Discovering dynamics and parameters of nonlinear oscillatory and chaotic systems from partial observations

George Stepaniants ⬤,[1,*] Alasdair D. Hastewell ⬤,[2] Dominic J. Skinner ⬤,[2,3] Jan F. Totz ⬤,[2] and Jörn Dunkel ⬤[2]

[1]*Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, California 91125, USA*
[2]*Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[3]*NSF-Simons Center for Quantitative Biology, Northwestern University, Evanston, Illinois 60208, USA*

Despite rapid progress in data acquisition techniques, many complex physical, chemical, and biological systems remain only partially observable, thus posing the challenge to identify valid theoretical models and estimate their parameters from an incomplete set of experimentally accessible time series. Here, we combine sensitivity methods and ranked-choice model selection to construct an automated hidden dynamics inference framework that can discover predictive nonlinear dynamical models for both observable and latent variables from noise-corrupted incomplete data in oscillatory and chaotic systems. After validating the framework for prototypical FitzHugh-Nagumo oscillations, we demonstrate its applicability to experimental data from squid neuron activity measurements and Belousov-Zhabotinsky reactions, as well as to the Lorenz system in the chaotic regime.

Nonlinear oscillations and chaotic dynamics are ubiquitous in natural and man-made systems [1], from neurons [2,3] and biochemical networks [4] to power grids [5,6], lasers [7], and the Earth's climate [8]. In many complex physical, chemical, and biological systems exhibiting these kinds of dynamics, direct measurements are limited to a few experimentally accessible observables [9] while essential components of the underlying dynamical circuit stay hidden [10]. Limited observability has led to the emergence of competing theoretical models for neuronal networks [11], gene-regulatory dynamics [10], and chemical oscillators [12] among many other systems, and identifying valid models and their parameters from incomplete data remains a central challenge. Here, we combine sensitivity methods [13] for differential equations with ranked-choice voting [14,15] to construct a hidden dynamics inference (HDI) framework that can discover predictive nonlinear dynamical models for both observable and latent variables from noise-corrupted incomplete data in oscillatory and chaotic systems.

Driven by rapidly advancing data acquisition techniques [16–20], dynamical model inference from time-series data is becoming increasingly important [21,22] in climate physics [23–25], fluid mechanics [26,27], and biophysics [28–30]. In canonical form, data-driven dynamics discovery consists of fitting the parameters **p** of an ordinary differential equation (ODE) $\dot{x} = f(x, \mathbf{p})$ to a given data set. Serious consideration has been given to the problem of dynamical

system discovery ever since early computers were used to numerically solve differential equations [31]. Renewed interest in model discovery came with increasing computational power in the 1980s and 1990s [21]. With the rising prevalence of machine learning techniques and the explosion of high-resolution data acquisition, once again attention has turned to this problem in recent years [13,32–35]. One class of approaches, encompassing methods for "equation-based" discovery [22,36], has shown promise in learning interpretable dynamical models from partially observed data using physics-informed neural networks [32,34,37–39], manifold methods [40], or data assimilation [33], enabling prediction of nonlinear and chaotic dynamics in mechanical, electrical, and hydrodynamic systems.

Complementary approaches to account for unobserved variables have frequently relied on time-delay embeddings [41–43], recurrent neural networks [44], or autoencoders [45] to estimate hidden dimensions and forecast complex dynamics [46]. These "equation-free" approaches, however, often cannot reveal coupling mechanisms and their dependencies on experimental conditions. Despite substantial progress in dynamical systems inference techniques, applications to partially observed experimental data from nonlinear biophysical and biochemical systems still face many open problems, as existing methods require long time-series recordings with low noise (e.g., to construct time-delay embeddings or train neural networks) and do not ensure stability of learned models.

In Sec. I we review previous approaches to ODE model inference focusing on techniques that can be extended to partial observations in the context of the HDI approach introduced here. We then discuss several important dynamical system model families in Sec. II. Our HDI framework is presented in Sec. III, where it is illustrated on the inference problem of a prototypical neuronal oscillator. Model inference from partial observations is not unique, and in Sec. IV we study how to

---

*Contact author: gstepan@caltech.edu

compare different models learned from the same partial observations and test whether they are structurally equivalent under a change of coordinates. Lastly, in Sec. V we show how our HDI framework is capable of discovering the chaotic Lorenz model from one to two coordinates, and in Sec. VI we use it to infer models from experimental recordings of squid axon neuron spike trains and the oscillatory color change dynamics of the Belousov-Zhabotinsky chemical reaction.

I. OVERVIEW OF DYNAMICAL SYSTEMS INFERENCE APPROACHES

We begin here by reviewing several major classes of dynamical systems models, namely parametric, nonparametric, and higher-order models, along with the methods used to optimize and search through these model classes.

A. Parametric models

The simplest setting for learning a dynamical system model is when it takes a parametric form, where the ODE right-hand side $\dot{x} = f(x, \mathbf{p})$ is fixed ahead of time by some knowledge of the system. Frequently, $\mathbf{p}$ will be low-dimensional, consisting of only a few parameters, as for example in the nonlinear pendulum $(\dot{x}, \dot{v}) = (v, A\sin(\omega x))$ with parameters $\mathbf{p} = (A, \omega)$.

B. Nonparametric models

Parametric models are only applicable when we have some prior knowledge about the form of the model governing the dynamics in our data, which is frequently not the case for experimental systems. If we cannot assume a particular parametric form of the equations, a more expressive and less constrained flow map f in $\dot{x} = f(x, \mathbf{p})$ is needed. Taking $f(x, \mathbf{p})$ to be a general function with a large set of model parameters $||\mathbf{p}||_0 \gg 1$, for instance a neural network or Gaussian process, results in a so-called *nonparametric* model. Here, nonparametric means that the large set of parameters $\mathbf{p}$ has no inherent interpretation. Examples of such functions f within the context of dynamical inference include locally linear models [47,48], basis expansions [49] (sometimes referred to as atlases or libraries including polynomial [22], rational polynomial [50,51], wavelet [52], and radial basis functions [53]), discrete-time perceptrons [54,55], or neural networks (referred to as neural ODEs) [13].

C. Reduced higher-order models

When some coordinates of a dynamical system are not observed, another approach is to learn a higher-order model in the coordinates which can be observed. We call such a model a *reduced model*. This can be done directly by taking continuous derivatives [56] of the observed coordinates [57] or indirectly by constructing a state vector from successive time lags [32,42,43]. However, even simple multivariate systems give rise to complicated reduced higher-order equations, which often have many terms, are implicit, and contain fractional powers [58–66]. Learning reduced models directly requires learning dense implicit ODEs with many candidate terms [51,67], a challenging approach that can be ill-posed [68].

Derivative fitting methods [58] require computing higher-order numerical derivatives of observed coordinates, which is prohibitive for noisy or highly oscillatory data. All such methods face serious problems due to the appearance of rational terms that complicate the optimization as the dynamics of the learned model can frequently set the denominator of these terms close to zero. While reduced models for partially observed systems are a useful tool for model comparison, as we will demonstrate throughout this paper, they are difficult to directly learn from data.

D. Optimization methods

The methods for fitting parametric and nonparametric models are largely the same. One class of approaches fit an equation of the form $\dot{x} = f(x, \mathbf{p})$ to data $\{(t_i, x_i)\}_{i=1}^n$ by minimizing a data loss $\sum_i |\dot{x}_i - f(x_i, \mathbf{p})|^2$. Approximations of the time derivative can be calculated by finite differences, smooth derivative approximations [21,69,70], or weak formulations [71]. Minimizing the data loss to find $\mathbf{p}$ can be done using either linear [22,72] or nonlinear [70,73] least squares based on the structure of the dynamical equations. For instance, the parameters of the Lorenz system enter linearly on the right-hand side while the parameters of the nonlinear pendulum $\ddot{x} = -A\sin(\omega x)$ do not. For nonparametric models, a common choice is to discretize the ODE model as $x_{t+\Delta t} = f(x_t, \mathbf{p})$ to avoid computing derivatives of time-series data [49,50,53–55]. These regression approaches, however, cannot be used for partial observations, since some coordinates of the system are not available to compute time derivatives. Furthermore, such methods are not robust and can overfit noisy data, as they impose no causal constraints that the observed data at time $t + \Delta t$ are related to the data at time t through the evolution of an underlying differential equation.

To avoid these issues, a second class of model fitting techniques instead optimizes the parameters of $\dot{x} = f(x, \mathbf{p})$ by integrating the dynamics of $x(t)$ and matching it to observed data $\{(t_i, x_i)\}_{i=1}^n$. Such optimization techniques roughly fall into two categories: (i) gradient-free methods like root finding [74], multiple shooting [75], Nelder-Mead, particle swarm optimization [76], or Kalman filters [77–79]; and (ii) gradient-based methods like Newton's method, ADAM [80], and BFGS [81], where gradients with respect to parameters $\mathbf{p}$ are approximated either using finite differencing on the ODE solutions, sensitivity methods, or automatic differentiation [13,82]. These techniques have been applied successfully to learn periodic, chaotic, and transient dynamics [13]. When partial knowledge of the ODE model is available, a parametric interpretable ODE model can be added to a nonparametric model and jointly fit to the data, an approach called hybrid modeling [78]. The question of parameter identifiability, which studies whether model coefficients can be determined uniquely from the data, has also been extensively explored for parametric models [83,84].

E. Problem constraints

As mentioned above, in the case of partial observations or noise corrupted data, a causal formulation of the problem is needed, which encodes prior knowledge that observed data

are related through the evolution of a differential equation. When the integrated solution to the ODE model has a closed form such as with linear time-invariant systems, least-squares fitting can be used to match the integrated solution to the data [85]. For more complicated models, we need to use numerical approximations to the integrated ODE solution and fit the parameters by solving suitably chosen often nonlinear and nonconvex optimization problems which depend only on the observed variables. The causal nature of the problem can be encoded by adding additional penalty terms to the data loss enforcing an ODE structure, leading to a large class of approaches known as data assimilation methods [33,86]. These approaches typically enforce that the fit dynamics is consistent with a single time step of an ODE model, although generalizations to n steps have shown better stability and noise robustness [87–89]. While better suited for data with partial observations and noise, there is still no guarantee that the integrated model is stable over long times or fits the data when these finite steps are performed sequentially. A stronger constraint on the ODE structure is that the solution $\{x(t, \mathbf{p})\}$ comes from numerically integrating the model, which is then fitted to the data. We can achieve this hard constraint by directly optimizing through the numerical ODE solver. This is the technique that we will exploit in our HDI methodology presented here.

F. Model selection

Without prior knowledge about the form of the dynamics, nonparametric models can fit an ODE to complex dynamical data remarkably well. Yet the resulting model lacks physical interpretability. This drawback prompted the development of automated methods for learning physically interpretable ODE models without assuming prior knowledge of the dynamics. A popular approach is to first learn a nonparametric model given as a linear combination of basis terms (i.e., atlas, library), and to subsequently perform model selection keeping only a few basis terms in the expansion [22,49]. Various techniques for basis term selection include sparse identification [22], bootstrapping [90], information criteria (MLE, AIC, BIC, MDL) ranking [49,91,92], and hypothesis testing [93,94]. Recent approaches have also explored genetic algorithms [95] and formal logic [96] for learning interpretable dynamical system models that consist of *nonlinear* compositions of mathematical functions.

G. Our contribution

Our HDI methodology falls into the class of physically interpretable methods, and learns ODE models as a sparse combination of polynomial or trigonometric basis functions. However, HDI uses a novel robust procedure for model selection by sampling the space of possible ODE models, identifying the largest cluster of the best-fit models, and keeping those basis terms in our final model that have the least variation in their coefficients across all model fits. Conceptually, this procedure is akin to a community of theoreticians being given a time-series observation and each constructing their own candidate model to explain the data. Then a census is reached by finding the most popular model form (e.g.,

largest model cluster) and ranking its terms to remove those basis terms to which the dynamics of the model are insensitive (e.g., large changes in coefficients do not affect model fit). This allows HDI to robustly select a few important model terms from libraries containing on the order of 10–100 candidate terms, thus exploring a broad range of potential ODE models. HDI's core model fitting procedure uses gradient-based sensitivity methods which are robust to high levels of noise in the data, ensure the learned models accurately capture the experimental dynamics, and are stable over long-time integration.

II. DYNAMICAL SYSTEM MODELS

Many dynamical systems exhibit nonlinear and chaotic dynamics that can be modeled by polynomial nonlinearities, and their coordinates can exhibit evolution at different timescales, leading to relaxation oscillations and excitations. This motivates us to define the following class of models, with observed variables $x_1, \ldots, x_m$ and hidden variables $h_{m+1}, \ldots, h_M$, given by

$$\dot{x}_k = \tau_k \sum_{|\boldsymbol{\alpha}| \leqslant d_k} c_{\boldsymbol{\alpha}}^k x_1^{\alpha_1} \cdots h_M^{\alpha_M}, \quad 1 \leqslant k \leqslant m, \qquad \text{(1a)}$$

$$\dot{h}_k = \tau_k \sum_{|\boldsymbol{\alpha}| \leqslant d_k} c_{\boldsymbol{\alpha}}^k x_1^{\alpha_1} \cdots h_M^{\alpha_M}, \quad m < k \leqslant M, \qquad \text{(1b)}$$

which encompass a broad range of nonlinear oscillatory and chaotic dynamics. Here, we only use polynomial terms on the right-hand side of the equation, although this can be extended to any other nonlinearities, such as trigonometric functions [Sec. VII of the Supplemental Material (SM) [97]]. The power of each monomial term above is expressed in multi-index notation $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_M)$ where each entry $\alpha_r \in \mathbb{N}$ denotes the power of x_r and we write $|\boldsymbol{\alpha}| = \sum_{k=1}^M \alpha_k$ to denote the total degree of the monomial term. To avoid scaling ambiguities between τ_k and $\mathbf{c}^k = \{c_{\boldsymbol{\alpha}}^k\}$ in Eq. (1), we enforce that each $\mathbf{c}^k$ has unit norm. HDI models are determined by a parameter vector $\mathbf{p}$ containing the initial conditions of the variables $\{x_k^0\}_{k=1}^m$, $\{h_k^0\}_{k=m+1}^M$, frequencies $\{\tau_k\}_{k=1}^M$, and polynomial coefficients $\{\mathbf{c}^k\}_{k=1}^M$. While time-delay embeddings can be used to provide lower bound estimates on the number of hidden variables $M - m$, here we restrict to periodic models with $M = 2$ variables or chaotic models with $M = 3$ variables, which we find sufficient to explain the experimental data.

An alternative class of dynamical system models are derivative embedding or higher-order models, sometimes referred to as time delay embedding models when written in discrete time. They take the general form

$$x_k^{(p_k)} = \sum_{|\boldsymbol{\alpha}| \leqslant d_k} c_{\boldsymbol{\alpha}}^k (x_1)^{\alpha_1} \cdots \left(x_m^{(p_m-1)}\right)^{\alpha_M}, \quad 1 \leqslant k \leqslant m, \qquad \text{(2)}$$

where the notation $x^{(p)}$ denotes the pth time derivative of x, and as before, x^{α} denotes exponentiation. This model is nothing more than a polynomial dynamical system whose state consists of each observed variable x_k for $1 \leqslant k \leqslant m$ along with all of its derivatives up to order $p_k - 1$, where $p_k \geqslant 0$ is some integer. This model assumes that the highest p_kth-order derivative of each observed variable x_k can be expressed as a

function of all other observed variables and their derivatives. The total number of variables in the model is M as we assume that $\sum_{k=1}^{m} p_k = M$. We again remark that such higher-order models can also be constructed with time lags instead of time derivatives.

Note that higher-order models as written in (2) are a special case of the more general polynomial hidden variable models from (1) as every hidden variable h_k can be set to one of the higher-order derivatives of the observed variables $x_k^{(p)}$. This is the classical transformation of a higher-order differential equation into first-order form. The benefit of higher-order dynamical models is that all coordinates of such dynamical systems are observed, and therefore these models hold the promise of being easily fit to partially observed data. At this point, one might hope to avoid using hidden variables and their associated ambiguities by learning the reduced higher-order equation in the observed variable directly [98,99]. However, even simple multivariate systems can give rise to complex reduced higher-order equations that are often less sparse, implicit (e.g., nonpolynomial), and contain fractional powers [58–66]; for example,

$$\dot{x} = xy^3, \quad \dot{y} = x \tag{3}$$

reduces to

$$x\ddot{x} = \dot{x}^2 + 3x^{7/3}\dot{x}^{2/3}. \tag{4}$$

Working with reduced models directly would thus require learning dense implicit ODEs with more candidate terms [51,67], a challenging approach that can be ill-posed [68]. A general advantage of "first-order" hidden variable models is that they can find sparse and integrable differential equation models that agree with observational data, avoiding the above complications.

III. HIDDEN DYNAMICS INFERENCE

The HDI framework integrates the robustness of sensitivity methods [13] and ranked-choice model selection [14,15] with traditional library-based learning methods [22,90]. This enables us to learn physically interpretable models for partially hidden nonlinear systems from short, highly noisy data trajectories in a manner that ensures correct long time dynamics. Since the hidden-variable dynamical equation discovered from partial observations may not be unique, we develop systematic algebraic tests for comparing learned models. After validating the HDI framework on 50% noise corrupted simulations of the FitzHugh-Nagumo oscillator, we apply our approach to experimental measurements of squid neuron spike trains and video observations of Belousov-Zhabotinsky chemical reactions, demonstrating how HDI can be used to measure model parameters as a function of external experimental conditions.

To illustrate the general framework of HDI, we consider as a canonical example of a nonlinear oscillator the FitzHugh-Nagumo (FHN) model [11]

$$\dot{v} = v - \frac{v^3}{3} - w + I, \quad \dot{w} = \tau(v + a - bw), \tag{5}$$

a simplified model of a firing neuron where the membrane voltage v undergoes a rapid increase before being diminished

by the slow recovery variable w [100]. The rapid spiking and slow recovery arises from a separation in timescales $\tau \ll 1$ between variables. FHN has become a prototypical model of neuron spike trains, as it is stable and parsimonious, relying only on a small number of polynomial terms. Throughout this section, we detail the HDI framework in its full generality and show how it performs when learning the FHN oscillator from noisy partial recordings of its v coordinate [Fig. 1(a)]. The v coordinate is corrupted with 50% Gaussian noise (Sec. II of the SM) so we are performing model inference in the high noise regime.

A. Time-series data

To begin, our framework is given m observed variables $\{y_{i1}\}_{i=1}^{n}, \ldots, \{y_{im}\}_{i=1}^{n}$ at n distinct time points $\{t_i\}_{i=1}^{n}$. Our goal is to learn an interpretable and predictive first-order multivariate model (1) with $M \geqslant m$ variables (m observed and $M - m$ hidden) such that its first m coordinates match these observed data. We will often denote the vector $\mathbf{y}_i = (y_{i1}, \ldots, y_{im})$ as the stacked vector of observables at time t_i. On the example of the FHN model, if we only observe the v coordinate and do not observe w, then we have $m = 1$ observed coordinates and data y_{i1} at some observation time points $\{t_i\}_{i=1}^{n}$.

B. Model optimization

The first step of HDI performs a model sweep, where it initializes randomly many first-order hidden variable models (1) with random coefficients and varying degree combinations $(d_1, \ldots, d_M)$.

All initialized models are dense, which means they contain all polynomial terms up to degree d_k in the kth equation for $1 \leqslant k \leqslant M$. Each randomly initialized model is parametrized by the stacked vector

$$\mathbf{p} = \left(\{x_k(0)\}_{k=1}^{M}, \{\mathbf{c}_k\}_{k=1}^{M}, \{\tau_k\}_{k=1}^{M}\right) \tag{6}$$

of its initial conditions, coefficients, and timescales and is trained for one round of optimization to minimize the objective

$$L(\mathbf{p}) = \text{MSE}(\mathbf{p}) + \text{Reg}(\mathbf{p}), \tag{7}$$

which is a sum of the mean-squared error

$$\text{MSE}(\mathbf{p}) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} [x_j(t_i, \mathbf{p}) - y_{ij}]^2 \tag{8}$$

and a regularization term. The regularization term in the objective function is given by

$$\text{Reg}(\mathbf{p}) = \lambda \sum_{k=1}^{M} \sum_{|\boldsymbol{\alpha}| \leqslant d_k} \sqrt{1 + |\boldsymbol{\alpha}|} |c_{\boldsymbol{\alpha}}^{k}| + \gamma \sum_{k=1}^{M} (\|\mathbf{c}_k\|^2 - 1)^2. \tag{9}$$

The first term in the regularizer above penalizes the sparsity of the learned model, since polynomial terms $x_1^{\alpha_1} \cdots x_M^{\alpha_M}$ with a higher total degree $|\boldsymbol{\alpha}| = \alpha_1 + \cdots + \alpha_M$ are more actively down-weighted by the factor $\sqrt{1 + |\boldsymbol{\alpha}|}$. The choice of square root in the regularization function was empirically chosen over the conventional L^1 regularizer $|\boldsymbol{\alpha}|$, as it allowed more
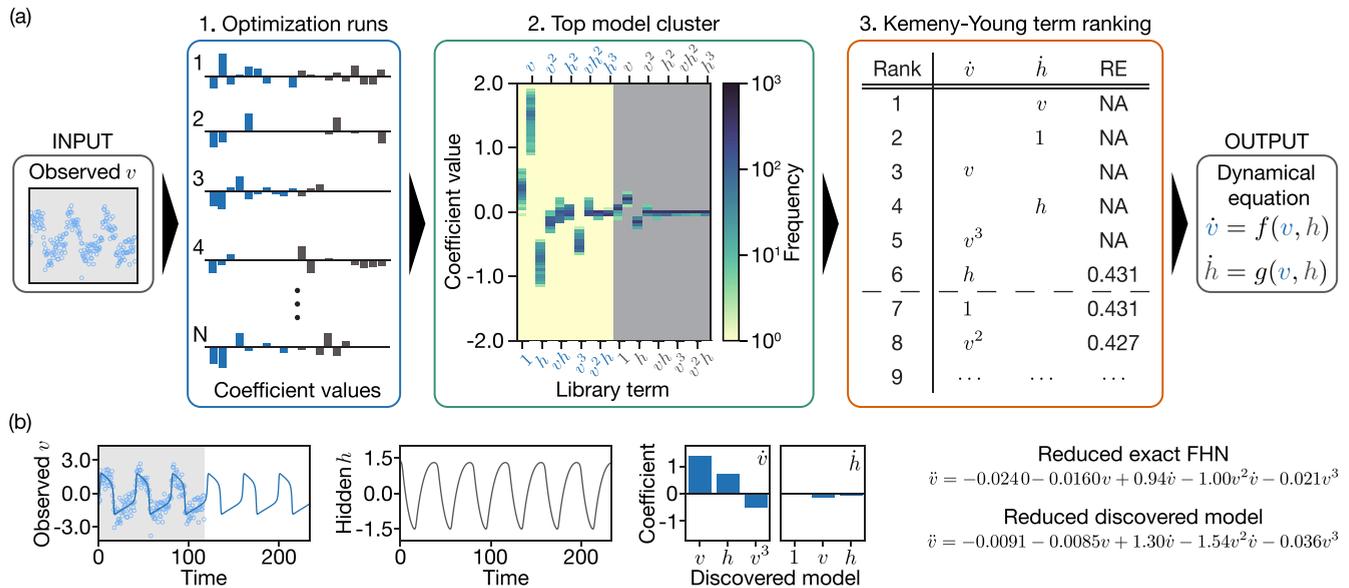
FIG. 1. General HDI framework illustrated for 50% noise-corrupted FHN simulation data. (a) Algorithm flow chart: (1) ODE sensitivity optimization [13] yields $N \sim 20\,000$ candidate models by tuning 20 parameters of dense two field cubic observed (blue) and hidden (dark-gray) variable equations from random initializations. Models are filtered for stability and fit quality. (2) The remaining $\sim$4000 models are hierarchically clustered using the cosine similarity between their parameter vectors (Sec. III of the SM; Model Clustering). Histograms of parameters in the largest cluster are used to rank the terms based on their coefficient of variation. (3) Kemeny-Young ranking produces a list of candidate models of decreasing sparsity. Models are refit at each sparsity level, and the user can select the model that best balances sparsity and relative error (RE). (b) Using data from only the v time series corrupted by 50% noise, HDI correctly discovers a sparse first-order system that reduces to the same second-order form as the FHN model.

flexibility in our optimization to place larger coefficient values at high-degree terms while still enforcing model sparsity. The second term in the regularizer enforces weakly that the norm of the coefficients in each equation is equal to unity, allowing us to optimize the coordinate timescales contained in $\{\tau_k\}_{k=1}^{M}$ independently of the polynomial coefficients. Our initial model sweep tests a range of regularization parameters $\lambda = 10^{-5}$–10^{0} and keeps the unity constraint weight at a sufficiently large value $\gamma = 5 \times 10^{4}$.

Each model is optimized for 100 000 AdaBelief [101] iterations followed by 50 000 iterations of BFGS [102] (large enough for our optimization to converge). In general, our method is not overly sensitive to how these values are set. Computing the gradient of these ODE models with respect to the parameter vector **p** is performed through fast forward sensitivity methods [13]. In Fig. 1(a,1), on the example of FHN with only its v coordinate observed, we show the typical list of models that result after training a batch of models to fit the partially observed data.

After one round of optimization, we remove in an automated fashion all models that do not pass the following dynamical criteria. Namely, for oscillatory data such as observed in neuronal and chemical dynamics, we only keep models whose fits to the data are periodic with amplitude and frequency commensurate with the recorded data. If the dynamics we aim to learn are chaotic instead of periodic such as the Lorenz attractor, we remove all models with periodic fits since periodicity is a very unlikely behavior for chaotic systems. In general, we also remove any models whose fits to the data converge to fixed points or diverge to infinity which

is easily checked by simulating these models for a sufficiently long time window.

For all remaining models, their relative errors (REs) on the training data $\{(t_i, \mathbf{y}_i)\}_{i=1}^{n}$ of the observed variables $\mathbf{x}(t) = \{x_k(t)\}_{k=1}^{m}$ are computed using the formula

$$\text{RE} = \sqrt{\frac{1}{m} \sum_{k=1}^{m} \frac{\sum_{i=1}^{n}(x_k(t_i) - y_{ik})^2}{\sum_{i=1}^{n}\left(y_{ik} - \frac{1}{n}\sum_{j=1}^{n}y_{jk}\right)^2}}. \quad (10)$$

Now we need to decide which models have a sufficiently good fit by keeping only those whose train RE is below a certain threshold θ_{RE}. This threshold is determined automatically through a histogram thresholding procedure (Fig. 2 of the SM).

C. Computing distances between ODE models

For all remaining models with RE below the threshold θ_{RE}, we rescale their coefficients c_α^k and frequencies τ_k such that the standard deviations of the trajectories of all hidden variables $h_{m+1}, \ldots, h_M$ are equal to the standard deviation of the trajectory of the first observed variable x_1. This normalization ensures that the standard deviations of the hidden variables across all learned models are on the same scale and hence the coefficients of these models can be compared to each other. After normalization, we extract from each model the stacked coefficient vector $\bar{\mathbf{c}} = \{\tau_k \mathbf{c}^k\}_{k=1}^{M}$ with timescales included. Now each model is fully described by its stacked
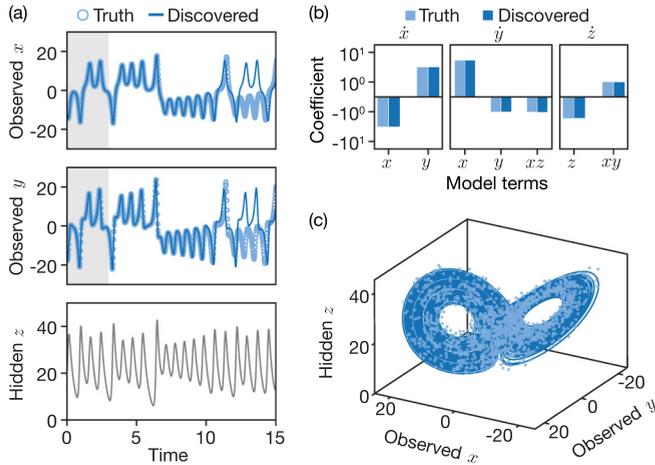
FIG. 2. HDI discovers true Lorenz system from observations of x and y coordinates. (a) Given observations of only the x and y co-ordinates (gray region), the learned model predicts the evolution for five to six Lyapunov timescales. (b) The correct Lorenz model terms are discovered by HDI search solely from data in the gray region of the previous panel. (c) Reconstructed attractor of the learned model closely agrees with the true Lorenz attractor (simulation parameters and noise robustness in Sec. VI of the SM).

vector of coefficients $\bar{\mathbf{c}}$, and this representation can be used to compare it to other models.

To compare the learned hidden variable models, we account for certain transformations of the hidden variables $(h_{m+1}, \dots, h_M)$, specifically sign flips and permutations, that result in equivalent models (Sec. III of the SM; Model Distance Matrix). Finally, we compute a pairwise distance matrix using the cosine distance between every pair of models (stacked coefficient vectors) $\bar{\mathbf{c}}, \bar{\mathbf{c}}'$ modulo these sign flip and permutation symmetries.

D. Model clustering

Using this distance matrix between all models, we perform single-linkage clustering and obtain a dendogram tree which specifies how models are sequentially grouped together into larger and larger clusters. We then determine where to cut this dendogram tree generated by hierarchical clustering to split our models into separate clusters. Choosing a cutoff too high or too low in the dendogram results in clusters that are overestimated (too many models) or underestimated (too few models), respectively. We develop a thresholding procedure to select cutoffs $n_{\min}, n_{\max}$, which give us the smallest and largest number of clusters, respectively, which we could reliably split our N models into. In Fig. 1(a,2) we show the largest model cluster that results when learning FHN solely from an observation of its v coordinate. Here we display a heatmap showing typical values of coefficients for each polynomial term across all models in this cluster.

E. Model term ranking

At every level $n_{\min} \leqslant n \leqslant n_{\max}$ of the dendogram tree, we study the largest cluster of models $C(n)$ and sort the terms in these models from most to least important. Namely, at a given
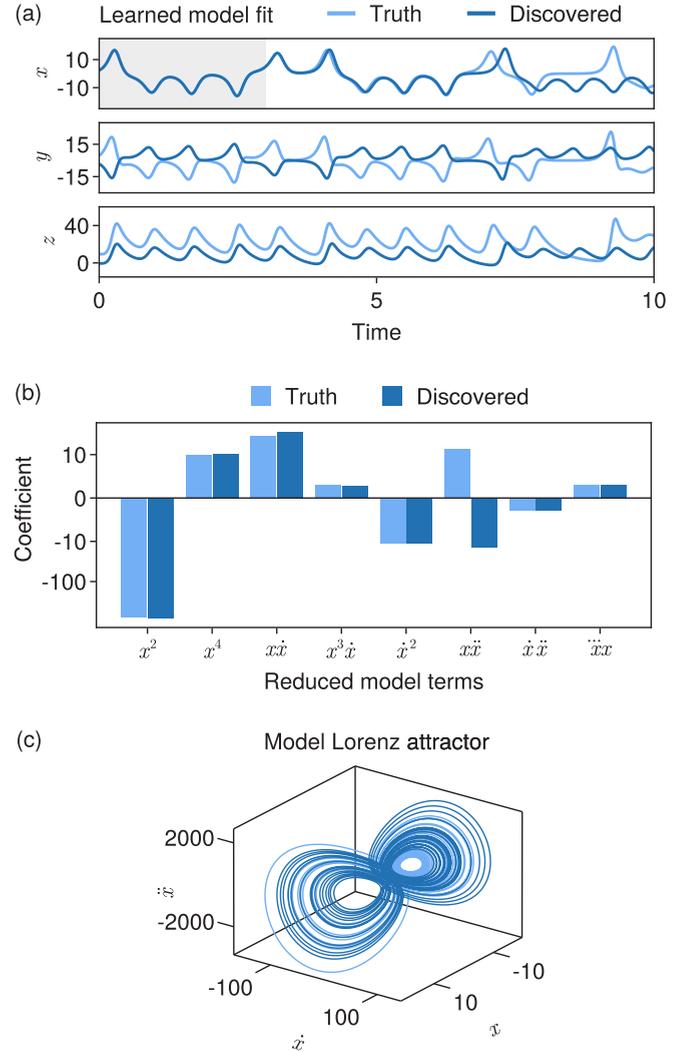
FIG. 3. HDI discovers a predictive model for the Lorenz system from just the x coordinate. (a) From observations of only the x coordinate (gray region), the learned model predicts the evolution of true Lorenz for three to four Lyapunov timescales. (b) The learned HDI model can be transformed into a reduced third-order model just in x where it agrees with the true x-reduced Lorenz model except at one term. (c) The attractor of the learned model in derivative embedding space $(x, \dot{x}, \ddot{x})$ matches the true Lorenz attractor.

level n we save the stacked coefficient vector $\bar{\mathbf{c}} = \{\tau_k \mathbf{c}^k\}_{k=1}^{M}$ for every polynomial model in this cluster (see Fig. 3 of the SM). Assume that in this cluster we have R models denoted by their stacked coefficient vectors $\bar{\mathbf{c}}_1, \dots, \bar{\mathbf{c}}_R$. We would like to understand which entries (model terms) across all r vectors are large in magnitude and have low variability.

The ith entry $\bar{\mathbf{c}}_{ir}$ in the stacked coefficient vector for models $r = 1, \dots, R$ corresponds to a polynomial term in one of the equations of our ODE system. To test whether this term is important for the dynamics learned by these R models, we compute its coefficient of variation

$$\frac{\text{interquartile range}\left(\{\bar{\mathbf{c}}_{ir}\}_{r=1}^{R}\right)}{\text{median}\left(\{\bar{\mathbf{c}}_{ir}\}_{r=1}^{R}\right)}, \tag{11}$$

which is the interquartile range of the coefficient values divided by their median value. The interquartile range of a set of numbers is defined as the difference between the 75th and 25th percentiles of their distribution. Intuitively, if a term i is consistent and tightly clustered around a large mean value for all models $\bar{\mathbf{c}}_{i1}, \ldots, \bar{\mathbf{c}}_{iR}$, then it receives a small coefficient of variation. Otherwise, if a term i has large spread or is close to zero across all R models, then it receives a large coefficient of variation.

From this we can obtain a ranking $r_1^n, \ldots, r_T^n$ of the T terms in our model by sorting the coefficients of variation from smallest to largest (breaking ties based on numeric order). Here $r_1^n, \ldots, r_T^n$ is simply a permutation of the set $\{1, \ldots, T\}$. To robustly determine the ranking of model terms from most to least important, we create a consensus from the individual rankings obtained at each cluster level $\{(r_1^n, \ldots, r_T^n)\}_{n=n_{\min}}^{n_{\max}}$. To aggregate these rankings into one, we use a traditional method for aggregating votes or ballots known as the Kemeny-Young or VoteFair popularity ranking algorithm [14,15]. This algorithm finds the *average* ranking that minimizes the sum of Kendall tau distances to the list of individual rankings at each cluster level $n = n_{\min}, \ldots, n_{\max}$. Using this method, we obtain the final aggregated ranking $r_1, \ldots, r_T$ of all T terms in our ODE model.

Returning to our FHN example, in Fig. 1(a,3) we see that this Kemeny-Young ranking of model terms returns a list where the top six terms give a sparse and predictive polynomial model for FHN of the form

$$\dot{v} = av + bh + cv^3,$$
$$\dot{h} = e + fv + gh. \tag{12}$$

Note that the top seven terms in this list from Fig. 1(a,3) exactly recover all the polynomial terms in the true FHN model. Next, we discuss why the six-term model above, which misses the constant term in the $\dot{v}$ equation, is mathematically identical to the true FHN equations.

IV. HIGHER-ORDER MODEL COMPARISONS

In our example on learning FHN from partial observations, we found several two-dimensional models that could generate the same v dynamics in their first coordinate. In fact, there exist many diffeomorphic transformations of the hidden w coordinate of FHN, leading to new dynamical systems that agree in simulations of the observed v coordinate. This ability to transform hidden coordinates is a general degree of freedom when inferring dynamical systems from partial observations. This necessitates methodologies allowing us to test whether the dynamical models we discover are indeed equivalent (under hidden coordinate diffeomorphisms) to the ground-truth dynamical systems from which we obtained partial observations. Below we propose two methods, one manual and one automated, to test for such model equivalences. We again illustrate these methods on the example of FHN.

A. Manually testing for model equivalence

Taking the true FHN model in Eq. (5), solving for w in terms of v, $\dot{v}$ and substituting into the $\dot{w}$ equation, we obtain a

second-order *reduced model* solely in v. Namely, we can write

$$w = -\dot{v} + v - cv^3 + I, \tag{13}$$

which can be substituted into the $\dot{w}$ equation of FHN (5) to obtain

$$\ddot{v} - \dot{v} + 3cv^2\dot{v} = -\frac{1}{\tau}v - \frac{a}{\tau} + \frac{b}{\tau}(-\dot{v} + v - cv^3 + I). \tag{14}$$

This equation can then be expanded into

$$\ddot{v} = -3cv^2\dot{v} + \left(1 - \frac{b}{\tau}\right)\dot{v} - \frac{bc}{\tau}v^3 + \frac{b-1}{\tau}v + \frac{bI - a}{\tau}. \tag{15}$$

We say that Eq. (15) is in "reduced form" because it depends solely on the value and derivatives of the observable v.

Performing a similar reduction of our learned six-term model

$$\dot{v} = av + bh + cv^3,$$
$$\dot{h} = e + fv + gh, \tag{16}$$

we can solve for h as a function of v and $\dot{v}$ and substitute in the $\dot{h}$ equation to find

$$\ddot{v} = 3cv^2\dot{v} + (a + g)\dot{v} - cgv^3 + (bf - ag)v + be. \tag{17}$$

We see that this v-reduced model has the same structure and similar coefficients as the true reduced FHN model [Fig. 1(b)]. In fact, it is easy to check that the coefficients (a, b, c, e, f, g) of model (16) can always be chosen appropriately so that (17) matches the coefficients of true reduced FHN (15). This confirms that HDI has recovered a two-variable model that is structurally equivalent to ground-truth FHN. In this example, one could argue that comparing our model to true FHN through such higher-order reductions was unnecessary as our learned model (16) was structurally similar to the true form of FHN and could be diffeomorphically transformed into true FHN by a simple shift in its hidden coordinate. However, models that are equivalent in a subset of their coordinates are not always so readily identified. We refer the reader to our SM, where we show how a very different polynomial model

$$\dot{v} = \left(1 + \frac{1}{\beta}\right)v - \frac{1}{\beta}h - cv^3 + \left(\frac{\alpha}{\beta} + I\right),$$
$$\dot{h} = -cv^3 - \left(\frac{b}{\tau} + \frac{1}{\beta}\right)h \tag{18}$$

is again equivalent to the true FHN model in its v coordinate, by transforming into a higher-order form. This equivalence can likewise be seen by properly transforming the h coordinate of (18) into the w coordinate of true FHN.

Because the evolution of a subset of coordinates of a dynamical system cannot be uniquely captured by one multivariate first-order differential equation, reductions of such multivariate first-order systems into a unique higher-order form are a useful procedure for performing model comparisons. In the model equivalence tests above, we described how two models can be shown to be equivalent in a subset of their coordinates (e.g., v coordinate) if reducing them to higher-order forms results in equations with the same terms and nearby coefficient values.

B. Robust algebraic test for model equivalence

We now discuss an alternative algebraic procedure that allows us to automatically verify when polynomial dynamical models are equivalent in a subset of their coordinates, using again the idea of higher-order reductions. This second test gives a necessary but not sufficient condition of model equivalence, but is more efficient and robust than the approach above and requires that only one of the equations being compared is reduced into a higher-order form.

To begin, suppose we wish to test whether a learned model such as (18) is equivalent to true FHN in its v coordinate. Recall the higher-order form of true FHN in its v coordinate derived in (15). We aim to test whether the model in (18) can possibly be reduced into this form. First, we rewrite (18) with fully general coefficients

$$\dot{v} = w_1 + w_2 v + w_3 h + w_4 v^3,$$
$$\dot{h} = w_5 h + w_6 v^3. \tag{19}$$

We can compute $\ddot{v}$ from this model by writing

$$\ddot{v} = w_2 \dot{v} + w_3 \dot{h} + 3 w_4 v^2 \dot{v}$$
$$= w_1 w_2 + w_2^2 v + (w_2 w_3 + w_3 w_5) h + 3 w_1 w_4 v^2$$
$$+ (4 w_2 w_4 + w_3 w_6) v^3 + 3 w_3 w_4 v^2 h + 3 w_4^2 v^5.$$

Now we substitute the values of $v, \dot{v}, \ddot{v}$ into the true higher-order form of FHN (15) and test whether it can possibly be satisfied by a model of the form (19). Specifically, we first write the higher-order form of FHN,

$$g(\ddot{v}, \dot{v}, v) = \ddot{v} + 3 c v^2 \dot{v} - \left(1 - \frac{b}{\tau}\right) \dot{v}$$
$$+ \frac{bc}{\tau} v^3 - \frac{b-1}{\tau} v - \frac{bI - a}{\tau} \tag{20}$$

and substituting $v, \dot{v}, \ddot{v}$ this turns into a polynomial $\overline{g}(v, h)$, which we aim to set exactly to zero,

$$g(\ddot{v}, \dot{v}, v) = \overline{g}(v, h; \{w_k\}_{k=1}^6) \equiv 0. \tag{21}$$

Note that because $\overline{g}$ is a polynomial in v, h, in order for it to exactly equal zero, the coefficients in front of each of its monomial terms must be zero. Enforcing that $\overline{g} \equiv 0$ gives us a (typically overcomplete) system of polynomial constraints on the $\{w_k\}_{k=1}^6$ model coefficients that must be satisfied and can be solved for through symbolic substitution or homotopy continuation methods [103]. We refer the reader to the SM (Sec. VIII) for more details and a general presentation of this model comparison test.

Hence, we can now test whether there exists any set of coefficients $\{w_k\}_{k=1}^6$ under which a model of the general form (19) could have the correct higher-order form as true FHN. This in turn gives us a necessary condition that the model learned in (18) is equivalent to FHN in its v coordinate. We show in the next section how these automatic tests are useful in proving that HDI accurately discovers the equations of the chaotic Lorenz attractor from partial observations of its coordinates.

V. CHAOTIC DYNAMICS: LORENZ SYSTEM

HDI straightforwardly extends to higher-dimensional nonlinear and chaotic systems. Here we study it in the three-variable chaotic Lorenz system

$$\dot{x} = \sigma(y - x), \tag{22a}$$
$$\dot{y} = x(\rho - z) - y, \tag{22b}$$
$$\dot{z} = xy - \beta z. \tag{22c}$$

We investigate the mathematical equivalence and predictive ability of the models that HDI learns when given one or two of the three coordinates of Lorenz as observations. The parameters of the ground-truth Lorenz model are set to $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$ for which the maximal Lyapunov timescale under which trajectories deviate is known to be approximately 0.9056. We will use this Lyapunov timescale as a reference time to measure the ability of our learned models to predict the trajectory of ground-truth Lorenz when both models are started from the same initial conditions.

A. Observations of *x* and *y* variables

We give our procedure observations of the x and y coordinates of the 3D Lorenz system for one or two lobe transitions [gray-shaded in Fig. 2(a)]. An HDI search over all polynomial three-variable ODEs in (x, y, z) with quadratic interactions recovers the exact Lorenz equations with correct coefficient values (modulo a trivial scaling of the hidden z variable) [Fig. 2(b); SM]. The learned model has the correct attractor dynamics and can predict the x, y, and z dynamics for a bit more than five Lyapunov timescales [Figs. 2(a) and 2(c)].

B. Observations of *x* variable

Even when only given observations of x, HDI learns a predictive model for Lorenz dynamics, albeit with reduced predictive power (Fig. 3). Our method finds a seven-term model

$$\dot{x} = w_1 x + w_2 y,$$
$$\dot{y} = w_3 x + w_4 x z,$$
$$\dot{z} = w_5 + w_6 z + w_7 x y, \tag{23}$$

which is strikingly similar to the true Lorenz system except that a linear term in y is missing from the $\dot{y}$ equation.

As shown in Fig. 3(a), this ODE model exactly fits the training data from the simulated Lorenz system (gray box) on which it was trained and is capable of predicting two additional branch switches of the Lorenz attractor which occur for approximately two Lyapunov timescales past the training window. We now test whether the model learned from just the x-coordinate of the Lorenz system exactly reduces to the true x-reduced Lorenz equation. To this end, we first compute the form of the true higher-order model of the x coordinate of Lorenz.

Taking (22a), we can rewrite it as

$$y = \frac{\dot{x}}{\sigma} + x, \quad \dot{y} = \frac{\ddot{x}}{\sigma} + \dot{x}.$$

Then substituting this and (22a) into (22b) gives us

$$z = -\frac{1}{\sigma}\frac{\ddot{x}}{x} - \left(1 + \frac{1}{\sigma}\right)\frac{\dot{x}}{x} + \rho - 1.$$

Given this expression for z, we can compute its derivative as

$$\dot{z} = -\frac{1}{\sigma}\frac{\dddot{x}}{x} + \frac{1}{\sigma}\frac{\dot{x}\ddot{x}}{x^2} - \left(1 + \frac{1}{\sigma}\right)\frac{\ddot{x}}{x} + \left(1 + \frac{1}{\sigma}\right)\frac{\dot{x}^2}{x^2}.$$

Finally, we can substitute the above two expressions for y, z, and $\dot{z}$ into (22c) to obtain

$$x\dddot{x} - \dot{x}\ddot{x} + (\sigma + \beta + 1)x\ddot{x} - (\sigma + 1)\dot{x}^2 + x^3\dot{x}$$
$$+ \beta(\sigma + 1)x\dot{x} + \sigma x^4 - \sigma\beta(\rho - 1)x^2 = 0. \quad (24)$$

The reduction of the Lorenz equations to a higher-order ODE in solely the x coordinate was previously derived in [58,61]. Using our automatic verification tool for model reductions described in Sec. IV, we find that our model (23) does not reduce to this form, no matter how we set the coefficients $w_1, \ldots, w_7$.

To see why this is, we reduce our learned model (23) in the x-coordinate and obtain the following equation:

$$x\dddot{x} - \dot{x}\ddot{x} + (w_1 + w_4 w_6)x\ddot{x} + w_1\dot{x}^2 - w_4 w_7 x^3\dot{x}$$
$$+ w_1 w_4 w_6 x\dot{x} + w_1 w_4 w_7 x^4 + w_2(w_3 w_6 - w_4 w_5)x^2$$
$$= 0.$$

The terms in this equation agree precisely with the x-reduced equation of the true Lorenz system in (24). Given that the true Lorenz system was simulated with parameters $\sigma = 10$, $\rho = 28$, $\beta = 8/3$, we now attempt to match the coefficient of our x-reduced model with that of Lorenz. This gives us the constraints

$$w_1 + w_4 w_6 = \sigma + \beta + 1,$$
$$w_1 = -\sigma - 1,$$
$$w_4 w_7 = -1,$$
$$w_1 w_4 w_6 = \beta(\sigma + 1),$$
$$w_1 w_4 w_7 = \sigma,$$
$$w_2(w_3 w_6 - w_4 w_5) = -\sigma\beta(\rho - 1), \quad (25)$$

from which we derive the contradiction that $w_1 = -\sigma$ and $w_1 = -\sigma - 1$. This explains why our automatic tool for model reductions could not find a set of coefficients $w_1, \ldots, w_7$ that would satisfy the true x-reduced Lorenz equation. In fact, notice that by setting $w_1 = -\sigma - 1$ we can set $w_4 w_6 = -\beta$ and $w_4 w_7 = -\sigma/(\sigma + 1)$ and all except the first and third constraint equations above will be satisfied, resulting in a sign flip of the coefficient in front of $x\ddot{x}$ and a slight shift in the coefficient of $x^3\dot{x}$. These are precisely the higher-order coefficients found by our learned Lorenz model [Fig. 3(b)]. Therefore, at the expense of estimating the wrong sign in the coefficient of $x\ddot{x}$, our model finds the correct coefficients for all other terms in the higher-order equation of Lorenz.

Due to the algebraic similarity of the learned model to the true Lorenz dynamics, this model closely matches the dynamics of the x coordinate of the Lorenz system. Another way to test this is to overlay the derivative embedded dynamics in $(x, \dot{x}, \ddot{x})$ of our learned model over that of true Lorenz

from which we see a close agreement of the model attractors [Fig. 3(c)]. We refer the reader to the SM (Sec. VI A) where we in fact show how this model is similarly predictive for several Lyapunov times when transformed into derivative embedding space.

We next apply HDI to identify quantitative models from experimental data for neuron activity and chemical reactions.

VI. EXPERIMENTAL APPLICATIONS

A. Squid axon voltage recordings

Figure 4(a) shows experimental measurements [104,105] of the membrane potential v in the giant axon of the North Atlantic longfin inshore squid (*Loligo pealeii*) in response to noisy stimulus input currents. Following previous spike train model formulations [100,107–109], we apply HDI to the time-series data for v to learn a sparse two-variable model [Figs. 4(a) and 4(b)]. Consistent with prior descriptions of neuron dynamics [11], the phase portrait of the discovered seven-term model is governed by a homoclinic orbit [Fig. 4(c)]. The limit cycle that is discovered by the HDI framework is valid in the regime of the available data. As seen by the model sweep (SM) there are multiple models that provide consistent limit cycle dynamics. To further distinguish between these models, additional transient data off the limit cycle would be needed. Importantly, the model generalizes to describe recordings from different squids, yielding consistent coefficients across all samples [Figs. 4(d) and 4(e)].

B. Belousov-Zhabotinsky color dynamics

For a second more challenging HDI application, we performed Belousov-Zhabotinsky (BZ) reaction experiments [110]. Over the course of the reaction, a substrate species is slowly consumed that fuels the periodic rise and decay of intermediary reagents far from thermodynamic equilibrium. The basic reaction scheme [111] involves more than 20 chemical species and 40 reaction steps. A plethora of different chemical models have been developed that capture the BZ reaction qualitatively [106,112–115]. In our experiments, the repeated oxidation and reduction of the metal catalyst ferroin produces a periodic change in color of the solution from red to light blue [Fig. 5(a)]. The recorded average color of the solution follows a 1D curve in color space which we map to our single observed coordinate $c(t)$ (Fig. 11 of the SM).

A standard three-variable ODE model for the BZ reaction is the Oregonator model,

$$\dot{u} = v - u, \quad (26a)$$

$$\dot{v} = \frac{1}{\varepsilon_1}(-w(v - \mu) - v^2 + v), \quad (26b)$$

$$\dot{w} = \frac{1}{\varepsilon_2}(fu + \phi - w(v + \mu)), \quad (26c)$$

where u, v, and w correspond to the concentrations of the oxidized catalyst M_{ox}, bromous acid $HBrO_2$, and bromide Br^-, respectively. Here μ, f, ϕ are all non-negative. Usually we have that $\varepsilon_2 \ll \varepsilon_1$ so the inhibitor species w can be *adiabatically eliminated* by setting $fu + \phi - w(v + \mu) = 0$.
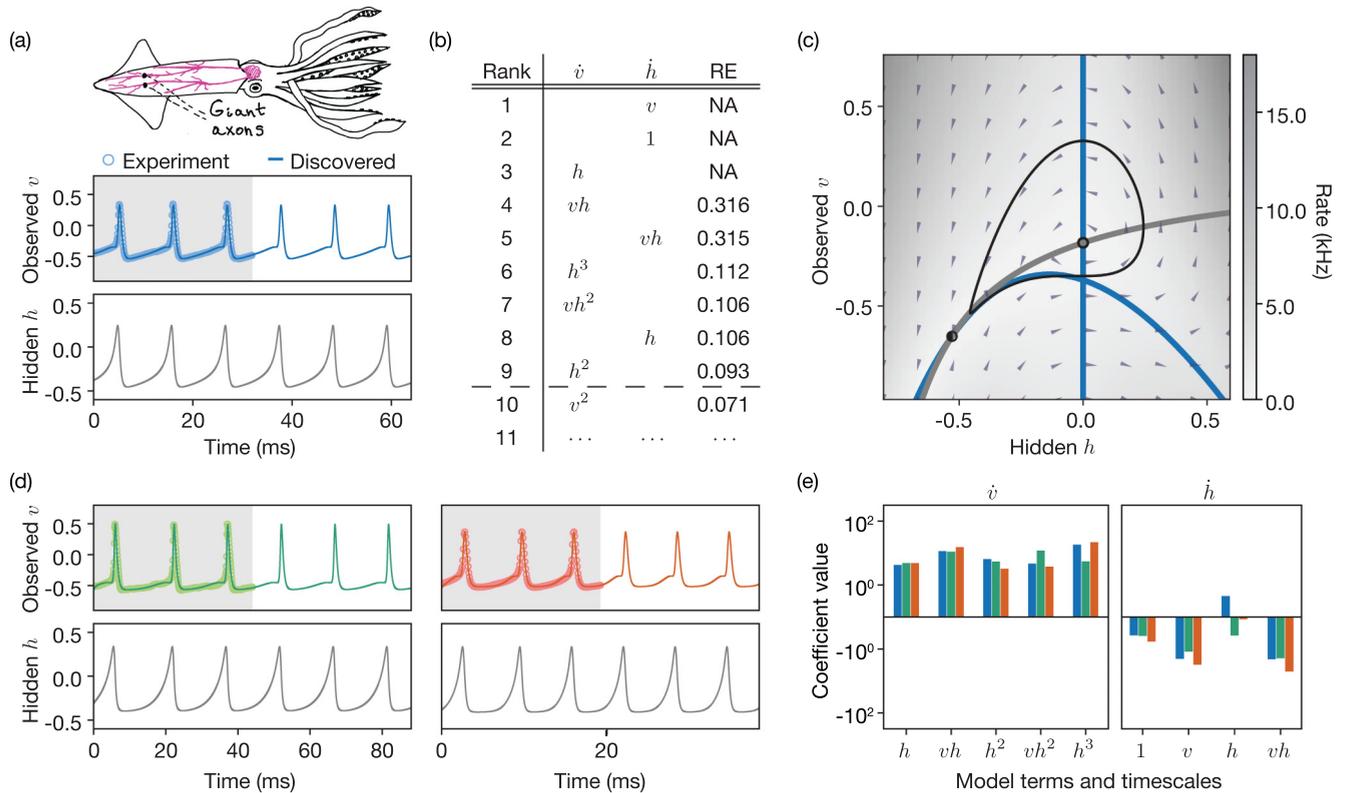
FIG. 4. HDI framework learns a parsimonious two-variable model from an experimental recording of the membrane potential in a squid giant axon and reproduces the dynamics in additional squid giant axons from the SGAMP database [104,105]. (a) North Atlantic longfin inshore squid (*Loligo pealeii*) with sketch of the nervous system and position of giant axons (top). Learned two-variable HDI model with nine terms accurately fits the membrane potential v (center, line) of an experimental squid giant axon (open circles) in response to a noisy stimulus input current. The hidden variable h (bottom) acts as a slow recovery variable. (b) Polynomial model terms in $\dot{v}$ and $\dot{h}$ equations ranked from most to least important based on their coefficient of variation in the largest model cluster. Training data losses of sparse models containing only top s ranked terms are shown, and the model with sparsity nine is chosen. (c) Limit cycle and fixed points (black) of learned model are consistent with prior models of regular spiking neurons [11] where the proximity of the saddle fixed point to the orbit likely arises from a homoclinic bifurcation. Nullclines of v, h plotted in blue and gray, respectively. (d) Selected nine-term model (line) generalizes to two additional squid axon recordings (open circles). (e) Coefficients of the nine-term model align across all three train and test squid axon experiments.

Solving for w, we get that

$$w = \frac{fu + \phi}{v + \mu},\qquad(27)$$

which leads to the two-component Oregonator model

$$\dot{u} = v - u,\qquad(28a)$$

$$\dot{v} = \frac{1}{\varepsilon_1}\left(v(1 - v) - \frac{v - \mu}{v + \mu}(fu + \phi)\right).\qquad(28b)$$

To build a two-variable polynomial HDI model that can match the dynamics of our experimental BZ recordings, we take inspiration from the two-component Oregonator model derived above. The first equation (28a) in $\dot{u}$ is already of polynomial form and is in fact linear. The second equation (28b) in $\dot{v}$ has a nonpolynomial rational term of the form $\frac{v - \mu}{v + \mu}$ which we must approximated through a polynomial expansion. A polynomial expansion is indeed possible because the coordinate trajectories of both Oregonator models (26) and (28) stay non-negative with $v > \mu$ if initialized in this way. The Taylor series expansion of this rational function contains monomials of all integer degrees, and hence we must decide where

to truncate its polynomial expansion. In the region $v > \mu$, since $\mu \ll 1$ the rational function $\frac{v - \mu}{v + \mu}$ plateaus quickly to 1, and hence is well-approximated by a cubic polynomial in v. Therefore, the entire right-hand side for the $\dot{v}$ equation (28b) can be approximated by a quartic polynomial in u, v.

The concentration of the oxidized catalyst u is most closely related to our observable, the color of the solution c, while the bromous acid concentration v is unobserved so we now denote it by the variable h. The two-component Oregonator model motivates us to search over all two-variable polynomial ODE models which are linear in their first equation and quartic in their second equation,

$$\dot{c} = w_1 + w_2 c + w_3 h, \quad \dot{h} = q_{\mathbf{w}}(c, h),\qquad(29)$$

where the first variable c predicts the color dynamics and the second variable h is hidden. Here $q_{\mathbf{w}}$ is notation for a quartic polynomial in c, h with learnable weights $\mathbf{w}$. From this class of models, HDI discovers a seven-term model that accurately fits the color dynamics $c(t)$ for BZ reactions [Figs. 5(a) and 5(d)] with parameters that vary smoothly across the different reactant concentrations in each experiment [Fig. 5(e)].
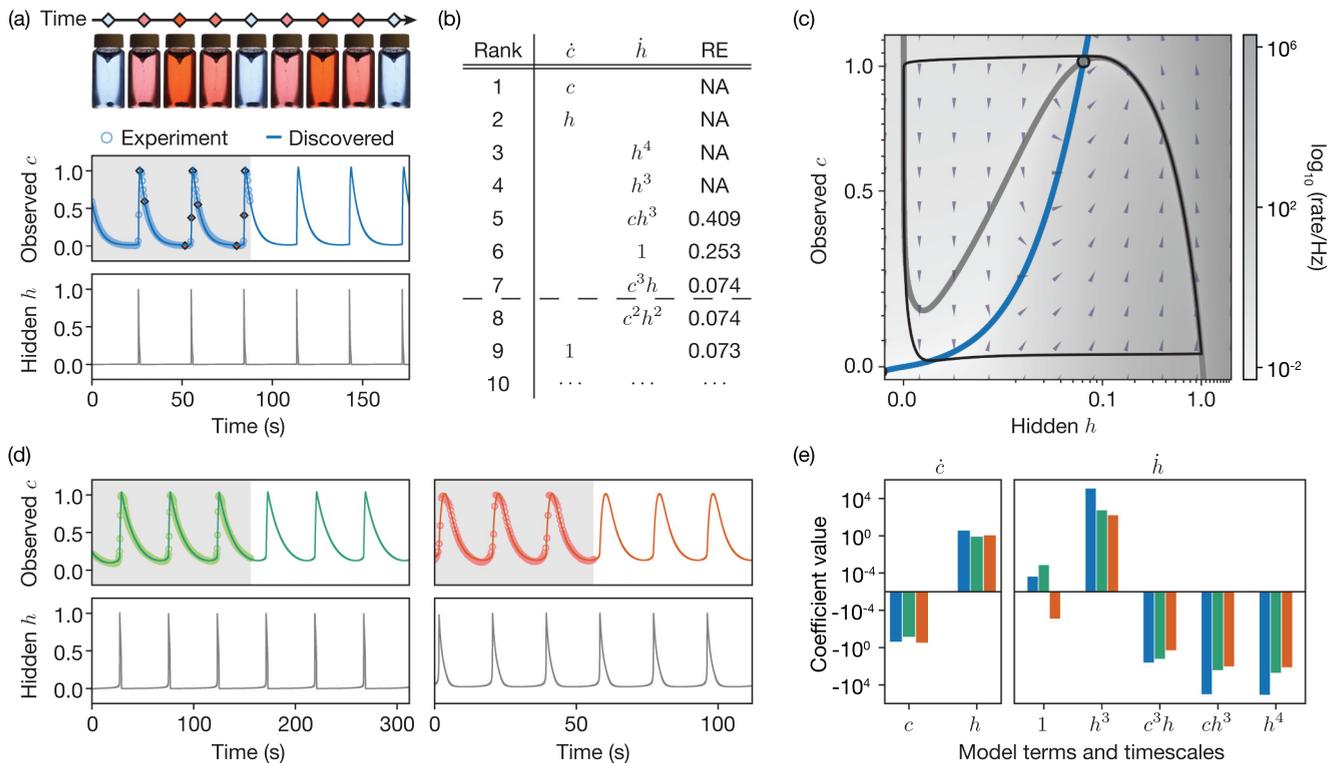
FIG. 5. HDI applied to our experimental BZ reaction data learns a two-variable linear-quartic model that generalizes under catalyst variations. (a) Experimental snapshots of the BZ reaction showing periodic color oscillations (top). Input data (open circles) and observed and hidden variables (solid line) integrated from the learned polynomial ODE model. Using data from three oscillations, the learning framework finds that a seven-term ODE can accurately describe the dynamics. (b) Polynomial ODE terms appearing in $\dot{c}$ and $\dot{h}$ equations ranked from most to least important based on their coefficient of variation. Model terms are added one-at-a-time in order of importance with the seventh term leading to a drop in the training loss. (c) Phase plane diagram of learned seven-term ODE from the previous panel contains crucial features found in most two-variable BZ models [106]. Limit cycle contains an unstable fixed point (black) with a monotonic x-nullcline (blue) and an h-nullcline (dark gray) in the form of a "cubic" curve as found in the FHN, Rovinsky, and ZBKE models. (d) Resulting seven-term model (solid line) accurately fits the dynamics of the color of the chemical solution (open circles) in two new BZ experiments. (e) Coefficients of the model remain consistent across all three experimental BZ reactions. Chemical concentrations: $0.20\,M\,H_2SO_4$, $0.11\,M\,NaBrO_3$, $0.05\,M\,CH_2(COOH)_2$, $0.03\,M\,NaBr$, $0.3\,mM$ ferroin (blue), $0.41\,M\,H_2SO_4$, $0.17\,M\,NaBrO_3$, $0.03\,M\,CH_2(COOH)_2$, $0.02\,M\,NaBr$, $0.3\,mM$ ferroin (green), $0.51\,M\,H_2SO_4$, $0.10\,M\,NaBrO_3$, $0.03\,M\,CH_2(COOH)_2$, $0.02\,M\,NaBr$, $0.3\,mM$ ferroin (red).

Furthermore, the phase portrait of the learned model correctly captures the dynamical properties of the BZ reaction [110], showing an unstable fixed point enclosed in a stable limit cycle with a typical cubic-shaped nullcline $\dot{h} = 0$ [Fig. 5(c)].

VII. CONCLUSION

By combining sensitivity methods and ranked choice voting, HDI can discover parsimonious predictive models from partial noisy observations of oscillatory and chaotic dynamics without extensive preprocessing of time-series data. The above framework can be directly applied to experimental observations of biophysical, ecological, and other systems, for which ODE models can inform the prediction, control, and optimal perturbations [116] of dynamical behavior. Extensions of HDI in the future to stochastic dynamics are possible due to recent advances in automatic differentiation of stochastic algorithms [117]. By mapping time series to ODE model coefficients, HDI can help facilitate the clustering of dynamical data, such as those appearing in health [118] and climate [8] studies.

The data that support the findings of this article are openly available. All source code for our HDI methodology and instructive examples are available [119]. Time-series data from the FHN oscillator and Lorenz attractor are generated by the authors' computer simulations with standard model and ODE solver parameters as described in the main text and the accompanying SM of this paper. Recordings from the giant axons of North Atlantic longfin inshore squids are taken from the publicly available SGAMP database [104,105]. Finally, the four BZ reactions studied in this paper were produced by coauthor J. Totz at MIT (reaction setup details in Sec. V of the SM [97]), and video recordings of the color change dynamics of these reactions are included along with this paper in the Supplemental Material, video file `SI_movie.mp4` [97].

[1] S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (CRC, Boca Raton, FL, 2018).

[2] K. M. Stiefel and G. B. Ermentrout, Neurons as oscillators, J. Neurophysiol. **116**, 2950 (2016).

[3] G. Buzsaki and A. Draguhn, Neuronal oscillations in cortical networks, Science **304**, 1926 (2004).

[4] N. Srinivas, J. Parkin, G. Seelig, E. Winfree, and D. Soloveichik, Enzyme-free nucleic acid dynamical systems, Science **358**, eaal2052 (2017).

[5] G. Filatrella, A. H. Nielsen, and N. F. Pedersen, Analysis of a power grid using a Kuramoto-like model, Eur. Phys. J. B **61**, 485 (2008).

[6] M. Rohden, A. Sorge, M. Timme, and D. Witthaut, Self-organized synchronization in decentralized power grids, Phys. Rev. Lett. **109**, 064101 (2012).

[7] G. Marty, S. Combrié, F. Raineri, and A. De Rossi, Photonic crystal optical parametric oscillator, Nat. Photon. **15**, 53 (2021).

[8] G. Vettoretti, P. Ditlevsen, M. Jochum, and S. O. Rasmussen, Atmospheric CO_2 control of spontaneous millennial-scale ice age climate oscillations, Nat. Geosci. **15**, 300 (2022).

[9] S. Xie and A. C. Martin, Intracellular signalling and intercellular coupling coordinate heterogeneous contractile events to facilitate tissue folding, Nat. Commun. **6**, 7161 (2015).

[10] A. Hilfinger, T. M. Norman, G. Vinnicombe, and J. Paulsson, Constraints on fluctuations in sparsely characterized biological systems, Phys. Rev. Lett. **116**, 058101 (2016).

[11] E. M. Izhikevich, *Dynamical Systems in Neuroscience* (MIT Press, Cambridge, MA, 2007).

[12] J. F. Totz, *Synchronization and Waves in Active Media* (Springer, Cham, Switzerland, 2019).

[13] C. Rackauckas, Y. Ma, J. Martensen, C. Warner, K. Zubov, R. Supekar, D. Skinner, A. Ramadhan, and A. Edelman, Universal differential equations for scientific machine learning, arXiv:2001.04385.

[14] J. G. Kemeny, Mathematics without numbers, Daedalus **88**, 577 (1959).

[15] P. Young, Optimal voting rules, J. Econ. Perspect. **9**, 51 (1995).

[16] T. Ling, K. C. Boyle, V. Zuckerman, T. Flores, C. Ramakrishnan, K. Deisseroth, and D. Palanker, High-speed interferometric imaging reveals dynamics of neuronal deformation during the action potential, Proc. Natl. Acad. Sci. U.S.A. **117**, 10278 (2020).

[17] A. A. Atanas, J. Kim, Z. Wang, E. Bueno, M. Becker, D. Kang, J. Park, C. Estrem, T. S. Kramer, S. Baskoylu, V. K. Mansingkha, and S. W. Flavell, Brain-wide representations of behavior spanning multiple timescales and states in c. elegans, bioRxiv (2022).

[18] H. Jeckel, E. Jelli, R. Hartmann, P. K. Singh, R. Mok, J. F. Totz, L. Vidakovic, B. Eckhardt, J. Dunkel, and K. Drescher, Learning the space-time phase diagram of bacterial swarm expansion, Proc. Natl. Acad. Sci. USA **116**, 1489 (2019).

[19] J. Alvelid, M. Damenti, C. Sgattoni, and I. Testa, Event-triggered sted imaging, Nat. Methods **19**, 1268 (2022).

[20] E. Frajka-Williams, I. J. Ansorge, J. Baehr, H. L. Bryden, M. P. Chidichimo, S. A. Cunningham, G. Danabasoglu, S. Dong, K. A. Donohue, S. Elipot *et al.*, Atlantic meridional overturning circulation: Observed transport and variability, Front. Mar. Sci. **6**, 260 (2019).

[21] L. A. Aguirre and C. Letellier, Modeling nonlinear dynamics and chaos: a review, Math. Probl. Eng. **2009**, 23896 (2009).

[22] 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. USA **113**, 3932 (2016).

[23] O. Talagrand and P. Courtier, Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, QJR. Meteorol. Soc. **113**, 1311 (1987).

[24] P. Courtier and O. Talagrand, Variational assimilation of meteorological observations with the adjoint vorticity equation. II: Numerical results, QJR. Meteorol. Soc. **113**, 1329 (1987).

[25] A. J. Majda, C. Franzke, and D. Crommelin, Normal forms for reduced stochastic climate models, Proc. Natl. Acad. Sci. USA **106**, 3649 (2009).

[26] M. Raissi and G. E. Karniadakis, Hidden physics models: Machine learning of nonlinear partial differential equations, J. Comput. Phys. **357**, 125 (2018).

[27] D. Kochkov, J. A. Smith, A. Alieva, Q. Wang, M. P. Brenner, and S. Hoyer, Machine learning-accelerated computational fluid dynamics, Proc. Natl. Acad. Sci. USA **118**, e2101784118 (2021).

[28] S. Yang, S. W. Wong, and S. Kou, Inference of dynamic systems from noisy and sparse data via manifold-constrained gaussian processes, Proc. Natl. Acad. Sci. USA **118**, e2020397118 (2021).

[29] R. Supekar, B. Song, A. Hastewell, G. P. T. Choi, A. Mietke, and J. Dunkel, Learning hydrodynamic equations for active matter from particle simulations and experiments, Proc. Natl. Acad. Sci. USA **120**, e2206994120 (2023).

[30] N. Romeo, A. Hastewell, A. Mietke, and J. Dunkel, Learning developmental mode dynamics from single-cell trajectories, eLife **10**, e68679 (2021).

[31] G. Box and G. Coutie, Application of digital computers in the exploration of functional relationships, Proc. IEE - Pt. B **103**, 100 (1956).

[32] J. Bakarji, K. Champion, J. N. Kutz, and S. L. Brunton, Discovering governing equations from partial measurements with deep delay autoencoders, arXiv:2201.05136.

[33] H. Ribera, S. Shirman, A. Nguyen, and N. Mangan, Model selection of chaotic systems from data with hidden variables using sparse data assimilation, Chaos **32**, 063101 (2022).

[34] M. Raissi, A. Yazdani, and G. E. Karniadakis, Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations, Science **367**, 1026 (2020).

[35] M. Heinonen, C. Yildiz, H. Mannerström, J. Intosalmi, and H. Lähdesmäki, Learning unknown ode models with gaussian processes, in *International Conference on Machine Learning* (PMLR, San Diego, California, 2018), pp. 1959–1968.

[36] W.-X. Wang, R. Yang, Y.-C. Lai, V. Kovanis, and C. Grebogi, Predicting catastrophes in nonlinear dynamical systems by compressive sensing, Phys. Rev. Lett. **106**, 154101 (2011).

[37] P. Y. Lu, J. Ariño Bernad, and M. Soljačić, Discovering sparse interpretable dynamics from partial observations, Commun. Phys. **5**, 206 (2022).

[38] S. Ouala, D. Nguyen, L. Drumetz, B. Chapron, A. Pascual, F. Collard, L. Gaultier, and R. Fablet, Learning latent dynamics for partially observed chaotic systems, Chaos **30**, 103121 (2020).

[39] I. Ayed, E. de Bézenac, A. Pajot, J. Brajard, and P. Gallinari, Learning dynamical systems from partial observations, arXiv:1902.11136.

[40] M. Cenedese, J. Axås, B. Bäuerlein, K. Avila, and G. Haller, Data-driven modeling and prediction of non-linearizable dynamics via spectral submanifolds, Nat. Commun. **13**, 872 (2022).

[41] F. Takens, Detecting strange attractors in turbulence, in *Dynamical Systems and Turbulence, Warwick, 1980* (Springer, Berlin, Heidelberg, Germany, 1981), pp. 366–381.

[42] G. Sugihara and R. M. May, Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, Nature (London) **344**, 734 (1990).

[43] G. Sugihara, R. May, H. Ye, C.-h. Hsieh, E. Deyle, M. Fogarty, and S. Munch, Detecting causality in complex ecosystems, Science **338**, 496 (2012).

[44] H. Haehne, J. Casadiego, J. Peinke, and M. Timme, Detecting hidden units and network size from perceptible dynamics, Phys. Rev. Lett. **122**, 158301 (2019).

[45] B. Chen, K. Huang, S. Raghupathi, I. Chandratreya, Q. Du, and H. Lipson, Discovering state variables hidden in experimental data, arXiv:2112.10755.

[46] H. Hewamalage, C. Bergmeir, and K. Bandara, Recurrent neural networks for time series forecasting: Current status and future directions, Int. J. Forecast. **37**, 388 (2021).

[47] J. D. Farmer and J. J. Sidorowich, Predicting chaotic time series, Phys. Rev. Lett. **59**, 845 (1987).

[48] A. C. Costa, T. Ahamed, and G. J. Stephens, Adaptive, locally linear models of complex dynamics, Proc. Natl. Acad. Sci. U.S.A. **116**, 1501 (2019).

[49] J. P. Crutchfield and B. McNamara, Equations of motion from a data series, Complex Syst. **1**, 417 (1987).

[50] M. V. Corrêa, L. A. Aguirre, and E. M. Mendes, Modeling chaotic dynamics with discrete nonlinear rational models, Int. J. Bifurcation Chaos **10**, 1019 (2000).

[51] 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 (2016).

[52] H.-L. Wei and S. A. Billings, Identification and reconstruction of chaotic systems using multiresolution wavelet decompositions, Int. J. Syst. Sci. **35**, 511 (2004).

[53] J. Suontausta, V. Ruoppila, and H. Koivo, Modelling of nonlinear systems using radial basis function networks, IFAC Proc. Vol. **27**, 627 (1994).

[54] A. Albano, A. Passamante, T. Hediger, and M. E. Farrell, Using neural nets to look for chaos, Physica D **58**, 1 (1992).

[55] R. Bakker, J. C. Schouten, C. L. Giles, F. Takens, and C. M. Van den Bleek, Learning chaotic attractors by neural networks, Neural Comput. **12**, 2355 (2000).

[56] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Geometry from a time series, Phys. Rev. Lett. **45**, 712 (1980).

[57] M. Casdagli, S. Eubank, J. D. Farmer, and J. Gibson, State space reconstruction in the presence of noise, Physica D **51**, 52 (1991).

[58] G. Gouesbet and C. Letellier, Global vector-field reconstruction by using a multivariate polynomial L2 approximation on nets, Phys. Rev. E **49**, 4955 (1994).

[59] H. Gottlieb, Question# 38. what is the simplest jerk function that gives chaos? Am. J. Phys. **64**, 525 (1996).

[60] J. Sprott, Simplest dissipative chaotic flow, Phys. Lett. A **228**, 271 (1997).

[61] S. J. Linz, Nonlinear dynamical models and jerky motion, Am. J. Phys. **65**, 523 (1997).

[62] R. Eichhorn, S. J. Linz, and P. Hänggi, Transformations of nonlinear dynamical systems to jerky motion and its application to minimal chaotic flows, Phys. Rev. E **58**, 7151 (1998).

[63] R. Eichhorn, S. J. Linz, and P. Hänggi, Classes of dynamical systems being equivalent to a jerky motion, Appl. Math. Mech. **79**, 287 (1999).

[64] R. Eichhorn, S. J. Linz, and P. Hänggi, Simple polynomial classes of chaotic jerky dynamics, Chaos Solitons Fractals **13**, 1 (2002).

[65] C. Letellier, L. A. Aguirre, and J. Maquet, Relation between observability and differential embeddings for nonlinear dynamics, Phys. Rev. E **71**, 066213 (2005).

[66] E. M. Mendes, C. Lainscsek, and C. Letellier, Diffeomorphical equivalence vs topological equivalence among sprott systems, Chaos **31**, 083126 (2021).

[67] K. Kaheman, J. N. Kutz, and S. L. Brunton, Sindy-pi: a robust algorithm for parallel implicit sparse identification of nonlinear dynamics, Proc. R. Soc. A. **476**, 20200279 (2020).

[68] P. Kunkel and V. Mehrmann, *Differential-Algebraic Equations: Analysis and Numerical Solution* (European Mathematical Society, Zürich, Switzerland, 2006), Vol. 2.

[69] F. Van Breugel, J. N. Kutz, and B. W. Brunton, Numerical differentiation of noisy data: A unifying multi-objective optimization framework, IEEE Access **8**, 196865 (2020).

[70] J. M. Varah, A spline least squares method for numerical parameter estimation in differential equations, SIAM J. Sci. Stat. Comput. **3**, 28 (1982).

[71] D. A. Messenger and D. M. Bortz, Weak sindy for partial differential equations, J. Comput. Phys. **443**, 110525 (2021).

[72] R. Brown, N. F. Rulkov, and E. R. Tracy, Modeling and synchronizing chaotic systems from experimental data, Phys. Lett. A **194**, 71 (1994).

[73] I. Dattner, H. Ship, and E. O. Voit, Separable nonlinear least-squares parameter estimation for complex dynamic systems, Complexity **2020**, 6403641 (2020).

[74] J. L. Breeden and A. Hübler, Reconstructing equations of motion from experimental data with unobserved variables, Phys. Rev. A **42**, 5817 (1990).

[75] E. Baake, M. Baake, H. G. Bock, and K. M. Briggs, Fitting ordinary differential equations to chaotic data, Phys. Rev. A **45**, 5524 (1992).

[76] J. M. Krueger, Parameter estimation methods for ordinary differential equation models with applications to microbiology, Ph.D. thesis, Virginia Tech, 2017.

[77] L. A. Aguirre, B. O. S. Teixeira, and L. A. B. Tôrres, Using data-driven discrete-time models and the unscented kalman filter to estimate unobserved variables of nonlinear systems, Phys. Rev. E **72**, 026226 (2005).

[78] M. Levine and A. Stuart, A framework for machine learning of model error in dynamical systems, Commun. Am. Math. Soc. **2**, 283 (2022).

[79] T. Schneider, A. M. Stuart, and J.-L. Wu, Ensemble kalman inversion for sparse learning of dynamical systems from time-averaged data, J. Comput. Phys. **470**, 111559 (2022).

[80] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980.

[81] D. C. Liu and J. Nocedal, On the limited memory bfgs method for large scale optimization, Math. Program. **45**, 503 (1989).

[82] Y. Cao, S. Li, L. Petzold, and R. Serban, Adjoint sensitivity analysis for differential-algebraic equations: The adjoint dae system and its numerical solution, SIAM J. Sci. Comput. **24**, 1076 (2003).

[83] A. F. Villaverde, A. Barreiro, and A. Papachristodoulou, Structural identifiability of dynamic systems biology models, PLoS Comput. Biol. **12**, e1005153 (2016).

[84] G. Massonis, A. F. Villaverde, and J. R. Banga, Distilling identifiable and interpretable dynamic models from biological data, PLoS computational biology **19**, e1011014 (2023).

[85] A. E. Cohen, A. D. Hastewell, S. Pradhan, S. W. Flavell, and J. Dunkel, Schrödinger dynamics and berry phase of undulatory locomotion, Phys. Rev. Lett. **130**, 258402 (2023).

[86] K. Law, A. Stuart, and K. Zygalakis, *Data Assimilation* (Springer, Cham, Switzerland, 2015), Vol. 214, p. 52.

[87] S. H. Rudy, J. N. Kutz, and S. L. Brunton, Deep learning of dynamics and signal-noise decomposition with time-stepping constraints, J. Comput. Phys. **396**, 483 (2019).

[88] J. Fullana, M. Rossi, and S. Zaleski, Parameter identification in noisy extended systems: a hydrodynamic case, Physica D **103**, 564 (1997).

[89] H. D. I. Abarbanel, R. Brown, and J. B. Kadtke, Prediction in chaotic nonlinear systems: Methods for time series with broadband fourier spectra, Phys. Rev. A **41**, 1782 (1990).

[90] U. Fasel, J. N. Kutz, B. W. Brunton, and S. L. Brunton, Ensemble-sindy: Robust sparse model discovery in the low-data, high-noise limit, with active learning and control, Proc. Math. Phys. Eng. Sci. **478**, 20210904 (2022).

[91] M. Small and C. K. Tse, Minimum description length neural networks for time series prediction, Phys. Rev. E **66**, 066701 (2002).

[92] X. Dong, Y.-L. Bai, Y. Lu, and M. Fan, An improved sparse identification of nonlinear dynamics with akaike information criterion and group sparsity, Nonlin. Dyn. **111**, 1485 (2023).

[93] Z. Wang, B. Wu, K. Garikipati, and X. Huan, A perspective on regression and bayesian approaches for system identification of pattern formation dynamics, Theor. Appl. Mech. Lett. **10**, 188 (2020).

[94] Z. Wang, X. Huan, and K. Garikipati, Variational system identification of the partial differential equations governing the physics of pattern-formation: Inference under varying fidelity and noise, Comput. Methods Appl. Mech. Eng. **356**, 44 (2019).

[95] M. Cranmer, Interpretable machine learning for science with pysr and symbolicregression. jl, arXiv:2305.01582.

[96] C. Cornelio, S. Dash, V. Austel, T. R. Josephson, J. Goncalves, K. L. Clarkson, N. Megiddo, B. El Khadir, and L. Horesh, Combining data and theory for derivable scientific discovery with ai-descartes, Nat. Commun. **14**, 1777 (2023).

[97] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.6.043062 for in-depth explanations of the algorithms presented in the main text along with further numerical tests and details of simulations and experimental data collection.

[98] C. Lainscsek, C. Letellier, and I. Gorodnitsky, Global modeling of the rössler system from the z-variable, Phys. Lett. A **314**, 409 (2003).

[99] A. Somacal, Y. Barrera, L. Boechi, M. Jonckheere, V. Lefieux, D. Picard, and E. Smucler, Uncovering differential equations from data with hidden variables, Phys. Rev. E **105**, 054209 (2022).

[100] R. FitzHugh, Impulses and physiological states in theoretical models of nerve membrane, Biophys. J. **1**, 445 (1961).

[101] J. Zhuang, T. Tang, Y. Ding, S. C. Tatikonda, N. Dvornek, X. Papademetris, and J. Duncan, Adabelief optimizer: Adapting stepsizes by the belief in observed gradients, Adv. Neural Inf. Proc. Syst. **33**, 18795 (2020).

[102] R. Fletcher, *Practical Methods of Optimization* (Wiley, Chichester, West Sussex, England, 2013).

[103] P. Breiding and S. Timme, HomotopyContinuation.jl: A Package for Homotopy Continuation in Julia, in *International Congress on Mathematical Software* (Springer, Cham, Switzerland, 2018), pp. 458–465.

[104] D. Paydarfar, D. B. Forger, and J. R. Clay, Noisy inputs and the induction of on–off switching behavior in a neuronal pacemaker, J. Neurophysiol. **96**, 3338 (2006).

[105] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals, Circulation **101**, e215 (2000).

[106] J. J. Tyson, Scaling and reducing the Field-Koros-Noyes mechanism of the Belousov-Zhabotinskii reaction, J. Phys. Chem. **86**, 3006 (1982).

[107] C. Morris and H. Lecar, Voltage oscillations in the barnacle giant muscle fiber, Biophys. J. **35**, 193 (1981).

[108] A. L. Hodgkin and A. F. Huxley, Currents carried by sodium and potassium ions through the membrane of the giant axon of loligo, J. Physiol. **116**, 449 (1952).

[109] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski, *Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition* (Cambridge University Press, Cambridge, England, 2014).

[110] I. R. Epstein and J. A. Pojman, *An Introduction to Nonlinear Chemical Dynamics: Oscillations, Waves, Patterns, and Chaos: Oscillations, Waves, Patterns, and Chaos* (Oxford University Press, New York, 1998).

[111] A. F. Taylor, Mechanism and Phenomenology of an Oscillating Chemical Reaction, Prog. React. Kinet. Mech. **27**, 247 (2002).

[112] R. J. Field, E. Koros, and R. M. Noyes, Oscillations in chemical systems. II. Thorough analysis of temporal oscillation in the bromate-cerium-malonic acid system, J. Am. Chem. Soc. **94**, 8649 (1972).

[113] A. B. Rovinskii and A. M. Zhabotinskii, Mechanism and mathematical model of the oscillating bromate-ferroin-bromomalonic acid reaction, J. Phys. Chem. **88**, 6081 (1984).

[114] A. M. Zhabotinsky, F. Buchholtz, A. B. Kiyatkin, and I. R. Epstein, Oscillations and waves in metal-ion-catalyzed bromate oscillating reactions in highly oxidized states, J. Phys. Chem. **97**, 7578 (1993).

[115] L. Ren, B. Fan, Q. Gao, Y. Zhao, H. Luo, Y. Xia, X. Lu, and I. R. Epstein, Experimental, numerical, and mechanistic analysis of the nonmonotonic relationship between oscillatory frequency and photointensity for the photosensitive Belousov–Zhabotinsky oscillator, Chaos **25**, 064607 (2015).

[116] G. Stepaniants, B. W. Brunton, and J. N. Kutz, Inferring causal networks of dynamical systems through transient dynamics and perturbation, Phys. Rev. E **102**, 042309 (2020).

[117] E. Krieken, J. Tomczak, and A. Ten Teije, Storchastic: A framework for general stochastic automatic differentiation, Adv. Neural Inf. Proc. Syst. **34**, 7574 (2021).

[118] K. Umetani, D. H. Singer, R. McCraty, and M. Atkinson, Twenty-four hour time domain heart rate variability and heart rate: Relations to age and gender over nine decades, J. Am. Coll. Cardiol. **31**, 593 (1998).

[119] https://github.com/adh18/HiddenDynamicsInference.

[120] A. Reuther, J. Kepner, C. Byun, S. Samsi, W. Arcand, D. Bestor, B. Bergeron, V. Gadepally, M. Houle, M. Hubbell, M. Jones, A. Klein, L. Milechin, J. Mullen, A. Prout, A. Rosa, C. Yee, and P. Michaleas, Interactive supercomputing on 40,000 cores for machine learning and data analysis, in *2018 IEEE High Performance extreme Computing Conference (HPEC)* (IEEE, Piscataway, NJ, 2018), pp. 1–6.