}(\mathbf{A}). \tag{10}$$

QSSA is a very classical model reduction approach and has been widely used in the study of chemical reactions, see e.g. Ref. 24 for details.

III. RESULTS AND DISCUSSION

In this section, through two concrete examples – the single proliferative compartment model (SPCM) of IFE (interfollicular epidermis) maintenance and a gene network with autoregulatory negative feedback, we are going to show how machine learning and multiscale modeling help each other in the study of chemical reactions.

A. Single proliferative compartment model

1. The basic model

In the first example, the SPCM of IFE maintenance considered by Clayton *et al.*[25] is adopted to illustrate how multiscale modeling helps to reduce the computational cost of machine learning during inferring the detailed reaction mechanisms and reaction rates. According to the observations by Clayton *et al.*,[25] the clone fate of proliferating epidermal progenitor cells (EPCs) plays an essential role in adult epidermal homeostasis. Also, the key clone size distribution is modeled by chemical master equations.

Consider two reactant species in the single-proliferative compartment model, including proliferating EPCs (denoted as A) and post-mitotic cells in the basal layer (B). There are four reactions which involve symmetric and asymmetric cell division. As shown in Fig. 2(a), A has a unlimited self-reproduction potential at a rate $r_1\lambda$ in order to maintain the epidermis, where λ is the integrated division rate of proliferating EPCs A. A can also differentiate into $A + B$ or $B + B$ at the rate of $(1 - r_1 - r_2)\lambda$ or $r_2\lambda$, respectively. In the transfer process, B cells in the basal layer leak from the clone-size distribution at rate Γ. The above reaction system reduces to the one studied by Clayton *et al.*[25] with symmetric division rates $r_1 = r_2$. With respect to the SPCM and coefficients given in Fig. 2, 10^6 times independent stochastic simulations are performed by using the Gillespie algorithm, which generate the training dataset to feed into our following ODENet-based machine learning procedure.

2. Learning mass-action equations by ODENet

Here, our major goal is to obtain the SPCM in Fig. 2(a) and the explicit rate constants. However, a direct application of ODENet to learn the time evolution of $p(n_A, n_B)$ (or the chemical master equations) from the training data generated by stochastic simulations is prohibited due to heavy computational cost. Therefore, by taking advantage of the knowledge of multiscale modeling in chemical reactions, especially the explicit correspondence between mesoscopic chemical master equations and macroscopic mass-action equations in the Kurtz's limit (see the Section of Methods), we turn to learn low-order moments instead of the probability distribution function governed by chemical mass-action equations. A similar idea has been previously applied by one of the authors to investigate the kinetics of amyloid aggregation, but without referring to machine learning.[26,27]

With respect to the training data of $\langle n_A\rangle = \sum n_A p(n_A, n_B)$ and $\langle n_B\rangle = \sum n_B p(n_A, n_B)$ by averaging the stochastic trajectories generated through Gillespie algorithms, we need to determine the exact types of chemical reactions involving these two reactants, their reaction orders, and reaction rate constants. Without loss of generality, here, we make a cutoff on the chemical reactions up to the second order, corresponding to a combination of A, B, $A + A$, $A + B$, $B + B$, which reads

$$\begin{cases} \frac{d\langle n_A\rangle}{dt} = \alpha_{11}\langle n_A\rangle + \alpha_{12}\langle n_B\rangle + \alpha_{13}\langle n_A\rangle^2 + \alpha_{14}\langle n_A\rangle\langle n_B\rangle + \alpha_{15}\langle n_B\rangle^2, \\ \frac{d\langle n_B\rangle}{dt} = \alpha_{21}\langle n_A\rangle + \alpha_{22}\langle n_B\rangle + \alpha_{23}\langle n_A\rangle^2 + \alpha_{24}\langle n_A\rangle\langle n_B\rangle + \alpha_{25}\langle n_B\rangle^2. \end{cases} \tag{11}$$

Now, we implement the ODENet to learn the dynamics in (11). Clearly, not all reaction rate constants will appear in the final model. Those redundant coefficients will be picked out by ODENet and removed through sparse identification. Besides the L1 norm in the loss function, in practice, the sparse identification is carried out in the following way. After every thousands of iterations during training, those parameters, whose absolute values are at least 10^{-3} smaller than the others, will be picked out and forced to be zero. In this way, we have fewer and fewer free parameters until the loss function converges. The above procedure will be repeated for hundreds of times independently, in order to avoid the model from trapping into local minima. In the current case, after training and regression, only three non-zero coefficients $\alpha_{11} = 0.0079$, $\alpha_{21} = 1.0903$, and $\alpha_{22} = -0.3094$ are kept in the final results.

3. Learning high-order moment equations

In the last step, since all coefficients in front of quadratic terms in (11) are removed, we can make a conclusion that only first-order reactions are present in the current system. Then, with respect to above learned dynamics and coefficients, the desired single proliferative compartment model as shown in Fig. 2(a) is

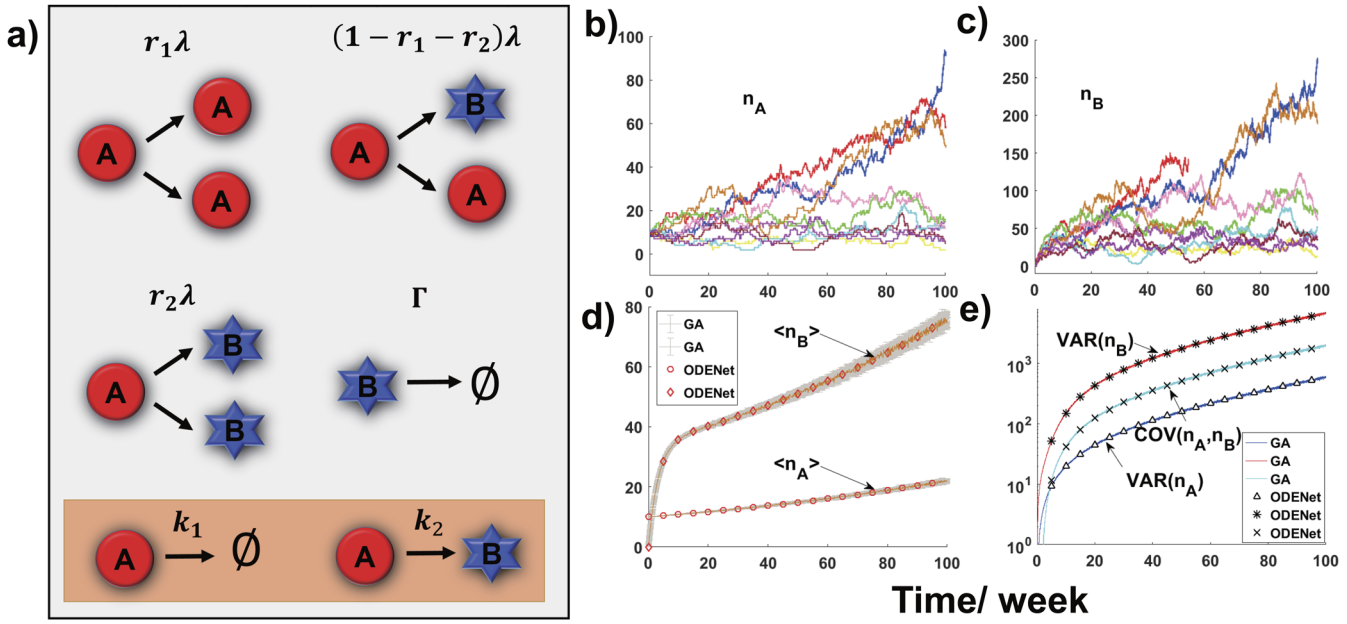

FIG. 2. Single proliferative compartment model. (a) The mechanism of the single proliferative compartment model. EPCs (red circles) have an unlimited self-division potential to maintain the epidermis at a rate of $r_1\lambda$. Proliferating EPC cells divide into two post-mitotic basal cells (blue stars) at a rate of $r_2\lambda$. Asymmetric divisions of EPCs into itself and post-mitotic basal cells are at a rate of $1 - (r_1 + r_2)\lambda$. After mitosis in the basal layer, the post-mitotic basal cells leak at a rate of Γ. k_1 and k_2 represent the rate constants for two additional possible reactions inferred by the ODENet. Ten representative stochastic trajectories for (b) n_A and (c) n_B are generated by GA. The learned results of ODENet are compared with the training data generated by GA on the (d) mean and (e) variance of cell numbers. Here, the rate constants in SPCM are set as λ = 1.1, r_1 = 0.0836, r_2 = 0.0764, Γ = 0.31 per week in accordance with experimental observations and modeling results of Ref. 25. The initial PDF is taken as a delta distribution with $p_0(10, 0) = 1$.

reconstructed by ODENet. However, the existence of two additional first-order reactions, $A \xrightarrow{k_1} \phi$ and $A \xrightarrow{k_2} B$ [see orange box in Fig. 2(a)], could not be excluded in principle, which means at the moment the probability distribution of A-type and B-type cells follows

$$\begin{aligned}\frac{d}{dt}p(n_A,n_B) =& k_1[(n_A+1)p(n_A+1,n_B) - n_A p(n_A,n_B)]\\ &+ k_2[(n_A+1)p(n_A+1,n_B-1) - n_A p(n_A,n_B)]\\ &+ \lambda[r_1(n_A-1)p(n_A-1,n_B)\\ &+ r_2(n_A+1)p(n_A+1,n_B-2)\\ &+(1-r_1-r_2)n_A p(n_A,n_B-1) - n_A p(n_A,n_B)]\\ &+ \Gamma[(n_B+1)p(n_A,n_B+1) - n_B p(n_A,n_B)]. \end{aligned} \quad (12)$$

Here, the same notations are borrowed just for simplicity. It should be noted that, at the moment, we still have no precise knowledge on all six reaction rate constants k_1, k_2, r_1, r_2, λ, and Γ.

To determine the unknown coefficients, we further go to the second-order moments of PDF, the variance of cell numbers to be exact. Based on (12), the first-order (mean) and second-order moments (variance) of n_A and n_B evolve according to

$$\begin{cases} \dfrac{d\langle n_A\rangle}{dt} = \alpha_{11}\langle n_A\rangle, \dfrac{d\langle n_B\rangle}{dt} = \alpha_{21}\langle n_A\rangle + \alpha_{22}\langle n_B\rangle,\\ \dfrac{dV_A}{dt} = \beta_{11}\langle n_A\rangle + \beta_{12}V_A,\\ \dfrac{dV_B}{dt} = \beta_{21}\langle n_A\rangle + \beta_{22}\langle n_B\rangle + \beta_{23}V_B + \beta_{24}Cov,\\ \dfrac{dCov}{dt} = \beta_{31}\langle n_A\rangle + \beta_{32}V_A + \beta_{33}Cov, \end{cases} \quad (13)$$

where $V_A = \langle n_A^2\rangle - \langle n_A\rangle^2$, $V_B = \langle n_B^2\rangle - \langle n_B\rangle^2$, and $Cov = \langle n_A n_B\rangle - \langle n_A\rangle\langle n_B\rangle$. Furthermore, we have $\alpha_{11} = (r_1 - r_2)\lambda - k_1 - k_2$, $\alpha_{21} = (1 - r_1 + r_2)\lambda + k_2$, $\alpha_{22} = -\Gamma$, $\beta_{11} = k_1 + k_2 + (r_1 + r_2)\lambda$, $\beta_{12} = -2(k_1 + k_2) + 2(r_1 - r_2)\lambda$, $\beta_{21} = k_2 + (1 - r_1 + 3r_2)\lambda$, $\beta_{22} = \Gamma$, $\beta_{23} = -2\Gamma$, $\beta_{24} = 2((1 - r_1 + r_2)\lambda + k_2)$, $\beta_{31} = -(2r_2\lambda + k_2)$, $\beta_{32} = (1 - r_1 + r_2)\lambda + k_2$, $\beta_{33} = (r_1 - r_2)\lambda - k_1 - k_2 - \Gamma$. Now, following the same procedure, the unknown values of β's could be learned by ODENet and are summarized in the supplementary material. Note that (13) is actually a multiscale model, since the ratio between the variance and the mean goes to zero with the increase of cell numbers.

4. Deriving chemical master equations

The relations among desired rate constants k_1, k_2, r_1, r_2, λ, Γ, and those learned parameters α's and β's are stated

TABLE I. Comparison on the rate constants for (12) learned by ODENet with the true values.

	k_1	k_2	r_1	r_2	λ	Γ
True value	0	0	0.0836	0.0764	1.1	0.31
Learned value	0.0123	−0.0197	0.0831	0.0824	1.1059	0.3074
Relative errors (%)	...	...	0.60	7.85	0.54	0.84

through the following matrix, i.e.,

$$\underbrace{\begin{bmatrix} 1 & -1 & -1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 2 & -2 & -2 & -2 & 0 & 0 \\ -1 & 3 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 2 & 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 1 & 1 & 0 \\ 1 & -1 & -1 & -1 & 0 & -1 \end{bmatrix}}_{V} \underbrace{\begin{bmatrix} r_1\lambda \\ r_2\lambda \\ k_1 \\ k_2 \\ \lambda \\ \Gamma \end{bmatrix}}_{\hat{u}} = \underbrace{\begin{bmatrix} \alpha_{11} \\ \alpha_{21} \\ -\alpha_{22} \\ \beta_{11} \\ \beta_{12} \\ \beta_{21} \\ \beta_{22} \\ -\beta_{31} \\ \beta_{32} \\ \beta_{33} \end{bmatrix}}_{\hat{b}}. \tag{14}$$

Direct calculations show that the rank of the augmented matrix $[rank(V|\hat{b}) = 7]$ is larger than that of the coefficient matrix $[rank(V) = 6]$, meaning the linear equations in (14) constitute an overdetermined system, which can be solved through the Least-Square Method. The unique least-square solution is given by $\hat{u} = \left(V^T V\right)^{-1} V^T \hat{b}$, whose relative errors with respect to the true values are less than 8% (see Table I).

Even though k_1 and k_2 are not exactly identified as zero, their values are about one order of magnitude smaller than the others. In this sense, we have successfully reconstructed the correct SPCM based on the stochastic time trajectories of n_A and n_B in the training dataset. As further validated in Fig. 3, the joint probability distribution of the desired single-proliferative compartment model is honestly reproduced (see the supplementary material for the marginal probability distribution), which highlights the accuracy and efficiency of our integrated approach of ODENet with multiscale modeling during the study of chemical reactions.

B. A gene network with autoregulatory negative feedback

1. The basic model

In the second example, we plan to show how machine learning can be used for model reduction, an important aspect of multiscale modeling with vast applications in chemical reactions. To illustrate our ideas, let us consider a gene network with autoregulatory negative feedback,[28] which includes five reactants – the gene

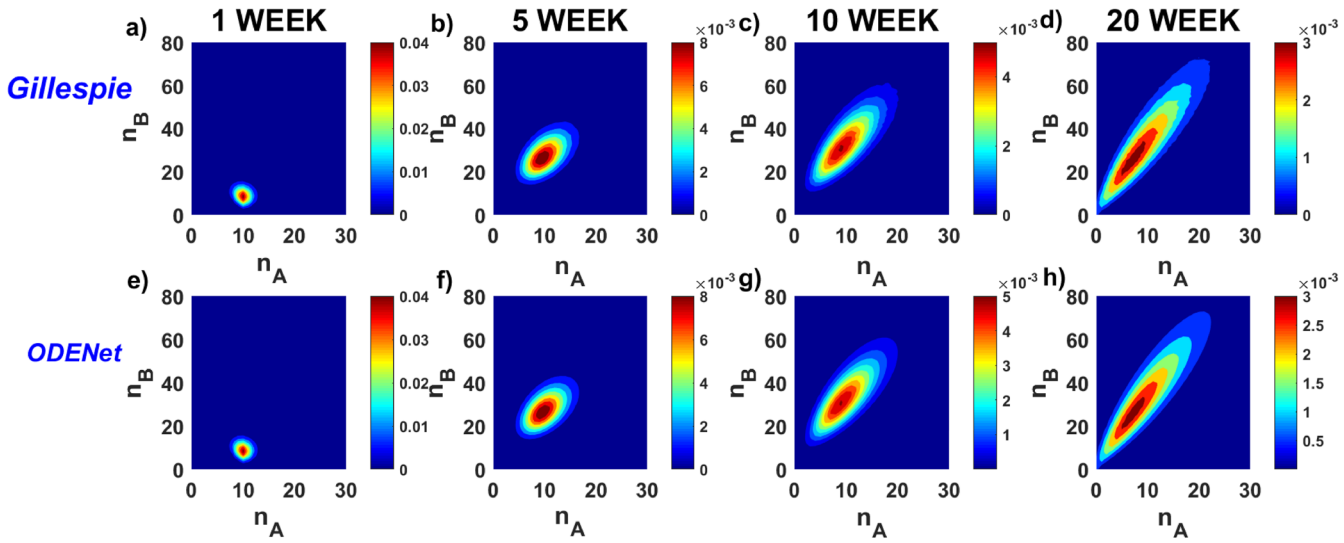

FIG. 3. Comparison of PDF generated by GA with the learned results of ODENet. Joint probability distributions $p(n_A, n_B)$ are shown in [(a) and (e)] 1, [(b) and (f)] 5, [(c) and (g)] 10, and [(d) and (h)] 20 weeks, respectively. The average relative errors between the PDF learned by ODENet and the exact PDF of CMEs are less than 3%.

(G), mRNA (M), protein (P), and two gene–protein complexes (GP, GP_2). Among them, there are eight reactions [see Fig. 4(a)]. k_0, k_s, k_{dm} are rate constants of transcription from the gene G, translation into the protein P, and mRNA degradation, respectively. The gene can bind with either one or two proteins, whose forward and backward reaction rate constants are denoted as k_1, k_{-1}, k_2, and k_{-2} separately. Furthermore, it is assumed that GP produces mRNA at the same rate k_0 as the transcription rate of G alone.

Macroscopically, the gene network in Fig. 4(a) is described by chemical mass-action equations,

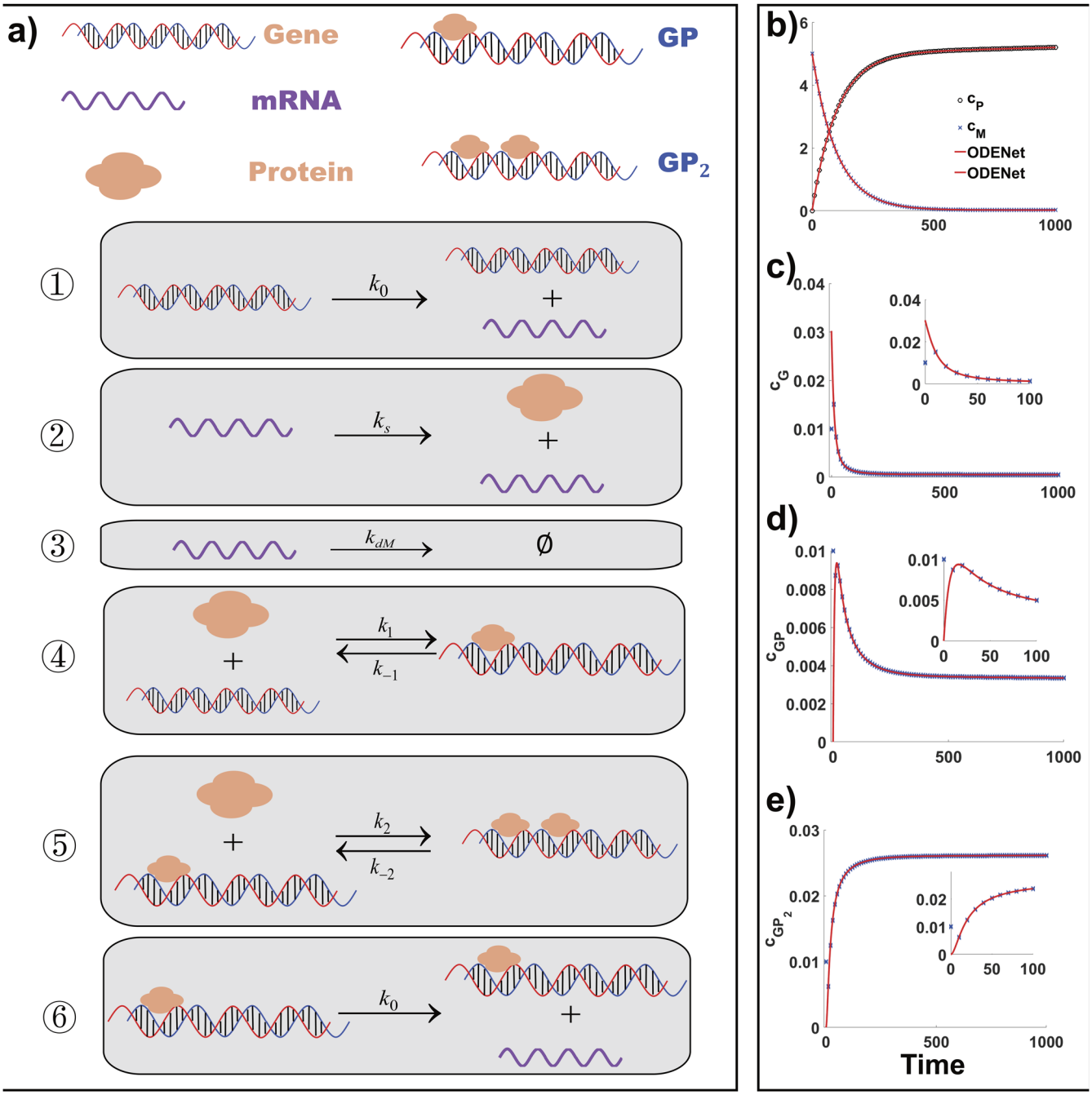

FIG. 4. Validation of ODENet aided model reduction. (a) A cartoon illustration of the gene network with negative feedback, including transcription, translation, degradation, and a negative feedback loop.[28] Predictions of the reduced model in Eq. (16) (blue crosses) are compared with the original model in Eq. (15) (red solid lines) on concentrations of (b) protein and mRNA in the slow time scale, and (c) gene and [(d)–(e)] gene–protein complexes in the fast time scale.

$$\begin{cases} \frac{d}{dt}c_P = k_s c_M - k_1 c_G c_P + k_{-1} c_{GP} - k_2 c_P c_{GP} + k_{-2} c_{GP_2}, \\ \frac{d}{dt}c_M = k_0 c_G + k_0 c_{GP} - k_{dM} c_M, \\ \frac{d}{dt}c_G = -k_1 c_G c_P + k_{-1} c_{GP}, \\ \frac{d}{dt}c_{GP} = k_1 c_G c_P - k_{-1} c_{GP} - k_2 c_P c_{GP} + k_{-2} c_{GP_2}, \\ \frac{d}{dt}c_{GP_2} = k_2 c_P c_{GP} - k_{-2} c_{GP_2}. \end{cases} \tag{15}$$

It is noted that the total gene concentration is a constant due to the conservation law, i.e., $c_G + c_{GP} + c_{GP_2} = c_{total}$. To produce a time-scale separation of reactions, we set $k_1 = 3$, $k_{-1} = 2.4$, $k_2 = 9$, $k_{-2} = 6$, $k_0 = 0.05$, $k_s = 0.01$, $k_{dm} = 0.01$, meaning the concentrations of G, GP, GP_2 can quickly reach a dynamical balance in comparison with those of mRNA and protein. So, G, GP, GP_2 are identified as fast variables, while the concentrations of mRNA and protein are slow variables in contrast. The initial conditions are set as $c_G = c_{GP} = c_{GP_2} = 0.01$, $c_P = 0$, $c_M = 5$.

2. Model reduction by ODENet

Due to the existence of time-scale separation, it is possible to make a simplification of the reaction system in (15). Classically, this is done by analytical methods, like QSSA and Partial Equilibrium Assumption.[24,29] Here, we are going to show how the simplification procedure can be carried out automatically by ODENet.

At the first step, with the help of traditional classification algorithms, like the correlation analysis based on the Pearson's coefficient among concentration derivatives (see Table IV in the supplementary material), the fast and slow variables can be easily separated into two groups. Inspired by the classical results of Michaelis–Menton kinetics, we suppose three fast variables c_G, c_{GP}, c_{GP_2} [see Figs. 4(b)–4(e)] are characterized by fractional functions, whose numerator and denominator are polynomials of c_P (up to the second-order in the current study). In contrast, c_M does not appear in the fractional functions, since the last three formulas in (15) contain no terms of c_M. Consequently, the simplified model we are seeking for is given by

$$\begin{cases} \frac{d}{dt}c_P = k_s c_M \underline{- k_1 c_P H(c_G) + k_{-1} H(c_{GP}) - k_2 c_P H(c_{GP}) + k_{-2} H(c_{GP_2})}, \\ \frac{d}{dt}c_M = -k_{dM} c_M + k_0 H(c_G) + k_0 H(c_{GP}), \\ H(c_G) = \frac{\Omega_1}{\Omega}, H(c_{GP}) = \frac{\Omega_2}{\Omega}, H(c_{GP_2}) = \frac{\Omega_3}{\Omega}, \end{cases} \tag{16}$$

where $\Omega = \beta_1 + \beta_2 c_p + \beta_3 c_P^2$, $\Omega_1 = \alpha_{11} + \alpha_{21} c_P + \alpha_{31} c_P^2$, $\Omega_2 = \alpha_{12} + \alpha_{22} c_P + \alpha_{32} c_P^2$, and $\Omega_3 = \alpha_{13} + \alpha_{23} c_P + \alpha_{33} c_P^2$.

$\beta_1, \ldots, \beta_3$ and $\alpha_{11}, \ldots, \alpha_{33}$ are 12 free parameters to be specified. As summarized in Table II, the simplified model learned by ODENet is very close to that by QSSA (see Sec. III B 3). In particular, the terms of $\alpha_{21} c_P$, $\alpha_{31} c_P^2$, α_{12}, $\alpha_{32} c_P^2$, α_{13}, and $\alpha_{23} c_P$ are removed by sparse identification during the learning procedure. A major difference between two simplification methods lies in the extra four underlined terms on the right-hand side of the first equation in (16). In QSSA, these four terms are exactly canceled by each other. While in our simplification procedure aided by ODENet, we can only conclude that their sum is very small instead of exactly zero (see the supplementary material).

TABLE II. Comparison on the parameters for (16) learned by ODENet with those by QSSA. All values are normalized by β_3.

	β_1	β_2	β_3	α_{11}	α_{21}	α_{31}
QSSA	0.53	0.67	1	0.0159	0	0
ODENet	0.54	0.66	1	0.0163	0	0
Relative error (%)	1.89	1.49	...	2.52	...	...
	α_{12}	α_{22}	α_{32}	α_{13}	α_{23}	α_{33}
QSSA	0	0.0201	0	0	0	0.03
ODENet	0	0.020	0	0	0	0.03
Relative error (%)	...	0.50	...	...	...	0

The idea of model reduction considered here is quite similar to that of an encoder-decoder architecture.[30] In particular, the fractional functions $H(c)$ motivated by QSSA and adopted in (16) can be regarded as a special kind of explicit encoder, while the usual encoder–decoder architecture will take layers of neural networks as a black box to approximate the unknown functions or variables. The encoder–decoder architecture has been widely used in the field of machine learning, like the famous Generative Adversarial Networks (GAN) for image recognition,[31] or SciNet for deriving new physical models or principles from the data.[32,33]

3. Comparison with QSSA

Our above ODENet aided model reduction is consistent with the classical QSSA. Since G, GP, and GP_2 are considered as the fast intermediates, in contrast to the slow species P and M, a direct application of QSSA to (15) leads to

$$\frac{c_G}{c_{total}} = \frac{K_3}{\Omega}, \quad \frac{c_{GP}}{c_{total}} = \frac{K_2 c_P}{\Omega}, \quad \frac{c_{GP_2}}{c_{total}} = \frac{{c_P}^2}{\Omega}, \tag{17}$$

where $\Omega = K_3 + K_2 c_P + {c_P}^2$, $K_1 = k_{-1}/k_1$, $K_2 = k_{-2}/k_2$, $K_3 = k_{-1}k_{-2}/(k_1 k_2)$, and $c_{total} = c_G + c_{GP} + c_{GP_2}$ is a constant. The corresponding reduced model is

$$\begin{cases} \frac{d}{dt}c_P = k_s c_M, \\ \frac{d}{dt}c_M = -k_{dM} c_M + k_0 (K_3 + K_2 c_P) c_{total}/\Omega, \\ \frac{c_G}{c_{total}} = \frac{K_3}{\Omega}, \quad \frac{c_{GP}}{c_{total}} = \frac{K_2 c_P}{\Omega}, \quad \frac{c_{GP_2}}{c_{total}} = \frac{{c_P}^2}{\Omega}, \end{cases} \tag{18}$$

which has been used to evaluate the performance of our ODENet aided model reduction.

IV. CONCLUSION

Nowadays, various machine learning algorithms, like deep learning and reinforcement learning, have found their applications

in diverse fields with great success. While in the field of chemical reactions, related studies begin to emerge yet are still quite few. In the current paper, through two concrete biochemical examples – the single proliferative compartment model and a gene network with autoregulatory negative feedback, we present our key ideas on how machine learning and multiscale modeling can help each other during the study of chemical reactions. And, as we believe, an effective integration of two approaches will be crucial for the success of related studies in this direction.

Potential generalizations of our current work include but are not limited to the following:

(1) **Spacial heterogeneity of chemical reactions.** In the current study, all reactions are assumed to proceed under well-mixed conditions, which means we can adopt a relatively simple ODE-based description. However, it is well-known that the spatial heterogeneity can produce far more complicated and also interesting phenomena,[1] like the Turing pattern, phase separation, active matter, etc. So, how to generalize our results to PDEs would be of general interest. Recently, PDE-based machine learning algorithms[34,35] shed light on this aspect.

(2) **Bistability, oscillation, and bifurcation of chemical reactions.** Even restricted to ODEs, a chemical reaction system can possess very complex dynamical behaviors, like bistability, oscillation, bifurcation, blow-up, etc., than one can imagine.[36,37] In the presence of noise, the situation becomes even more complicated. The high nonlinearity of chemical reactions puts forward great challenges to our ODENet-based model derivation and model reduction. Besides simple fixed-point dynamics discussed here, limit-cycle solutions of Lotka–Volterra equations and chaotic Lorenz equations have been explored in our previous studies within the ODENet framework too,[4] though further detailed examinations are still required.

(3) **Extension to other model reduction methods.** Here, we test the possibility and accuracy of ODENet aided model reduction with respect to the QSSA method. Extension of our ideas to partial equilibrium approximation,[29] maximum entropy principle,[38] maximum likelihood estimation,[39] as well as other statistics or probability based approximations would be worthy of further studies.

(4) **Non-unique correspondence between reaction mechanisms and mathematical models.** The essence of this study is to use the ODENet to infer the unknown architecture of chemical reaction networks and their reaction rate constants. However, it should be noted that, in general, the corresponding relationship between a reaction network and a set of ODEs is not unique. This point has already been seen in the first example, in which the existence of two unwanted reactions cannot be excluded, when only referring to the mathematical model and its coefficients. We call great attention of readers to this issue.

SUPPLEMENTARY MATERIAL

See the supplementary material for the setup of hyperparameters in the ODENet, the comparison on the marginal probability distribution of SPCM, the identification of fast and slow variables in the second example by Pearson's correlation coefficients, etc.

AUTHORS' CONTRIBUTIONS

W.Y. and L.P. contributed equally to this work.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation of China (Grant Nos. 21877070 and 11871299), the Startup Research Funding of Minjiang University (Grant No. mjy19033), and the Special Project of COVID-19 Epidemic Prevention and Control by Fuzhou Science and Technology Bureau (Grant No. 2020-XG-002). The authors would like to thank the helpful discussions from Dr. Pipi Hu.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] T. Janos and E. Peter, *Mathematical Models of Chemical Reactions: Theory and Applications of Deterministic and Stochastic Models* (Princeton University Press, Princeton, 1989).
[2] K. H. Johnson, "Quantum chemistry," Annu. Rev. Phys. Chem. **26**, 39–57 (2003).
[3] N. M. Mangan, S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Inferring biological networks by sparse identification of nonlinear dynamics," IEEE Trans. Mol., Biol. Multi-Scale Commun. **2**, 52–63 (2016).
[4] P. Hu, W. Yang, Y. Zhu, and L. Hong, "Revealing hidden dynamics from time-series data by ODENet," arXiv:2005.04849 (2020).
[5] W. Zhang, S. Klus, T. Conrad, and C. Schütte, "Learning chemical reaction networks from trajectory data," SIAM J. Appl. Dyn. Syst. **18**, 2000–2046 (2019).
[6] Z. Costello and H. G. Martin, "A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data," npj Syst. Biol. Appl. **4**, 19 (2018).
[7] W. Yang, P. Tan, X. Fu, and L. Hong, "Prediction of amyloid aggregation rates by machine learning and feature selection," J. Chem. Phys. **151**, 084106 (2019).
[8] M. A. Kayala, C.-A. Azencott, J. H. Chen, and P. Baldi, "Learning to predict chemical reactions," J. Chem. Inf. Model. **51**, 2209–2222 (2011).
[9] Q. Song, Q. Zhang, and Q. Meng, "Revisiting the Gaussian process regression for fitting high-dimensional potential energy surface and its application to the OH + HO_2 → O_2 + H_2O reaction," J. Chem. Phys. **152**, 134309 (2020).
[10] A. F. Villaverde and J. R. Banga, "Reverse engineering and identification in systems biology: Strategies, perspectives and challenges," J. R. Soc., Interface **11**, 20130505 (2014).
[11] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," Proc. Natl. Acad. Sci. U. S. A. **113**, 3932–3937 (2016).
[12] L. Boninsegna, F. Nüske, and C. Clementi, "Sparse learning of stochastic dynamical equations," J. Chem. Phys. **148**, 241723 (2018).
[13] K. Choi, "Robust approaches to generating reliable predictive models in systems biology," in *Systems Biology*, RNA Technologies, edited by N. Rajewsky, S. Jurga, and J. Barciszewski (Springer, Cham, 2018), pp. 301–312.
[14] B. C. Daniels and I. Nemenman, "Automated adaptive inference of phenomenological dynamical models," Nat. Commun. **6**, 8133 (2015).
[15] W. E, *Principles of Multiscale Modeling* (Cambridge University Press, Cambridge, 2011).

[16]J. Han, C. Ma, Z. Ma, and W. E, "Uniformly accurate machine learning-based hydrodynamic models for kinetic equations," Proc. Natl. Acad. Sci. U. S. A. **116**, 21983–21991 (2019).
[17]W. E, J. Han, and L. Zhang, "Integrating machine learning with physics-based modeling," arXiv:2006.02619 (2020).
[18]R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, "Neural ordinary differential equations," in *Advances in Neural Information Processing Systems*, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Curran Associates, Inc., 2018), Vol. 31, pp. 6571–6583.
[19]K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[20]Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," Neural Comput. **1**, 541–551 (1989).
[21]H. G. Othmer, *Analysis of Complex Reaction Networks*, Lecture Notes (School of Mathematics, University of Minnesota, 2003).
[22]D. T. Gillespie, "Exact stochastic simulation of coupled chemical reactions," J. Phys. Chem. **81**, 2340–2361 (1977).
[23]T. G. Kurtz, "The relationship between stochastic and deterministic models for chemical reactions," J. Chem. Phys. **57**, 2976–2978 (1972).
[24]L. A. Segel and M. Slemrod, "The quasi-steady-state assumption: A case study in perturbation," SIAM Rev. **31**, 446–477 (1989).
[25]E. Clayton, D. P. Doupé, A. M. Klein, D. J. Winton, B. D. Simons, and P. H. Jones, "A single type of progenitor cell maintains normal epidermis," Nature **446**, 185–189 (2007).
[26]L. Hong and W.-A. Yong, "Simple moment-closure model for the self-assembly of breakable amyloid filaments," Biophys. J. **104**, 533–540 (2013).
[27]P. Tan and L. Hong, "Modeling fibril fragmentation in real-time," J. Chem. Phys. **139**, 084904 (2013).
[28]P. Thomas, A. V. Straube, and R. Grima, "The slow-scale linear noise approximation: An accurate, reduced stochastic description of biochemical networks under timescale separation conditions," BMC Syst. Biol. **6**, 39 (2012).
[29]Y.-J. Huang, L. Hong, and W.-A. Yong, "Partial equilibrium approximations in apoptosis. II. The death-inducing signaling complex subsystem," Math. Biosci. **270**, 126–134 (2015).
[30]I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, 2016).
[31]I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2014), pp. 2672–2680.
[32]R. Iten, T. Metger, H. Wilming, L. Del Rio, and R. Renner, "Discovering physical concepts with neural networks," Phys. Rev. Lett. **124**, 010508 (2020).
[33]P. G. Breen, C. N. Foley, T. Boekholt, and S. P. Zwart, "Newton versus the machine: Solving the chaotic three-body problem using deep neural networks," Mon. Not. R. Astron. Soc. **494**, 2465–2470 (2020).
[34]S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Data-driven discovery of partial differential equations," Sci. Adv. **3**, e1602614 (2017).
[35]M. Raissi, P. Perdikaris, and G. Karniadakis, "Numerical Gaussian processes for time-dependent and nonlinear partial differential equations," SIAM J. Sci. Comput. **40**, A172–A198 (2017).
[36]H. Qian, P.-Z. Shi, and J. Xing, "Stochastic bifurcation, slow fluctuations, and bistability as an origin of biochemical complexity," Phys. Chem. Chem. Phys. **11**, 4861 (2009).
[37]L. M. Bishop and H. Qian, "Stochastic bistability and bifurcation in a mesoscopic signaling system with autocatalytic kinase," Biophys. J. **98**, 1–11 (2010).
[38]E. T. Jaynes, "Information theory and statistical mechanics," Phys. Rev. **106**, 620 (1957).
[39]F. E. Harrell, "Overview of maximum likelihood estimation," in *Regression Modeling Strategies*, With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis (Springer International Publishing, Cham, 2015), pp. 181–217.