Machine Learning Dynamic Correlation in Chemical Kinetics

Changhae Andrew Kim, Nathan D. Ricke, and Troy Van Voorhis*

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

E-mail: tvan@mit.edu

Abstract

The kinetics of surface reactions are often described using a lattice model. Since it is expensive to propagate the configuration probabilities of the entire lattice, it is practical to consider the occupation probabilities of a typical site or a cluster of sites instead. This amounts to a moment closure approximation of the chemical master equation (CME). Unfortunately, simple closures, such as the mean-field (MF) and the pair approximation (PA), exhibit weaknesses in systems with significant long-range correlation. In this paper, we show that machine learning (ML) can be used to construct accurate moment closures in chemical kinetics, using the lattice Lotka-Volterra model (LLVM) as a model system. We trained feed-forward neural networks (FFNNs) to estimate the three-site occupation probabilities, using kinetic Monte Carlo (KMC) results at select values of rate constants and initial conditions. Given the same level of input as PA, the machine learning moment closure (MLMC) gave drastic improvements in the simulated dynamics and descriptions of the dynamical regimes throughout the parameter space. In this way, MLMC is a promising tool to interpolate KMC simulations or construct pre-trained closures that would enable researchers to extract useful insight at a fraction of the computational cost.

Introduction

Machine learning (ML) is an important tool in computational chemistry. On one hand, it has been used to accelerate the discovery of drugs and materials by deducing the electronic properties of molecules,[1–5] reactivities of organic compounds,[6–8] and secondary structures of proteins[9–11] using just the topologies of the molecules. On the other hand, it has been used to improve simulations by replacing the approximate right hand sides, such as density functionals,[12,13] electron densities,[14,15] and force fields,[16–19] with ML models. Meanwhile, there have been few applications of ML in chemical kinetics. We believe that ML might provide a way to improve the solution of the chemical master equation (CME).

There is a duality in what various authors mean by CME. First, there are homogeneous systems, where the state of the system is defined by the numbers of molecules,[20–24] and the positions of the molecules are not explicit variables. Second, there are heterogeneous systems, where the state of the system is defined by the configuration of a lattice.[25–27] Hence, not only the numbers of molecules, but also their positions are explicit variables. In both cases, CME is a system of ordinary differential equations (ODEs) that propagates the probabilties of all possible states of the system, and the computational costs are exponential. Only, the homogeneous case scales as the number of molecules raised to the number of species, whereas the heterogeneous case scales as the number of species raised to the number of molecules. In this paper, we focus on the heterogeneous case. Nonetheless, due to the mathematical commonalities, results that hold in heterogeneous systems might have counterparts in homogeneous systems, and vice versa.

Since the shear size of state space often makes it impractical to solve the full CME, moment closure approximations have been considered an affordable approach to extract qualitative insight. The kinetic equations are written in terms of n-species subsystems (homogeneous case) or n-site clusters (heterogeneous case), and the higher-order moments, which describe the interactions of the n-th order moments with the rest of the system, are approximated using a moment closure. In homogeneous systems, the most popular clo-

sures are stochastic closures, such as the normal,[28,29] Poisson,[30] and log-normal closures.[31] Recently, Smadbeck and Kaznessis proposed an alternative scheme that computes the higher-order moments and their probability distribution by maximizing information entropy.[23] In heterogeneous systems, the most popular closures are the mean-field (MF) and the pair approximation (PA).[32–35] There have been attempts to go to higher-order moments, such as the triple approximation,[36] the approximate master equations,[37,38] and the cluster mean-field approximation.[26] In principle, moment closure approximations become more accurate as higher-order moments are used as the basis. However, an increase in the order is accompanied by a steep rise in the computational costs.

Recently, Mjolsness *et al.* demonstrated that moment closures based on deep Boltzmann machines (DBMs) can obtain accurate dynamics of the Lotka-Volterra model on the lattice.[27] In analogy to the empirical construction of density functionals, we believe that ML could provide a breakthrough in overcoming the complexity-accuracy trade-off of moment closure approximations, provided that it can be formulated in a way that is intuitive and accessible to the chemical community.

In this paper, we show that a simple ML architecture can be used to construct an accurate moment closure for chemical kinetics. Our choice of feed-forward neural networks (FFNNs) has both theoretical and practical relevance. On the theoretical side, FFNNs are the simplest neural networks. They are oblivious of time, nor do they have memory of the previous inputs and outputs. Indeed, they are functions that approximate the instantaneous values of the higher-order moments using the instantaneous values of the lower-order moments. On the practical side, FFNNs are fast to train and evaluate. They might scale better to larger numbers of species and higher orders of moments. Moreover, FFNNs are already available in popular software libraries, such as TensorFlow[39] and scikit-learn.[40] Hence, it is an architecture with which many chemists are already familiar.

The remainder of this paper is organized as follows. We begin by introducing the lattice Lotka-Volterra model (LLVM) as the model system, and we explain the origin, strengths,

and weaknesses of MF and PA. Then, we train FFNNs to estimate the three-site occupation probabilities, using the results of kinetic Monte Carlo (KMC) simulations at select values of rate constants and initial conditions. At the same level of input as PA, the machine learning moment closure (MLMC) can reduce the absolute and relative errors in the three-site probabilities by an order of magnitude. Furthermore, MLMC gives drastic improvements in the simulated dynamics and improved descriptions of the dynamical regimes throughout the parameter space of the model system.

Theory

Model System

The Lokta-Volterra model is a classic model system in which the activities of competing components lead to the emergence of oscillations. Originally, it was devised to describe autocatalytic chemical reactions,[41] but its application has been extended to biological systems,[42–44] where it had another intuitive interpretation. Our implemtnation of LLVM is given by

$$O + A \xrightarrow{k_1} A + A \tag{1}$$

$$A + B \xrightarrow{k_2} B + B \tag{2}$$

$$B \xrightarrow{k_3} O \tag{3}$$

where one often interprets O as the vacancy, A as the prey, and B as the predator. To our knowledge, there is no chemical reaction that follows this mechanism *per se*. However, it can be regarded as a coarse-grain approximation. In the SI, we discuss how the $NO + CO$ reaction on the Pt(100)-(1×1) surface[45] might be coarse-grained on to LLVM.

Variations of LLVM have been a subject of interest in the physics community.[46–48] They are known to display a number of features that are insensitive to the implementation.[49] In
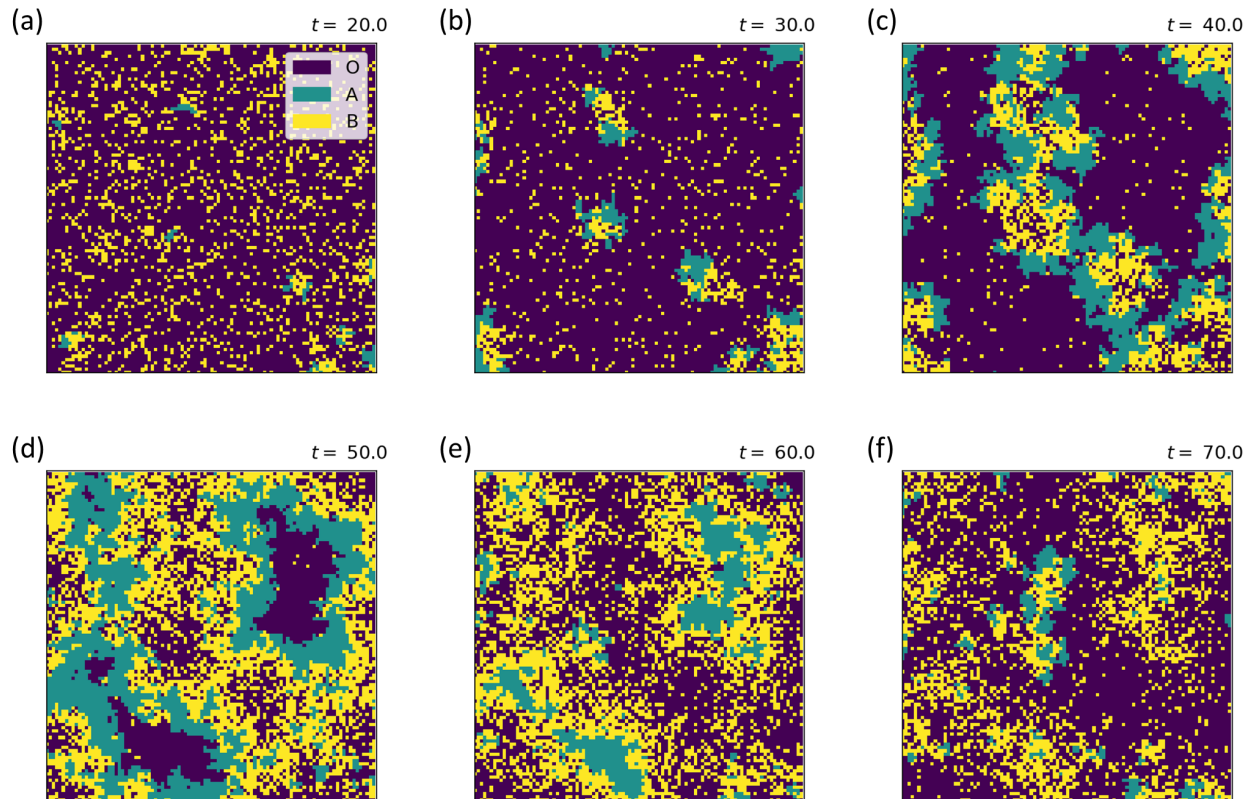
Figure 1: Snapshots of a 100×100 sublattice of a 500×500 lattice in a KMC simulation. The rate constants were $(k_1, k_2, k_3) = (0.5, 0.3, 0.1)$, and the initial conditions were $([O], [A], [B]) = (0.2, 0.3, 0.5)$.

particular, the collective activities of the prey and the predator give rise to spatiotemporal patterns,[50] along with density oscillations[51] that average out in the thermodynamic limit.[47,48] Some of the spatiotemporal patterns in our implementation are shown in Figure 1. It is interesting to see traveling wave patterns emerge. First, A grows into islands (Figures 1a and 1b). As the islands of A expand, B begins to grow (Figure 1c) and then proceeds to overrun the islands (Figures 1d and 1e). The cycle resets as B gives way to O (Figure 1f). We emphasize that the individual molecules are immobile in our implementation, but their formation and consumption as a collective give rise to the apparent pursuit-evasion behavior.

Self-organization of reactants is not uncommon in surface reactions. Even the benign $A + B \rightarrow AB \uparrow$ model is known to result in monospecific islands of A and B,[52,53] and numerous real-world examples can be found in heterogeneous catalysis.[54] Indeed, the NO + CO/Pt(100) reaction is one of them. Similar to LLVM, the NO + CO reaction gives rise to spatiotemporal patterns, and the dephasing of local oscillations[45,55] leads to the damping of global oscillations.[56] The types of correlations in LLVM are quite relevant to chemical systems.

Chemical Master Equation and Moment Closure Approximation

Consider a chemical reaction on a lattice. The molecules can adsorb on a vacant site, desorb, diffuse to a neighboring site, or react with another molecule. Let p_Ψ denote the probability of finding the lattice in configuration Ψ. The chemical master equation (CME) is given by

$$\frac{dp_\Psi}{dt} = -\sum_\Phi k_{\Psi \rightarrow \Phi} p_\Psi + \sum_\Phi k_{\Phi \rightarrow \Psi} p_\Phi \tag{4}$$

where $k_{\Psi \rightarrow \Phi}$ is a sum of the elementary rate constants, if any, that would take a lattice in configuration Ψ to configuration Φ. Assuming Markovian processes and an *a priori* knowledge of the rate constants, CME gives an exact treatment of both static disorder (site-to-site variations that are reflected in the rate constants, $k_{\Psi \rightarrow \Phi}$)[57,58] and dynamic correlation (segre-

gation and self-organization of reactants that manifest on the explicit lattice configurations, Ψ).[52,53] Unfortunately, the dimensionality of CME scales as S^L, where S is the number of species and L is the number of sites on the lattice, making CME intractable in many systems of practical relevance.

In princple, the kinetic Monto Carlo (KMC)[20,25,59] can recover the static disorder and dynamic correlation. A stochastic simulation of the lattice amounts to sampling a trajectory through the configuration space. By averaging over multiple simulations, one can approach the full CME results. However, the computational cost of the simulations can be formidable, especially if rapid equilibrium or diffusion is involved.[25,59,60]

Since the desired outcome in chemical kinetics is often an ensemble average, such as the surface coverage or the reaction rate, we are motivated to rewrite the kinetic equations in the occupation probabilities of n-site clusters (n-site probabilities)

$$[X] \equiv [X]_i = \sum_{\Psi} \delta_{\psi_i X} p_{\Psi} \tag{5}$$

$$[XY] \equiv [XY]_{ij} = \sum_{\Psi} \delta_{\psi_i X} \delta_{\psi_j Y} p_{\Psi} \tag{6}$$

$$[XYZ] \equiv [XYZ]_{ijk} = \sum_{\Psi} \delta_{\psi_i X} \delta_{\psi_j Y} \delta_{\psi_k Z} p_{\Psi} \tag{7}$$

where i, j, and k are a string of typical sites on the lattice; and $\delta_{\psi_i X} = 1$ if the occupant of the site i is $\psi_i = X$ and 0 otherwise. These n-site probabilities are special cases of moments.

The kinetic equations of one-site clusters are given by

$$\frac{d[X]}{dt} = -\sum_{Y} k_{X \to Y}[X] + \sum_{Y} k_{Y \to X}[Y]$$
$$- N \sum_{W,Y,Z} k_{WX \to YZ}[WX] + N \sum_{W,Y,Z} k_{YZ \to WX}[YZ] \tag{8}$$

where $k_{XY \to UV}$ is the rate constant of the elementary step, $X + Y \to U + V$, if it exists; and N is the number of nearest neighbors. Observe that the equations are not closed. Unless the

elementary steps consist of unimolecular reactions only, the equations of n-site clusters are going to depend on information about $(n+1)$-site clusters. In order to create a closed system of equations, we need a prescription to approximate the higher-order moments using just the information about the lower-order moments – hence, a moment closure approximation. The simplest and the most popular closure is the mean-field approximation (MF), $[XY] = [X][Y]$, which neglects any correlation that might exist between the sites.

The next simplest closure, which incorporates some site-to-site correlation, is the pair approximation (PA). Consider the kinetic equations of two-site clusters

$$
\begin{aligned}
\frac{d[XY]}{dt} = & -\sum_{U,V} k_{XY \to UV}[XY] + \sum_{U,V} k_{UV \to XY}[UV] \\
& - (N-1)\sum_{U,V,W} k_{WX \to UV}[WXY] - (N-1)\sum_{U,V,Z} k_{YZ \to UV}[XYZ] \\
& + (N-1)\sum_{U,V,W} k_{UV \to WX}[UVY] + (N-1)\sum_{U,V,Z} k_{UV \to YZ}[XUV]
\end{aligned}
\tag{9}
$$

With the two-site probabilities, $[XY]$, as the variables, the kinetic equations now depend on the three-site probabilities, $[XYZ]$, which must be approximated in terms of the two-site probabilities. PA estimates the three-site probabilities using the definition of conditional probability

$$
[XYZ] = \frac{[XY][YZ]}{[Y]}
\tag{10}
$$

Due to its simple rationale, it has been invented many times by independent workers in chemistry,[32] population biology,[33,34] and epidemiology.[35]

In principle, the closed system of equations would become more accurate as higher-order moments are used as the basis of the moment closure approximation. Formally, one can interpret MF and PA as special cases of product approximations,[61,62] so it should be possible to generalize PA to n-site probabilities. Examples of these attempts include the triple approximation,[36] the approximate master equations,[37,38] and the cluster mean-field approximation.[26] However, the derivation and computation of the kinetic equations become

challenging, as the geometries of the clusters come into play, and the number of moments grows with the order as S^n.

A subset of the readers might be more familiar with stochastic closures, such as the normal, [28,29] Poisson, [30] and log-normal closures, [31] which are often used in homogeneous systems. The physical and mathematical arguments behind these closures are quite different, since the moments in the homogeneous case are expected numbers of molecules ($\in \mathbb{R}$) as opposed to occupation probabilities ($\in [0,1]$). In short, stochastic closures assume that the numbers of molecules have a certin probability distribution. Applying stochastic closures to heterogeneous systems amounts to considering a probability distribution of probabilities. Although it can be done with suitable adjustments, our experience indicates that stochastic closures encounter instabilities and unphysical values in heterogeneous systems that exhibit strong spatial correlation, such as LLVM. PA is robust as long as the removable discontinuity at zero is assigned an appropriate value. Thus, we focus on PA.

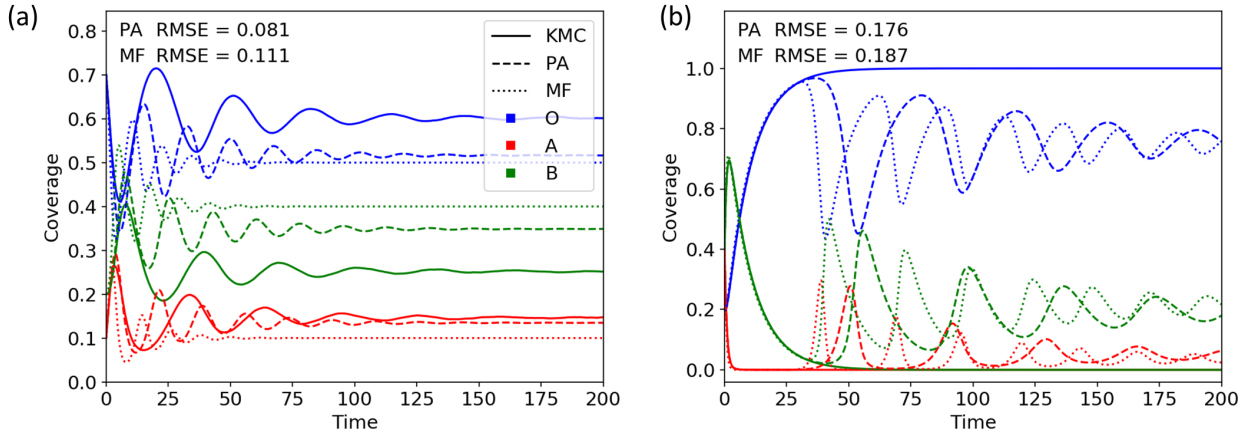## Pair Approximation and Machine Learning Moment Closure

Figure 2: Time-dependent coverages according to MF (dotted lines), PA (dashed lines), and KMC (solid lines). (a) The rate constants were $(k_1, k_2, k_3) = (0.4, 0.5, 0.2)$, and the initial conditions were $([O], [A], [B]) = (0.7, 0.1, 0.2)$. (b) The rate constants and the initial conditions were $(0.2, 0.7, 0.1)$ and $(0.2, 0.4, 0.4)$, respectively. Clearly, PA provides significant but not satisfactory improvement over MF in both cases.

Figure 2 demonstrates the strengths and weaknesses of PA. Although PA (dashed lines)

9

does give a noticeable improvement over MF (dotted lines), it still looks too much like MF as opposed to KMC (solid lines). In the oscillatory regime (Figure 2a), PA underestimates the amplitude and period. In the non-oscillatory regime where $[A]$ and $[B]$ go to zero (Figure 2b), PA can still predict damped oscillations. Moreover, the steady-state coverages are mispredicted. The source of the error is the long-range correlation. Because PA only knows about the short-range (two-site) correlation, it is not able to anticipate the formation of islands or traveling waves that span a large number of sites. That is, not without information about the mechanism and the nature of the correlation built into the approximation. The goal of MLMC is a smarter closure that still takes low-order moments as the input, but uses mechanism-specific information to give a more accurate output.
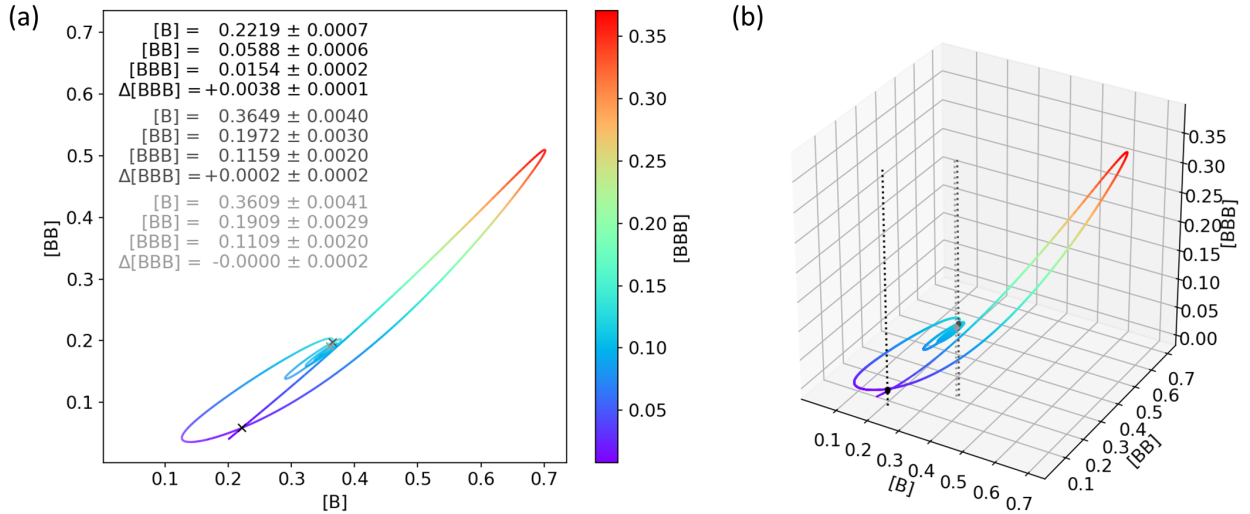

Figure 3: Self-crossings of the time-dependent one-, two-, and three-site probabilities. The curve was obtained as an average over 20 KMC simulations on a 1000×1000 lattice. (a) On the 2D plot, the self-crossings on the xy-plane are indicated by crosses, and the coordinates and the standard deviations are provided on the upper left. (b) On the 3D plot, the self-crossings on the xy-plane are indicated by bold dots, and the vertical dotted lines are visual aids. The rate constants were $(k_1, k_2, k_3) = (0.5, 0.3, 0.1)$, and the initial conditions were $([O], [A], [B]) = (0.2, 0.6, 0.2)$.

But what is the lowest order at which the moments can be expect to be predictive of the correlation? For the case of LLVM, Figure 3 gives some insight. Since the system is undergoing non-equilibrium dynamics, it is not possible in priciple to write the n-site

probabilities as functions of the k-site probabilities ($k < n$). Indeed, Figure 3a shows that it is not plausible to write the two-site probability, $[BB]$, as a function of the one-site probability, $[B]$. Due to the oscillatory nature of the dynamics, the system returns to the same value of $[B]$ multiple times, and it is clear that the relationship between $[B]$ and $[BB]$ is not one-to-one. Hence, there is no function such that $[BB] = f([B])$

On the other hand, there are far fewer points where the system returns to the same values of $[B]$ and $[BB]$ at the same time. These appear as self-crossings of the curve in Figure 3a. Of course, there are still differences in the three-site probability, $[BBB]$, each time the curve crosses itself, so it is not possible to write $[BBB] = f([B], [BB])$ in strict terms. However, the numbers on the upper left of Figure 3a and the three-dimensional plot in Figure 3b indicate that the differences are small. In many cases, they are on the same order of magnitude as the statistical noise in the KMC simulations, and they can be negligible if the curve crosses itself with minimal time for the correlationn to evolve. Similar results were obtained using other initial conditions and reference species (Figures S1–S3). Therefore, we conjecture that the three-site probabilities can be written to a good approximation as functions of the two-site probabilities. A viable ML model of the three-site probabilities could be constructed using no more than the two-site probabilities as the input.

Results and Discussion

As detailed in Methods, we trained FFNNs to predict the three-site probabilities in terms of the two-site probabilities

$$[XYZ] = f([XY], [XZ], [YZ], \cdots) \tag{11}$$

Figure 4 demonstrates the accuracy of MLMC. Based on the spreads of the training sets and the test set, there does not appear to have been serious over-fitting. Moreover, there are qualitative improvements compared to PA (Figure 5). Not only is the off-diagonal spread
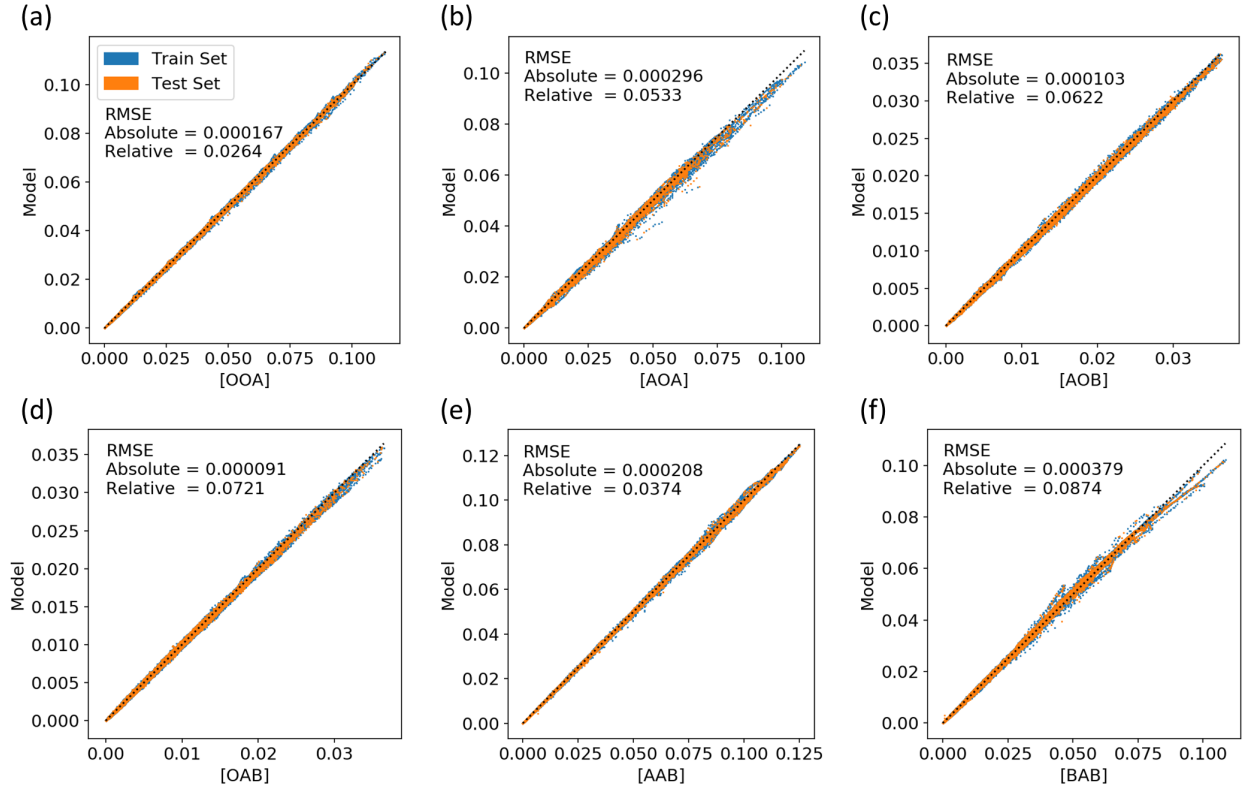
Figure 4: Scatter plots of the ML estimates against the KMC three-site probabilities. The training set is plotted in blue dots, and the test set is plotted in orange dots.
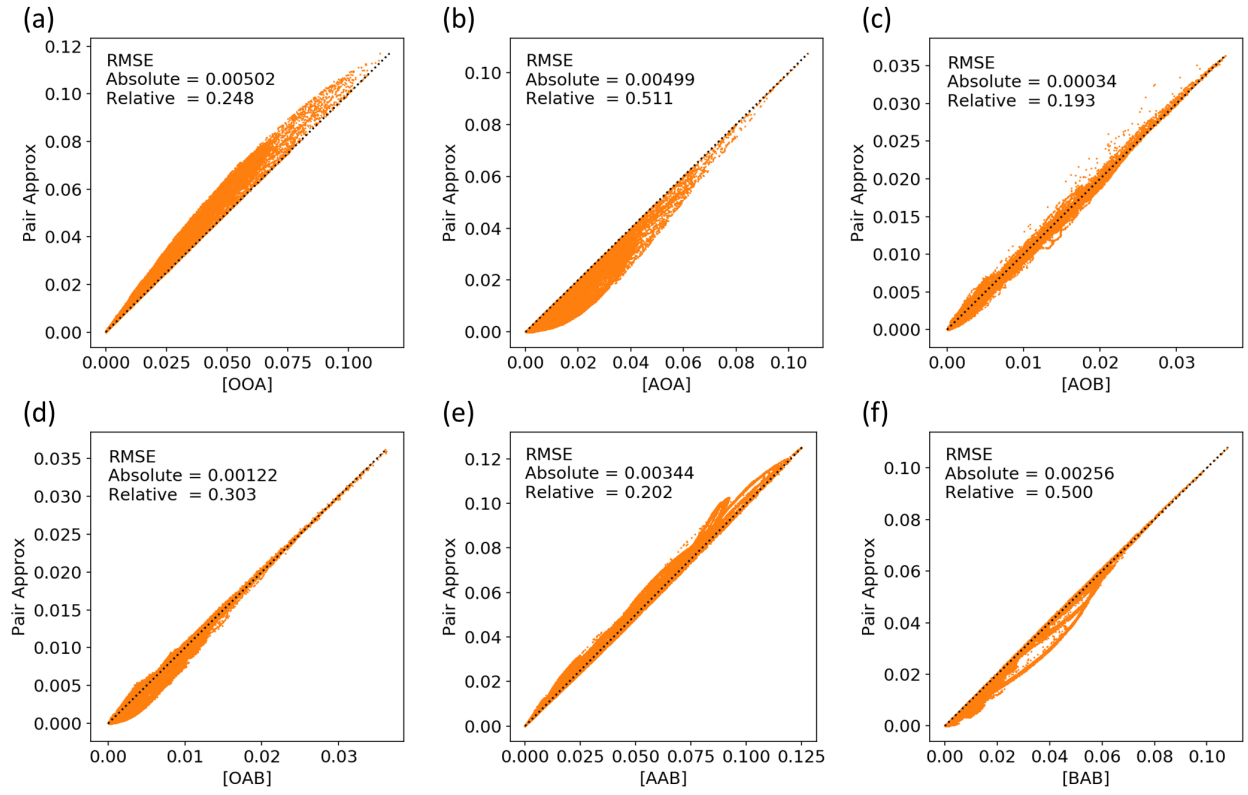
Figure 5: Scatter plots of the PA estimates against the KMC three-site probabilities. Only the test set is plotted, since PA is not a trained model.

of the data points narrower in general, but also the spread is narrower towards the smaller values in particular. As desired, MLMC has reduced the relative error. For a quantitative evaluation, we computed two types of root mean squared error (RMSE) on the test set: the root mean squared *absolute error* and the root mean squared *relative error*. The values are shown on the upper left of Figures 4a-f and 5a-f. Indeed, MLMC gives an order-of-magnitude improvement over PA in both critera.

Remarkably, the kinetic equations (Equation 9) employing MLMC can be integrated using a standard ODE solver.[63-66] The numerical integrability is non-trivial in a number of ways. Even though our FFNNs should be continuous functions, there are a number of pathological properties that can make numerical integration difficult, such as wild oscillations and jaggedness. Moreover, the FFNNs have not been encoded with even basic laws of probability, such as the sum rule $\sum_Z [XYZ] = [XY]$ and the inequality $[XYZ] \leq [XY]$. Indeed, numerical integration can be problematic if the inequality is violated, because overestimation of the three-site probabilities and hence the reaction rates can cause the two-site probabilities to overshoot zero and take negative values. The replacement of zero-valued probabilities in the training data with PA estimates helps to mitigate the problem by guiding the FFNNs to go to zero as the probabilties go to zero.

Figure 6 shows some of the dynamics that were obtained using MLMC. Additional examples are shown in Figure S6. Overall, MLMC gives qualitative improvements over PA. In the ocillatory regime, MLMC gives accurate predictions of the amplitude and period (Figures 6a and S6a), though it can underestimate the damping (Figure S6b). The transition to the non-oscillatory regime also appears to be predicted with good accuracy (Figures 6b, S6c, and S6d). A challenging situation is when one or both of $[A]$ and $[B]$ are near zero. The minority species has a chance to recover or vanish altogether. In the recovering case, KMC has a hard time converging on the dynamics, because the dynamics hinges on a small number of seed molecules. Since the accuracy of MLMC is limited by the training data and hence KMC, MLMC also gives eratic dynamics (Figure 6c). Nonetheless, it is an improvement
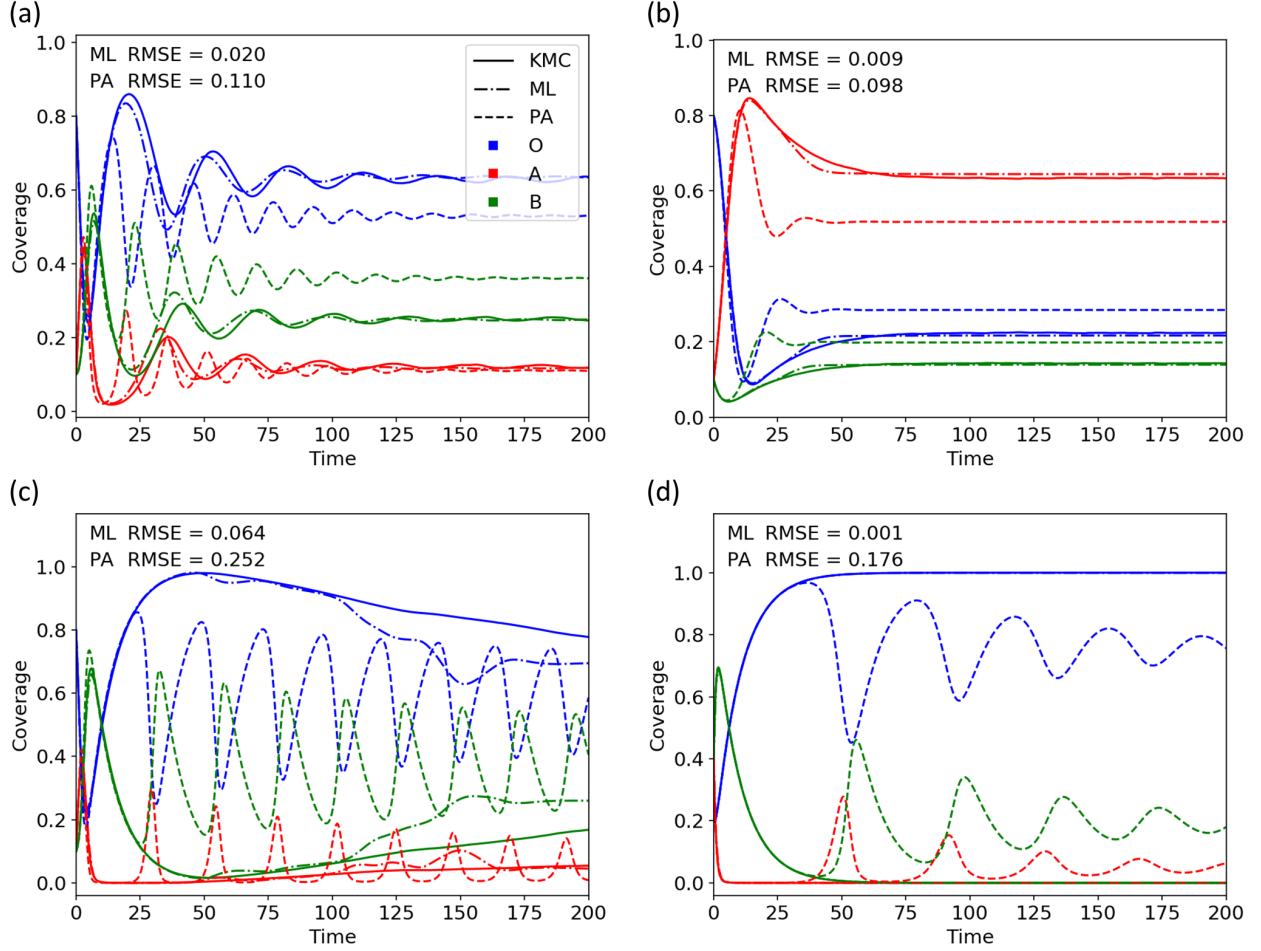
Figure 6: Time-dependent coverages according to PA (dashed lines), ML (dash-dot lines), and KMC (solid lines). (a) The rate constants were $(k_1, k_2, k_3) = (0.5, 0.6, 0.2)$, and the initial conditions were $([O], [A], [B]) = (0.8, 0.1, 0.1)$. (b) The rate constants and the initial conditions were $(0.2, 0.3, 0.4)$ and $(0.8, 0.1, 0.1)$, respectively. (c) The rate constants and the initial conditions were $(0.8, 0.4, 0.1)$ and $(0.2, 0.1, 0.7)$, respectively. (d) The rate constants and the initial conditions were $(0.2, 0.7, 0.1)$ and $(0.2, 0.4, 0.4)$, respectively. In general, MLMC gives both qualitative and quantitative improvements over PA.

over the PA prediction of periodic oscillations. In the vanishing case, MLMC can capture the depletion of a species (Figures 6d and S6e), but sometimes it predicts fictitious recovery with erratic oscillations (Figure S6f). The source of the error might be the replacement of zero-valued probabilities in the training data with PA estimates. Hence, there might be some trade-off between accuracy and robustness.
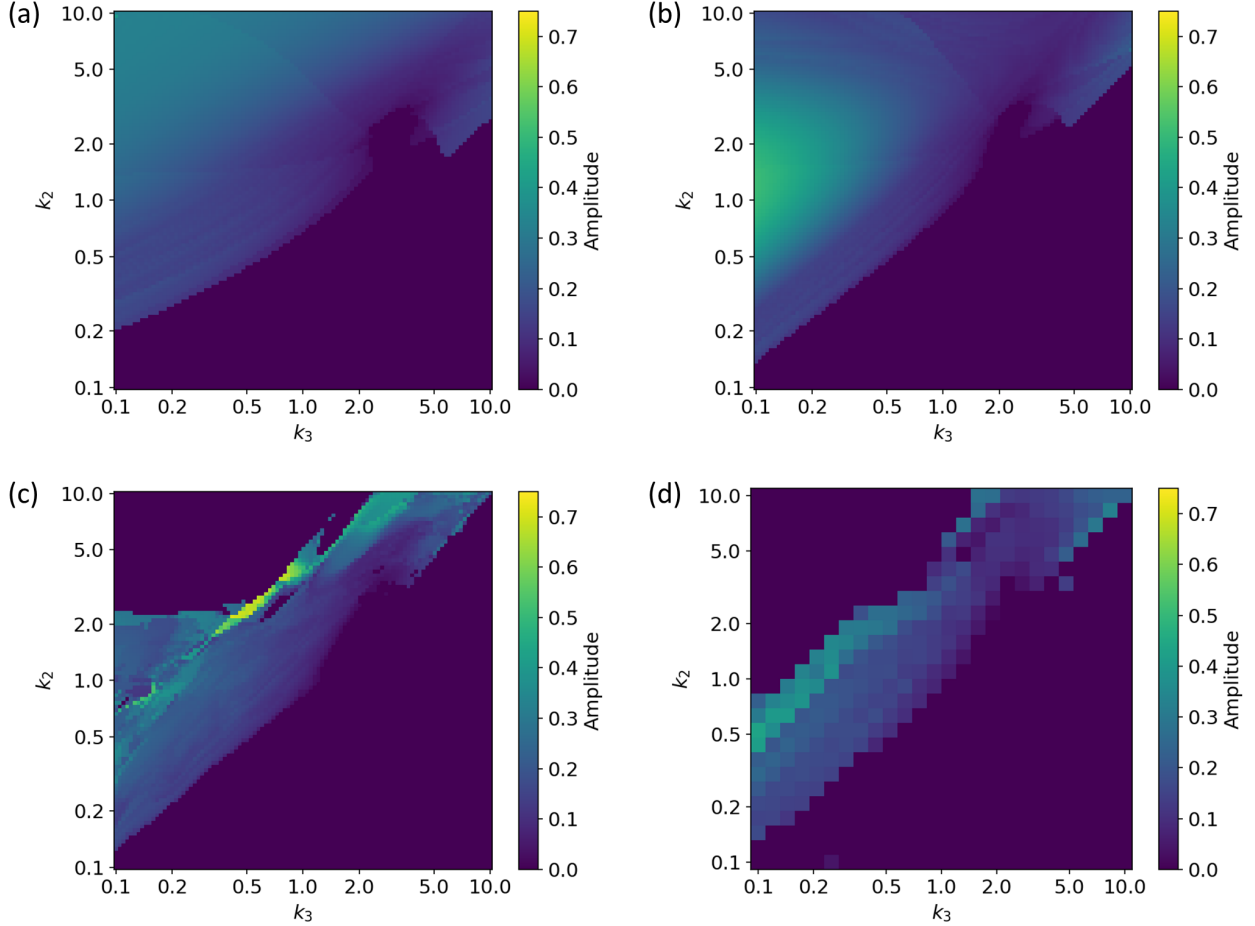

Figure 7: Contour plots of the nonlinear oscillation amplitudes according to the (a) MF, (b) PA, (c) ML, and (d) KMC. Given the nonlinear oscillation of the coverages, the amplitude was defined as the average crest-trough and trough-crest separation over the first two cycles (or three if resolvable). The rate constants were normalized to $k_1 = 1.0$, and the initial conditions were $([O], [A], [B]) = (0.5, 0.3, 0.2)$.

For a comprehensive overview of the method performances, we conducted a systematic survey of the rate constant space by fixing one of the rate constants ($k_1 = 1.0$) and varying the remaining rate constants ($k_2, k_3 \in [0.1, 10.0]$). Figures 7, S7, and S8 map the nonlinear oscil-

lation amplitudes, frequencies, and damping ratios, respectively. Due to the time-consuming nature of KMC simulations, the plots of the KMC data are sparser than the others. Again, MLMC (Figure 7c) has qualitative resemblance to KMC (Figure 7d), whereas PA (Figure 7b) has more in common with MF (Figure 7a). MLMC predicts the accurate position of the lower boundary on the oscillatory regime ($k_2 \gtrsim k_3$) and an approximate position of the upper boundary ($k_2 \lesssim 5k_3$). Notice a triangular region on the center left, where MLMC predicts oscillations in contrast to KMC. The dynamics in this region corresponds to the fictitious recovery discussed earlier (Figure S6f). Because the fictitious ocillations are slow and sustained, the triangular region is almost invisible in the frequency plot (Figure S7) and highlighted in the damping ratio plot (Figure S8). We expect that more training data in that part of the parameter space would have mitigated the erroneous dynamics.

We emphasize that MLMC has not been trained to predict the dynamics. The FFNNs were optimized to predict the instantaneous three-site probabilities using the instantaneous two-site probabilities. On one hand, it is not surprising that accurate estimates of the three-site probabilities could improve the dynamics, On the other hand, it is not obvious that using the best estimates of the three-site probabilities at each instant would yield the best estimate of the overall dynamics. MLMC has conceptual similarities to proposals of ML differential equations, where a ML model is trained on time-dependent data to extract the right hand side of the underlying equation.[67–69] The kinetic equations in terms of moments can be regarded as a special case, where the linear terms are known and the nonlinear terms have well known properties. In the future, it would be interesting to investigate training the FFNNs to output values that yield the best estimate of the dynamics, rather than the best estimate of instantaneous values.

A weakness of MLMC is that it needs KMC data to be trained in the first place. However, we have shown that MLMC can be applied to combinations of rate constants and initial conditions to which it has not been exposed. The internal transferability suggests two practical applications of MLMC.

In the short term, MLMC could be used to interpolate KMC simulations. Often, one needs to run numerous simulations in a small region of the parameter space in order to obtain a phase diagram, locate a critical point, or fit rate constants. Considerable savings in computational costs can result if a select subset of the simulations were performed using KMC and the gaps were filled using MLMC. Indeed, the 500×500 lattice simulations to create Figure 7d took 5 h per simulation. To make the plot at the same resolution as the other methods, KMC would have taken 50 000 h of CPU time, disregarding repetitions to average out stochastic noise. Using MLMC, we could complete the task in 600 h of CPU time: 500 h of KMC to prepare the data, 60 h of training, and 40 h of MLMC simulation. Of course, caution should be taken to detect and correct patches of poor performance, where MLMC might exhibit erroneous dynamics.

In the long term, one can imagine pre-trained closures that are aimed at classes of reaction mechanisms. Similar to the density functional theory (DFT), pre-trained closures can help researchers extract useful insight about typical systems of interest at a fraction of the computational cost of KMC.

Conclusion

We have explored the application of ML to derive moment closures for chemical kinetics. As demonstrated by the case of LLVM, PA exhibits weaknesses in systems with strong long-range correlation. In order to capture the long-range correlation at the same level of input as PA, we trained FFNNs to predict the three-site probabilities using the two-site probabilities. MLMC reduced the absolute and relative errors in the three-site probabilities by an order of magnitude. Furthermore, MLMC gave drastic improvements in the simulated dynamics. The amplitude and period in the oscillatory regime could be predicted to good accuracies, and the dynamical transitions to the non-oscillatory regime could be located to a reasonable precision. Based on these outcomes, we propose that MLMC could be used to interpolate

KMC simulations or construct pre-trained closures to avoid KMC in certain systems.

In the future, we want to demonstrate MLMC on realistic models of specific chemical systems. Given the proposal of pre-trained closures, it would be desirable to construct MLMCs that are transferrable among reactions that share similar mechanisms. Further development of the ML model design might be necessary, such as data transforms and neural network architectures that enforce physical symmetries and the laws of probability. Finally, it might be interesting to find ways to incorporate the effects of static disorder. It would yield a method that can capture both static disorder and dynamic correlation at a fraction of the computational cost of KMC.

Methods

The stochastic simulations were performed using rejection-free KMC[25,59] with periodic boundary conditions (PBC). Unless otherwise mentioned, the KMC coverages in the figures were obtained as averages over 10 simulations on a 500×500 lattice. The kinetic equations were integrated using the variable-coefficient multistep backward differentiation formula (BDF) method[63–65] as implemented in the GNU Scientific Library 2.5.[66] Unless otherwise mentioned, the equations were integrated using relative and absolute tolerances of 10^{-5} and 10^{-10}, respectively.

For the training data, we used the outcomes of KMC simulations at select values of initial conditions and rate constants. We ran the simulation only once at each combination of parameters. First, the the initial conditions were fixed at $([O], [A], [B]) = (0.4, 0.3, 0.3)$, and the rate constants were varied in $k_1, k_2, k_3 \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. The simulations were performed on a 200×200 lattice. At intervals of 0.1 unit time up to 200.0 unit time, we sampled the lattice configurations and counted the two- and the three-site clusters. Then, we chose seven combinations of rate constants that gave distinct dynamics (Table S1). For each combination, the initial conditions were varied in $[O], [A], [B] \in \{0.1, 0.2, ..., 0.9\}$ with

the constraint $[O]+[A]+[B] = 1.0$. This time, the simulations were performed on a 500×500 lattice, and the configurations were sampled at intervals of 0.1 unit time up to 500.0 unit time. The $125 + 7 \times 36 = 377$ simulations yielded a total of 1,171,229 data points.

We found it useful to apply a log transform $y' = \log y$ to both the two-site probabilities, $[XY]$, and the three-site probabilities, $[XYZ]$, so that the ML models minimize the relative error, as opposed to the absolute error. Constraining the relative error ensures that the reaction rates and hence the dynamics would remain accurate when one or more of the probabilities are small. In applying the log transform, we replaced the zero-valued two-site probabilities with small values $\delta = 10^{-p}/(2 \cdot 500^2)$ where $p = 1, 2, 3$ and the zero-valued three-site probabilities with PA estimates. Even though the replacement somewhat reduces the accuracy of the ML models in themselves, we found that it improves the robustness of the chemical kinetic simulations as one or more of the probabilities approach zero. The treatment of the zeros increased the number of data points to 1,232,155. Finally, we standardized the input and output data with the mean $\mu = 0$ and standard deviation $\sigma = 1$.

The ML models were constructed and trained using TensorFlow 1.13.[39] We trained a separate FFNN for each of the six three-site probabilities that appear in the kinetic equations. The FFNNs shared a simple architecture: $6 \rightarrow 100 \rightarrow 100 \rightarrow 75 \rightarrow 50 \rightarrow 25 \rightarrow 12 \rightarrow 1$ units on each layer, which were fully connected and had a rectified linear unit (ReLU) activation, except the linear input and output. As a means of regularization, the data was shuffled and split into six sets, one of which was set aside as the test set, and the other five sets were utilized in a 5-fold cross-validation (CV). At each cycle of the CV, a FFNN was trained by stochastic gradient descent (SGD) on the mean squared error (in the log space) via the Adam optimizer.[70] The SGD used a batch size of 10000 and stopped at 150 iterations. At the end of the cycles, the FFNN that yielded the smallest loss on the test set was chosen as the final model. Beside the CV, no further regularization methods were employed.

Acknowledgement

Supporting Information Available

Coarse-graining of NO + CO / Pt(100)-(1 × 1) on to LLVM; 2D and 3D plots of three-site probabilities as functions of one-site and two-site probabilities; table of rate constants and initial conditions that were represented in the training data; scatter plots of the ML and PA estimates of the three-site probabilities in the log scale; additional examples of time-dependent coverages according to PA, ML, and KMC; and contour plots of nonlinear oscillation frequencies and amplitude ratios.

References

(1) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.

(2) Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; von Lilienfeld, O. A.; Tkatchenko, A.; Mller, K.-R. Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. *J. Chem. Theory Comput.* **2013**, *9*, 3404–3419, PMID: 26584096.

(3) Montavon, G.; Rupp, M.; Gobre, V.; Vazquez-Mayagoitia, A.; Hansen, K.; Tkatchenko, A.; Mller, K.-R.; von Lilienfeld, O. A. Machine learning of molecular electronic properties in chemical compound space. *New J. Phys.* **2013**, *15*, 095003.

(4) Pyzer-Knapp, E. O.; Li, K.; Aspuru-Guzik, A. Learning from the Harvard Clean Energy

Project: The Use of Neural Networks to Accelerate Materials Discovery. *Adv. Funct. Mater.* **2015**, *25*, 6495–6502.

(5) Janet, J. P.; Kulik, H. J. Predicting electronic structure properties of transition metal complexes with neural networks. *Chem. Sci.* **2017**, *8*, 5137–5152.

(6) Gelernter, H.; Rose, J. R.; Chen, C. Building and refining a knowledge base for synthetic organic chemistry via the methodology of inductive and deductive machine learning. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 492–504.

(7) Zhang, Q.-Y.; Aires-de Sousa, J. Structure-Based Classification of Chemical Reactions without Assignment of Reaction Centers. *J. Chem. Inf. Model.* **2005**, *45*, 1775–1783, PMID: 16309284.

(8) Coley, C. W.; Barzilay, R.; Jaakkola, T. S.; Green, W. H.; Jensen, K. F. Prediction of Organic Reaction Outcomes Using Machine Learning. *ACS Cent. Sci.* **2017**, *3*, 434–443.

(9) Di Lena, P.; Nagata, K.; Baldi, P. Deep architectures for protein contact map prediction. *Bioinformatics* **2012**, *28*, 2449–2457.

(10) Eickholt, J.; Cheng, J. Predicting protein residueresidue contacts using deep networks and boosting. *Bioinformatics* **2012**, *28*, 3066–3072.

(11) Faraggi, E.; Zhang, T.; Yang, Y.; Kurgan, L.; Zhou, Y. SPINE X: Improving protein secondary structure prediction by multistep learning coupled with prediction of solvent accessible surface area and backbone torsion angles. *J. Comput. Chem.* **2012**, *33*, 259–267.

(12) Snyder, J. C.; Rupp, M.; Hansen, K.; Müller, K.-R.; Burke, K. Finding Density Functionals with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 253002.

(13) Li, L.; Snyder, J. C.; Pelaschier, I. M.; Huang, J.; Niranjan, U.-N.; Duncan, P.;

Rupp, M.; Müller, K.-R.; Burke, K. Understanding machine-learned density functionals. *Int. J. Quantum Chem.* **2016**, *116*, 819–833.

(14) Brockherde, F.; Vogt, L.; Li, L.; Tuckerman, M. E.; Burke, K.; Müller, K.-R. Bypassing the Kohn-Sham equations with machine learning. *Nat. Commun.* **2017**, *8*, 872.

(15) Grisafi, A.; Fabrizio, A.; Meyer, B.; Wilkins, D. M.; Corminboeuf, C.; Ceriotti, M. Transferable Machine-Learning Model of the Electron Density. *ACS Cent. Sci.* **2019**, *5*, 57–64.

(16) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.

(17) Seko, A.; Takahashi, A.; Tanaka, I. Sparse representation for a potential energy surface. *Phys. Rev. B* **2014**, *90*, 024101.

(18) Li, Z.; Kermode, J. R.; De Vita, A. Molecular Dynamics with On-the-Fly Machine Learning of Quantum-Mechanical Forces. *Phys. Rev. Lett.* **2015**, *114*, 096405.

(19) Chmiela, S.; Sauceda, H. E.; Müller, K.-R.; Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. *Nat. Commun.* **2018**, *9*, 3887.

(20) Gillespie, D. T. A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. *J. Comput. Phys.* **1976**, *22*, 403–434.

(21) Munsky, B.; Khammash, M. The finite state projection algorithm for the solution of the chemical master equation. *J. Chem. Phys.* **2006**, *124*, 044104.

(22) Hellander, A.; Ltstedt, P. Hybrid method for the chemical master equation. *J. Comput. Phys.* **2007**, *227*, 100–122.

(23) Smadbeck, P.; Kaznessis, Y. N. A closure scheme for chemical master equations. *Proceedings of the National Academy of Sciences* **2013**, *110*, 14261–14265.

(24) Schnoerr, D.; Sanguinetti, G.; Grima, R. Comparison of different moment-closure approximations for stochastic chemical kinetics. *J. Chem. Phys.* **2015**, *143*, 185101.

(25) Chatterjee, A.; Vlachos, D. G. An overview of spatial microscopic and accelerated kinetic Monte Carlo methods. *J. Comput. Aided Mater. Des.* **2007**, *14*, 253–308.

(26) Pineda, M.; Stamatakis, M. Beyond mean-field approximations for accurate and computationally efficient models of on-lattice chemical kinetics. *J. Chem. Phys.* **2017**, *147*, 024105.

(27) Ernst, O. K.; Bartol, T.; Sejnowski, T.; Mjolsness, E. Deep Learning Moment Closure Approximations using Dynamic Boltzmann Distributions. 2019, arXiv:1905.12122. arXiv.org e-Print archive. https://arxiv.org/abs/1905.12122 (accessed Nov 23, 2020).

(28) Whittle, P. On the Use of the Normal Approximation in the Treatment of Stochastic Processes. *J. R. Stat. Soc. B* **1957**, *19*, 268–281.

(29) Gómez-Uribe, C. A.; Verghese, G. C. Mass fluctuation kinetics: Capturing stochastic effects in systems of chemical reactions through coupled mean-variance computations. *J. Chem. Phys.* **2007**, *126*, 024109.

(30) Nåsell, I. An extension of the moment closure method. *Theor. Popul. Biol.* **2003**, *64*, 233–239.

(31) Keeling, M. J. Multiplicative Moments and Measures of Persistence in Ecology. *J. Theor. Biol.* **2000**, *205*, 269–281.

(32) Dickman, R. Kinetic phase transitions in a surface-reaction model: Mean-field theory. *Phys. Rev. A* **1986**, *34*, 4246–4250.

(33) Matsuda, H. In *Animal Societies: Theories and Facts*; Ito, Y., Brown, J. P., Kikkawa, J., Eds.; Japan Scientific Societies Press: Tokyo, 1987; pp 67–80.

(34) Matsuda, H.; Ogita, N.; Sasaki, A.; Satō, K. Statistical Mechanics of Population: The Lattice Lotka-Volterra Model. *Prog. Theor. Phys.* **1992**, *88*, 1035–1049.

(35) Keeling, M. J.; Rand, D. A.; Morris, A. J. Correlation models for childhood epidemics. *Proc. R. Soc. Lond. B* **1997**, *264*, 1149–1156.

(36) Hiebeler, D. E.; Millett, N. E. Pair and triplet approximation of a spatial lattice population model with multiscale dispersal using Markov chains for estimating spatial autocorrelation. *J. Theor. Biol.* **2011**, *279*, 74 – 82.

(37) Marceau, V.; Noël, P.-A.; Hébert-Dufresne, L.; Allard, A.; Dubé, L. J. Adaptive networks: Coevolution of disease and topology. *Phys. Rev. E* **2010**, *82*, 036116.

(38) Lindquist, J.; Ma, J.; van den Driessche, P.; Willeboordse, F. H. Effective degree network disease models. *J. Math. Biol.* **2011**, *62*, 143–164.

(39) Abadi, M. et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. 2016, arXiv:1603.04467. arXiv.org e-Print archive. https://arxiv.org/abs/1603.04467 (accessed Oct 14, 2020).

(40) Pedregosa, F. et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

(41) Lotka, A. J. Contribution to the Theory of Periodic Reactions. *J. Phys. Chem.* **1910**, *14*, 271–274.

(42) Lotka, A. J. Analytical Note on Certain Rhythmic Relations in Organic Systems. *Proc. Nat. Acad. Sci.* **1920**, *6*, 410–415.

(43) Lotka, A. J. *Elements of Physical Biology*; Williams & Wilkins, 1925.

(44) Volterra, V. Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. *Mem. Acad. Lincei Roma.* **1926**, *2*, 31–113.

(45) Imbihl, R.; Fink, T.; Krischer, K. Bifurcation analysis of the threevariable model for the NO+CO reaction on Pt surfaces. *J. Chem. Phys.* **1992**, *96*, 6236–6248.

(46) Lipowski, A.; Lipowska, D. Nonequilibrium phase transition in a lattice preypredator system. *Physica A* **2000**, *276*, 456 – 464.

(47) Antal, T.; Droz, M. Phase transitions and oscillations in a lattice prey-predator model. *Phys. Rev. E* **2001**, *63*, 056119.

(48) Kowalik, M.; Lipowski, A.; Ferreira, A. L. Oscillations and dynamics in a two-dimensional prey-predator system. *Phys. Rev. E* **2002**, *66*, 066107.

(49) Mobilia, M.; Georgiev, I. T.; Täuber, U. C. Phase Transitions and Spatio-Temporal Fluctuations in Stochastic Lattice Lotka-Volterra Models. *J. Stat. Phys.* **2007**, *128*, 447–483.

(50) Keitt, T. H.; Johnson, A. R. Spatial Heterogeneity and Anomalous Kinetics: Emergent Patterns in Diffusion-Limited Predatory-Prey Interaction. *J. Theor. Biol.* **1995**, *172*, 127–139.

(51) Provata, A.; Nicolis, G.; Baras, F. Oscillatory dynamics in low-dimensional supports: A lattice LotkaVolterra model. *J. Chem. Phys.* **1999**, *110*, 8361–8368.

(52) Kopelman, R. Fractal Reaction Kinetics. *Science* **1988**, *241*, 1620–1626.

(53) Moiny, F.; Dumont, M.; Dagonnier, R. Fractal kinetics and surface reactions. *J. Chem. Phys.* **1998**, *108*, 4572–4581.

(54) Imbihl, R.; Ertl, G. Oscillatory Kinetics in Heterogeneous Catalysis. *Chem. Rev.* **1995**, *95*, 697–733.

(55) Veser, G.; Imbihl, R. Synchronization and spatiotemporal selforganization in the NO+CO reaction on Pt(100). I. Unsynchronized oscillations on the 11 substrate. *J. Chem. Phys.* **1994**, *100*, 8483–8491.

(56) Fink, T.; Dath, J.; Imbihl, R.; Ertl, G. Kinetic oscillations in the NO + CO reaction on Pt(100): Experiments and mathematical modeling. *J. Chem. Phys.* **1991**, *95*, 2109–2126.

(57) Geva, N.; Vaissier, V.; Shepherd, J.; Van Voorhis, T. Mean field treatment of heterogeneous steady state kinetics. *Chem. Phys. Lett.* **2017**, *685*, 185 – 190.

(58) McIsaac, A. R.; Vaissier Welborn, V.; Einzinger, M.; Geva, N.; Weir, H.; Baldo, M. A.; Van Voorhis, T. Investigation of External Quantum Efficiency Roll-Off in OLEDs Using the Mean-Field Steady-State Kinetic Model. *J. Phys. Chem. C* **2020**, *124*, 14424–14431.

(59) Voter, A. F. In *Radiation Effects in Solids*; Sickafus, K. E., Kotomin, E. A., Uberuaga, B. P., Eds.; Springer Netherlands: Dordrecht, 2007; pp 1–23.

(60) Gillespie, D. T. Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.* **2001**, *115*, 1716–1733.

(61) Chow, C.; Liu, C. Approximating discrete probability distributions with dependence trees. *IEEE Trans. Inf. Theory* **1968**, *14*, 462–467.

(62) Kang, H.-J.; Kim, K.; Kim, J. H. Optimal approximation of discrete probability distribution with kth-order dependency and its application to combining multiple classifiers. *Pattern Recognit. Lett.* **1997**, *18*, 515 – 523.

(63) Byrne, G. D.; Hindmarsh, A. C. A Polyalgorithm for the Numerical Solution of Ordinary Differential Equations. *ACM Trans. Math. Softw.* **1975**, *1*, 7196.

(64) Brown, P. N.; Byrne, G. D.; Hindmarsh, A. C. VODE: A Variable-Coefficient ODE Solver. *SIAM J. Sci. Stat. Comput.* **1989**, *10*, 1038–1051.

(65) Hindmarsh, A. C.; Brown, P. N.; Grant, K. E.; Lee, S. L.; Serban, R.; Shumaker, D. E.;

Woodward, C. S. SUNDIALS: Suite of Nonlinear and Differential/Algebraic Equation Solvers. *ACM Trans. Math. Softw.* **2005**, *31*, 363–396.

(66) Galassi, M.; Gough, B.; Rossi, F.; Theiler, J.; Jungman, G.; Booth, M.; Davies, J. *GNU Scientific Library Reference Manual*; Network Theory Ltd., 2001.

(67) Long, Z.; Lu, Y.; Ma, X.; Dong, B. PDE-Net: Learning PDEs from Data. 2017, arXiv:1710.09668. arXiv.org e-Print archive. https://arxiv.org/abs/1710.09668 (accessed Oct 23, 2020).

(68) Raissi, M.; Karniadakis, G. E. Hidden physics models: Machine learning of nonlinear partial differential equations. *J. Comput. Phys.* **2018**, *357*, 125 – 141.

(69) Rackauckas, C.; Ma, Y.; Martensen, J.; Warner, C.; Zubov, K.; Supekar, R.; Skinner, D.; Ramadhan, A.; Edelman, A. Universal Differential Equations for Scientific Machine Learning. 2020, arXiv:2001.04385. arXiv.org e-Print archive. https://arxiv.org/abs/2001.04385 (accessed Oct 21, 2020).

(70) Kingma, D. P.; Ba, J. Adam: A Method for Stochastic Optimization. 2014, arXiv:1412.6980. arXiv.org e-Print archive. https://arxiv.org/abs/1412.6980 (accessed Oct 14, 2020).

Graphical TOC Entry