Supplementary information for "Automatically discovering ordinary differential equations from data with sparse regression"

Kevin Egan[1], Weizhen Li[1], Rui Carvalho[1,2*]

[1]Department of Engineering, Durham University, Lower Mountjoy, South Road, Durham, DH1 3LE, United Kingdom.
[2]Institute for Data Science, Durham University, South Road, Durham, DH1 3LE, United Kingdom.

*Corresponding author(s). E-mail(s): rui.carvalho@dur.ac.uk;

Supplementary Note 1. Algorithms

Algorithm 1 Automatic Savitzky-Golay Filter

Input: $\tilde{\mathbf{X}} \in \mathbb{R}^{n \times m}$, dt.
Output: Savitzky-Golay optimally smoothed $\mathbf{X}$ and $\dot{\mathbf{X}}$.

determine lower and upper bounds of (odd) window length l:
$l_{min} = 13$,
$l_{max} = \max(13,\ \min(n - (n-1) \bmod 2, 101))$;
build $L = (l_{min}, \ldots, l_{max})$;
▷ v = degree of derivative
for $j = 1, \ldots, m$ **do**
$l^* = \arg\min_L \|SG(\tilde{\mathbf{x}}_j, o = 4, l = L, v = 0, dt) - \tilde{\mathbf{x}}_j\|_2^2$,
$\mathbf{x}_j = SG(\tilde{\mathbf{x}}_j, o = 4, l = l^*, v = 0, dt)$,
$\dot{\mathbf{x}}_j = SG(\tilde{\mathbf{x}}_j, o = 4, l = l^*, v = 1, dt)$;
end for
consolidate $\mathbf{X}, \dot{\mathbf{X}} \in \mathbb{R}^{n \times m}$ with each $\mathbf{x}_j$ and $\dot{\mathbf{x}}_j$, respectively.

Algorithm 2 Automatic Regression for Governing Equations (ARGOS)

Input: $\mathbf{X} \in \mathbb{R}^{n\times m}$, $\dot{\mathbf{x}}_j \in \mathbb{R}^n$, d, $\alpha = 0.05$.

▷ STEP ONE: Initial design matrix

1: $p^{(0)} = \binom{m+d}{d}$;

2: create $\boldsymbol{\Theta}^{(0)}(\mathbf{X}) \in \mathbb{R}^{n\times p^{(0)}}$ with basis functions up to order d of the columns of $\mathbf{X}$;

▷ STEP TWO: Trim design matrix

▷ Variable selection with the lasso or adaptive lasso

▷ λ^* : Optimal λ from 10-fold cross-validation

▷ lasso: $w = 1$

▷ adaptive lasso: $w =$ ridge regression coefficients

3: $\hat{\beta}^{(0)} = \arg\min_{\beta} \left\| \dot{\mathbf{x}}_j - \boldsymbol{\Theta}^{(0)}(\mathbf{X})\beta \right\|_2^2 + \lambda^* \sum_{k=1}^{p^{(0)}} w_k |\beta_k|$;

4: extract $\boldsymbol{\Theta}^{(1)}(\mathbf{X})$ to contain columns of $\boldsymbol{\Theta}^{(0)}(\mathbf{X})$ up to the largest order $d^{(1)}$ of the selected variables in $\hat{\beta}^{(0)}$;

▷ STEP THREE: Final point estimates

▷ Repeat sparse regression algorithm from STEP TWO

5: $p^{(1)} = \binom{m+d^{(1)}}{d^{(1)}}$;

6: $\hat{\beta}^{(1)} = \arg\min_{\beta} \left\| \dot{\mathbf{x}}_j - \boldsymbol{\Theta}^{(1)}(\mathbf{X})\beta \right\|_2^2 + \lambda^* \sum_{k=1}^{p^{(1)}} w_k |\beta_k|$;

▷ Apply threshold values

7: $\eta = [10^{-8}, 10^{-7}, \ldots, 10^{1}]$;

8: **for** $i = 1, \ldots,$ **card**(η) **do**

▷ Ordinary least squares regression (OLS) estimate after variable selection

$$\hat{\beta}^{\mathrm{OLS}}[i] = \arg\min_{\beta_{\mathcal{K}_i}} \left\| \dot{\mathbf{x}}_j - \boldsymbol{\Theta}^{(1)}_{\mathcal{K}_i}(\mathbf{X})\beta_{\mathcal{K}_i} \right\|_2^2 \text{ where } \mathcal{K}_i = \{k : |\hat{\beta}^{(1)}_k| \geq \eta_i\},$$

$$\mathrm{BIC}_i = \mathrm{BIC}(\hat{\beta}^{\mathrm{OLS}}[i]);$$

9: **end for**

10: $\hat{\beta} = \left\{ \hat{\beta}^{\mathrm{OLS}}[i] \middle| i : \arg\min(\mathrm{BIC}) \right\}$;

▷ STEP FOUR: Bootstrap estimates for confidence intervals

▷ $B = 2000$ bootstrap samples

11: bootstrap Statements 6 – 10 to approximate confidence interval bounds: $CI_{\mathrm{lo}} = [B\alpha/2]$ and $CI_{\mathrm{up}} = B - CI_{\mathrm{lo}} + 1$;

12: construct bootstrap confidence intervals for $\hat{\beta}$: $\hat{\beta}_k \in \left[\hat{\beta}_k^{\mathrm{OLS}\{CI_{\mathrm{lo}}\}}, \hat{\beta}_k^{\mathrm{OLS}\{CI_{\mathrm{up}}\}}\right]$, and $0 < \hat{\beta}_k^{\mathrm{OLS}\{CI_{\mathrm{lo}}\}}$ or $0 > \hat{\beta}_k^{\mathrm{OLS}\{CI_{\mathrm{up}}\}}$.

Supplementary Note 2. Linear systems

A. Two-dimensional damped oscillator with linear dynamics

We examined the two-dimensional linear system as [1]

$$\begin{aligned} \dot{x}_1 &= -0.1x_1 + 2x_2, \\ \dot{x}_2 &= -2x_1 - 0.1x_2. \end{aligned} \tag{S1}$$

Table S1 Minimum number of observations (n) needed for each method to obtain 80% accuracy in identifying governing equations of dynamical systems. Top-performing algorithms are in red, and three-dimensional systems have a shaded background.

System	Algorithm	n
Two-dimensional linear	ARGOS-Lasso	$10^{2.6}$ (399)
	ARGOS-Adaptive Lasso	$10^{2.6}$ (399)
	SINDy with AIC	$10^{3.3}$ (1996)
Three-dimensional linear	ARGOS-Lasso	$10^{2.9}$ (795)
	ARGOS-Adaptive Lasso	$10^{3.2}$ (1585)
	SINDy with AIC	NA
Two-dimensional cubic	ARGOS-Lasso	$10^{3.2}$ (1585)
	SINDy with AIC	$10^{3.3}$ (1996)
	ARGOS-Adaptive Lasso	$10^{4.1}$ (12590)
Lotka-Volterra	ARGOS-Adaptive Lasso	$10^{3.2}$ (1585)
	SINDy with AIC	$10^{3.2}$ (1585)
	ARGOS-Lasso	$10^{3.3}$ (1996)
Rossler	ARGOS-Adaptive Lasso	$10^{2.9}$ (795)
	ARGOS-Lasso	$10^{3.2}$ (1585)
	SINDy with AIC	$10^{3.2}$ (1585)
Lorenz	ARGOS-Adaptive Lasso	$10^{3.8}$ (6310)
	ARGOS-Lasso	$10^{3.9}$ (7944)
	SINDy with AIC	NA
Van der Pol	ARGOS-Adaptive Lasso	$10^{2.9}$ (795)
	SINDy with AIC	$10^{2.9}$ (795)
	ARGOS-Lasso	$10^{3.0}$ (1000)
Duffing	ARGOS-Lasso	$10^{2.6}$ (399)
	SINDy with AIC	$10^{2.9}$ (795)
	ARGOS-Adaptive Lasso	$10^{3.0}$ (1000)

For $x_1(t)$ and $x_2(t)$, we generated a random uniform distribution of 100 values between $[10^{-1}, 10^3]$.

In Fig. S1, we observe the performance of our approach and SINDy with AIC in discovering the two-dimensional damped oscillator with linear dynamics. We found that with less than 300 observations and low SNR, our method identified overly sparse models and struggled to represent the underlying equations of the system accurately. As the length of the time series n increased and the data became less contaminated with noise, however, the performance of our method improved in extracting the true terms. Conversely, SINDy with AIC demonstrated a tendency to produce dense models, which contained numerous erroneous variables, particularly with less than 1000 observations and low to medium SNR.

Table S2 Minimum signal-to-noise ratio (SNR) tolerated by each method to achieve 80% accuracy in identifying the governing equations of the dynamical systems. Top-performing algorithms are in red, and three-dimensional systems have a shaded background.

System	Algorithm	SNR
Two-dimensional linear	ARGOS-Lasso	25
	ARGOS-Adaptive Lasso	25
	SINDy with AIC	37
Three-dimensional linear	ARGOS-Lasso	31
	ARGOS-Adaptive Lasso	40
	SINDy with AIC	∞
Two-dimensional cubic	ARGOS-Lasso	43
	SINDy with AIC	46
	ARGOS-Adaptive Lasso	NA
Lotka-Volterra	ARGOS-Adaptive Lasso	16
	SINDy with AIC	22
	ARGOS-Lasso	28
Rossler	ARGOS-Adaptive Lasso	31
	ARGOS-Lasso	34
	SINDy with AIC	NA
Lorenz	ARGOS-Adaptive Lasso	46
	ARGOS-Lasso	55
	SINDy with AIC	∞
Van der Pol	SINDy with AIC	16
	ARGOS-Adaptive Lasso	19
	ARGOS-Lasso	25
Duffing	ARGOS-Lasso	28
	ARGOS-Adaptive Lasso	28
	SINDy with AIC	34

B. Three-dimensional linear system

We evaluated a three-dimensional system [1]:

$$\begin{aligned} \dot{x}_1 &= -0.1x_1 + 2x_2, \\ \dot{x}_2 &= -2x_1 - 0.1x_2, \\ \dot{x}_3 &= -0.3x_3. \end{aligned} \tag{S2}$$

For $x_1(t)$, $x_2(t)$, and $x_3(t)$, we again developed a random uniform distribution containing 100 values between $[10^{-1}, 10^3]$.

Figure S2 demonstrates the efficacy of employing the lasso within our framework, as the method more consistently identified the three-dimensional linear system than the other two algorithms studied here. With greater than 1000 observations and medium SNR, our approach accurately represented these simple dynamics. We also found that, with low SNR and less than 300 observations, our method discovered models that were overly sparse and did not fully represent the dynamics, while SINDy with AIC often

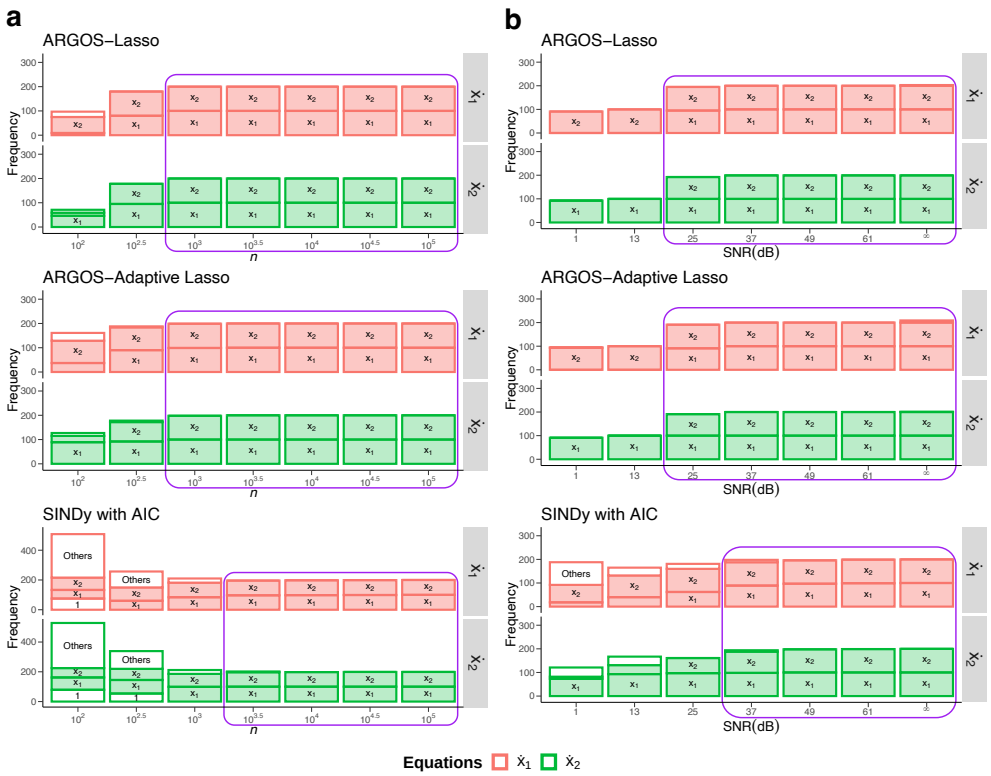

Fig. S1 Frequency of identified variables for the two-dimensional damped oscillator with linear dynamics across algorithms. Colors correspond to each governing equation; filled boxes indicate correctly identified variables, while white boxes denote erroneous terms. Panels show the frequency of identified variables for data sets with (**a**) increasing time-series length n (signal-to-noise ratio (SNR) = 49 dB), and (**b**) SNR ($n = 5000$). Purple-bordered regions demarcate model discovery above 80%.

failed to discover a parsimonious representation of the system by identifying dense models with superfluous terms.

Supplementary Note 3. Nonlinear systems

A. Two-dimensional damped oscillator with cubic dynamics

We examined a two-dimensional system with cubic dynamics as [1]

$$\begin{aligned} \dot{x}_1 &= -0.1x_1^3 + 2x_2^3, \\ \dot{x}_2 &= -2x_1^3 - 0.1x_2^3. \end{aligned} \tag{S3}$$

For $x_1(t)$ and $x_2(t)$, we generated a random uniform distribution containing 100 values between $[-2, 2]$.

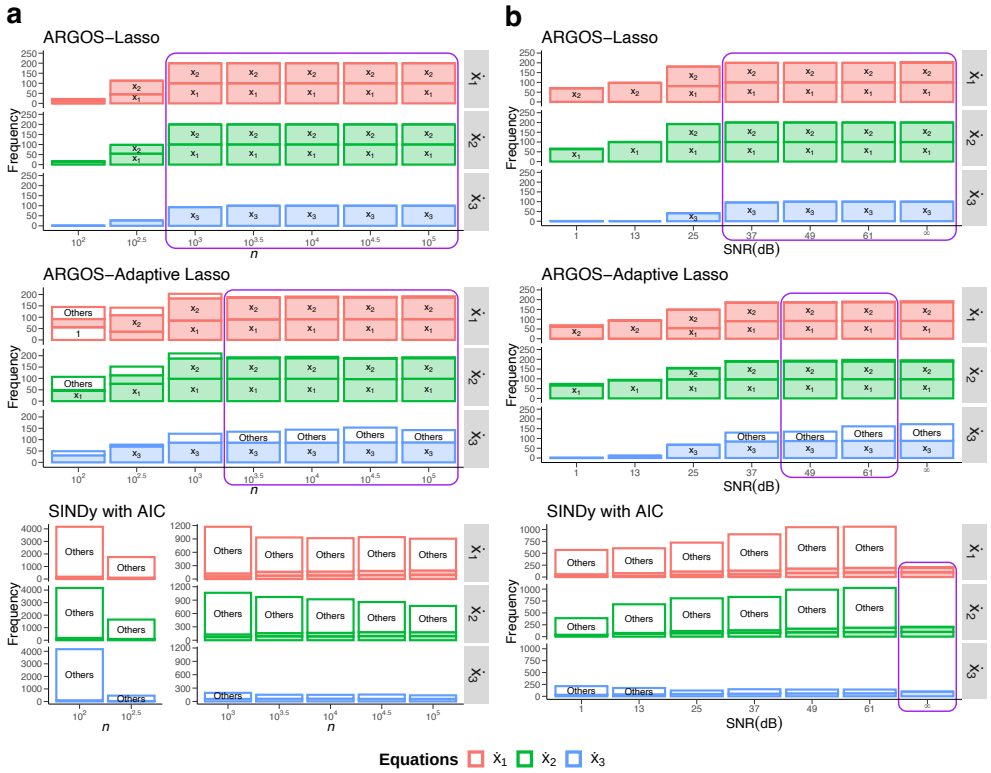

Fig. S2 Frequency of identified variables for the three-dimensional linear system across algorithms. Colors correspond to each governing equation; filled boxes indicate correctly identified variables, while white boxes denote erroneous terms. Panels show the frequency of identified variables for data sets with (**a**) increasing time-series length n (signal-to-noise ratio (SNR) = 49 dB), and (**b**) SNR ($n = 5000$). Purple-bordered regions demarcate model discovery above 80%.

Figure S3 further demonstrates the effectiveness of the lasso algorithm for identifying the two-dimensional damped harmonic oscillator with cubic dynamics. Here, SINDy with AIC performed model discovery with similar accuracy to our approach, while both the lasso and SINDy with AIC ultimately outperformed the adaptive lasso.

B. Lotka-Volterra system

The Lotka-Volterra system is described by two first-order nonlinear differential equations commonly used to depict the interaction dynamics between two species in biological systems, with one being the predator and the other the prey [2]. The predator-prey equations are represented as

$$\begin{aligned} \dot{x}_1 &= \alpha x_1 - \zeta x_1 x_2, \\ \dot{x}_2 &= \delta x_1 x_2 - \gamma x_2, \end{aligned} \tag{S4}$$

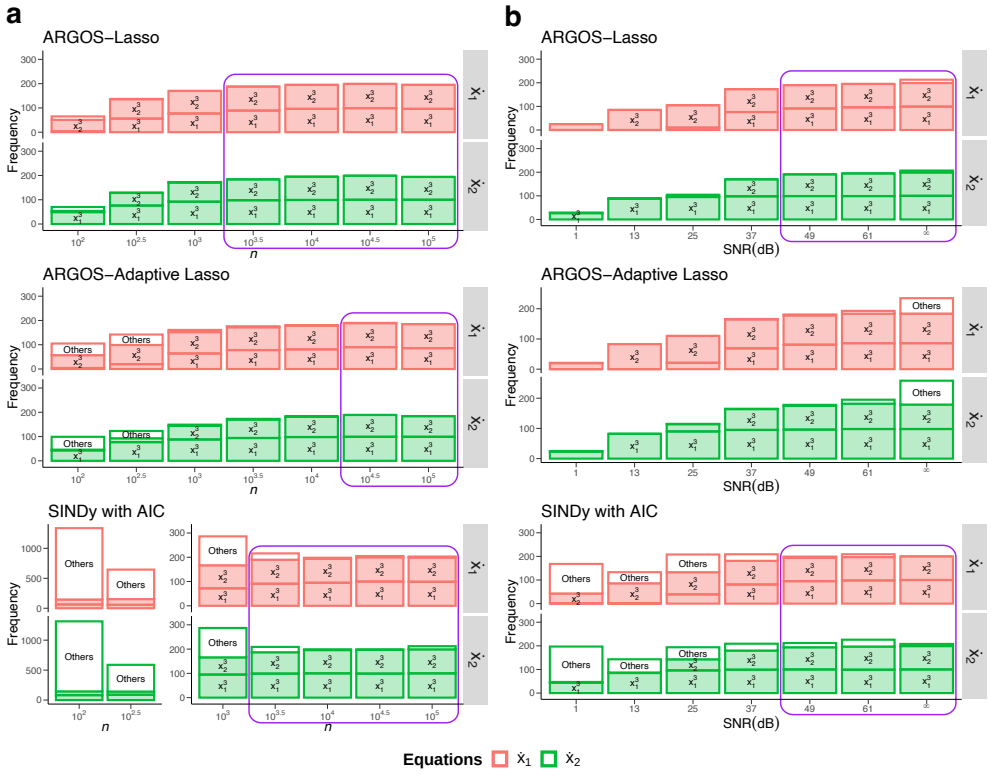

Fig. S3 Frequency of identified variables for the two-dimensional damped oscillator with cubic dynamics across algorithms. Colors correspond to each governing equation; filled boxes indicate correctly identified variables, while white boxes denote erroneous terms. Panels show the frequency of identified variables for data sets with (**a**) increasing time-series length n (signal-to-noise ratio (SNR) = 49 dB), and (**b**) SNR ($n = 5000$). Purple-bordered regions demarcate model discovery above 80%.

where $\alpha = 1$ represents the prey birth rate and $\delta = -1$ is the predator death rate, and $\zeta = -1$ and $\gamma = 1$ are the interaction parameters [3]. Since the population cannot be negative, we used a random uniform distribution with 100 positive values between $[1, 10]$ for the initial conditions of both $x_1(t)$ and $x_2(t)$.

Figure S4 illustrates that, as n and SNR increased, we most consistently identified the true governing terms of the equations using the adaptive lasso within our framework. In contrast, SINDy with AIC tended to discover numerous erroneous terms when data contained fewer than 3000 observations and low SNR.

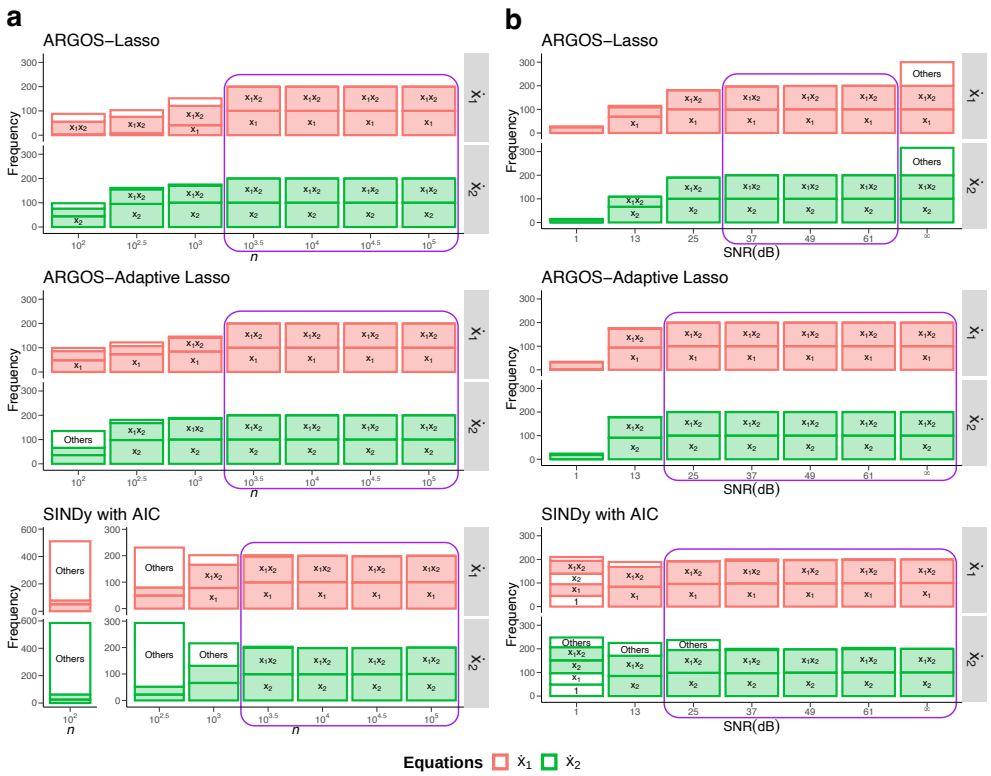

Fig. S4 Frequency of identified variables for the Lotka-Volterra system across algorithms. Colors correspond to each governing equation; filled boxes indicate correctly identified variables, while white boxes denote erroneous terms. Panels show the frequency of identified variables for data sets with (**a**) increasing time-series length n (signal-to-noise ratio (SNR) = 49 dB), and (**b**) SNR ($n = 5000$). Purple-bordered regions demarcate model discovery above 80%.

C. Rossler system

We examined the Rossler system, a three-dimensional chaotic system represented as

$$\begin{aligned} \dot{x}_1 &= -x_2 - x_3, \\ \dot{x}_2 &= x_1 + a x_2, \\ \dot{x}_3 &= b + x_3(x_1 - c), \end{aligned} \tag{S5}$$

where $a = 0.2$, $b = 0.2$, and $c = 5.7$ [4]. For $x_1(t)$, $x_2(t)$, and $x_3(t)$, we generated a random uniform distribution containing 100 values between $[-10, 10]$, $[-10, 10]$, and $[0, 20]$.

Figure S5 demonstrates the effectiveness of our approach in accurately representing the Rossler system, provided that sufficient data is available. Here, our method consistently identified the underlying dynamics while SINDy with AIC failed to surpass 80% success for any SNR value, emphasizing the limitations of the sequential thresholding procedure.

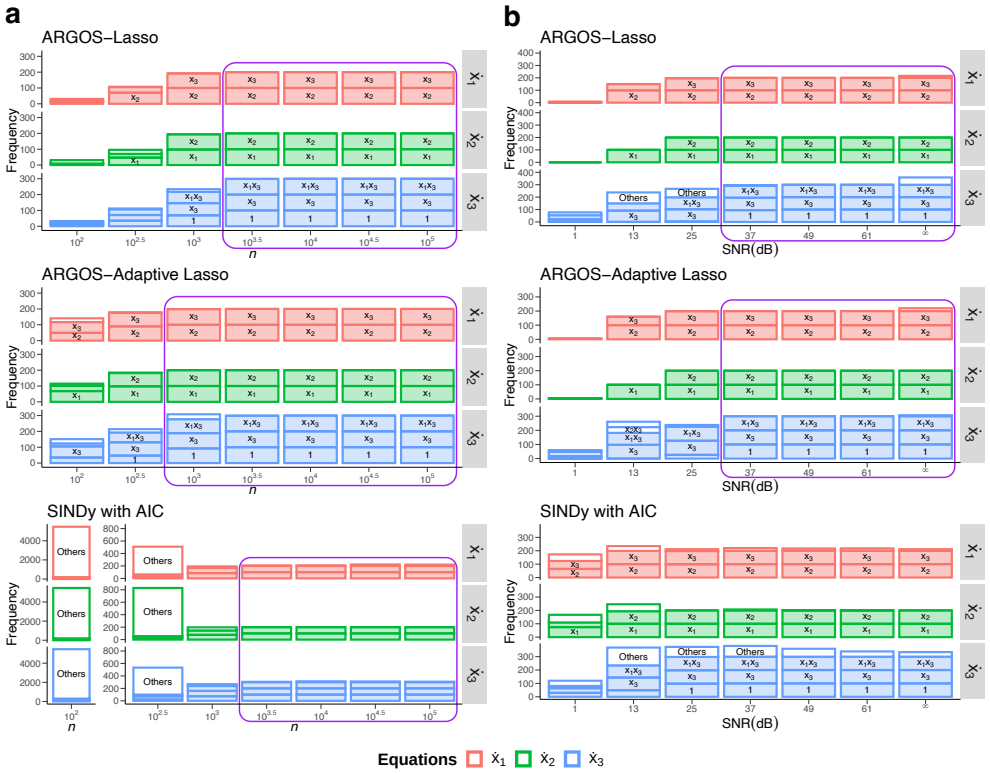

Fig. S5 Frequency of identified variables for the Rossler system across algorithms. Colors correspond to each governing equation; filled boxes indicate correctly identified variables, while white boxes denote erroneous terms. Panels show the frequency of identified variables for data sets with (**a**) increasing time-series length n (signal-to-noise ratio (SNR) = 49 dB), and (**b**) SNR ($n = 5000$). Purple-bordered regions demarcate model discovery above 80%.

D. Lorenz system

We examined the Lorenz chaotic system, a low-dimensional nonlinear structure originally a simple model for atmospheric convection. The Lorenz systems are modeled using the following equations:

$$\begin{aligned} \dot{x}_1 &= \sigma(x_2 - x_1), \\ \dot{x}_2 &= x_1(\rho - x_3) - x_2, \\ \dot{x}_3 &= x_1 x_2 - \zeta x_3, \end{aligned} \tag{S6}$$

with the values of the original parameters $\sigma = 10$, $\rho = 28$, and $\zeta = 8/3$ [1]. For $x_1(t)$, $x_2(t)$, and $x_3(t)$, we developed a random uniform distribution containing 100 values between $[-15, 15]$, $[-15, 15]$, and $[10, 40]$. The Results section provides more detail regarding each method's performance in discovering the Lorenz system.

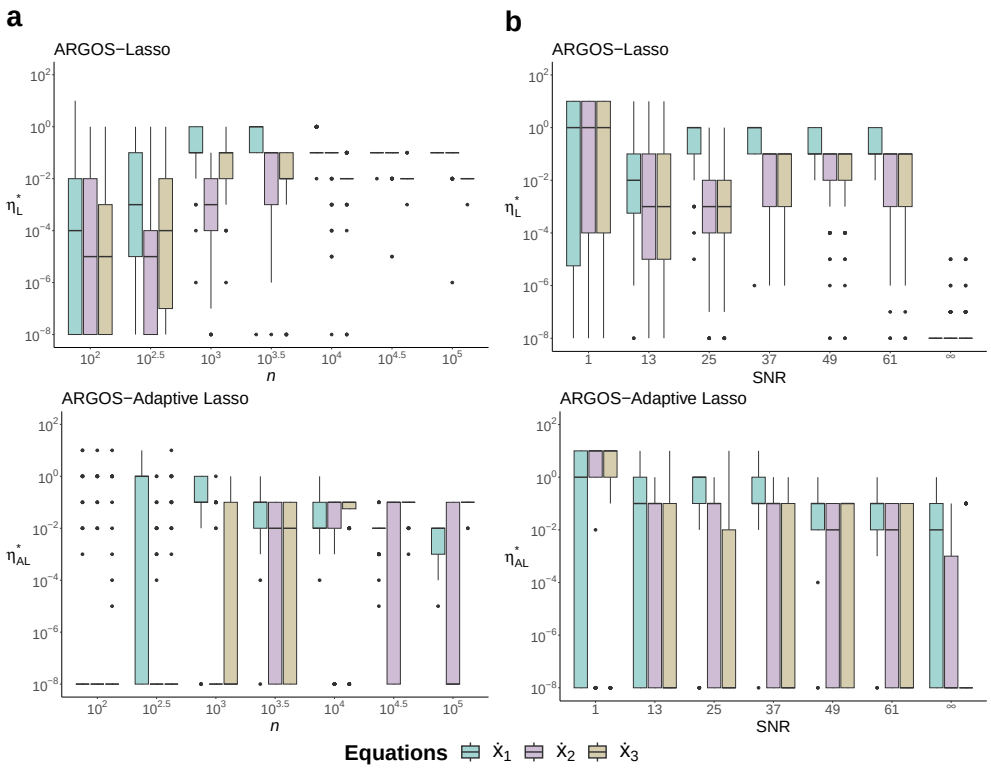

Fig. S6 Distribution of the Optimal Thresholding Parameter η^* Under Various Conditions for the Lorenz system. Boxplots display distributions of η^* from 100 initial conditions, optimized using a BIC-selected grid of thresholds. Whiskers extending from each box show 1.5 times the interquartile range. Data points beyond the end of the whiskers are outlying points. Colors denote governing equations for data sets with increasing (**a**) time-series length n (signal-to-noise ratio (SNR) = 49 dB), and (**b**) SNR ($n = 5000$).

E. Van der Pol oscillator

We examined the Van der Pol oscillator, introduced in 1922 as a nonlinear circuit model with a triode tube, represented as

$$\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= \mu(1 - x_1^2)x_2 - x_1, \end{aligned} \tag{S7}$$

where $\mu = 1.2$ controls the nonlinear damping level of the system [5]. For $x_1(t)$ and $x_2(t)$, we developed a random uniform distribution containing 100 values between $[-4, 4]$.

Figure S8 further illustrates the tendencies of each algorithm when faced with a limited number of observations and low SNR. Under these conditions, our method developed overly sparse models, while SINDy with AIC produced dense models that did not accurately represent the Van der Pol oscillator. However, as n and SNR

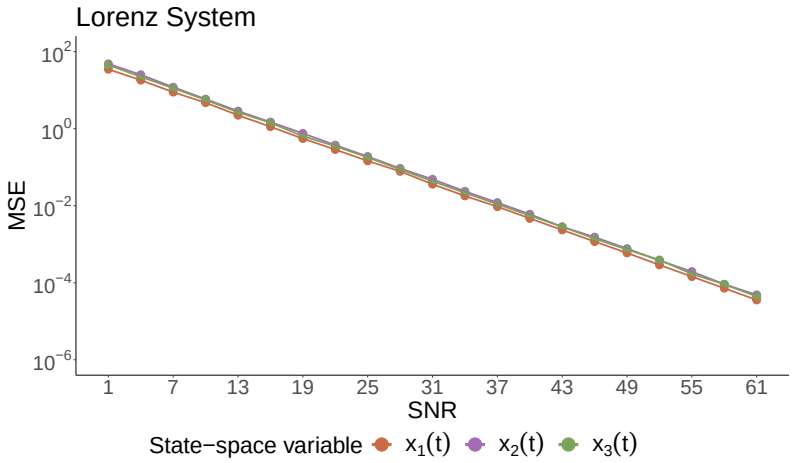

Fig. S7 Mean squared error (MSE) between each Savitzky-Golay smoothed $SG(\tilde{\mathbf{x}}_j)$ and original $\tilde{\mathbf{x}}_j$ for a single instance of the Lorenz system with time-series length $n = 5000$ as the signal-to-noise ratio (SNR) increases. The lines show exponential fits.

increased, our approach demonstrated a marked improvement in accurately discovering the underlying equations of the oscillator.

F. Duffing oscillator

We examined the Duffing oscillator as an alternative cubic nonlinear system that can represent chaos. The Duffing oscillator models a spring-damper-mass system that contains a spring with a restoring force of $f(\zeta) = -\kappa\zeta - \epsilon\zeta^3$, where $\epsilon > 0$ represents a hard spring [5]. However, when $\epsilon < 0$, it represents a soft spring and is given by

$$\ddot{\zeta}_1 + \gamma\dot{\zeta} + (\kappa + \epsilon\zeta^2)\zeta = 0. \tag{S8}$$

We converted $x = \zeta$ and $y = \dot{\zeta}$ and transformed Eq. (S8) to

$$\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= -\gamma x_2 - \kappa x_1 - \epsilon x_1^3. \end{aligned} \tag{S9}$$

Here, we used parameter values for which the Duffing oscillator does not represent chaotic behavior: $\kappa = 1$, $\gamma = 0.1$, and $\epsilon = 5$ [5]. For $x_1(t)$ and $x_2(t)$, we developed a random uniform distribution containing 100 values between $[-2, 2]$, $[-6, 6]$.

Figure S9 shows that our method consistently represented the Duffing oscillator with high accuracy as n and SNR increased. In this example, with less than 1000 observations and low SNR, our approach developed overly sparse models that inadequately captured the dynamics of the system, while SINDy with AIC again developed dense models that misrepresented the dynamics.

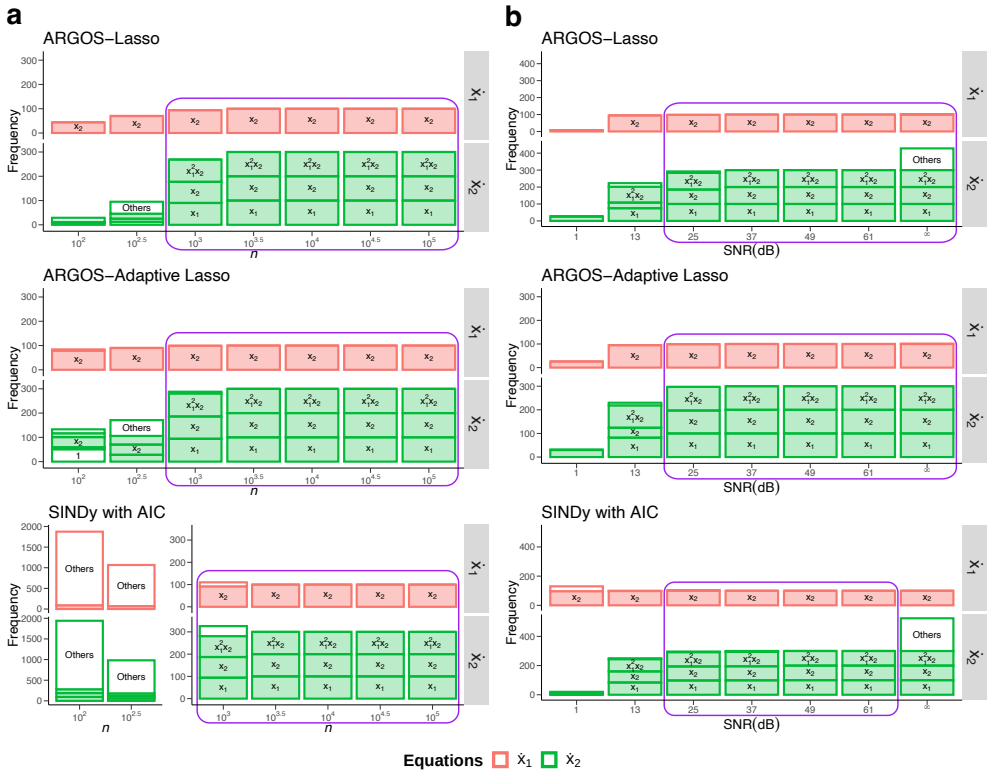

Fig. S8 Frequency of identified variables for the Van der Pol oscillator across algorithms. Colors correspond to each governing equation; filled boxes indicate correctly identified variables, while white boxes denote erroneous terms. Panels show the frequency of identified variables for data sets with (**a**) increasing time-series length n (signal-to-noise ratio (SNR) = 49 dB), and (**b**) SNR ($n = 5000$). Purple-bordered regions demarcate model discovery above 80%.

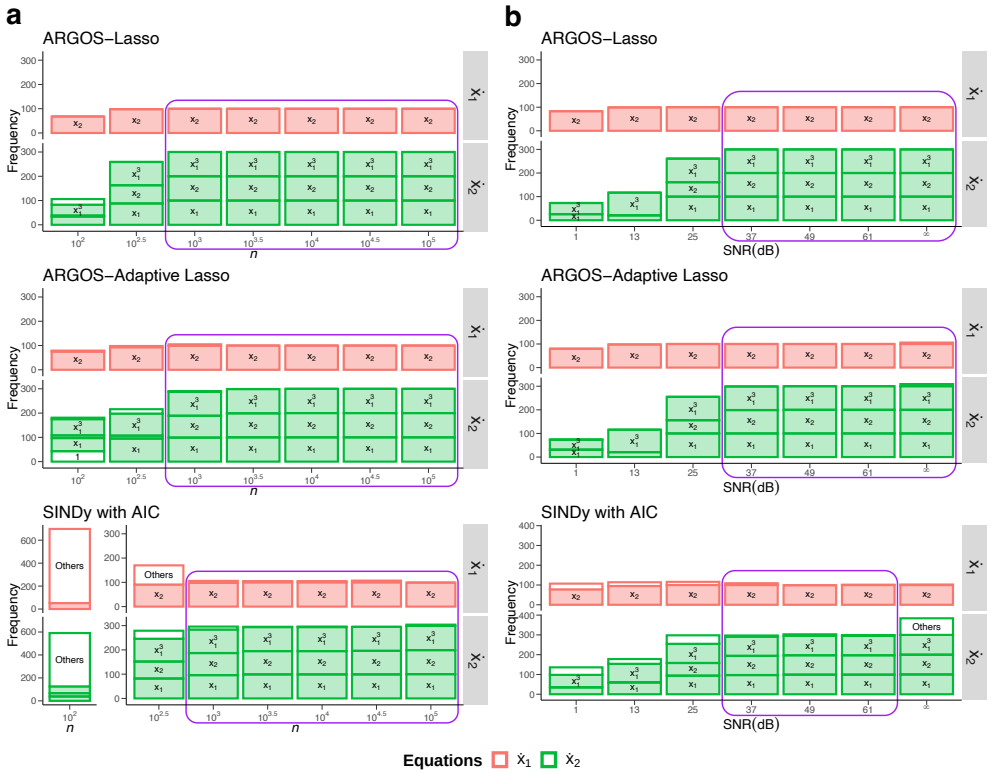

Fig. S9 Frequency of identified variables for the Duffing oscillator across algorithms. Colors correspond to each governing equation; filled boxes indicate correctly identified variables, while white boxes denote erroneous terms. Panels show the frequency of identified variables for data sets with (**a**) increasing time-series length n (signal-to-noise ratio (SNR) = 49 dB), and (**b**) SNR ($n = 5000$). Purple-bordered regions demarcate model discovery above 80%.

Supplementary Note 4. Interplay between Data Size, Dictionary Matrix, and Noise Level

The results highlight an intricate relationship between the number of given data, the size of the dictionary matrix, and the noise level. Theoretically, as the number of observations increases, our method can handle more variables in the design matrix while still successfully identifying the underlying system. However, our results show that, with minimal observations and a large number of predictors, our method cannot uncover the underlying system.

Regarding noise levels, our observations suggest that, as expected, increased noise tends to obscure the advantages brought by a larger data set, and this problem becomes more challenging with more variables in the design matrix. However, with sufficient data and low to medium noise levels, our method automatically discovers the underlying equations from data, as shown in Fig. 2 of the main text.

Supplementary References

[1] Brunton, S.L., Proctor, J.L. & Kutz, J.N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences* **113**, 3932–3937 (2016).

[2] Lotka, A.J. Contribution to the theory of periodic reactions. *The Journal of Physical Chemistry* **14**, 271–274 (1910).

[3] Naozuka, G.T., Rocha, H.L., Silva, R.S. & Almeida, R.C. SINDy-SA framework: Enhancing nonlinear system identification with sensitivity analysis. *Nonlinear Dynamics* **110**, 2589–2609 (2022).

[4] Tran, G. & Ward, R. Exact recovery of chaotic systems from highly corrupted data. *Multiscale Modeling & Simulation* **15**, 1108–1129 (2017).

[5] Cortiella, A., Park, K.-C. & Doostan, A. Sparse identification of nonlinear dynamical systems via reweighted ℓ_1-regularized least squares. *Computer Methods in Applied Mechanics and Engineering* **376**, 113620 (2021).