



# Benchmarking sparse system identification with low-dimensional chaos

Alan A. Kaptanoglu · Lanyue Zhang ·

Zachary G. Nicolaou · Urban Fasel ·

Steven L. Brunton

Received: 4 February 2023 / Accepted: 19 April 2023  
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

**Abstract** Sparse system identification is the data-driven process of obtaining parsimonious differential equations that describe the evolution of a dynamical system, balancing model complexity and accuracy. There has been rapid innovation in system identification across scientific domains, but there remains a gap in the literature for large-scale methodological comparisons that are evaluated on a variety of dynamical systems. In this work, we systematically benchmark sparse regression variants by utilizing the dysTs standardized database of chaotic systems introduced by Gilpin (in: Advances in neural information processing systems (NeurIPS), 2021. [arXiv:2110.05266](https://arxiv.org/abs/2110.05266)). In particular, we demonstrate how this open-source tool can be used to quantitatively compare different methods of system identification. To illustrate how this benchmark can be utilized, we perform a large comparison of four algorithms for solving the sparse identification of nonlinear dynamics (SINDy) optimization problem, finding strong performance of the original algorithm and

a recent mixed-integer discrete algorithm. In all cases, we used ensembling to improve the noise robustness of SINDy and provide statistical comparisons. In addition, we show very compelling evidence that the weak SINDy formulation provides significant improvements over the traditional method, even on clean data. Lastly, we investigate how Pareto-optimal models generated from SINDy algorithms depend on the properties of the equations, finding that the performance shows no significant dependence on a set of dynamical properties that quantify the amount of chaos, scale separation, degree of nonlinearity, and the syntactic complexity.

**Keywords** System identification · Dynamical systems · Chaos · Nonlinear systems · Sparse regression · SINDy

## 1 Introduction

The governing equations of a dynamical system have traditionally been derived from first principles or phenomenology and confirmed with experimentation. However, the availability of enormous volumes of data in the modern era is facilitating *data-driven* model discovery, which is now widely applied to many areas such as mathematics, physics, chemistry, biology, engineering, and economics. A large number of such approaches have been developed in recent years, such as the dynamic mode decomposition [2–4], Koopman theory [5, 6], nonlinear autoregressive algorithms [7], neu-

A. A. Kaptanoglu  
IREAP, University of Maryland, College Park, MD, USA

A. A. Kaptanoglu (✉) · S. L. Brunton  
Department of Mechanical Engineering, University of Washington, Seattle, WA, USA  
E-mail: akaptano@uw.edu

L. Zhang · Z. G. Nicolaou  
Department of Applied Mathematics, University of Washington, Seattle, WA, USA

U. Fasel  
Department of Aeronautics, Imperial College, London, UK

ral networks [8–10], Gaussian process regression [11], operator inference and reduced-order modeling [12–14], symbolic regression [15–17], divide-and-conquer strategies [18], and sparse regression [19]. Even within each of these approaches, there has been prolific methodological innovation.

There are several factors that must be considered when choosing a data-driven modeling strategy, including the quantity and quality of the data and the ultimate use of the model. A key trade-off for all of these methods is balancing accuracy with model complexity, since a naïve determination of many free, nonzero parameters will likely result in a model that is overfit to the training data, especially as real-world data will contain noise of various types. Despite the variety of new methods for data-driven model discovery, there is a relative lack of comparisons between such methods on a large set of nontrivial dynamical systems. It is this gap in the literature that we intend to address by providing a systematic comparison of sparse system identification variants on a recently developed large, curated database of known, chaotic, dynamical systems [1].

Sparse system identification is an umbrella term for system identification methods that address this key trade-off between model complexity and accuracy by promoting sparsity in the parameterization of the model. This facilitates a priori parsimonious descriptions that model the data with as few terms as necessary. The sparse identification of nonlinear dynamics (SINDy) [19] is a leading method for parsimonious modeling, and it and its variants will be investigated here. It is based on sparsity-regularized linear regression, so it can be computed quickly and robustly. Moreover, identified models and optimization results can be readily understood by researchers, as there is a wealth of literature on generalized linear regression and sparse optimization. Another compelling reason to investigate variations of SINDy is that it has seen rapid application and extension since its introduction in 2016. SINDy has been widely applied for model identification in applications such as chemical reaction dynamics [20], chemical processes [21,22], air pollutant modeling [23], biological transport [24], stellar dynamics [25], disease transmission [26], convective heat transfer [27], nonlinear optics [28], power systems [29,30], hydraulics [31], scientific computing [32], human behavior models [33], fluid dynamics [34–45], turbulence modeling [46–51], plasma physics [52–54], structural modeling [55], among oth-

ers [31,56–62]. It has also been extended to handle more complex modeling scenarios such as PDEs [63, 64], delay equations [65], dynamical bifurcations [66], stochastic differential equations [67–71], Bayesian modeling [72], dimensional analysis [73], systems with inputs or control [74–76], systems with implicit dynamics [77,78], hybrid systems [79,80], to enforce physical constraints [34,53,81], to incorporate information theory [82] or group sparsity [83] or global stability [84], to identify models from corrupt or limited data [85–89], to identify models with partial measurements of the state space [90–93], to identify models with clever subsampling strategies [94], and ensembles of initial conditions [95], to perform cross-validation with ensemble methods [96,97], and extending to related objective functions [98], a weak or integral formulation [99–109], tensor representations [110,111], and stochastic forcing [112].

Most of the work in this field demonstrates performance on just a few canonical dynamical systems, such as the Lorenz63 system [113]. Moreover, there has been little systematic investigation of the comparative performance of different sparse system identification techniques. We posit two plausible reasons for the relative lack of these investigations in the literature: (1) for more sophisticated variants of system identification, it can be computationally intensive to compute thousands or hundreds of thousands of dynamical models to draw statistical conclusions like those presented here, and (2) an analysis of the performance of system identification is more straightforward if the true governing equations are actually known. This is clearly not an option for real experimental data. With benchmarks, proper comparisons can be made between new and existing methods to understand the way in which a new algorithm produces novel benefits and functionality. Moreover, essentially all system identification methods require hyperparameter tuning and a proper comparison between methods necessitates that only the optimally tuned models from each method are compared on a large set of examples.

At first glance, there may seem to be too many confounding variables for a thorough comparison of methods. In principle, sparse system identification performance can depend on the equation complexity, the searchable space of functional forms, the size and diversity of the training data, the amount of computing power available for hyperparameter scans, the nonlinearity of the underlying dynamics, the amount and type of noise in the data, the quality of the algorithm used to solve

the problem (and its associated hyperparameters), and potentially many more factors.

To control for these parameters, we have carefully designed our experiments in a number of ways. We use a fixed functional space for fitting all systems, containing all possible terms necessary for successful model identification. Moreover, the dynamics and size of the training data are standardized by using all of the polynomial dynamical systems in the dysts database [1], and the amount of computing power is negligible by using sparse regression algorithms with minimal hyperparameters. For generating robust conclusions, the following steps have been taken: a training trajectory  $X_{\text{train}}$  is generated by integrating the true dynamics, and then  $d\%$  noise is sampled from a zero-mean Gaussian  $\mathcal{N}(0, \sigma = d\|X_{\text{train}}\|_2/100)$  and added to every data point in the training data. This noise is subsequently amplified when using the baseline SINDy method, since it requires computing  $\dot{X}_{\text{train}}$  from  $X_{\text{train}}$ , e.g., by naïve finite differences. Moreover, a large ensemble of SINDy models are generated by subsampling the same data [96], hyperparameter scans are performed, and finally, only statistical model performance metrics are reported. This work begins by detailing this standardization and methodology.

The entirety of the methods and results in the present work can be reproduced with the [PySINDy code](#) [114, 115]. Moreover, we have included a discussion in “Appendix A” about how to effectively use this tool in future work, as researchers design new algorithms, libraries, databases, or other functionality. The present work illustrates the scope and flexibility of this tool by performing a few large-scale investigations regarding SINDy optimization and performance.

First, we generate a statistical comparison of different optimization algorithms for solving the SINDy model identification problem across a large database. The sequentially thresholded least-squares (STLSQ) [19], sparse relaxed regularized regression (SR3) [116], Lasso [117], mixed-integer optimization sparse regression (MIOSR) [118], and weak form [119] algorithms are compared. To our knowledge, this is the first large-scale and statistical comparison of these algorithms. We find that the original STLSQ algorithm [19] holds up well for performing SINDy optimization and the recent MIOSR algorithm is high performance particularly for low-dimensional or constrained problems. Moreover, we find compelling evidence that the weak formulation of the SINDy problem provides significant

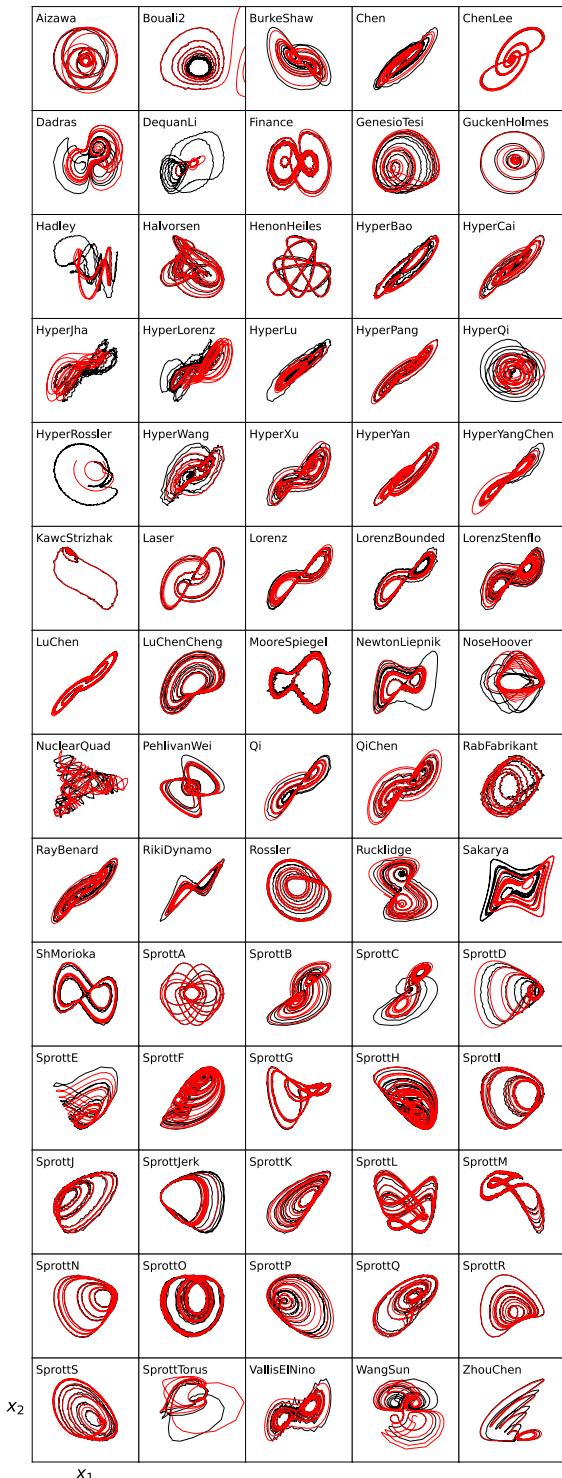
performance improvements, even in the clean data setting. Next, we classify the Pareto-optimal model performance against properties of the true underlying dynamics. We find that the SINDy method has no significant dependence on the amount of chaos, scale separation, degree of nonlinearity, and even the syntactical complexity of the equations, albeit with some interesting deviations that could merit additional future investigation.

## 2 Methodology

Robust conclusions about the performance of system identification methods require that only standardized comparisons are made. We now proceed by explaining the dataset, metrics for measuring sparse system identification performance, how we use hyperparameter scanning to find Pareto-optimal models for each system, and explicitly define the dynamical properties to be investigated. This section is intended to clarify how we have carefully managed confounding variables in the system identification in order to draw conclusions with respect to different optimization algorithms and with respect to the dynamical properties of the database.

### 2.1 Dysts database

The dysts database [1], introduced by Gilpin in 2021, is ideal for evaluating system identification performance, because it is a large and highly standardized set of chaotic systems with known equations of motion. Chaotic systems are both ubiquitous and challenging dynamical systems to analyze, and are subsequently a natural choice for benchmarking many system identification and forecasting-related methods. Many of the chaotic systems exhibit syntactically simple equations, e.g., Lorenz63 is a three-state system with quadratic nonlinearity and only seven terms in the differential equation. However, this database generates far more complicated and dynamical data than that of many existing databases, such as the Nguyen symbolic regression database [120] used previously for benchmarking symbolic regression algorithms [121]. The database also provides data, equations, and dynamical properties for over 100 chaotic systems exhibiting strange attractors and coming from disparate scientific



**Fig. 1** A training trajectory in the  $(x_1, x_2)$  subspace with 1% added noise (black), along with a clean testing trajectory (red) for the systems in the present work, taken from [1]. Some of the dynamical systems have multiple attractors, and the data may not be sufficient to fully “cover” the attractor. (Color figure online)

fields. It can be used for testing sparse system identification because the dynamical properties have been pre-computed and there is a high degree of standardization across the systems. For instance, phase surrogate significance testing is used to select optimal integration time steps and sampling rates for all the systems, so that dynamics across systems are aligned with respect to their smallest and largest timescales [122]. This is a critical step for generating similar time series for each model; the minimum significant timescale in this dataset varies by more than four orders of magnitude. Most importantly, the true governing equations are available for evaluating the model performance.

To narrow the scope of our comparison, we consider 70 systems in the database, representing systems of ordinary differential equations (ODEs) that have polynomial nonlinearities, with degree no more than four. We exclusively consider these systems and define a single “feature library” (space of functions that can be used to fit the data) for every model fit. Polynomial models are often the natural choice for capturing leading-order dynamics for general nonlinear systems near an attracting set or fixed point, where the dynamics might be reasonably expanded in a Taylor series.

Unlike some other dynamical systems, chaotic systems exhibit a continuous spectrum of frequencies, making the dynamical data particularly challenging to systematically distinguish from noise, e.g., by using frequency filters. Our chosen systems exhibit a specific subclass of chaos, as they are all bounded and are characterized by strange attractor(s) with the default system parameters used in the database. Moreover, 54 of the systems have a state space dimension of three, the remaining 16 have a dimension of four, and there is a single Hamiltonian system in the data. 51 of the systems are “hyperchaotic” systems that have been estimated to exhibit two or more positive Lyapunov exponents. A training and testing trajectory of 1000 data points are each visualized in Fig. 1 so that the strange attractors are apparent.

To generate statistical conclusions, we generate five training trajectories and five testing trajectories representing ten periods of motion, all sampled at 100 points per period for a total of 1000 data points for each trajectory. Each trajectory is generated from a different initial condition on the attractor.

To summarize, the underlying dynamics are chaotic, and we assume that the dynamical equations can be represented as systems of ODEs with polynomial non-

linearities up to fourth order. Additionally, full-state measurements are available that, at various points in the coming analysis, contain added zero-mean Gaussian noise.

## 2.2 The SINDy method

SINDy utilizes sparse regression to discover governing equations from data [19]. The aim is to identify differential equation models of the form

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) \approx \Theta(\mathbf{x}) \mathbf{\Xi}, \quad (1)$$

where  $\mathbf{f}(\mathbf{x})$  is a nonlinear function of the state  $\mathbf{x}$ ,  $\Theta$  is a library of nonlinear functions, and  $\mathbf{\Xi}$  is a matrix of constant coefficients. Together  $\Theta \mathbf{\Xi}$  represents a linear combination of nonlinear functions that are chosen to best approximate  $\mathbf{f}(\mathbf{x})$ . The choice of nonlinear functions in  $\Theta$  is clearly important for successful approximation. Before moving from continuous state space to discrete measurement data, recall that high-order differential equations can always be reduced to first-order coupled differential equations, so this form is fairly general. Extensions of Eq. (1) for control inputs [74], partial differential equations [64], implicit terms [78], and other variations are available with PySINDy [123].

In order to identify Eq. (1) from data, we need measurements of the state  $\mathbf{x}$ . First, we uniformly sample the full-state data on a set of training trajectories and a data matrix  $\mathbf{X} \in \mathbb{R}^{M \times d}$  is formed from the time series data of the state,  $\mathbf{x}(t_1), \mathbf{x}(t_2), \dots, \mathbf{x}(t_M)$ :

$$\mathbf{X} = \begin{array}{c} \xrightarrow{\text{state}} \\ \left[ \begin{array}{cccc} x_1(t_1) & x_2(t_1) & \cdots & x_d(t_1) \\ x_1(t_2) & x_2(t_2) & \cdots & x_d(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(t_M) & x_2(t_M) & \cdots & x_d(t_M) \end{array} \right] \end{array} \xrightarrow{\text{time}}. \quad (2)$$

$M$  is the number of time points and  $d$  is the size of the state space, i.e., three or four for our dataset. A matrix of derivatives in time,  $\dot{\mathbf{X}}$ , is defined similarly and can be numerically computed from  $\mathbf{X}$ , for instance by finite differences or more sophisticated differentiation [88, 124, 125]. Then we form a feature library  $\Theta(\mathbf{X})$  of possible terms that could describe the evolution of the system, and attempt to find a sparse vector of coefficients  $\xi$  (representing the flattened matrix of coefficients  $\xi = \mathbf{\Xi}[:, :]$ ) representing the best linear

combination of terms in the library. In general, the feature library  $\Theta$  can include a wide range of generalized functions, and designing this library often requires domain knowledge; however, the systems of interest in the present work are all polynomial. Therefore, we fix  $\Theta$  to be a polynomial library in  $\mathbf{X}$ , up to fourth order. This is an important point; in general system identification methods rely on adequate functional forms in the feature library, and here we have completely bypassed this issue, isolating and disambiguating the optimization performance from the choice of library. We know definitively that the feature library contains the necessary terms in order to describe all 70 chaotic systems considered in the present work.

The vector of coefficients  $\xi$  is determined via the following sparse optimization problem:

$$\arg \min_{\xi} \left[ \frac{1}{2} \|\Theta \xi - \dot{\mathbf{X}}\|^2 + \lambda \|\xi\|_0 \right]. \quad (3)$$

The first term in the SINDy optimization problem in Eq. (3) is a least-squares fit of a system of ODEs  $\Theta \xi$  to the given data in  $\dot{\mathbf{X}}$ . The  $l_0$  loss,  $\|\xi\|_0$ , is a function that counts the number of nonzero elements of  $\xi$ . Using the  $l_0$  loss or alternative regularization to promote sparsity in the coefficients of an identified model tends to improve robustness to numerical or experimental noise, addressing the trade-off between accuracy and generalizability and avoiding overfitting to the training data. The  $l_0$  loss is nonconvex and nonsmooth, and therefore many convex relaxations of this problem have been introduced [19, 64, 81]. The first major result presented in this work compares a number of different algorithms for solving Eq. (3), including the STLSQ algorithm from the original SINDy paper [19], the traditional convex relaxation of Eq. (3) using the Lasso [117], the SR3 algorithm [81, 116], and a recent mixed-integer discrete algorithm called MIOSR [118].

As much as is feasible, we use default PySINDy parameters to reduce the number of hyperparameters that must be scanned. For instance, in the present work  $\dot{\mathbf{X}}$  is calculated by simple finite differences. Improved performance is immediately available for future work using more sophisticated differentiation methods [88, 124, 125], many of which are already available in the PySINDy code. Weak or integral formulations of SINDy [99] are also known to mitigate the noise amplification inherent to finite differencing, and we illustrate in Sect. 3.2 that the weak formulation

implemented in PySINDy produces substantial performance benefits.

### 2.3 Metrics for performance

Methodological comparisons require appropriate metrics for determining the relative performance quality of different methods. Unlike real-world data, we have access to the true equations of motion of all 70 systems to be tested. Therefore we can compute a number of useful metrics to characterize the performance of the resulting sparse symbolic models on these examples. We define the normalized coefficient and root-mean-square errors

$$E_{\text{coef}} = \frac{\|\boldsymbol{\xi}_{\text{True}} - \boldsymbol{\xi}_{\text{SINDy}}\|_2}{\|\boldsymbol{\xi}_{\text{True}}\|_2}, \quad (4)$$

$$E_{\text{RMSE}} = \frac{\|\dot{\boldsymbol{X}}_{\text{True}} - \dot{\boldsymbol{X}}_{\text{SINDy}}\|_{\text{F}}}{\|\dot{\boldsymbol{X}}_{\text{True}}\|_{\text{F}}}, \quad (5)$$

where  $\|\dot{\boldsymbol{X}}_{\text{True}}\|_{\text{F}}$  denotes the Frobenius norm. Since these are normalized metrics, errors below a few percent can be considered strong model performance. Other works have also explored the  $\|\cdot\|_1$  and  $\|\cdot\|_0$  norm versions of Eq. (4) to quantify model mismatch [78, 121]. Obtaining very small values for  $E_{\text{RMSE}}$  (or variants such as the mean absolute error and regular mean absolute percent error) is simple with a basic system identification algorithm applied to clean data from our chosen database. This is because models with many parameters can overfit using many small but nonzero terms and potentially produce very accurate testing trajectories if we only investigate regions that are nearby to the strange attractor(s). Despite this caveat, for forecasting, prediction, reconstruction of the strange attractor, calculation of the Lyapunov spectrum, and other important metrics for capturing the behavior of a dynamical system, models with low  $E_{\text{RMSE}}$  are often more than adequate. Moreover, the usual situation of interest is one in which the dynamical equations and therefore the coefficient errors are unknown. We comment further on errors and stability in “Appendix B.”

Given that we know the true coefficients, we can check for overfitting by considering the coefficient errors, which can be incorrect even when  $E_{\text{RMSE}}$  is extremely small\*. Indeed, on the strange attractor, the dynamics may be significantly reduced compared to the dynamics on the full state space. A connection to

invariant manifolds and the reduced effective dimension and entropy of the dynamical system is discussed in “Appendix C.”<sup>1</sup>

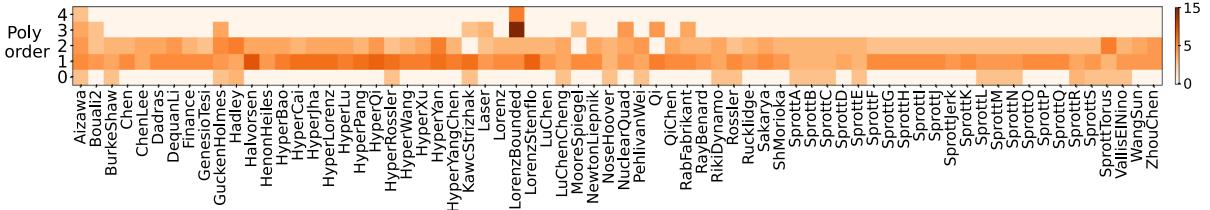
### 2.4 Automated hyperparameter scanning with ensembles

Before any conclusions can be drawn about system identification performance, each algorithm’s hyperparameters must be tuned. Sweeping the sparsity-related hyperparameters generates a spectrum ranging from very sparse models with limited accuracy to dense models with high prediction accuracy, which is typically referred to as a Pareto front [126]. The dataset is too large to manually tune hyperparameters for optimal system identification performance on each chaotic trajectory, so for hyperparameter scanning, we use the following procedure. For all the algorithms, we use ensembling [96]. We subsample the five training trajectories by choosing 50% of the total trajectory data without replacement in order to generate 10 models for each value of the  $\lambda$  hyperparameter. For STLSQ, we then sweep over 300 values of increasingly large values of  $\lambda$  to generate increasingly sparse models, and determine the best model in this Pareto front by finding the  $\lambda$  that produces the minimal, finite-sample-size-corrected, Akaike information criterion (AIC) [82] averaged over the 10 models generated by subsampling,

$$\begin{aligned} \text{AIC} = & M \log(\|\dot{\boldsymbol{X}}_{\text{True}} - \dot{\boldsymbol{X}}_{\text{SINDy}}\|_2^2) + 2k \\ & + \frac{2k(k+1)}{(M-k-1)}, \end{aligned} \quad (6)$$

where  $k = \|\boldsymbol{\xi}_{\text{SINDy}}\|_0$  and  $\dot{\boldsymbol{X}}_{\text{True}}$  and  $\dot{\boldsymbol{X}}_{\text{SINDy}}$  are both computed from the test trajectories. The AIC is a well-known statistical metric for comparing models while balancing sparsity and accuracy [82]. This systematic process requires the identification of  $300 \times 10 \times 70 = 210,000$  models. Despite the large number of models to generate, the entire procedure for STLSQ takes

<sup>1</sup> Another common metric is the recovery rate [121], which returns a discrete 0 or 1 value for each correctly identified term in the equations. However, small errors may result from insufficient hyperparameter tuning, noise in the data, insufficient optimization convergence, and many other sources. Moreover, it is often useful to penalize large coefficient errors more than small ones. Thus, we omit the recovery rate metric in this work but mention it for future investigation.



**Fig. 2** Total number of polynomial terms in each equation, for each polynomial degree. It can be seen that quadratic nonlinearities are dominantly represented

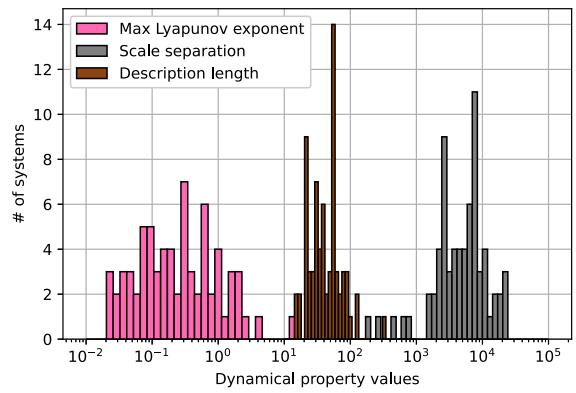
about 10 min to run on single Intel Core i9-7920X 2.90GHz CPU. After the best  $\lambda$  is determined via this Pareto sweep, the corresponding ten models allow us to characterize a distribution of identified coefficients and errors. The other SINDy algorithms are tuned similarly for their respective hyperparameters. The SR3 algorithm has an extra parameter  $v$  that we set to either 1 or 0.1 since these values tend to work well on clean data and it avoids additional hyperparameter scanning.

The use of SINDy, rather than more sophisticated system identification methods, facilitates very fast calculation of many models. For instance, previous work [127] investigating mostly genetic programming algorithms for symbolic regression required roughly 200 core-hours per algorithm per dynamical system in order to generate a best model from the Pareto front. In contrast, using STLSQ we require roughly  $\sim 10^{-3}$  core-hours per dynamical system to generate our Pareto-optimal model, consistent with findings elsewhere [1, 128]. This speed is required because the systematic process we have outlined requires generating  $\mathcal{O}(10^5)$  SINDy models per algorithm.

## 2.5 Dynamical properties

Systems of differential equations can often be effectively classified by various dynamical properties. Now that the dynamical systems used in the present work have been introduced, we define a set of dynamical properties for testing sparse system identification performance: the degree of chaos, the degree of scale separation, the syntactic complexity of the underlying equations, and the amount of nonlinearity. In the present work, the degree of chaos is measured by the largest Lyapunov exponent. The degree of scale separation is taken as the system's dominant timescale divided by the system's smallest significant timescale as determined

by the method in Gilpin [1]. The syntactic complexity is quantified as the description length, representing the number of bits to describe every object in the equations. Lastly, the amount of nonlinearity is taken as sum of the total number of terms in the equations, weighted by each term's polynomial degree, so it is a mix between the degree of nonlinearity and the number of equation terms. For more precise definitions and discussion of the metric choices, see “Appendix C.” A summary of the nonlinearities appearing in the dynamical equations is shown in Fig. 2, indicating that most of the systems exhibit strong quadratic nonlinearity. Quadratic nonlinearity is common in reduced-order models of fluids and many other dynamical systems, e.g., the Lorenz or Rossler systems, because of the quadratic nonlinearity coming from the convective term of the material derivative. Ten systems exhibit up to cubic nonlinearity and only two systems have quartic nonlinearity. Figure 3 summarizes the distribution of the remaining dynamical properties.



**Fig. 3** Distributions of the different properties of the dynamical systems used in this work. The maximum Lyapunov exponent is a measure of the degree of chaos and the description length is a measure of the syntactic complexity of the equations. Note the log scale on the x-axis; all of these distributions vary by at least an order of magnitude

cal properties, indicating a wide range of values taken by the dataset.

### 3 Results

We now demonstrate the utility of combining dysts with PySINDy by illustrating a series of benchmark experiments to better understand performance trade-offs of sparse system identification variants. Now that we have explained the characteristics of the database, the sparse system identification algorithm, the error metrics, and the dynamical properties of interest, we illustrate the results of our investigations with the standardized, Pareto-optimal SINDy models. We emphasize that the [entirety of these results](#) are reproduced as examples in the PySINDy code and the following investigations use Pareto-optimal identified models that have been chosen by producing the minimal AIC in a hyper-parameter scan.

#### 3.1 Comparison of SINDy algorithms

As an illustration of this methodology and database, we compare Pareto-optimal models for a class of SINDy algorithms: STLSQ, Lasso, SR3 (fixing the  $\nu$  hyperparameter to 1 or 0.1), and MIOSR. The features of these algorithms are briefly described in Table 1. Additionally, we use the STLSQ optimizer with the weak form implementation in the PySINDy code in order to compare with the traditional method. Although not shown in this work, the weak formulation can also be used with any of the other PySINDy optimizers with minimal modification. Weak form runs were performed with the other algorithms and similar trends were seen between algorithms, so we have omitted these results. The weak form uses 200 subdomains, which becomes 200 points to use in the optimization, rather than the 1000 trajectory points used in the traditional optimization. This many subdomains is probably excessive; it was found empirically that, after 100 subdomains, there is essentially no further improvement (in fact just 10 subdomains is much faster to compute and still performs quite well). The MIOSR optimizer is set to use no more than 5 s to solve and prove optimality per state variable. The limit of five seconds is rarely invoked with clean data but more often invoked once the data has 1% added noise.

The results were generated by the process outlined earlier in Sect. 2.4. Figure 4a illustrates the Pareto-optimal model performances for each of the SINDy algorithms. Results are shown for both clean data and data with 1% added noise. These results illustrate the known insufficiency of the Lasso for generating high-performance system identification results, although the Lasso degrades less than some of the other optimizers as noise is added. On the clean data, the remainder of the optimizers regularly produce models with  $E_{\text{coef}} < 1\%$  and  $E_{\text{RMSE}} < 10\%$ .

For additional evidence that these are accurate dynamical models, Fig. 5 illustrates the Pareto-optimal STLSQ model simulations of ten trajectories from new initial conditions starting slightly off each of the strange attractors. An increasingly gray-tinted background indicates an increasing number of simulated trajectories that went unstable in the ODE solver. Overall, the major conclusion is that, without noise, Pareto-optimal SINDy models regularly produce new trajectories that recreate the strange attractor of each chaotic system, meaning that the dynamical features such as the Lyapunov spectrum can be accurately computed.

#### 3.2 Weak formulation results

The weak or integral form of SINDy [99, 100, 102] subdivides the time span of the trajectories and considers integrals of the ODEs over these subdomains. This method then avoids the calculation of derivatives on noisy data and subsequent noise amplification by integrating by parts when appropriate.

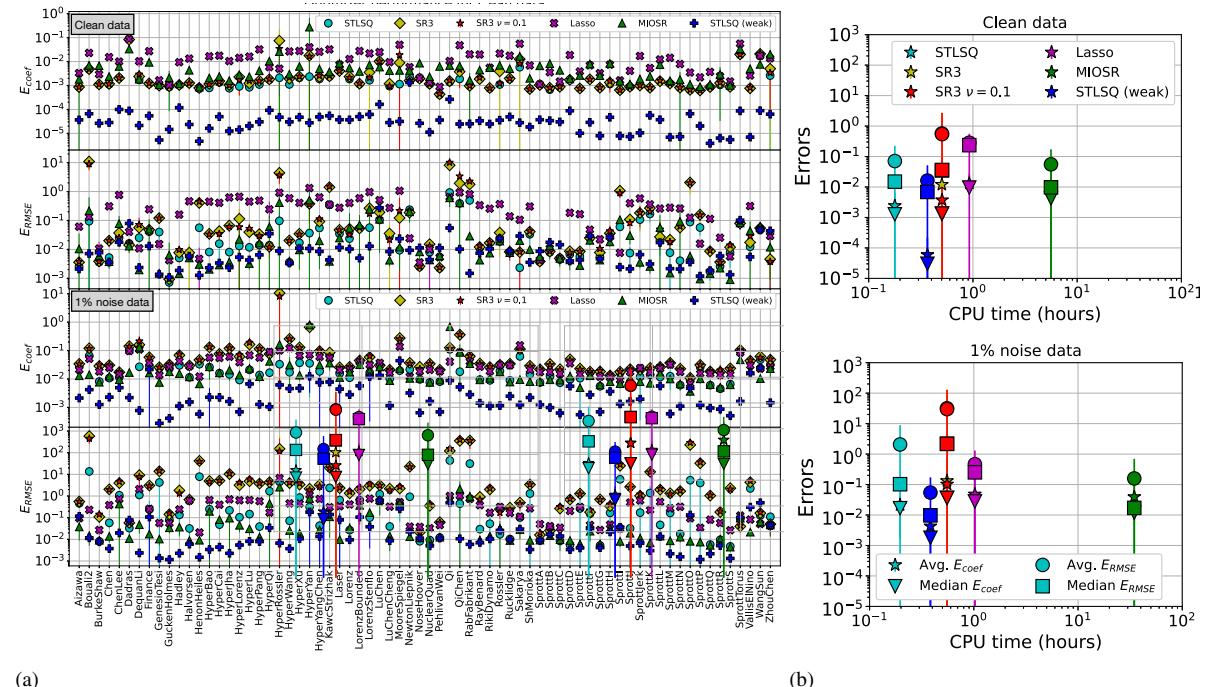
The results in Fig. 4a indicate that the weak formulation performs extraordinarily well in producing the correct equation coefficients. The weak form  $E_{\text{RMSE}}$  results appear slightly less impressive for a rather technical reason. We are computing the pointwise  $E_{\text{RMSE}}$  defined in Eq. (5) in order to compare against the other optimizers. Unlike the other optimizers,  $E_{\text{RMSE}}$  (of the training data) is not directly minimized in the weak formulation of the problem; the weak form solves its optimization problem by minimizing an error metric that is integrated over the subdomains. Despite not directly minimizing these quantities, the weak form still generally obtains the smallest values of  $E_{\text{RMSE}}$ .

These results provide compelling evidence for the considerable strength of the weak formulation. Interestingly, the typical justification for the high perfor-

**Table 1** Summary of the PySINDy optimizers tested in this work, including sequentially thresholded least-squares (STLSQ), Lasso, sparse relaxed regularized regression (SR3), and a mixed-integer optimization sparse regression (MIOSR)

Optimizers	Constraints	Regularization	Params	Type
Lasso	×	$l_1$	$\lambda$	Exact
STLSQ	×	$l_0$	$\lambda$	Greedy
SR3	✓	$l_0$ or $l_1$	$\lambda, \nu$	Exact
MIOSR	✓	$l_0$	$N_{\text{nonzero}}$	Exact

Here  $N_{\text{nonzero}}$  refers to a parameter that specifies the number of nonzero terms that the identified model should exhibit. Lasso and MIOSR can typically find global optima while STLSQ and SR3 can typically find only local optima



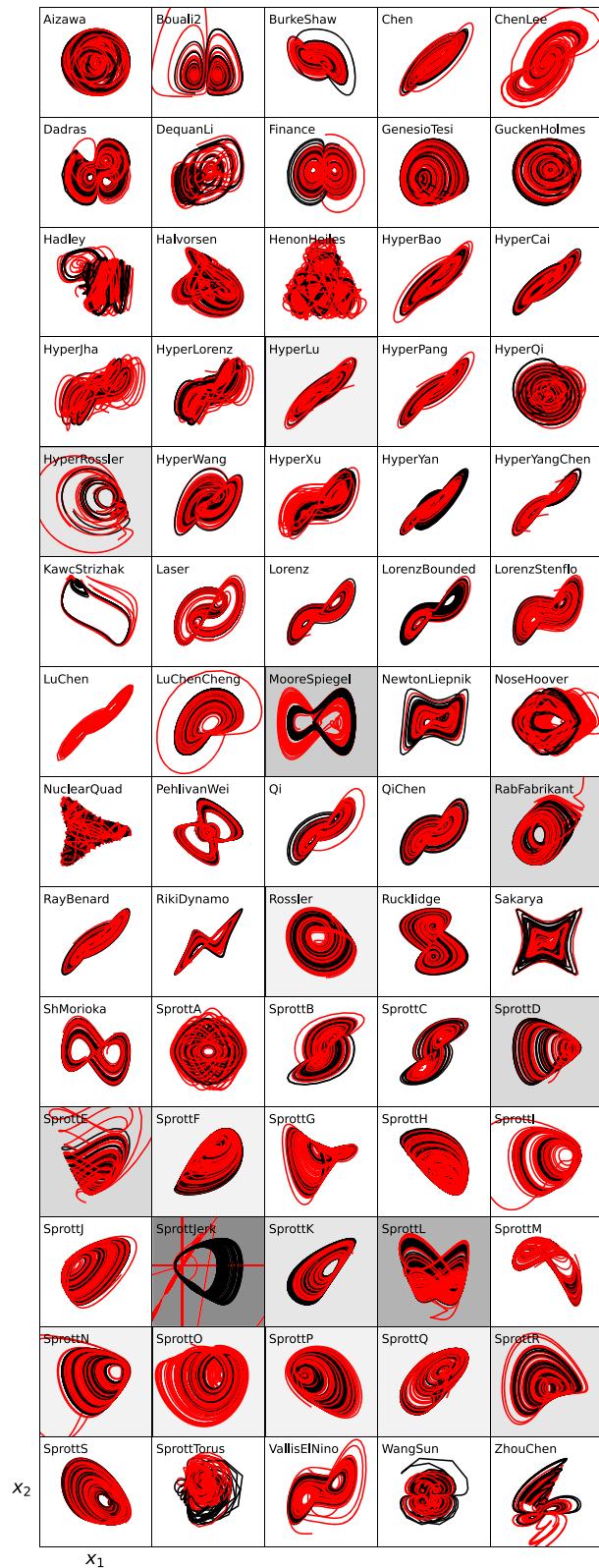
**Fig. 4** **a** Summary of the average errors for each optimization algorithm, without noise (top two rows) and with 1% noise (bottom two rows), on all 70 chaotic systems. STLSQ and MIOSR generally produce the best models of the traditional optimizers. Lasso has weak performance, and SR3 seems to require additional hyperparameter scanning to avoid substantial errors when noise is present. The weak formulation with STLSQ identifies

the coefficients extremely well and demonstrates the substantial advantages of using the weak form. **b** Algorithm average and median performance aggregated over all the dynamical systems. Marker colors indicate different optimizers, and the marker shapes indicate the type of error. All runs were performed serially with an Intel Core i9-7920X 2.90GHz CPU

mance of weak formulations of sparse regression is that it allows the user to avoid amplifying noise when computing derivatives of the data. However, we find significant improvements with the weak formulation even when there is no added noise—only the intrinsic and unavoidable “noise” of finite sampling rates. Even further, the weak formulation results are not truly Pareto-optimal; the threshold value for STLSQ is still

scanned but the additional hyperparameters available in the weak form are just fixed to a set of plausible values. Combined, we consider these results persuasive for using the weak formulation whenever possible. For flexible usage, the weak formulation is entirely integrated into the feature library functionality of the PySINDy code, and therefore can be used with any of

**Fig. 5** Pareto-optimal SINDy STLSQ model simulations (red) vs true trajectories (black) in the  $(x_1, x_2)$  subspace using ten new initial conditions. Grayer backgrounds indicate more unstable trajectories, which are otherwise not pictured. (Color figure online)



the current or future optimizers, or with any generic library terms of interest.

### 3.3 Greedy vs exact algorithms

MIOSR is an exact mixed-integer algorithm that can often solve the  $l_0$ -regularized (i.e., nonconvex) SINDy problem to optimality. In contrast, the STLSQ algorithm solves this problem greedily, meaning it chooses the best model on a subset of the coefficients at each iteration, and it cannot recover from thresholding mistakes that occur in earlier iterations.

On the noisy data, the MIOSR algorithm seems to outperform the others, while with clean data all of the nonweak algorithms perform similarly well except for Lasso. With clean data, the SR3 algorithms perform very well. Some variation is seen with SR3 between the  $\nu = 1$  and  $\nu = 0.1$  cases but a clear trend is not distinguishable. The SR3 errors become significantly larger as noise is added and these results seem to indicate that hyperparameter scans over  $\nu$  are required for high performance when noise is present. All of the optimizer performances degrade somewhat when noise is added. These overall takeaways are summarized further in Fig. 4b.

At 1% noise, STLSQ and MIOSR show performance advantages over the other optimizers. As measured by the average coefficient errors, MIOSR leads the group. STLSQ at 1% noise exhibits a small number of very large RMSE errors and this tends to reduce the average performance. This is consistent with the explanation that greedy methods can increasingly make catastrophic mistakes as the noise levels increase. This trend can also be seen to an extent in the weak formulation results since the optimization is still being performed greedily with STLSQ.

One takeaway is that the MIOSR algorithm performs well, as we might expect for an exact mixed-integer optimization. However, Fig. 4b shows that, at larger levels of noise, the MIOSR algorithm requires much more computing time than the other algorithms. This is true despite putting a time limit on the computation, but it is reassuring that strong performance is seen even if the optimizer does not finish finding a solution to optimality. Indeed, as is pointed out in detail in Bertsimas et al. [118], MIOSR runs slower when the regression is harder, e.g., in the noisy setting, but the computational efficiency scales very well with the amount of data and

additionally sees substantial speedups with parallelization or better CPUs than used in this work. However, our results appear to show MIOSR does not significantly dominate the other algorithms in our chosen performance metrics. This difference with the results in Bertsimas et al. [118] may be because their performance metric focus is primarily on the true positive rate of the model coefficients, and their MIOSR results are most impressive in the low-data limit.

In fact, these results also highlight the strong performance of the STLSQ algorithm; STLSQ appears relatively robust to noise, very fast to compute (regardless of the noise level), and, for the most part, generates models of comparable performance to MIOSR. Even at 1% noise, the errors that accumulate from the greedy nature of STLSQ appear tolerable. These optimizers were not tested at larger values of noise. It is well known that weak or integral formulations of the optimization problem are required for retaining high performance in the high-noise setting [96, 99, 101]. This fact is clearly reinforced by the impressive weak form performance in Fig. 4a.

### 3.4 Dynamical properties

A pressing question for practitioners of system identification is whether certain types of dynamical systems are more difficult to identify than others. For instance, some methods are specifically designed to search for Hamiltonian systems [129–131]. Another example is that traditional recurrent neural networks should be modified for modeling chaotic systems in order to avoid unbounded gradients [132].

We now present an investigation into the importance of the dynamical properties with respect to the performance of the various SINDy optimizers. The required data and computations for this analysis have already been generated in the previous section. For each chaotic system in our dataset, the errors of the Pareto-optimal models are sorted by the dynamical property in question, and some simple data fits are attempted: linear, log-linear, and log-log. Linear fits between  $\log(E_{\text{RMSE}})$  and the scale separation seemed to produce the largest  $R^2$  coefficient of determination for all the optimizers, and similarly for log-log fits between  $\log(E_{\text{RMSE}})$  and the logarithm of the remaining dynamical properties. We make no claims that these simple fits to the data best capture the correlations, but rather they are used

to roughly quantify if there are correlations between increasing dynamical property values and worsening system identification performance.

Visual inspection of the different dynamical property fits for the weak form STLSQ optimizer at 0.1% noise in Fig. 6 does seem to indicate that, on average, the optimizer performance rapidly drops as the scale separation increases. Similarly, visual inspection of the different dynamical property fits for the MIOSR optimizer at 1% noise in Fig. 6 indicates that log-log fits to the data weakly capture the optimizer performance for all of the dynamical properties except the scale separation, which shows no clear trend.

All of the best  $R^2$  coefficients of determination are illustrated in Fig. 7 for each optimizer, each dynamical property, and for noise levels of 0, 0.1, and 1%. Most of the  $R^2$  values are below 0.25, indicating that the dominant behavior is that the optimizers produce results that are approximately independent of the dynamical properties. Nonetheless, there appear to be some intriguing trends; for instance, the weak formulation  $E_{RMSE}$  shows substantial correlation with the scale separation. MIOSR seems to increasingly correlate with the dynamical properties (except that of scale separation) at larger values of noise.

These results suggest that there are some relatively persistent but quite weak correlations with the underlying dynamical properties of the equations. Moreover, which correlations dominate appears to depend on the level of noise in the data and the choice of optimization algorithm.

Our large experiments are a strong affirmation of the “Task 4” result in Gilpin indicating that SINDy model results were independent of the level of chaos [1]. Our metric for the degree of chaos in the system is a rather coarse measure, but similar results were found in both Gilpin and the present work when alternative metrics for chaos were used.

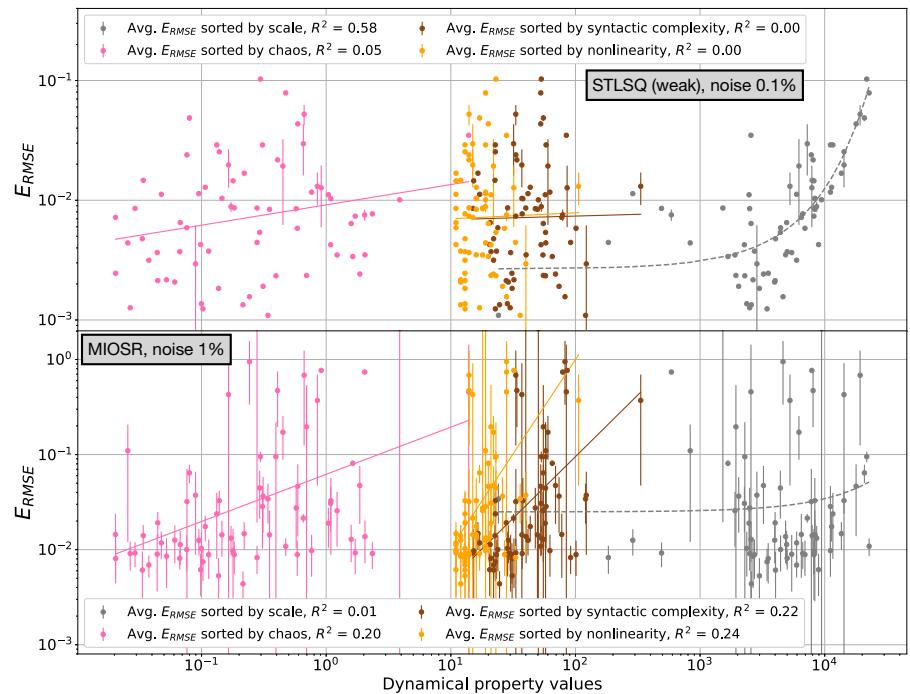
There is some substantial evidence that deep neural networks [133] for multi-step forecasting, recurrent neural networks [132], and related methods see degraded forecasting performance at higher levels of chaos. One possible explanation is that forecasting methods based on deep neural networks typically use the multi-time-step forecast of the data as an optimization target, while this work was limited to the traditional single-step prediction with the SINDy method. In indirect support of this thesis, deep symbolic regression [121] also uses only the single-time-step predic-

tion, and it was found that the Spearman correlation with chaos is quite small and not consistent across different metrics for the level of chaos [1]. We have noted elsewhere that optimizing the single-time-step is advantageous from a perspective of speed; the SINDy optimization can be solved much faster (and therefore we can quickly compute 1 million models) than more complex methods. There are some recent variations of the SINDy method that do use the multi-step prediction as a loss function [89, 91]. Future work could resolve if the multi-time-step objective is the source of the chaos dependence by investigating if these methods exhibit stronger correlation with the degree of chaos. Another partial explanation comes from an additional difference between the SINDy method and the existing literature. The SINDy optimization problem is solving for a set of coefficients that parametrize an explicit set of equations for the dynamics from a predefined library. In contrast, the literature showing dependence on chaos typically is solving for the weights in a high-dimensional mapping between the inputs and outputs – these parameter estimation problems are somewhat different. Lastly, another partial explanation for this difference with the literature may come from the phase surrogate significance testing used in the dysts database to scale the dynamical systems to similar time series. The dominant significant frequency is calculated and the associated time step used as the sampling rate. This timescale may be somewhat related to the largest Lyapunov exponent, and this rescaling might account for some of the independence with respect to chaos.

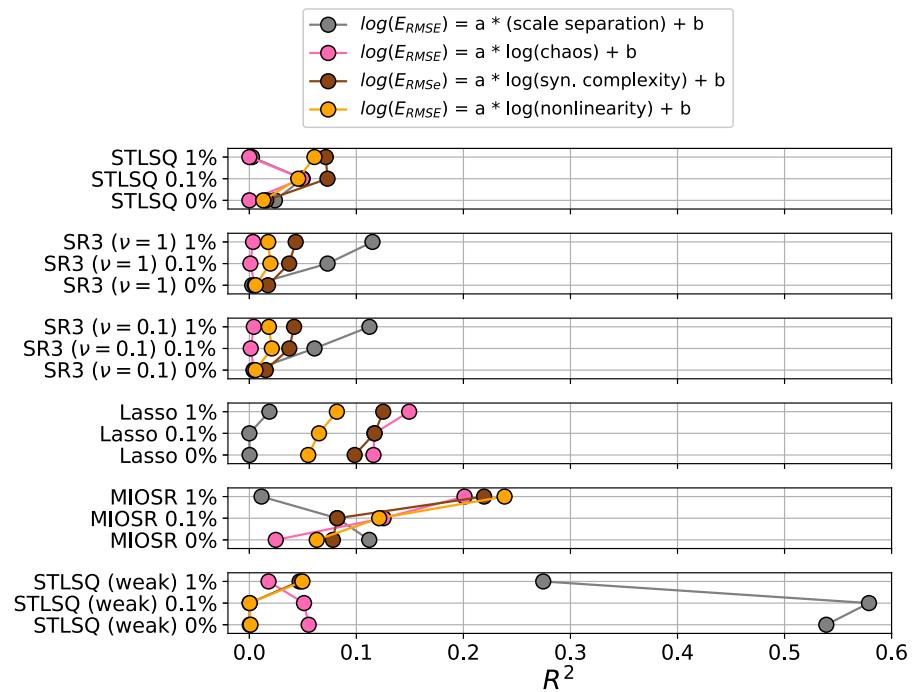
The lack of correlation with the description length is an interesting and potentially counter-intuitive discovery. To be more specific, if the dynamical terms live in the subspace of the feature library, the quality of Pareto-optimal models generated by sparse regression onto data seems to be approximately uncorrelated with the length or complexity of the underlying dynamical equations. This could be seen as a positive result. If adequate functional terms are available to describe the underlying dynamics, and the data is high-quality, the complexity of the equation factors out. This motivates large and expansive libraries, although this will generally increase the condition number of the feature library and therefore require additional regularization and shift the Pareto front.

At 1% noise levels, the quality of the Pareto-optimal models generally drops; for instance, a typical result with STLSQ produces only a few models with less than

**Fig. 6** Two illustrations of the  $E_{RMSE}$  results for the Pareto-optimal STLSQ optimizer (weak form) on data with 0.1% noise and for the MIOSR optimizer at 1% noise. The STLSQ result suggests that  $E_{RMSE}$  grows quickly with the scale separation. The MIOSR algorithm produces weak log-log trends with the other three dynamical properties tested in this work, although these trends strengthen as noise increases



**Fig. 7** Summary of the weak  $R^2$  coefficients of determination found between each optimizer's Pareto-optimal  $E_{RMSE}$  and the dynamical properties, choosing the best fit from linear, log-linear, and log-log regressions. Some differing trends can be observed for each optimizer. Even weaker correlations are found between  $E_{coef}$  and the dynamical properties (not pictured)



10%  $E_{\text{RMSE}}$  and eight of the models are not able to produce any new trajectories (starting from slightly off the attractor) that either stay bounded or remotely match the qualitative behavior of the strange attractor.

## 4 Conclusions

The methodology and analysis in the present work has shown that sparse system identification can be used to identify the dynamics of 70 polynomial chaotic systems, reproducing the strange attractors and therefore reproducing the fundamental dynamics. We extensively used subsampling to generate large ensembles of models, utilized the AIC to identify ensembles at the optimal hyperparameter values, and relied on the computational speed of sparse linear regression to compute over one million dynamical models in just two days worth of CPU time (mostly expended during the MIOSR algorithm).

This methodology was used to benchmark a variety of different SINDy optimizers, finding that: STLSQ & MIOSR produce the best performance, Lasso is relatively robust to noise but otherwise lower performance, and large-scale fitting with SR3 requires additional hyperparameter scanning for high performance. As significant noise is added to the data, STLSQ retains its fast computational speed but starts making “greedy” mistakes that cannot be recovered from, while conversely, MIOSR slows down considerably but can retain its strong performance for most systems. Lastly, we find very persuasive evidence that the weak formulation provides significant performance improvements across the database, even in the zero-added-noise setting. The weak form can be used with any of the PySINDy optimizers and we recommend using the weak formulation for almost all cases. Similarly, subsampling the data to create an ensemble of SINDy models [96] is highly recommended, especially in the presence of noise.

We have also presented one of the first large-scale and quantitative analyses into how sparse regression depends on the dynamical properties of the underlying governing equations. Overall, we found very weak correlations between the dynamical properties and the performance of the Pareto-optimal SINDy models, with some noteworthy deviations, providing a foundation for more sophisticated investigations in future work.

There are many opportunities for future work that builds on the database and analysis presented here.

Examples include further improvements to recent and important work addressing partial measurements [91, 134], the low-data high-noise regime [96, 118], extended feature libraries (which can already be built in PySINDy), and algorithm comparisons with convex constraints. For instance, most of the optimizers, STLSQ, SR3, Lasso, and MIOSR, can in principle accommodate general convex constraints. However, future PySINDy work is required to implement these options. Currently, constraints with STLSQ [34] and Lasso are not supported, while SR3 constraints in PySINDy (except linear equalities) rely on a CVXPY [135] backend which significantly slows down the algorithm. See, e.g., Kaptanoglu et al. [136] for a much faster SR3 algorithm with quadratic inequality constraints.

There are also numerous other dynamical metrics, sparse system identification algorithms, subsampling methods, types of dynamical systems (i.e., not just those defined by polynomial dynamics), and other parameters to systematically test. Although it is beyond the scope of this work, the dysts database facilitates another very interesting line of inquiry—building data-driven models for predicting chaotic bifurcations in global dynamical behavior, e.g., between exhibiting a strange attractor and a stable periodic solution. Additional future work could perform a quantitative analysis of how invariant manifolds affect the performance and usefulness of system identification; this topic is discussed in “Appendix C.” Lastly, many dynamical or engineering-focused metrics, such as the description length of the equations, can be directly minimized during the optimization and this may change the qualitative conclusions found in this work with the baseline SINDy optimization problem, especially since the present work found that the correlations are weak between model performance and the metrics tested.

**Acknowledgements** AAK would like to acknowledge useful conversations with William Gilpin, Chris Hansen, and Nathan Kutz. ZGN is a Washington Research Foundation Postdoctoral Fellow.

**Author contributions** All authors contributed to the conception and design of the present study. Data generation and analysis were performed by Lanyue Zhang and Alan Kaptanoglu. Weak formulation PySINDy runs were optimized by Zachary Nicolaou. The first draft of the manuscript was written by Alan Kaptanoglu and all authors helped with manuscript revisions. Urban Fasel and Steven Brunton provided critical high-level guidance. All authors read and approved the final manuscript.

**Funding** The authors acknowledge support from the National Science Foundation AI Institute in Dynamic Systems under grant number 2112085.

**Data availability** All of the code and data used in the current study are available in the open-source [PySINDy](#) code. The database and new datasets can be generated using the open-source [dysts](#) repository.

## Declarations

**Conflict of interest** The authors have no relevant financial or nonfinancial interests to disclose.

## Appendix A: How to use this benchmark

The goal of the present work is to provide a systematic approach to benchmark new and existing innovations in the field of sparse system identification. In order to do so, we have endeavored to provide simple, intuitive scripts, built on the open-source PySINDy code. A script for performing the type of Pareto-optimal scans in this work can be found in one of the PySINDy code [examples](#). It contains options for the user to specify (1) the data (including the number of points, sampling rates, etc. of the training and testing trajectories), (2) the optimization algorithm, and (3) the feature library (including whether to normalize the feature library, how to subsample the data for making model ensembles, and the number of models to generate for making statistical conclusions). For the [dysts](#) database, it shows additionally how to aggregate all the equations and their associated dynamical properties. The entirety of the results in this work can be reproduced by running the `run_all.py` script in the same directory.

The PySINDy code is regularly updated with new methods from the literature and now contains many SINDy variations and advanced functionality. New algorithms, feature libraries, subsampling strategies, and more can be readily added into the PySINDy code and then immediately used with the scripts mentioned above. Much of the backend for forming general SINDy libraries and new optimizers is already pre-made in the code and should significantly reduce the overhead designing and testing new algorithms.

## Appendix B: Small model errors and associated instability

True sparse regression is a nonconvex problem because of the  $l_0$  loss term in the objective function. The  $l_1$  loss term is often used in place of the  $l_0$  since, under certain conditions, it will produce the same solution with high probability [137]. When these conditions are not met, the Lasso is known to weakly choose irrelevant features [138]. Furthermore, even the optimally solved  $l_0$ -regularized problem can fail to predict exactly the correct features when the data is sufficiently corrupted by noise. In the context of system identification, small model errors can generate significant issues.

For instance, very small errors in the model identification can cause finite-time and unphysical instability when generating new trajectories from the identified models (although this is not an issue if the user merely seeks to find an approximation of the equations rather than to *use* them for generating new trajectories). In the case of fourth-order polynomials, we can use the following heuristic. We would expect that small coefficient errors on the third- and fourth-order terms will almost inevitably produce instability on *some* trajectory since these terms will be negligible near the origin (or near the attractor) but an initial condition can always be chosen far enough from the origin in order to make high-order terms dominant. Only in the constrained [34] and stability-promoted [84] settings should we expect that most (or all) trajectories generated from the identified models are stable. Finite-time blow up of the solution obviously prevents a reasonable comparison of the RMSE error of  $X$  (rather than  $\dot{X}$ ) and additionally prevents calculation of the Lyapunov spectrum and other dynamical properties of interest.

There are two half-remedies: only test the model with trajectories with initial conditions reasonably close to the strange attractor(s) and avoid rescaling the SINDy matrix. The first remedy comes from the intuition that small coefficient errors on high-order terms will only become dominant in regions of the state space sufficiently far from the origin. The latter remedy is also intuitive if we consider the STLSQ algorithm. It has been previously noted, e.g., in Delahunt et al. [87], that if  $x(t) \sim 10$  on the strange attractor, the two terms  $x(t)$  and  $0.1x^2(t)$  will contribute similarly to the time evolution, yet the latter term will be the first term to be truncated by an algorithm that relies on some version of hard-thresholding the smallest coefficients in the equa-

tions. Rather than seeing this as a problem of the algorithm, we can view this instead as an algorithmic bias towards describing each system with low-order polynomial terms, which we expect should contribute indirectly to boundedness. Unfortunately, the opposite is true if  $x(t) \sim 0.1$ , so it can be useful to rescale the different dynamical systems so that this thresholding mismatch is all in the same direction for the entire database. Although not explored in this work, minimizing an objective function related to the  $N_t$ -step SINDy model prediction as in, e.g., Kaheman et al. [89] or Bakarji et al. [91], rather than the traditional single time-step prediction, should generally improve the model stability but at the expense of nonconvexity (it is no longer sparse linear regression).

Lastly, even when the identified models are essentially perfect and reproduce the strange attractor(s), reporting the  $X$  RMSE error for chaotic systems can often be profoundly misleading, since inevitably there will be (initially exponentially large and then bounded) errors in the trajectory RMSEs from tiny errors in the integration, identified coefficients, and so on. We found it much more convincing to simply show visually that, with new initial conditions, the identified models can approximately reproduce the correct attractors, and therefore accurate calculations of the Lyapunov spectrum and other dynamical properties are possible.

## Appendix C: Dynamical property definitions

In this section we provide precise definitions for the dynamical properties computed and analyzed in the present work.

1. *The amount of chaos* can be measured in numerous ways [1], and a common choice is the largest Lyapunov exponent. The largest Lyapunov exponent measures the rate at which trajectories from two initially infinitesimally close points exponentially separate as time evolves [139]. This exponent can have very large variation over a trajectory, so it is usually considered as a global average.
2. *The degree of scale separation* could be defined through a ratio of the largest timescale to smallest timescale in the underlying equations. However, these timescales are not typically well defined even for polynomially nonlinear systems, so we could instead use the ratio of the largest to smallest Lyapunov exponents. As before, the Lyapunov

spectrum can often vary tremendously over a trajectory and therefore this is quite a coarse measure of scale separation. For a much improved metric to measure scale separation, we use the dominant timescale divided by the smallest significant timescale as defined in Gilpin [1]. To see why this measure of scale separation might be dynamically relevant to symbolic regression, consider the following simplified model

$$\begin{aligned}\dot{x}_1 &= f(x_1, x_2), \\ \dot{x}_2 &= \frac{1}{\epsilon}g(x_1, x_2),\end{aligned}\quad (\text{C1})$$

for  $\epsilon \ll 1$  and assuming the existence of a strange attractor, as in the case for our dataset. Following the standard derivation for invariant manifolds [140], on long time intervals (compared to the  $\epsilon$  timescale)  $x_2(t)$  quickly finds an equilibrium parametrized as  $x_2(t) \sim \eta(x_1(t))$ , and is functionally controlled by the  $x_1(t)$  evolution. The conclusion is the existence of an invariant manifold defined by

$$\begin{aligned}\dot{x}_1 &= f(x_1, \eta(x_1)), \\ \dot{x}_2 &= \partial_{x_1} \eta(x_1) \dot{x}_1 = \partial_{x_1} \eta(x_1) f(x_1, \eta(x_1)),\end{aligned}\quad (\text{C2})$$

which is a manifold in the  $(x_1, x_2)$  state space. If few data points are available for the period when  $t \sim \epsilon$  (or, equivalently, if the initial value  $x_2(0)$  starts close to the equilibrium or strange attractor), we cannot reasonably expect that Eqs. (C1) and (C2) can be distinguished. In other words, we might expect that for the best symbolic regression results (i.e., results that correctly capture the dynamical terms necessary to converge to the strange attractor from initial conditions starting substantially off the attractor) we need to use initial conditions far from the equilibrium, and resolved at the smallest timescale. This intuition is given some additional practical and theoretical justification in Bucci et al. [141]. It is shown explicitly that trajectories from linearly unstable fixed points contain less entropy, i.e., less information, than trajectories starting on the attractors, since near the fixed points only the linear terms are active and information about the nonlinearities cannot be gleaned. Similarly, on the attractor there is a balance of terms that reduces the

information available about the equation terms, relative to some generic trajectory in the phase space. Some chaotic systems can have effective dimensions on the attractor that are significantly smaller than the state space dimension.

There is another important note to make here. We know that Eqs. (C1) are discoverable with the pre-defined, fourth-order polynomial feature library used for symbolic regression. This is not guaranteed for the equations defining the invariant manifold, and thus we may still find the correct equations simply because the more specific invariant manifold equations are not accessible in the regression. Proceeding down this line of analysis is complex and further complicated by the fact that often polynomial terms are very close to linearly dependent and therefore multiple sparse models may fit the data well [87].

In order to sample trajectories while they are off the attractor, i.e., when  $t \sim \epsilon$ , we need to sample each period such that  $\sim \epsilon^{-1}$  points are generated, greatly increasing the required sampling resolution. Alternatively, a more sophisticated weighting strategy may be employed. There has been some recent symbolic regression work [142–145] connecting system identification and invariant manifolds. An interesting future investigation could attempt to identify the map  $\eta(x_1(t))$  by searching for Eq. (C2) with the constraint that library terms for  $\dot{x}_2$  must be proportional to  $\dot{x}_1$ . These types of constraints are already implemented in PySINDy.

3. *The syntactic complexity* of the model has been considered in earlier work as useful for both direct optimization and as a post-fit metric [146]. There are many possible definitions of syntactic complexity, but we adopt the approximate description-length metric [147] since the description length has been used in a number of contexts for direct optimization and has a natural interpretation as approximating the number of bits to describe each object in the equations.
4. *The amount of nonlinearity* is also not typically well defined. The highest degree of nonlinearity can be defined as the largest polynomial appearing in the equations governing each system, and previous work has used this definition for a complexity measure in sparse regression [148]. However, some systems have many quadratic nonlinear terms, and we might consider such systems very nonlinear despite

the lack of higher-order polynomials. As a basic metric to trade-off between these considerations, we consider a weight vector  $w_i = [1, 2, 3, 4, 5]$  and for each system simply report

$$N_{\text{nonlinearity}}$$

$$= \sum_{j=1}^d \sum_{i=1}^5 w_i \|x^{i-1} \text{ terms in equation } j\|_0. \quad (\text{C3})$$

In other words, we sum all of the terms present in the coupled set of ODEs, weighted by the polynomial degree of each term. This includes the linear and constant terms, so it is a mix between the degree of nonlinearity and the number of equation terms. Because this definition takes into account the number of terms, there is presumably some overlap with the information captured by the description length metric.

We make no claims that any of these metrics are the most compelling choice for capturing the particular dynamical information, but each metric seems to capture an important aspect. A substantial discussion on some of these points can be found in Murdoch et al. [149]. Future work could investigate more sophisticated metrics such as the stiffness or integrability of the identified ODE systems, or investigating the corresponding stability features of the models. For generic nonlinear ODE systems, investigating stability seems intractable, but there is a significant volume of literature for finding Lyapunov functions for polynomially nonlinear systems [150, 151]. Optimizing stiffness [152] and stability would be additionally useful for finding models that can be used for applying chaos control [153].

## References

1. Gilpin, W.: Chaos as an interpretable benchmark for forecasting and data-driven modelling. In: Advances in Neural Information Processing Systems (NeurIPS). [arXiv:2110.05266](https://arxiv.org/abs/2110.05266) (2021)
2. Schmid, P.J.: Dynamic mode decomposition of numerical and experimental data. J. Fluid Mech. **656**, 5 (2010)
3. Rowley, C.W., Mezić, I., Bagheri, S., Schlatter, P., Henningson, D.: Spectral analysis of nonlinear flows. J. Fluid Mech. **641**, 115 (2009)

4. Kutz, J.N., Brunton, S.L., Brunton, B.W., Proctor, J.L.: Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems. SIAM, New Delhi (2016)
5. Mezić, I.: Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dyn.* **41**, 309 (2005)
6. Brunton, S.L., Budišić, M., Kaiser, E., Kutz, J.N.: Modern Koopman theory for dynamical systems. *SIAM Rev.* **64**, 229 (2022)
7. Billings, S.A.: Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains. Wiley, Hoboken (2013)
8. Pathak, J., Hunt, B., Girvan, M., Lu, Z., Ott, E.: Model-free prediction of large spatiotemporally chaotic systems from data: a reservoir computing approach. *Phys. Rev. Lett.* **120**, 024102 (2018)
9. Vlachas, P.R., Byeon, W., Wan, Z.Y., Sapsis, T.P., Koumoutsakos, P.: Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks. *Proc. R. Soc. A* **474**, 20170844 (2018)
10. Raissi, M., Perdikaris, P., Karniadakis, G.: Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686 (2019)
11. Raissi, M., Perdikaris, P., Karniadakis, G.E.: Machine learning of linear differential equations using Gaussian processes. *J. Comput. Phys.* **348**, 683 (2017)
12. Benner, P., Gugercin, S., Willcox, K.: A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Rev.* **57**, 483 (2015)
13. Peherstorfer, B., Willcox, K.: Data-driven operator inference for nonintrusive projection-based model reduction. *Comput. Methods Appl. Mech. Eng.* **306**, 196 (2016)
14. Qian, E., Kramer, B., Peherstorfer, B., Willcox, K.: Lift and Learn: physics-informed machine learning for large-scale nonlinear dynamical systems. *Phys. D* **406**, 132401 (2020)
15. Bongard, J., Lipson, H.: Automated reverse engineering of nonlinear dynamical systems. *Proc. Natl. Acad. Sci.* **104**, 9943 (2007). <https://doi.org/10.1073/pnas.0609476104>
16. Schmidt, M., Lipson, H.: Distilling free-form natural laws from experimental data. *Science* **324**, 81 (2009)
17. Tenachi, W., Ibata, R., Diakogiannis, F. I.: Deep symbolic regression for physics guided by units constraints: toward the automated discovery of physical laws. arXiv preprint [arXiv:2303.03192](https://arxiv.org/abs/2303.03192) (2023)
18. Udrescu, S.-M., Tegmark, M.: AI Feynman: a physics-inspired method for symbolic regression. *Sci. Adv.* **6**, eaay2631 (2020)
19. Brunton, S.L., Proctor, J.L., Kutz, J.N.: Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci.* **113**, 3932 (2016)
20. Hoffmann, M., Fröhner, C., Noé, F.: Reactive SINDy: discovering governing reactions from concentration data. *J. Chem. Phys.* **150**, 025101 (2019)
21. Bhadriraju, B., Bangi, M.S.F., Narasingam, A., Kwon, J.S.-I.: Operable adaptive sparse identification of systems: application to chemical processes. *AIChE J.* **66**, e16980 (2020)
22. Scheffold, L., Finkler, T., Piechottka, U.: Gray-box system modeling using symbolic regression and nonlinear model predictive control of a semibatch polymerization. *Comput. Chem. Eng.* **146**, 107204 (2021)
23. Rubio-Herrero, J., Marrero, C.O., Fan, W.-T.L.: Modeling atmospheric data and identifying dynamics temporal data-driven modeling of air pollutants. *J. Clean. Prod.* **333**, 129863 (2022)
24. Lagergren, J.H., Nardini, J.T., Michael Lavigne, G., Rutter, E.M., Flores, K.B.: Learning partial differential equations for biological transport models from noisy spatio-temporal data. *Proc. Royal Soc. A* **476**, 20190800 (2020)
25. Pasquato, M., Abbas, M., Trani, A.A., Nori, M., Kwieciński, J.A., Trevisan, P., Braga, V.F., Bono, G., Macciò, A.V.: Sparse identification of variable star dynamics. *Astrophys. J.* **930**, 161 (2022)
26. Jiang, Y.-X., Xiong, X., Zhang, S., Wang, J.-X., Li, J.-C., Du, L.: Modeling and prediction of the transmission dynamics of COVID-19 based on the SINDy-LM method. *Nonlinear Dyn.* **105**, 2775 (2021)
27. Zucatti, V., Lui, H.F., Pitz, D.B., Wolf, W.R.: Assessment of reduced-order modeling strategies for convective heat transfer. *Numerical Heat Transf. Part A: Appl.* **77**, 702 (2020)
28. Sorokina, M., Sygletos, S., Turitsyn, S.: Sparse identification for nonlinear optical communication systems: SINO method. *Opt. Exp.* **24**, 30433 (2016)
29. Stanković, A. M., Sarić, A. A., Sarić, A. T., Transtrum, M. K.: Data-driven symbolic regression for identification of nonlinear dynamics in power systems, in *2020 IEEE Power & Energy Society General Meeting (PESGM)* (IEEE, 2020) pp. 1–5
30. Cai, Y., Wang, X., Joos, G., Kamwa, I.: An online data-driven method to locate forced oscillation sources from power plants based on sparse identification of nonlinear dynamics SINDy). *IEEE Trans. Power Syst.* **24**, 1001 (2022)
31. Narasingam, A., Kwon, J.S.-I.: Data-driven identification of interpretable reduced-order models using sparse regression. *Comput. Chem. Eng.* **119**, 101 (2018)
32. Thaler, S., Paehler, L., Adams, N.A.: Sparse identification of truncation errors. *J. Comput. Phys.* **397**, 108851 (2019)
33. Dale, R., Bhat, H.S.: Equations of mind: Data science for inferring nonlinear dynamics of socio-cognitive systems. *Cogn. Syst. Res.* **52**, 275 (2018)
34. Loiseau, J.-C., Brunton, S.L.: Constrained sparse Galerkin regression. *J. Fluid Mech.* **838**, 42 (2018)
35. Loiseau, J.-C., Noack, B.R., Brunton, S.L.: Sparse reduced-order modeling: sensor-based dynamics to full-state estimation. *J. Fluid Mech.* **844**, 459 (2018)
36. Loiseau, J.-C.: Data-driven modeling of the chaotic thermal convection in an annular thermosyphon. *Theoret. Comput. Fluid Dyn.* **34**, 339 (2020)
37. El Sayed M, Y., Semaan, R., Radespiel, R.: Sparse modeling of the lift gains of a high-lift configuration with periodic coanda blowing, in *2018 AIAA Aerospace Sciences Meeting* (2018) p. 1054
38. Chang, H., Zhang, D.: Machine learning subsurface flow equations from data. *Comput. Geosci.* **23**, 895 (2019)

39. Deng, N., Noack, B.R., Morzyński, M., Pastur, L.R.: Low-order model for successive bifurcations of the fluidic pinball. *J. Fluid Mech.* **884**, A37 (2020)
40. Fukami, K., Murata, T., Zhang, K., Fukagata, K.: Sparse identification of nonlinear dynamics with low-dimensionalized flow representations. *J. Fluid Mech.* **926**, 106623 (2021)
41. Callaham, J.L., Brunton, S.L., Loiseau, J.-C.: On the role of nonlinear correlations in reduced-order modelling. *J. Fluid Mech.* **938**, A1 (2022)
42. Khoo, Z.C., Chan, C.H., Hwang, Y.: A sparse optimal closure for a reduced-order model of wall-bounded turbulence. *J. Fluid Mech.* **939**, A11 (2022)
43. Deng, N., Noack, B.R., Morzyński, M., Pastur, L.R.: Galerkin force model for transient and post-transient dynamics of the fluidic pinball. *J. Fluid Mech.* **918**, A4 (2021)
44. Xiao, Q., Wang, J., Yang, X., Jiang, B.: Construction of a reduced-order model of an electroosmotic micromixer and discovery of attractors for petal structure. *Phys. Fluids* **210**, 4632 (2023)
45. Foster, J. A., Decuyper, J., De Troyer, T., Runacres, M.: Estimating a sparse nonlinear dynamical model of the flow around an oscillating cylinder in a fluid flow using SINDy. In: Conference on Noise and Vibration Engineering ISMA 2022 (ISMA 2022, 2022) p. tba
46. Schmelzer, M., Dwight, R.P., Cinnella, P.: Discovery of algebraic Reynolds-stress models using sparse symbolic regression. *Flow Turbul. Combust.* **104**, 579 (2020)
47. Beetham, S., Capecelatro, J.: Formulating turbulence closures using sparse regression with embedded form invariance. *Phys. Rev. Fluids* **5**, 084611 (2020)
48. Beetham, S., Capecelatro, J.: Multiphase turbulence modeling using sparse regression and gene expression programming, arXiv preprint [arXiv:2106.10397](https://arxiv.org/abs/2106.10397) (2021)
49. Beetham, S., Fox, R.O., Capecelatro, J.: Sparse identification of multiphase turbulence closures for coupled fluid-particle flows. *J. Fluid Mech.* **914**, A11 (2021)
50. Callaham, J.L., Rigas, G., Loiseau, J.-C., Brunton, S.L.: An empirical mean-field model of symmetry-breaking in a turbulent wake. *Sci. Adv.* **8**, eabm4786 (2022)
51. Sansica, A., Loiseau, J.-C., Kanamori, M., Hashimoto, A., Robinet, J.-C.: System identification of two-dimensional transonic buffet. *AIAA J.* **60**, 3090 (2022)
52. Dam, M., Brøns, M., JuulRasmussen, J., Naulin, V., Hesthaven, J.S.: Sparse identification of a predator-prey system from simulation data of a convection model. *Phys. Plasmas* **24**, 022310 (2017)
53. Kaptanoglu, A.A., Morgan, K.D., Hansen, C.J., Brunton, S.L.: Physics-constrained, low-dimensional models for magnetohydrodynamics: first-principles and data-driven approaches. *Phys. Rev. E* **104**, 015206 (2021)
54. Alves, E.P., Fiuza, F.: Data-driven discovery of reduced plasma physics models from fully kinetic simulations. *Phys. Rev. Res.* **4**, 033192 (2022)
55. Lai, Z., Nagarajaiah, S.: Sparse structural system identification method for nonlinear dynamic systems with hysteresis/inelastic behavior. *Mech. Syst. Signal Process.* **117**, 813 (2019)
56. de Silva, B.M., Higdon, D.M., Brunton, S.L., Kutz, J.N.: Discovery of physics from data: universal laws and discrepancies. *Front. Artif. Intell.* **3**, 25 (2020)
57. Pan, S., Arnold-Medabalimi, N., Duraisamy, K.: Sparsity-promoting algorithms for the discovery of informative Koopman-invariant subspaces. *J. Fluid Mech.* **917**, A18 (2021)
58. Subramanian, R., Moar, R.R., Singh, S.: White-box machine learning approaches to identify governing equations for overall dynamics of manufacturing systems: A case study on distillation column. *Mach. Learn. Appl.* **3**, 100014 (2021)
59. Brenner, M., Hess, F., Mikhaeil, J. M., Bereska, L. F., Monfared, Z., Kuo, P.-C., Durstewitz, D.: Tractable dendritic RNNs for reconstructing nonlinear dynamical systems, In: International Conference on Machine Learning (PMLR, 2022) pp. 2292–2320
60. Zhang, S., Ahmed, F., Song, H.-S.: Knowledge-informed data-driven modeling for sparse identification of governing equations for microbial inactivation processes in food. *Front. Food Sci. Technol.* **29**, 213 (2022)
61. Golden, M., Grigoriev, R., Nambisan, J., Fernandez-Nieves, A.: Physically-informed data-driven modeling of active nematics, arXiv preprint [arXiv:2202.12853](https://arxiv.org/abs/2202.12853) (2022)
62. Joshi, C., Ray, S., Lemma, L.M., Varghese, M., Sharp, G., Dogic, Z., Baskaran, A., Hagan, M.F.: Data-driven discovery of active nematic hydrodynamics. *Phys. Rev. Lett.* **129**, 258001 (2022)
63. Schaeffer, H.: Learning partial differential equations via data discovery and sparse optimization. *Proc. R Soc. A* **473**, 20160446 (2017)
64. Rudy, S.H., Brunton, S.L., Proctor, J.L., Kutz, J.N.: Data-driven discovery of partial differential equations. *Sci. Adv.* **3**, e1602614 (2017)
65. Sandoz, A., Ducret, V., Gottwald, G.A., Vilmar, G., Perron, K.: SINDy for delay-differential equations: application to model bacterial zinc response. *Proc. Royal Soc. A* **479**, 20220556 (2023)
66. Nicolaou, Z. G., Huo, G., Chen, Y., Brunton, S. L., Kutz, J. N.: Data-driven discovery and extrapolation of parameterized pattern-forming dynamics, arXiv preprint [arXiv:2301.02673](https://arxiv.org/abs/2301.02673) (2023)
67. Klimovskaya, A., Ganscha, S., Claassen, M.: Sparse regression based structure learning of stochastic reaction networks from single cell snapshot time series. *PLoS Comput. Biol.* **12**, e1005234 (2016)
68. Brückner, D.B., Ronceray, P., Broedersz, C.P.: Inferring the dynamics of underdamped stochastic systems. *Phys. Rev. Lett.* **125**, 058103 (2020)
69. Dai, M., Gao, T., Lu, Y., Zheng, Y., Duan, J.: Detecting the maximum likelihood transition path from data of stochastic dynamical systems. *Chaos: Interdiscip. J. Nonlinear Sci.* **30**, 113124 (2020)
70. Callaham, J.L., Loiseau, J.-C., Rigas, G., Brunton, S.L.: Nonlinear stochastic modelling with Langevin regression. *Proc. Royal Soc. A* **477**, 20210092 (2021)
71. Huang, Y., Mabrouk, Y., Gompper, G., Sabass, B.: Sparse inference and active learning of stochastic differential equations from data. *Sci. Rep.* **12**, 21691 (2022)

72. Hirsh, S.M., Barajas-Solano, D.A., Kutz, J.N.: Sparsifying priors for Bayesian uncertainty quantification in model discovery. *Royal Soc. Open Sci.* **9**, 211823 (2022)
73. Bakarji, J., Callaham, J., Brunton, S.L., Kutz, J.N.: Dimensionally consistent learning with Buckingham pi. *Nat. Comput. Sci.* **1**, 1–11 (2022)
74. Brunton, S.L., Proctor, J.L., Kutz, J.N.: Sparse identification of nonlinear dynamics with control (SINDYc). *IFAC-PapersOnLine* **49**, 710 (2016)
75. Kaiser, E., Kutz, J.N., Brunton, S.L.: Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. *Proc. Royal Soc. London A* **474**, 20180335 (2018)
76. Kaiser, E., Kutz, J. N., Brunton, S. L.: Discovering conservation laws from data for control, In: 2018 IEEE Conference on Decision and Control (CDC) ( IEEE, 2018) pp. 6415–6421
77. Mangan, N.M., Brunton, S.L., Proctor, J.L., Kutz, J.N.: Inferring biological networks by sparse identification of nonlinear dynamics. *IEEE Trans. Mole. Biol. Multi-Scale Commun.* **2**, 52 (2016)
78. Kaheman, K., Kutz, J.N., Brunton, S.L.: SINDy-PI: a robust algorithm for parallel implicit sparse identification of nonlinear dynamics. *Proc. Royal Soc. A* **476**, 20200279 (2020)
79. Mangan, N.M., Askham, T., Brunton, S.L., Kutz, J.N., Proctor, J.L.: Model selection for hybrid dynamical systems via sparse regression. *Proc. Royal Soc. A* **475**, 20180534 (2019)
80. Thiele, G., Fey, A., Sommer, D., Krüger, J.: System identification of a hysteresis-controlled pump system using SINDy, In: 2020 24th International Conference on System Theory, Control and Computing (ICSTCC) ( IEEE, 2020) pp. 457–464
81. Champion, K., Zheng, P., Aravkin, A.Y., Brunton, S.L., Kutz, J.N.: A unified sparse optimization framework to learn parsimonious physics-informed models from data. *IEEE Access* **8**, 169259 (2020)
82. Mangan, N.M., Kutz, J.N., Brunton, S.L., Proctor, J.L.: Model selection for dynamical systems via sparse regression and information criteria. *Proc. Royal Soc. A Math. Phys. Eng. Sci.* **473**, 20170009 (2017)
83. Dong, X., Bai, Y.-L., Lu, Y., Fan, M.: An improved sparse identification of nonlinear dynamics with Akaike information criterion and group sparsity. *Nonlinear Dyn.* **1**, 1485 (2022)
84. Kaptanoglu, A.A., Callaham, J.L., Aravkin, A., Hansen, C.J., Brunton, S.L.: Promoting global stability in data-driven models of quadratic nonlinear dynamics. *Phys. Rev. Fluids* **6**, 094401 (2021)
85. Tran, G., Ward, R.: Exact recovery of chaotic systems from highly corrupted data. *Multiscale Model. Simul.* **15**, 1108 (2017). <https://doi.org/10.1137/16m1086637>
86. Schaeffer, H., Tran, G., Ward, R.: Extracting sparse high-dimensional dynamics from limited data. *SIAM J. Appl. Math.* **78**, 3279 (2018)
87. Delahunt, C.B., Kutz, J.N.: A toolkit for data-driven discovery of governing equations in high-noise regimes. *IEEE Access* **10**, 31210 (2022)
88. Wentz, J., Doostan, A.: Derivative-based SINDY (DSINDY): Addressing the challenge of discovering governing equations from noisy data. *arXiv preprint arXiv:2211.05918* (2022)
89. Kaheman, K., Brunton, S.L., Kutz, J.N.: Automatic differentiation to simultaneously identify nonlinear dynamics and extract noise probability distributions from data. *Mach. Learn.: Sci. Technol.* **3**, 015031 (2022)
90. Somacal, A., Barrera, Y., Boechi, L., Jonckheere, M., Lefieux, V., Picard, D., Smucler, E.: Uncovering differential equations from data with hidden variables. *Phys. Rev. E* **105**, 054209 (2022)
91. Bakarji, J., Champion, K., Kutz, J. N., Brunton, S. L.: Discovering governing equations from partial measurements with deep delay autoencoders, *arXiv preprint arXiv:2201.05136* (2022b)
92. Conti, P., Gobat, G., Fresca, S., Manzoni, A., Frangi, A.: Reduced order modeling of parametrized systems through autoencoders and SINDy approach: continuation of periodic solutions, *arXiv preprint arXiv:2211.06786* (2022)
93. Gao, L., Kutz, J. N.: Bayesian autoencoders for data-driven discovery of coordinates, governing equations and fundamental constants, *arXiv preprint arXiv:2211.10575* (2022)
94. Zhao, Z., Li, Q.: Adaptive sampling methods for learning dynamical systems, In: Mathematical and Scientific Machine Learning (PMLR, 2022) pp. 335–350
95. Wu, K., Xiu, D.: Numerical aspects for approximating governing equations using data. *J. Comput. Phys.* **384**, 200 (2019)
96. Fasel, U., Kutz, J.N., Brunton, B.W., Brunton, S.L.: Ensemble-SINDy: robust sparse model discovery in the low-data, high-noise limit, with active learning and control. *Proc. Royal Soc. A* **478**, 20210904 (2022)
97. Gao, L., Fasel, U., Brunton, S. L., Kutz, J. N.: Convergence of uncertainty estimates in ensemble and Bayesian sparse model discovery, *arXiv preprint arXiv:2301.12649* (2023)
98. Jiang, F., Du, L., Yang, F., Deng, Z.-C.: Regularized least absolute deviation-based sparse identification of dynamical systems. *Chaos: Interdiscip. J. Nonlinear Sci.* **33**, 013103 (2023)
99. Schaeffer, H., McCalla, S.G.: Sparse model selection via integral terms. *Phys. Rev. E* **96**, 023302 (2017)
100. Gurevich, D.R., Reinbold, P.A., Grigoriev, R.O.: Robust and optimal sparse regression for nonlinear PDE models. *Chaos: Interdiscip. J. Nonlinear Sci.* **29**, 103113 (2019)
101. Reinbold, P.A., Gurevich, D.R., Grigoriev, R.O.: Using noisy or incomplete data to discover models of spatiotemporal dynamics. *Phys. Rev. E* **101**, 010203 (2020)
102. Messenger, D.A., Bortz, D.M.: Weak SINDy for partial differential equations. *J. Comput. Phys.* **443**, 110525 (2021)
103. Kageorge, L. M., Grigoriev, R. O., Schatz, M. F.: Data-driven detection of drifting system parameters, *arXiv preprint arXiv:2111.12114* (2021)
104. Reinbold, P.A., Kageorge, L.M., Schatz, M.F., Grigoriev, R.O.: Robust learning from noisy, incomplete, high-dimensional experimental data via physically constrained symbolic regression. *Nat. Commun.* **12**, 1 (2021)
105. Gurevich, D. R., Reinbold, P. A., Grigoriev, R. O.: Learning fluid physics from highly turbulent data using sparse physics-informed discovery of empirical relations (SPI-DER), *arXiv preprint arXiv:2105.00048* ( 2021)
106. Russo, B., Laiu, M. P.: Convergence of weak-SINDy surrogate models, *arXiv preprint arXiv:2209.15573* ( 2022)

107. Messenger, D. A., Bortz, D. M.: Asymptotic consistency of the WSINDy algorithm in the limit of continuum data, arXiv preprint [arXiv:2211.16000](https://arxiv.org/abs/2211.16000) (2022a)
108. Messenger, D.A., Bortz, D.M.: Learning mean-field equations from particle data using WSINDy. *Phys. D* **439**, 133406 (2022)
109. Messenger, D. A., Dall’Anese, E., Bortz, D.: Online weak-form sparse identification of partial differential equations, In: Mathematical and Scientific Machine Learning (PMLR, 2022) pp. 241–256
110. Gelß, P., Klus, S., Eisert, J., Schütte, C.: Multidimensional approximation of nonlinear dynamical systems. *J. Comput. Nonlinear Dyn.* **14**, 452 (2019)
111. Goebmann, A., Götte, M., Roth, I., Sweke, R., Kutyniok, G., Eisert, J.: Tensor network approaches for learning nonlinear dynamical laws, arXiv preprint [arXiv:2002.12388](https://arxiv.org/abs/2002.12388) (2020)
112. Boninsegna, L., Nüske, F., Clementi, C.: Sparse learning of stochastic dynamical equations. *J. Chem. Phys.* **148**, 241723 (2018)
113. Lorenz, E.N.: Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 130 (1963)
114. de Silva, B., Champion, K., Quade, M., Loiseau, J.-C., Kutz, J.N., Brunton, S.: PySINDy: a Python package for the sparse identification of nonlinear dynamical systems from data. *J. Open Source Softw.* **5**, 1 (2020)
115. Kaptanoglu, A.A., de Silva, B.M., Fasel, U., Kaheman, K., Goldschmidt, A.J., Callaham, J., Delahunt, C.B., Nicolaou, Z.G., Champion, K., Loiseau, J.-C., Kutz, J.N., Brunton, S.L.: PySINDy a comprehensive Python package for robust sparse system identification. *J. Open Source Softw.* **7**, 3994 (2022). <https://doi.org/10.21105/joss.03994>
116. Zheng, P., Askham, T., Brunton, S.L., Kutz, J.N., Aravkin, A.Y.: A unified framework for sparse relaxed regularized regression: SR3. *IEEE Access* **7**, 1404 (2019)
117. Tibshirani, R., Wainwright, M., Hastie, T.: Statistical Learning with Sparsity: The Lasso and Generalizations. Chapman and Hall/CRC, Boca Raton (2015)
118. Bertsimas, D., Gurnee, W.: Learning sparse nonlinear dynamics via mixed-integer optimization. *Nonlinear Dyn.* **1**, 1–20 (2023)
119. Schaeffer, H., Tran, G., Ward, R.: Learning dynamical systems and bifurcation via group sparsity, arXiv preprint [arXiv:1709.01558](https://arxiv.org/abs/1709.01558) (2017)
120. Uy, N.Q., Hoai, N.X., O’Neill, M., McKay, R.I., Galván-López, E.: Semantically-based crossover in genetic programming: application to real-valued symbolic regression. *Genet. Program Evolvable Mach.* **12**, 91 (2011)
121. Petersen, B. K., Larma, M. L., Mundhenk, T. N., Santiago, C. P., Kim, S. K., Kim, J. T.: Deep symbolic regression: Recovering mathematical expressions from data via risk-seeking policy gradients, in <https://openreview.net/forum?id=m5Qsh0kBQG>. In: International Conference on Learning Representations (2021)
122. Kantz, H., Schreiber, T.: Nonlinear time series analysis. Cambridge University Press, Cambridge (2004)
123. Kaptanoglu, A.: An Exploration of Data-Driven System Identification and Machine Learning for Plasma Physics. University of Washington, Seattle (2021)
124. Van Breugel, F., Kutz, J.N., Brunton, B.W.: Numerical differentiation of noisy data: a unifying multi-objective optimization framework. *IEEE Access* **8**, 196865 (2020)
125. Van Breugel, F., Liu, Y., Brunton, B.W., Kutz, J.N.: Pynumdiff: a python package for numerical differentiation of noisy time-series data. *J. Open Source Softw.* **7**, 4078 (2022)
126. Blasco, X., Herrero, J.M., Sanchis, J., Martínez, M.: A new graphical visualization of n-dimensional Pareto front for decision-making in multiobjective optimization. *Inf. Sci.* **178**, 3908 (2008)
127. La Cava, W., Orzechowski, P., Burlacu, B., de Franca, F. O., Virgolin, M., Jin, Y., Kommenda, M., Moore, J. H.: Contemporary symbolic regression methods and their relative performance. In: Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 1)
128. Orzechowski, P., La Cava, W., Moore, J. H.: Where are we now? A large benchmark study of recent symbolic regression methods. In: Proceedings of the Genetic and Evolutionary Computation Conference (2018) pp. 1183–1190
129. Bhat, H. S.: Learning and interpreting potentials for classical Hamiltonian systems. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases (Springer, 2019) pp. 217–228
130. Chu, H.K., Hayashibe, M.: Discovering interpretable dynamics by sparsity promotion on energy and the Lagrangian. *IEEE Robot. Autom. Lett.* **5**, 2154 (2020)
131. Bertalan, T., Dietrich, F., Mezić, I., Kevrekidis, I.G.: On learning Hamiltonian systems from data. *Chaos: Interdiscip. J. Nonlinear Sci.* **29**, 121107 (2019)
132. Mikhalei, J. M., Monfared, Z., Durstewitz, D.: On the difficulty of learning chaotic dynamics with RNNs, arXiv preprint [arXiv:2110.07238](https://arxiv.org/abs/2110.07238) (2021)
133. Sangiorgio, M., Dericole, F., Guariso, G.: Forecasting of noisy chaotic systems with deep neural networks. *Chaos, Solitons Fractals* **153**, 111570 (2021)
134. Ouala, S., Brunton, S.L., Chapron, B., Pascual, A., Collard, F., Gaultier, L., Fablet, R.: Bounded nonlinear forecasts of partially observed geophysical systems with physics-constrained deep learning. *Phys. D: Nonlinear Phenomena* **442**, 133630 (2023)
135. Diamond, S., Boyd, S.: CVXPY: A Python-embedded modeling language for convex optimization. *J. Mach. Learn. Res.* **17**, 2909 (2016)
136. Kaptanoglu, A.A., Qian, T., Wechsung, F., Landreman, M.: Permanent-magnet optimization for stellarators as sparse regression. *Phys. Rev. Appl.* **18**, 044006 (2022)
137. Wainwright, M.J.: Sharp thresholds for high-dimensional and noisy sparsity recovery using  $l - 1$ -constrained quadratic programming (Lasso). *IEEE Trans. Inf. Theory* **55**, 2183 (2009)
138. Bertsimas, D., Pauphilet, J., Parys, B.V.: Sparse Regression: Scalable Algorithms and Empirical Performance. *Stat. Sci.* **35**, 555 (2020). <https://doi.org/10.1214/19-STS701>
139. Sommerer, J.C., Ott, E.: Particles floating on a moving fluid: A dynamically comprehensible physical fractal. *Science* **259**, 335 (1993)

140. Pavliotis, G., Stuart, A.: Multiscale Methods: Averaging and Homogenization. Springer Science & Business Media, New York (2008)
141. Bucci, A., Semeraro, O., Allauzen, A., Chibbaro, S., Mathelin, L.: Curriculum learning for data-driven modeling of dynamical systems. <https://doi.org/10.48550/ARXIV.2112.08458> (2021)
142. Bramburger, J.J., Dylewsky, D., Kutz, J.N.: Sparse identification of slow timescale dynamics. *Phys. Rev. E* **102**, 022204 (2020)
143. Cenedese, M., Axås, J., Bäuerlein, B., Avila, K., Haller, G.: Data-driven modeling and prediction of non-linearizable dynamics via spectral submanifolds. *Nat. Commun.* **13**, 1 (2022)
144. Szalai, R.: Data-driven reduced order models using invariant foliations, manifolds and autoencoders, arXiv preprint [arXiv:2206.12269](https://arxiv.org/abs/2206.12269) (2022)
145. Axås, J., Cenedese, M., Haller, G.: Fast data-driven model reduction for nonlinear dynamical systems. *Nonlinear Dyn.* **1**, 1–12 (2022)
146. Udrescu, S.-M., Tan, A., Feng, J., Neto, O., Wu, T., Tegmark, M.: AI Feynman 2.0: Pareto-optimal symbolic regression exploiting graph modularity. *Adv. Neural. Inf. Process. Syst.* **33**, 4860 (2020)
147. Grünwald, P.D., Myung, I.J., Pitt, M.A.: Advances in Minimum Description Length: Theory and Applications. MIT press, Cambridge (2005)
148. Vladislavleva, E.J., Smits, G.F., Den Hertog, D.: Order of nonlinearity as a complexity measure for models generated by symbolic regression via pareto genetic programming. *IEEE Trans. Evol. Comput.* **13**, 333 (2008)
149. Murdoch, W.J., Singh, C., Kumbier, K., Abbasi-Asl, R., Yu, B.: Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci.* **116**, 22071 (2019)
150. Tesi, A., Villoresi, F., Genesio, R.: On the stability domain estimation via a quadratic Lyapunov function: convexity and optimality properties for polynomial systems. *IEEE Trans. Autom. Control* **41**, 1650 (1996)
151. Ahmadi, A. A., Majumdar, A., Tedrake, R.: Complexity of ten decision problems in continuous time dynamical systems. In: 2013 American Control Conference ( IEEE, 2013) pp. 6376–6381
152. Dikeman, H. E., Zhang, H., Yang, S.: Stiffness-reduced neural ODE models for data-driven reduced-order modeling of combustion chemical kinetics. In: AIAA SCITECH 2022 Forum ( 2022) p. 0226
153. Ott, E., Grebogi, C., Yorke, J.A.: Controlling chaos. *Phys. Rev. Lett.* **64**, 1196 (1990)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.