Journal of Nonlinear Modeling and Analysis http://jnma-online.com
Volume 6, Number 2, June 2024, 333–359 DOI:10.12150/jnma.2024.333

Model Selection of Dnamical Systems via Entropic Regression and Bayesian Information Criteria*

Jinhui Li[1] and Aiyong Chen[2,†]

Abstract Recovering system model from noisy data is a key challenge in the analysis of dynamical systems. Based on a data-driven identification approach, we develop a model selection algorithm called Entropy Regression Bayesian Information Criterion (ER-BIC). First, the entropy regression identification algorithm (ER) is used to obtain candidate models that are close to the Pareto optimum and combine as a library of candidate models. Second, BIC score in the candidate models library is calculated using the Bayesian information criterion (BIC) and ranked from smallest to largest. Third, the model with the smallest BIC score is selected as the one we need to optimize. Finally, the ER-BIC algorithm is applied to several classical dynamical systems, including one-dimensional polynomial and RC circuit systems, two-dimensional Duffing and classical ODE systems, three-dimensional Lorenz 63 and Lorenz 84 systems. The results show that the new algorithm accurately identifies the system model under noise and time variable t, laying the foundation for nonlinear analysis.

Keywords Data-driven, system identification, model selection, ER algorithm, BIC

MSC(2010) 37M05, 34A34, 93E24.

1. Introduction

Nonlinear dynamical systems have been one of the hot topics of fundamental theoretical research in recent years, both at home and abroad. the usual dynamical systems include Lorenz 84 system [5, 12, 34, 35], Lorenz 63 system [11, 21], Duffing system [23, 28, 29], RC circuit system [15] with trigonometric term, etc. The task that follows is to recover the periodic dynamical equations in order to study their properties. Thus, the data-driven approach [16, 24] that has emerged in recent years–obtaining dynamical behavior from data–can be an excellent solution to this problem.

For data-driven methods, common approaches include dynamic mode decomposition (DMD) [6, 9, 32, 33], Koopman theory [8, 17, 19], sparse representation, etc. Considering that most dynamical systems only contain a few terms, Brunton et

†the corresponding author.
Email address: lijinhuixlq@163.com(J. Li), aiyongchen@163.com(A. Chen)
[1]School of Mathematics and Computing Science, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China
[2]Department of Mathematics, Hunan First Normal University, Changsha, Hunan 410205, China
*The authors were supported by Natural Science Foundation of Hunan Province (No. 2021JJ30166).

al. [7] proposed a nonlinear sparse regression algorithm (SINDy) in 2016. This algorithm obtains the sparse matrix based on the sequential least squares method and obtains the system model from. AlRahman et al. [2] proposed a new information-theoretic regression method (called entropy regression (ER)) to infer the sparse structure of the general model and its parameters based on the cost function of the metric to determine the error quantification that makes the existing methods fail under large noise and outliers. The ER method is robust in the presence of noise and outliers in the data. The ER algorithm [2] emphasizes information relevant to selecting models according to the entropy standard without models. The underlying terms will be incorporated into the model simply because they are correlated and thus together constitute the sparsest model.

For parameter estimation problems, most scholars choose the likelihood function as the objective function in their study, that is, the best model fit is achieved by maximizing the likelihood function. In this case, making a judgment requires the use of a criterion that can balance the accuracy and complexity of the model. The AIC criterion and the BIC criterion are commonly used [18]. Japanese statistician Akaike [1] proposed the AIC criterion for model selection in 1974. Schwarz [31] proposed the Bayesian information criterion (BIC) in 1978. It is also used for model selection and normally selects the model with the lowest BIC value. Thus AIC can be considered when the objective of the study is predictive and BIC can be considered when the objective of the study is descriptive [4].

Mangan et al. [26] proposed a new mathematical framework to evaluate a large set of candidate models for combinatorics using Akaike information criteria (AIC) with sparse regression for model selection. Sparse regression for nonlinear system identification was first performed by SINDy, and model selection was then performed via an information criterion. SINDy-AIC [26] sparsely selects dynamic models from information theoretic criteria, ranks the candidate models, and further shows that the correct model is strongly supported by the AIC score. Moreover, it utilizes data to uncover governing equations that are data-driven among the candidate models, where the data has undergone cross-validation and ranking.

In this work, we propose a model selection algorithm (ER-BIC algorithm) based on entropy regression (ER) and BIC, which are similar to SINDy-AIC [26] algorithm. First, we use the ER algorithm to obtain near-Pareto optimal models for nonlinear systems with a relatively small number of models. Second, the BIC scores of a limited number of the obtained models are calculated and sorted from small to large. Finally, the model with the lowest BIC score is chosen as the optimal model for the system. Our algorithm is robust to identify nonlinear systems, and can select a smaller number of models near Pareto optimality than the SINDy-AIC algorithm, but with higher accuracy. In particular, when the trigonometric function term is included, the model accuracy of the ER-BIC algorithm is higher than that of the SINDy-AIC algorithm.

The paper is divided into the following subsections. In section 2, we introduce the framework of the ER-BIC algorithm, where the theory of the ER algorithm, the notion of the BIC score, and the specific algorithmic procedure are presented. In section 3, some key applications and results of the ER-BIC algorithm are presented. The algorithm is applied to ordinary dynamical system models and shows robustness even under noisy data. In Section 4, we consider the accuracy of the Lorenz 63 system at different noise amplitudes and different amounts of data for the ER-BIC algorithm selection. The conclusions and discussion are provided in section 5.

2. ER-BIC framework

2.1. Entropic regression algorithm (ER)

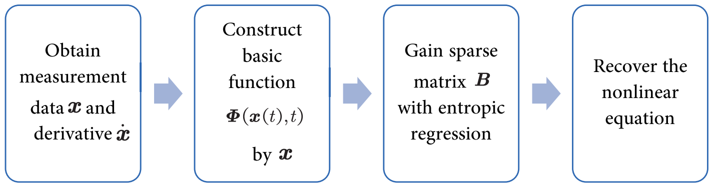


Figure 1. ER algorithm process.

In 2020, AIRahman et al. [2] proposed method to identify nonlinear dynamic system, which is called Entropic Regression (ER). Consider the dynamics equation

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t)), \tag{2.1}$$

where $\boldsymbol{x}(t) \in \boldsymbol{R}^n$ denotes the state variable of the system at time, $\dot{\boldsymbol{x}}(t)$ is the derivative of $\boldsymbol{x}(t)$, and $\boldsymbol{f}$ is the vector field. Each component function $f_i(\boldsymbol{z}(t))$ can be represented as follows:

$$\dot{x}_i(t) = f_i(\boldsymbol{x}(t)) = \sum_{k=0}^{\infty} b_{ik}\phi_k(\boldsymbol{x}(t),t). \tag{2.2}$$

The $\{\phi_k\}_{k=0}^{\infty}$ is basis functions, where

$$\boldsymbol{\Phi}(\boldsymbol{x}(t),t) = \begin{bmatrix} | & | & | & | & | & | & | \\ 1 & \phi_1(\boldsymbol{x}(t),t) & \phi_2(\boldsymbol{x}(t),t) & \ldots & \sin(t) & \ldots & \cos(\omega t) \\ | & | & | & | & | & | & | \end{bmatrix}. \tag{2.3}$$

The specific algorithm process is shown in Figure 1. According to the time series data $\boldsymbol{x}(t) = [x_1(t), x_2(t), \ldots, x_m(t)]^T$, the noise is added by the form of $\boldsymbol{x}(t) = \boldsymbol{x}(t) + \boldsymbol{\eta}(t)$, with $\boldsymbol{\eta}(t)$ representing the noise. Next, using fourth order central difference obtains the derivatives $\dot{\boldsymbol{x}}(t)$:

$$\dot{\boldsymbol{x}}(t-2) = \frac{-\boldsymbol{x}(t+2) + 8\boldsymbol{x}(t+1) - 8\boldsymbol{x}(t-1) + \boldsymbol{x}(t-2)}{12dt}. \tag{2.4}$$

And then, the basis function $\boldsymbol{\Phi}(\boldsymbol{x}(t),t)$ can be obtained from equation (2.3) via the data $\boldsymbol{x}(t)$. Then get the sparse matrix $\boldsymbol{B}$ by the entropic regression, with the procedure shown in the Algorithm 1. In particular, the Algorithm 1 contains two stages: forward ER and backward ER. In both steps, each model is obtained based on a retropy criterion and parameter (tol). We can refer to reference [2] for specific details. Finally, recover the equation of dynamic system via the (2.5).

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{\Phi}(\boldsymbol{x}(t),t)\boldsymbol{B}. \tag{2.5}$$

Algorithm 1 ER algorithm obtain $\boldsymbol{B}$.

Input: Basic function library ($\boldsymbol{\Phi}(\boldsymbol{x}(t), t)$) and derivative ($\dot{\boldsymbol{x}}$), and tolerance library $\{tol_0, tol_1, \ldots, tol_p\}$

Output: sparse matrix ($\boldsymbol{B}$)

```
tol ∈ {tol_1, tol_2, ..., tol_q}
Forward ER: (Φ(x(t), t), ẋ, tol)
val = ∞
while val > tol do
    implement Forward ER
return B_f
Backward ER: (Φ(x(t), t), ẋ, tol, B_f)
val = −∞
while val < tol do
    implement Backward ER
return B_b
return B = B_b
```

2.2. Pareto optimal

Pareto's law is derived from economics and mainly describes the theoretical impossibility of increasing the profit of one party without decreasing the profit of the other. In dynamical systems, the aim of the Pareto is mainly about choosing the optimal model. The most common approach is to obtain a large number of models via an algorithm and then select models using Pareto optimality [3,10]. Among the selected models, not only the one with the lowest error, but it should also be highly interpretable, generalization, and predictive. Figure 2 shows the resulting model and the Pareto-optimal region. Specifically, the models (black dots and red dots) generated by some algorithms (SINDy, ER) can have several models with the same number of terms. For a given number of terms, the model with Pareto boundary is the model with the lowest error (red dots), and the solid line is the approximation of the Pareto boundary. The shaded part of the figure is the Pareto-optimal region, and the model inside is the exact reduced model.

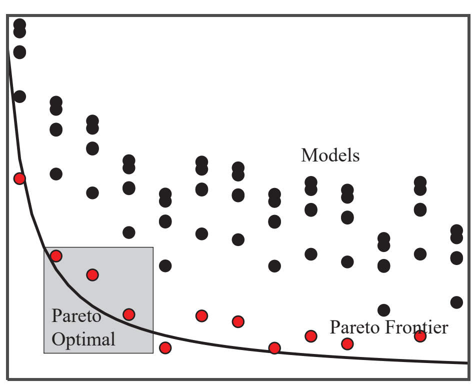


Figure 2. Pareto optimal selection.

The next goal is to select the optimal model from the Pareto optimal solution space according to certain rules. The model selection problem seeks the best balance between the complexity of the model and the ability of the model to describe the dataset (i.e., the likelihood function). Numerous information criteria, such as AIC and BIC, have been introduced to mitigate overfitting issues by incorporating a penalty term for model complexity. A lower value of this penalty term is indicative of a superior model. Commonly used methods are AIC and BIC, see Section 3.2 for details. In this paper, we first use the ER method to obtain the models in the Pareto optimal solution region by different tolerance values, and then use the BIC score ranking to select the model with the smallest score as the optimal model.

2.3. Bayesian information criteria (BIC)

In statistics, Bayesian Information Criterion (BIC) [14] is the criteria for selecting models from a limited set of models. In general, models with lower BIC are preferred. It is partly based on the likelihood function and is closely related to the Akaike Information Criterion (AIC) [30]. One form for calculating the BIC is given by

$$BIC = -2\ln(\hat{\boldsymbol{L}}) + k\ln(N), \tag{2.6}$$

where $\hat{\boldsymbol{L}} = p(\boldsymbol{x}|\hat{\mu})$ is the maximized value of the likelihood function of the model, and $\boldsymbol{x}$ is the observed data, where $\hat{\mu}$ is the parameter values that maximize the likelihood function. The k is the number of parameters estimated by the model and N is the number of data points in $\boldsymbol{x}$, the number of observations, or equivalently, the sample size.

Note that a penalty of $k\ln(N)$ imposes a lower bound on the relative BIC score. Similar to AIC, when training a model, increasing the number of parameters, that is, increasing the complexity of the model, enhances the likelihood function but also leads to overfitting. For this problem, AIC and BIC introduce a penalty term that depends on the number of model parameters. The penalty term of the BIC is larger than that of the AIC. Considering the number of samples effectively prevents the model complexity from becoming too large due to the over-accuracy of the model when the number of samples is too large. In practice, the BIC requires finite sample size corrections:

$$BIC_c = BIC + \frac{2k(k+1)}{N-k-1}. \tag{2.7}$$

In this work, BIC can be expressed as

$$BIC = N\ln(\frac{SSR}{N}) + k\ln(N), \tag{2.8}$$

where SSR is the sum of squared residuals, $SSR = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2$, and $\hat{y}_i$ is estimated value of the model.

2.4. ER-BIC algorithm

The main algorithm of this paper is a combination of the entropy regression method for dynamical system identification with model selection based on the BIC information criterion, which is called ER-BIC algorithm. This method is similar to the

SINDy-AIC [26], so we can identify a best-fit model from the entire model space. The specific ER-BIC algorithm process is shown in the Figure 3. First, different models with different tolerance values can be obtained using the ER algorithm. The BIC scores of the obtained models are then computed and the most informative model is ranked, which automatically selects the best model. Algorithm 2 uses BIC as our information criterion and is executed for model selection.

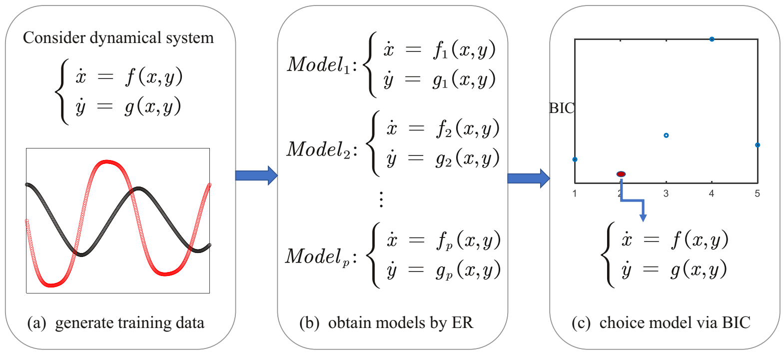


Figure 3. The ER-BIC algorithm process for system: (a) generate measurement data, (b) obtain the model library via ER algorithm, (c) choice model by the BIC score, with the lowest score.

Algorithm 2 ER-BIC.

Input: Basic function library ($\mathbf{\Phi}(\boldsymbol{x}(t), t)$), derivative ($\dot{\boldsymbol{x}}$), validation data ($\boldsymbol{y}$) and tolerance library $\{tol_0, tol_1, \ldots, tol_p\}$

Output: Best model

```
ER-BIC(Φ(x(t), t), ẋ, y)
for tol(i) ∈ {tol_1, tol_2, ..., tol_p} do
    Model(i) ← ER(Φ(x(t), t), ẋ, tol_i)
    ẋ_i ← Model_i
    I_i ← BIC(y, ẋ_i)
[ind,val] ← sort(I)
return model(ind(1))
```

The specific details of the ER-BIC algorithm can be learned from Algorithm 2. Firstly, a series of dynamic models are obtained by the ER algorithm at different tolerance values. We get all possible models using random permutation (for 1D system), or a minority of models via ER algorithm (for all system) under the different parameters (tol_i). Secondly, use the obtained model to reconstruct the derivative $\dot{\boldsymbol{x}}_i$. Finally, the BIC scores are calculated using the validation data and the derived quantities, sorted from small to large. Thus we can select the model with the lowest BIC score (i.e., the optimal model).

To be more specific, if the number of dimensions is relatively tiny (e.g. a one-dimensional polynomial model), we consider computing all possible models and selecting the optimal one by the BIC score. We also use the ER algorithm to select the model near the Pareto optimal, where the number of models is much smaller than the number of all models and the accuracy is high. So, for the one-dimensional polynomial system, we compute all possible and few models via ER. For other systems, we only consider few models by ER algorithm.

3. Application and result

In this section, we choose the best model of ordinary differential equations (ODEs) of nonlinear dynamical systems via ER-BIC method. The work consists of the following steps. Firstly, acquire the measurement data $\boldsymbol{x}$ utilizing the fourth-order Runge-Kutta method. Subsequently, introduce noise into the data and compute the derivative $\dot{\boldsymbol{x}}$ using the appropriate equation (2.4).

Secondly, we can receive the model library via the ER algorithm under different tolerance values $tol(i) \in \{tol_1, tol_2, \ldots, tol_p\}$. Specifically, at each tolerance value, we can obtain a library of candidate functions ($\boldsymbol{\Phi(x(t), t)}$) by noisy data using (3.1). Then, following the steps of Algorithm 1, model bases ($Model_i \in \{Model_1, Model_2, \ldots, Model_p\}$) for different model combinations are obtained via (3.2) with different tolerance values.

$$\boldsymbol{\Phi}(\boldsymbol{x}(t), t) = \begin{bmatrix} | & | & | & & | & & | \\ 1 & \phi_1(\boldsymbol{x}(t), t) & \phi_2(\boldsymbol{x}(t), t) & \ldots & \sin(x) & \ldots & \cos(\omega t) \\ | & | & | & & | & & | \end{bmatrix}. \tag{3.1}$$

$$Model_i = \boldsymbol{\Phi}(\boldsymbol{x}(t), t)\boldsymbol{B}_i. \tag{3.2}$$

Thirdly, the BIC scores of each model are computed and ranked using the Algorithm 2, and the model with the lowest score is the optimal model. In detail, derivatives are obtained by simulation using the model ($Model_i$) obtained by the ER algorithm, and then the BIC scores are computed using the BIC formula (2.8) with validation data ($\boldsymbol{y}$) and ranked. Finally, among the BIC scores sorted, the model with the lowest score (i.e. the best) is the model we need. And compared with the SINDy-AIC method, it is found that ER-BIC is more effective in identifying systems and takes less time.

The considered nonlinear dynamical systems are as follows: one-dimensional simple polynomial system, RC circuit system, two-dimensional Duffing system and classical ODE system, three-dimensional Lorenz 63 system and Lorenz 84 system. The RC circuit system, the classical ODE system and the Lorenz 84 system contain trigonometric terms.

3.1. 1D polynomial model

In the first example, we consider the one-dimensional simple polynomial system with two terms,

$$\dot{x} = x^2 - x^4. \tag{3.3}$$

The ER-BIC method is extended to consider a one-dimensional polynomial model. First, we simulated the system from $t = 0$ to $t = 5$ with a time-step of $dt = 0.25$ under the three initial values $x_0 = [0.2, 0.8, 2]$ to obtain three time series using the fourth-order Runge-Kutta method. Adding noise to each state variable to obtain the noise data $\boldsymbol{x}(t)$, we gain the derivative $\dot{\boldsymbol{x}}$ by the (2.4). The second step involves obtaining the basic function $\boldsymbol{\Phi}(\boldsymbol{x}(t), t)$ from noisy data $\boldsymbol{x}(t)$. We assume that the right-hand side of (3.3) has an extension of a degree $d = 5$, which is exactly described by 2 of the $p = 6$ monomials. In this case, since this system is one-dimensional, we compute all possible models, which are 63 possible models. However, as the number of dimensions and polynomials increases, there are a large number of possible models that are not favorable for each analysis. Therefore, we resort to the ER-BIC algorithm to select a small number of models near the Pareto optimum.

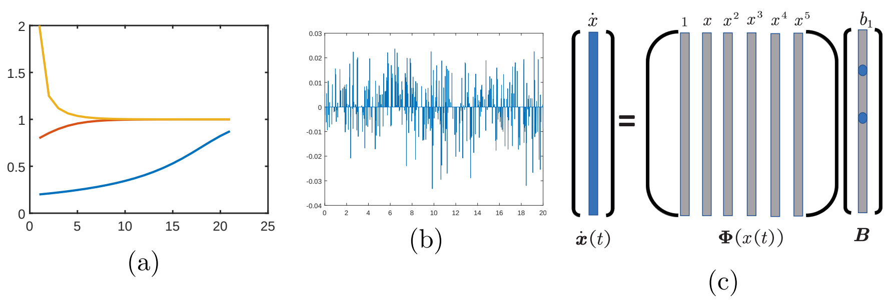


Figure 4. Recover the 1D polynomial model using the ER-BIC method. (a) Noise-free time-series; (b) Noise level; (c) The form of 1D polynomial system identification. The training data consists of (a) and (b).

The training data and the form of one-dimensional polynomial system are shown in Figure 4, where (a) shows the time-series of training data with noise-free, which contains a time series of three sets of test data. (b) displays the noise data with $\eta = 0.01$. (c) indicates the form of the system equation (3.3).

Figure 5 shows the model selection results of the one-dimensional polynomial system using ER-BIC algorithm. From Figure 5 (a) and (b), we can see that only 5 models are recovered from one-dimensional polynomial system via ER-BIC method, similar to SINDy-AIC. The BIC score in the figure shows that the model at the lowest point is the model we need (optimal), that is, the model with only two items has the lowest score. Figure 5 (c) demonstrates all possible models (which have 63 models, black dot) and identification using ER-BIC algorithm (red dot), (d) displays the identification of model by SINDy-AIC method (red dot). From (c) and (d), we conclude the following: (i) we are able to identify only two terms of one-dimensional systems, consistent with the expression of the system equation; (ii) ER-BIC identifies two fewer models than SINDy-AIC algorithm, and the models identified by SINDy are scattered.

From Table 1, we can know that the $Model_1$ is the best model via the ER-BIC algorithm, which can recover the model as follows:

$$\dot{x}(t) = 1.003373x^2 - 1.002969x^4.$$

Table 1 shows the three models via ER-BIC method, and selects the first model as the result based on the BIC scores. Obviously, using ER-BIC algorithm is a smaller number of identification models with SINDy-AIC.

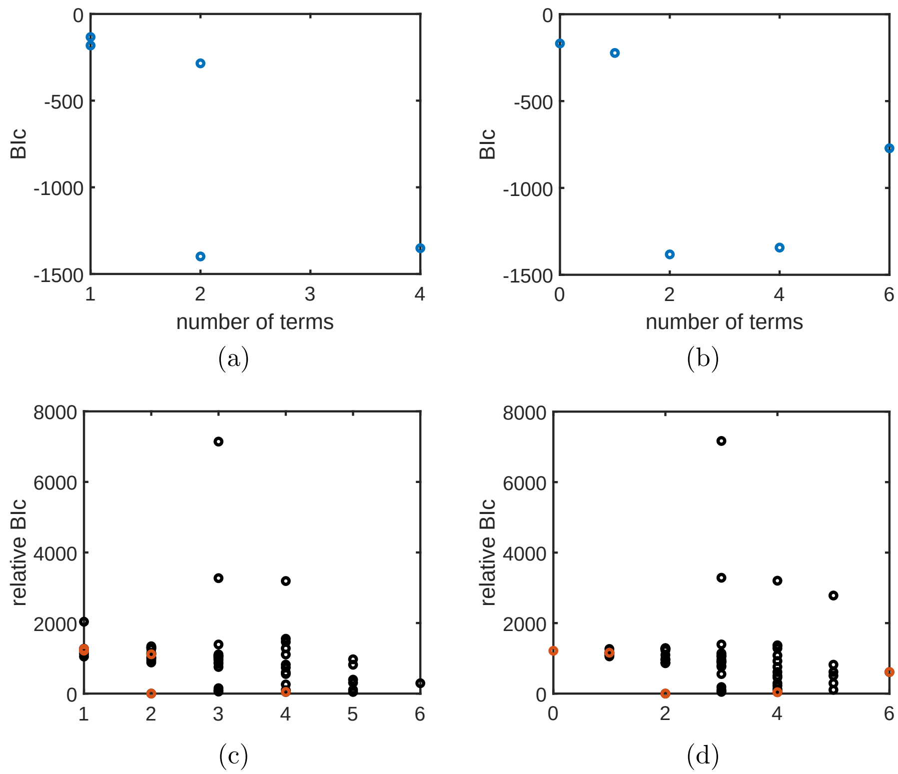


Figure 5. The model choice result of one-dimensional polynomial system by the ER-BIC algorithm: (a) shows the model results identified by ER-BIC, and the number of points represents the number of models; (b) displays the model results identified by SINDy-AIC. The optimal model is the point with the lowest BIC score, and the optimal number of items in the graph is two; (c) demonstrates the model of identification, where black dot denotes all the models and red dot indicates model via ER-BIC; and (d) shows the red dot indicates model by SINDy-AIC.

Table 1. 1D polynomial model selection result of the ER-BIC algorithm.

	1	x	x^2	x^3	x^4	x^5
$Model_1$	0	0	1.003373	0	-1.002969	0
$Model_2$	0	0.0061	0.9718	0.0280	-1.0055	0
$Model_3$	0	0	6.0560	-6.015	0	0
$Model_4$	0	0	-22.6449	0	0	0
$Model_5$	0	0	0	-4.7917	0	0

3.2. 1D RC circuit

In this case, we consider a simple system, RC circuit [13] with the 1-periodic forcing term. The RC circuit is modeled in a dimensionless form via non-autonomous ordinary differential equations.

$$\dot{x} = \sin(2\pi t) - x, \tag{3.4}$$

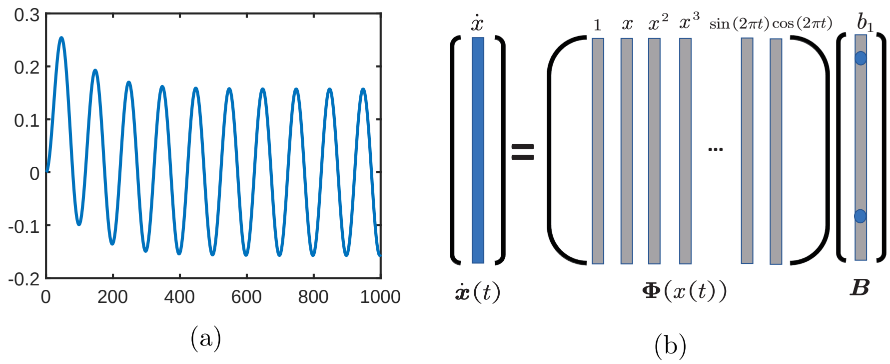


Figure 6. Restore the RC circuit model by the ER-BIC method: (a) training time series, which together with Figure 2 (b) comprises the experimental data, (b) form of the RC circuit equation.

Table 2. RC model selection result of the ER-BIC algorithm.

	1	x	x^2	x^3	$\sin(2\pi t)$	$\cos(2\pi t)$
$Model_1$	0	-1.000284	0	0	1.000452	0
$Model_2$	0	0	0	0	0.972527	0

where the initial condition is $x_0 = 0$. We simulate the RC circuit model from $t = 0.01$ to $t = 15$ with the time-step $dt = 0.01$ to obtain the data and add noise. Note that the derivative $\dot{\boldsymbol{x}}$ has the trigonometric term $\sin(2\pi t)$ in the state variable. Assuming a degree $d = 3$ and the trigonometric term denotes the right-hand-side of (3.4). We also compute all possible models based on the one-dimension, few models via the ER-BIC and SINDy-AIC. Figure 6 shows the time series of the RC system and the form of the equation of (3.4).

In this example, we compare ER-BIC with SINDy-AIC algorithm proposed in [26]. We recover the models via ER-BIC method based on different parameters tol, which are from 10^{-6} to 10. The model selection of the RC circuit system is shown in Figure 7. Figure (a) and (b) display the number of models via the ER-BIC with SINDy-AIC algorithms. We can see that, the ER-BIC has only 2 models of selection and the SINDy-AIC has 6 models. And we can also know that the 2 terms of the model have the lowest BIC score. Figure 7 (c) and (d) demonstrate all possible models (which have 63 models, black dot) and models of two algorithms (red dot). Comparing the results of the two algorithms, it is found that the ER algorithm identifies fewer models and the results are as accurate as SINDy.

Table 2 shows the models from the ER-BIC algorithm, where we can know that $Model_1$ is an accurate model for the RC system. According to the table, we can recover the RC circuit system as

$$\dot{x}(t) = -1.000284x + 1.000452\sin(2\pi t),$$

where compared with the original equation, the error is less than 0.05%.

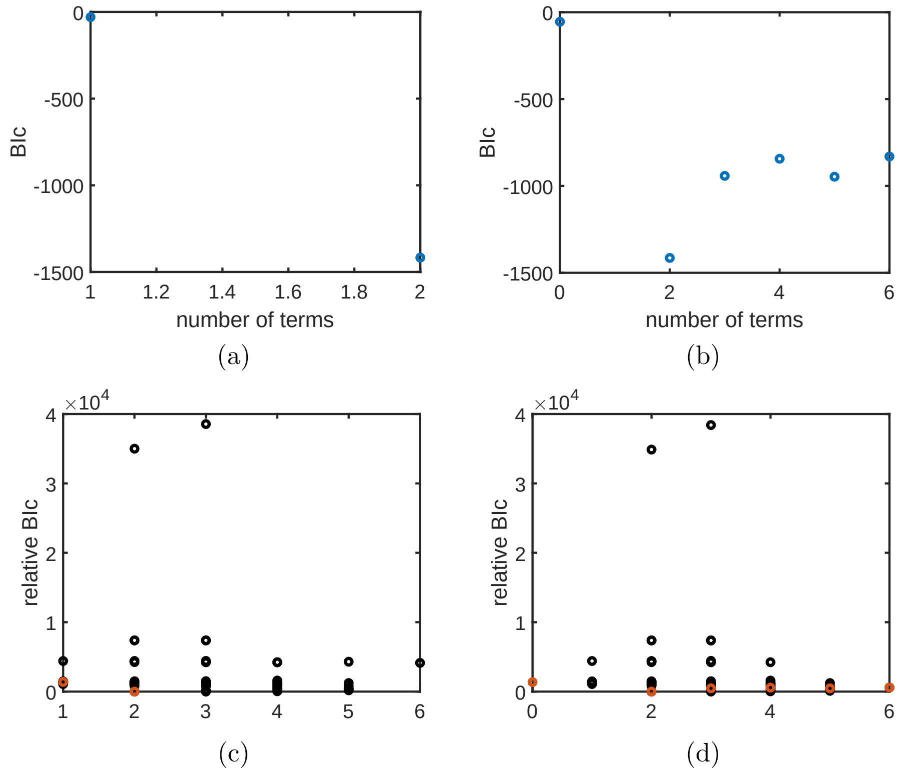


Figure 7. The model selection results of RC circuit system via the ER-BIC algorithm: (a) shows the time series; (b) demonstrates the few models of RC circuit by ER-BIC identification, and the points represents the number of models; (c) displays the model of identification via ER-BIC algorithm; (d) shows the dot indicates model by SINDy-AIC method.

3.3. 2D Duffing system

The motion of the system is constantly influenced by damping, with the damping force primarily depending on the velocity. Considering this damping effect, the equations of motion of an elastic system with a symmetric potential under periodic forces can be formulated as follows:

$$\ddot{x} + a\dot{x} + kx + \mu x^3 = 0, \tag{3.5}$$

which is the famous Duffing equation [27], and the Duffing equation is a representative differential equation commonly used in nonlinear theory, although it is a nonlinear oscillator model derived from a simple physical model.

We set $\dot{x} = y$, giving

$$\begin{cases} \dot{x} = y, \\ \dot{y} = -kx - ay - \mu x^3, \end{cases} \tag{3.6}$$

where the parameter values are set to $k = 1, a = 0.1, \mu = 5$, and the initial condition is $x_0 = [1, 0]^T$. Using (3.6), we determine that the variable n is equal to 2, and we set the degree of the polynomial basis to be $d = 5$. This results in a total of $p = 21$

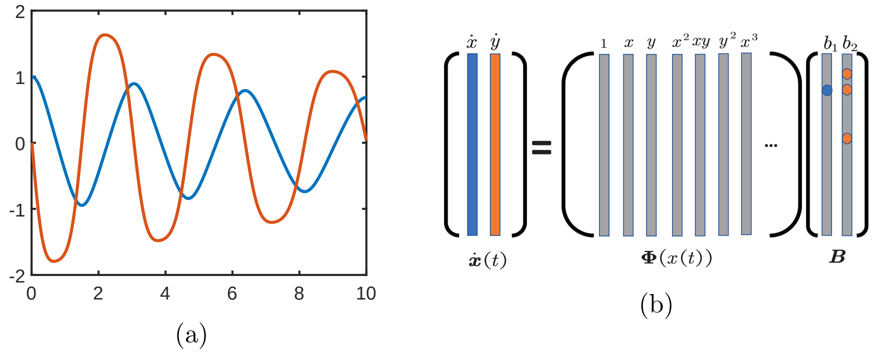


Figure 8. Recover model of the 2D Duffing system: (a) time-series of data, and add noise obtain train data, (b) shows the format of the Duffing equation (3.5).

Table 3. Duffing model choice result of the ER-BIC algorithm.

	$Model_1$		$Model_2$		$Model_3$		$Model_4$	
	$\dot{x}$	$\dot{y}$	$\dot{x}$	$\dot{y}$	$\dot{x}$	$\dot{y}$	$\dot{x}$	$\dot{y}$
1	0	0	0	0	0	0	0	0
x	0	-1.00007	0	1.14741	0	0	0	0
y	0.99997	-0.10014	0.99997	1.98214	0.99997	0	0.99997	0
x^2	0	0	0	0	0	0	0	0
xy	0	0	0	0	0	0	0	0
y^2	0	0	0	-4.18472	0	0	0	0
x^3	0	-4.99917	0	-2.07659	0	-6.74029	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

terms. In this example, we obtain the training data from $t = 0.01$ to $t = 25$ with time series $dt = 0.01$ using the fourth-order Runge-Kutta method. The time series of Duffing system and the form of equation (3.6) are shown in Figure 8.

In this case, Table 3 shows the identification of Duffing model, where $Model_1$ is the best model for the results. The $model_1$ recovers the equation, which is given by

$$\begin{cases} \dot{x} = 0.99997y, \\ \dot{y} = -1.00007x - 0.10014y - 4.99917x^3. \end{cases}$$

Finally, comparison between ER-BIC and SINDy-AIC is shown in Figure 9. Similar to the previous case, ER-BIC yields fewer models in the identification. Figure 9 (a) displays the Duffing system from $t = 0.01$ to $t = 25$. (b) displays the BIC scores of four models obtained through the ER-BIC algorithm. Among these models, the one with four terms has the lowest BIC score. This is consistent with our conjecture for equation. Figure 9 (c) and (b) exhibit the results of the ER-BIC and SINDy-AIC methods. We can know that both algorithms can identify

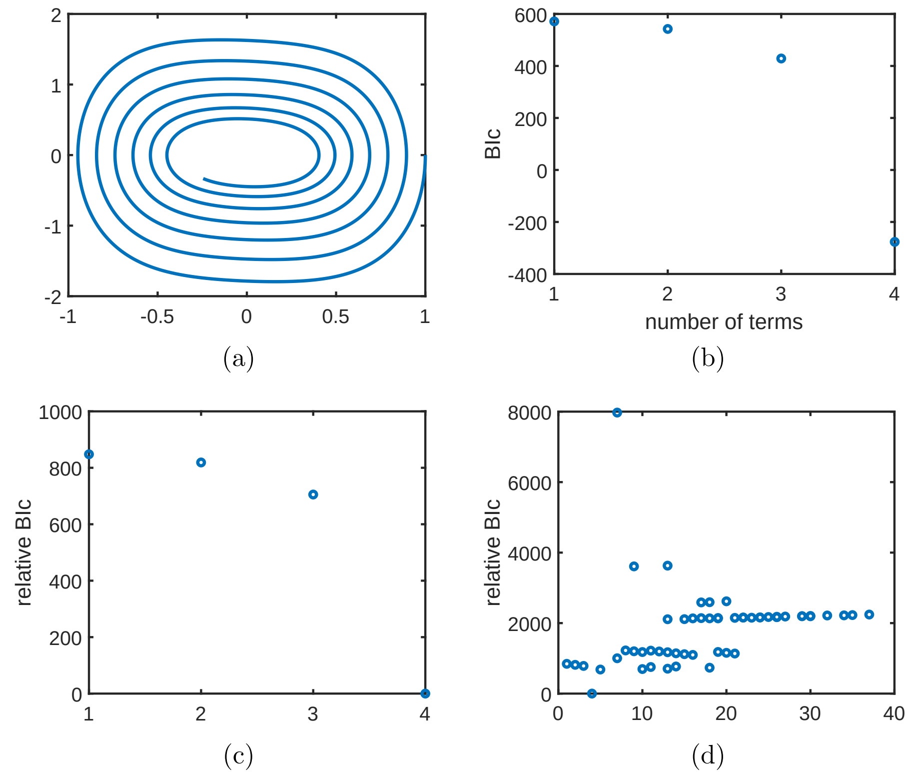


Figure 9. The model choice results of two-dimensional Duffing system by the ER-BIC algorithm: (a) Duffing system. (b) and (c) demonstrate the model of identification by ER-BIC, where black dot denotes all the models and red dot indicates model via ER-BIC. And (d) shows restore model by SINDy-AIC.

equations (four terms score is the smallest). By comparison, we can see that there are very few models of ER algorithms near the Pareto optimum point. In other words, ER takes less time to identify the system than SINDy.

3.4. 2D classic ODE system

The ER-BIC algorithm is extended to consider general models [36, 37]:

$$\ddot{x} + a^2 x + \varepsilon((k_1 x^2 + k_2 x^3) + \dot{x}(b_1 + b_2 x^2)) = \varepsilon F cos(\Omega t), \tag{3.7}$$

where ε is a damping coefficient, and we set $\varepsilon = 1$.

Assuming $\dot{x} = y$, we can get the equation:

$$\begin{cases} \dot{x} = y, \\ \dot{y} = -a^2 x - \varepsilon((k_1 x^2 + k_2 x^3) + y(b_1 + b_2 x^2)) + \varepsilon F cos(\Omega t), \end{cases} \tag{3.8}$$

where the parameters are set as $a = 1, b_1 = -1, b_2 = 2, k_1 = -2, k_2 = 4, F = 5$, and the initial value is $x_0 = [2; 0]$. As in 2D classic ODE system (3.7), the sparsity is seven, and the number of variables is $n = 2$. We set the degree of the function library as $d = 3$ and add the trigonometric term, yielding $p = 14$ monomial terms. In this case, we obtain the training data from $t = 0.01$ to $t = 15$ with time-series of

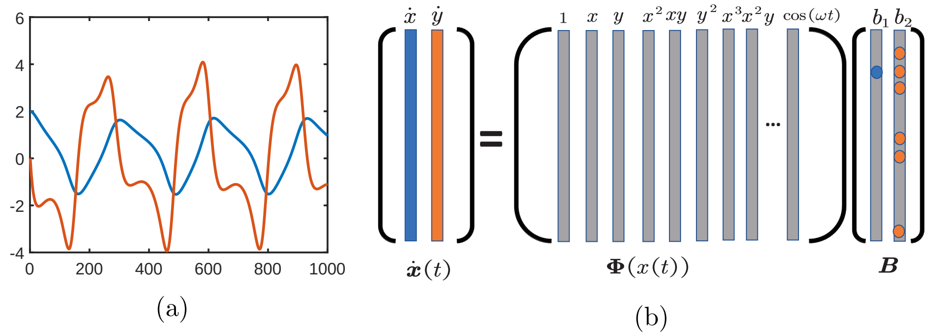


Figure 10. Restore the 2D classic ODE system via the ER-BIC algorithm: (a) training data time series, (b) shows the form of (3.8).

$dt = 0.01$. Then, get the different models using ER-BIC algorithm under different tolerance *tol*. Finally, we compute the BIC score for those models and sort it to obtain the identification results. The time-series of classic ODE system and the form of equation (3.8) are shown in Figure 10.

The results of model selection based on ER-BIC algorithm, are shown in Table 4. According to the models, we can recover the classic ODE system as follows:

$$\begin{cases} \dot{x} = 0.99996y, \\ \dot{y} = -0.99959x + 1.00011y + 1.99945x^2 - 3.99928x^3 + 4.99747\cos(\Omega t). \end{cases}$$

Since the system contains triangular terms, the results identified by SINDy are not as accurate as those of ER and the error is below 0.05%.

Table 4. ODE model selection result of the ER-BIC algorithm.

	$Model_1$		$Model_2$		$Model_3$		$Model_4$	
	$\dot{x}$	$\dot{y}$	$\dot{x}$	$\dot{y}$	$\dot{x}$	$\dot{y}$	$\dot{x}$	$\dot{y}$
1	0	0	0	0	0	0	0	0
x	0	-0.9996	0	1.1474	0	0	0	0
y	1.0000	1.0001	1.0000	1.9821	1.0000	0	1.0000	0
x^2	0	1.9995	0	0	0	0	0	0
xy	0	0	0	0	0	0	0	0
y^2	0	0	0	-4.1847	0	0	0	0
x^3	0	-3.9993	0	-2.0766	0	-2.3450	0	0
x^2y	0	-1.9998	0	0	0	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\sin(\omega t)$	0	0	0	4.0230	0	0	0	0
$\cos(\omega t)$	0	4.9975	0	0	0	0	0	0

In this sample, similar to the previous example, we observe from (b) and (c) panels of Figure 11 that the ER-BIC algorithm identifies only three models in the vicinity of the Pareto optimum. And also we can know the seven terms for which the model has the smallest BIC score. The (d) panel of Figure 11 illustrates the model results for SINDy-AIC identification. Obviously, the number of models identified by ER-BIC is much less than that of SINDy-AIC, and the computation time is faster and more accurate.

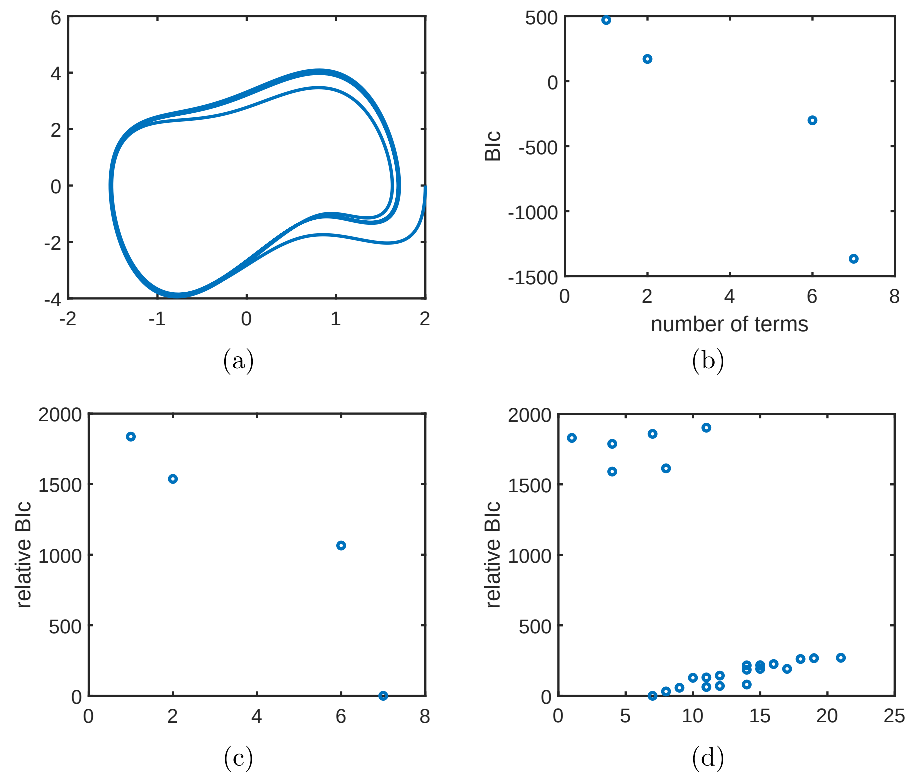


Figure 11. The model selection results of two-dimensional classic ODE system using the ER-BIC algorithm: (a) displays the model of system. (b) shows the model results only by ER-BIC, and the number of points represents the number of models, (c) demonstrates the model of identification via ER-BIC and (d) shows the model by SINDy-AIC method.

3.5. 3D Lorenz 63 system

In 1963, Lorenz [21] proposed a simplified three-variable model to study atmospheric convection. By making some simplifications to the buzzing approximation, the Lorenz model is derived for a monolayer fluid which is uniformly heated from below and cooled from above. The relatively simple but nonlinear system of ordinary differential equations is:

$$\begin{cases} \dot{x} = \gamma(y - x), \\ \dot{y} = x(\rho - z) - y, \\ \dot{z} = xy - \beta z, \end{cases} \tag{3.9}$$

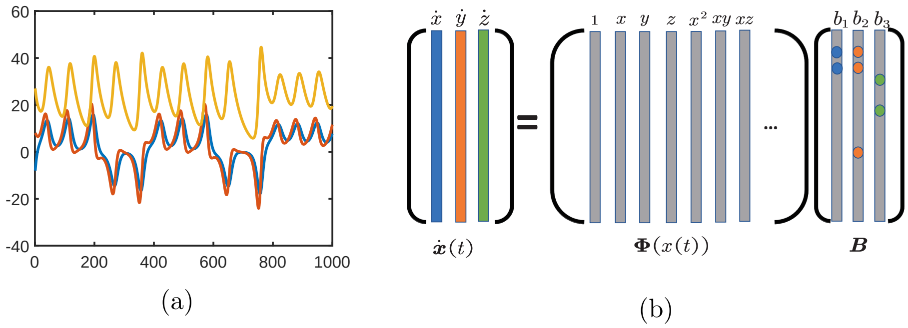


Figure 12. Recover the Lorenz 63 system: (a) training data time series, (b) the form of equation (3.9).

where x, y, z represent dynamical variables, and γ, ρ, β are parameters. Setting $\gamma = 10, \rho = 28, \beta = 8/3$ and the initial value is $x_0 = [-8; 8; 27]$. For those parameters, the Lorenz 63 system [20] exhibits chaotic behavior.

In this case, we obtain the training data from $t = 0.01$ to $t = 15$ with time-series $dt = 0.01$. Since the Lorenz 63 system is three-dimensional, we set the polynomial index of $d = 2$ and obtain a library of functions consisting of terms with $p = 10$. Next, we select different tolerance values $tol_i \in [10^{-6},\ 10]$ to gain the models, which are shown in Figure 13. Figure 12 displays the time series of three variables and the form of equation (3.9).

From Table 5, after sorting, $Model_1$ has the smallest BIC score and is the most consistent with the Lorenz 63 dynamical system, where the recovery equation is expressed as follows:

$$\begin{cases} \dot{x} = -9.99989x + 9.99982y, \\ \dot{y} = 27.9987x - 0.99971y - 0.99997xz, \\ \dot{z} = -2.6666z + 0.99997xy. \end{cases}$$

Figure 13 (a) show the dynamical model with Lorenz 63 system. Figure 13 (b) and (c) illustrate the number of identification using the ER-BIC algorithm. For small noise levels, the number of models identified by the ER algorithm near the Pareto optimum at different tolerance values is only eight. The model with the lowest BIC score is the 7-term model, which satisfies the equation. Figure 13 (d) displays the results via SINDy-AIC method, which includes 343 models and 7-term model has the lowest score.

3.6. 3D Lorenz 84 system

In this section, we describe the ER-BIC results for the Lorenz 84 atmospheric jet model. The model was first described and analyzed in 1984 [22], in which the thermal forcing parameters F and G are periodically forced. The statement is as follows:

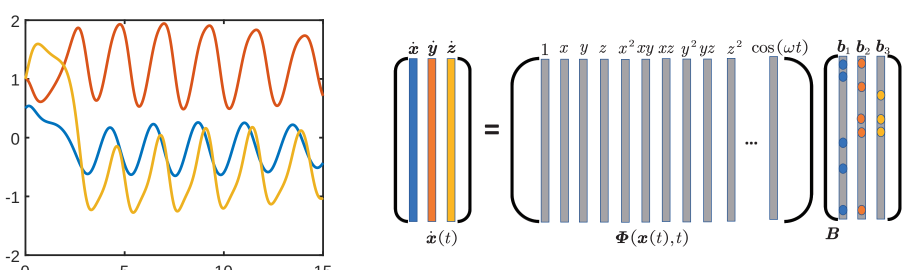
$$\begin{cases} \dot{x} = -ax - y^2 - z^2 + aF(1 + \epsilon\cos(\omega t)), \\ \dot{y} = -y + xy - bxz + G(1 + \epsilon\cos(\omega t)), \\ \dot{z} = -z + bxy + xz, \end{cases} \tag{3.10}$$

Table 5. Lorenz 63 model selection result of the ER-BIC algorithm.

	$Model_1$			$Model_2$			...	$Model_8$		
	$\dot{x}$	$\dot{y}$	$\dot{z}$	$\dot{x}$	$\dot{y}$	$\dot{z}$	...	$\dot{x}$	$\dot{y}$	$\dot{z}$
1	0	0	0	0	0	0	...	0	0	0
x	-9.99989	27.9987	0	-9.99989	27.9987	0	...	0	0	0
y	9.99982	-0.99971	0	9.99982	0.99971	0	...	0	0	0
z	0	0	-2.6666	0	0	0	...	0	0	0
z^2	0	0	0	0	0	0	...	0	0	0
xy	0	0	0.99997	0	0	0	...	0	0	0
xz	0	-0.99997	0	0	0.99997	0	...	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

where x, y, z represent dynamical variables of Lorenz 84 system, and $a, b, F, G, \omega, \epsilon$ are system parameters. The parameters are set as follows: $a = 0.25, b = 4, F = 7, G = 1.7, \epsilon = 0.5$, and the initial conditions are $x_0 = [0.5; 1; 1]$. $T = \frac{2\pi}{\omega} = 73$ is the period of the forcing and ϵ is the relative amplitude of the seasonal forcing. For these parameter values, the Lorenz-84 system shows present chaotic behavior.

In the Lorenz 84 system, we obtain the training data from $t = 0.01$ to $t = 15$ with $dt = 0.01$. Then we set the polynomial index to $d = 2, \omega = 1, 2$ and obtain a library of functions consisting of terms with $p = 14$. Next, Figure 14 shows the model of Lorenz 84 system from $t = 0.01$ to $t = 15$, and the form of the Lorenz 84 model (3.10).



(a) (b)

Figure 14. Restore the Lorenz 84 system: (a) time series of measurement data, (b) the form of equation (3.10).

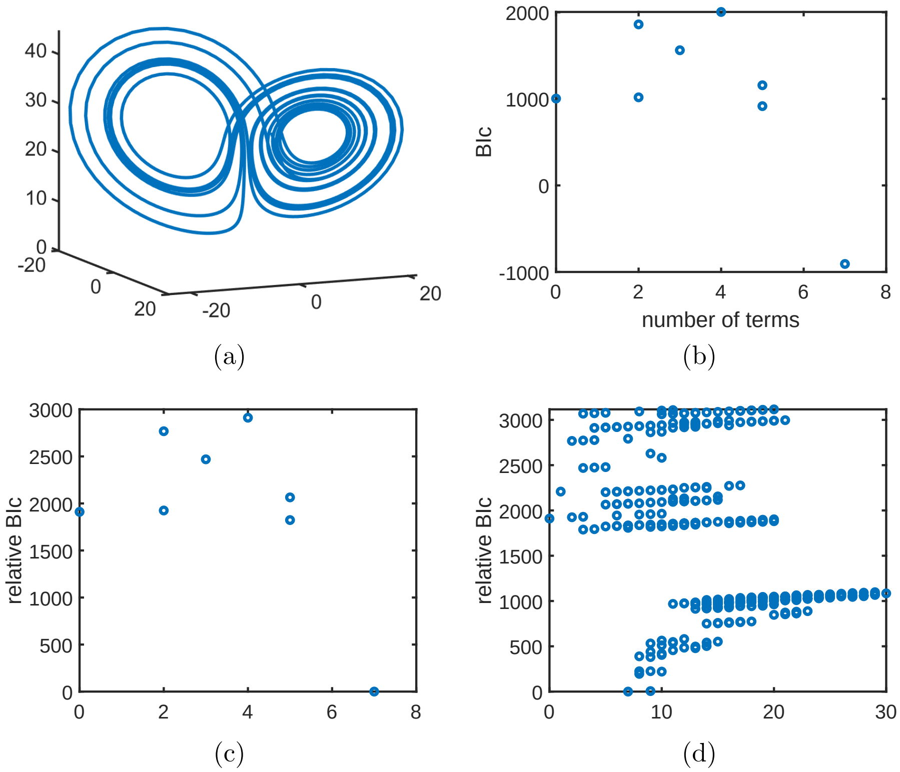


Figure 13. The model choice result of Lorenz 63 system by the ER-BIC algorithm: (a) displays the model. (b) shows the model results identified by ER-BIC, and the number of points represents the number of models. The optimal model is the point with the lowest BIC score, and the optimal number of items in the graph is seven, (c) and (d) demonstrate the model of identification via ER-BIC and SINDy-AIC algorithms.

Table 6. Lorenz 84 model selection result of the ER-BIC algorithm.

	$Model_1$			$Model_2$			...	$Model_{80}$		
	$\dot{x}$	$\dot{y}$	$\dot{z}$	$\dot{x}$	$\dot{y}$	$\dot{z}$	...	$\dot{x}$	$\dot{y}$	$\dot{z}$
1	1.7504	1.6983	0	1.7504	1.698	0	...	0	0	0
x	-0.2502	0	0	-0.2502	0	0	...	0	2.45	0
y	0	-0.9997	0	0	-1.000	0	...	0	0	0
z	0	0	-0.9995	0	0	0	...	0	0	0
z^2	0	0	0	0	0	0	...	0	0	0
xy	0	0.9981	3.9989	0	0.998	0	...	0	0	0
xz	0	-4.0003	0.9973	0	-4.000	0	...	0	0	0
yy	-1.0001	0	0	-1.0001	0	0	...	0	0	0
yz	0	0	0	0	0	0	...	0	0	0
zz	-0.9997	0	0	-0.9997	0	0	...	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
$\cos(\omega t)$	0.8750	0.8515	0	0.8750	0.852	0	...	0	0	0

From Table 6, after sorting, $Model_1$ has the smallest BIC score and is the most consistent with the Lorenz 84 dynamical system. The recovery equation is expressed as follows:

$$\begin{cases} \dot{x} = 1.7504 - 0.2502x - 1.0001y^2 - 0.9997z^2 + 0.8750\cos(\omega t), \\ \dot{y} = 1.6983 - 0.9997y + 0.9981xy - 4.0003xz + 0.8515\cos(\omega t), \\ \dot{z} = -0.9995z + 3.9989xy + 0.9973xz, \end{cases}$$

where the parameters are very similar to those of the original system, so it is known that the ER-BIC algorithm is able to accurately identify the Lorenz 84 atmospheric system.

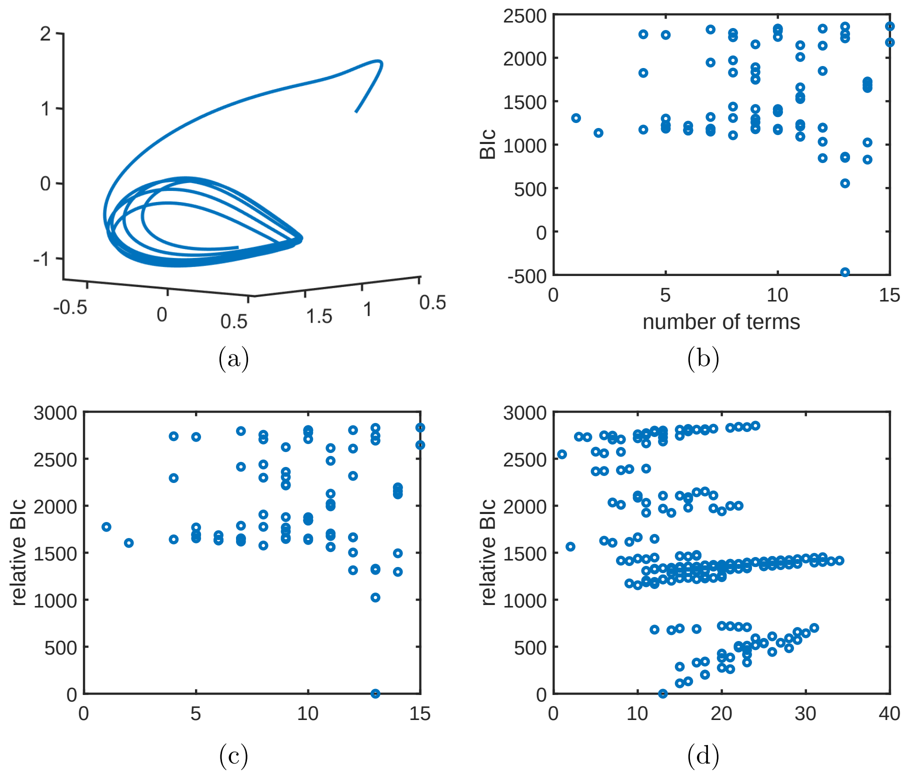


Figure 15. The model selection result of Lorenz 84 system via the ER-BIC algorithm: (a) Lorenz 84 model. (b) shows the model results only by ER-BIC. (c) demonstrate the model of identification via ER-BIC and (d) shows the model by SINDy-AIC.

From Figure 15, we can see that (a) shows the dynamical model with Lorenz 84 system. Figure 15 (b) and (c) display the number of identification using the ER-BIC algorithm. For small noise levels, the number of models identified by the ER algorithm near the Pareto optimum at different tolerance values is fewer than the models from the SINDy-AIC, which includes 210 models. The model with the lowest BIC score is the 13-term model, which satisfies Equation(3.10). Figure 15 (d) displays the results via SINDy-AIC method.

4. Potential impacts: noise and number of measurements

The ER algorithm with BIC can successfully select the correct model for a variety of known systems when the noise amplitude is sufficiently low and a low-dimensional data set is available for validation. In practice, the signal-to-noise ratio may be lower than expected and the amount of data available for validation may be limited. In the following, we consider separately the accuracy of the ER-BIC algorithm for selecting a model for the Lorenz 63 system at different noise amplitudes and with different amounts of data.

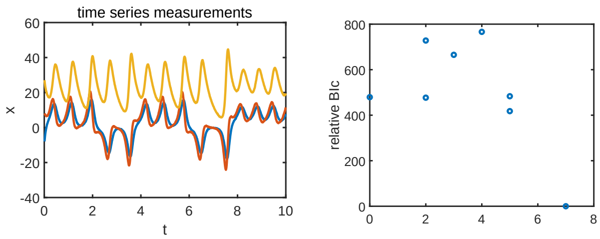
4.1. Different noise amplitude

In addition to the low noise amplitude ($\eta = 0.01$) used in this paper, we consider the effect of the ER-BIC algorithm on the selection of the Lorenz 63 system for several noise amplitudes ($\eta = 0.1, 0.2, 0.3, 0.5, 1, 5, 10$). In Figure 16, we show the effect of increasing the noise amplitude on the correct selection of the Lorenz 63 system. It can be seen from the figure (a) that for low noise amplitude ($\eta = 0.1$) the BIC score of the correct model is close to 0 and the BIC score of the other identified results is too large. As the noise increases, for example, at $\eta = 0.2$ and 0.3, the BIC fraction of the correctly identified model is the smallest and the model result we need. However, from the comparison in Figure 16, we know that the incorrect model BIC fraction becomes smaller as the noise increases, which gradually affects our selection results. From this, it can be seen that the relative BIC fraction of the incorrect model is very sensitive to random additive noise as the noise amplitude increases. For other examples of randomly generated measurement noise, the BIC scores also fluctuate despite the true model maintaining the lowest scores (data not shown).

From the figure we can see that above a certain noise level ($\eta = 0.5$), ER-BIC cannot robustly select the correct model and for even higher levels of noise ($\eta = 1, 5, 10$), a large number of incorrect models BIC scores are even lower. For the Lorenz 63 system used here, ER fails to select the correct model and thus BIC fails to evaluate the true model at these noise levels.

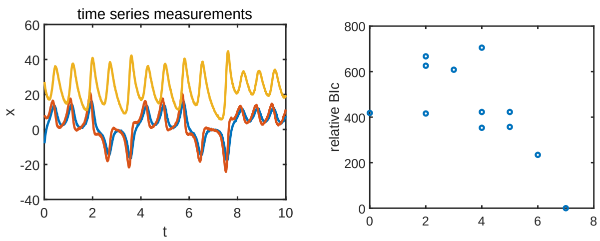
4.2. Different number of measurements

We next consider the effect of the number of validated time series on the choice of the exact model of ER-BIC compared to other models. In this subsection, we set the number of time series ($N = 2, 4, 20, 60, 80, 100$) and validate the effect of these time series numbers on the selection of the correct Lorenz 63 system by the ER-BIC algorithm. The results are shown in Figure 17, where the noise amplitude is $\eta = 0.2$. Figures (b)-(h) show the distribution of the BIC scores obtained by using the algorithm, from which we can see that the Lorenz 63 model is accurately selected for all the number of time series sets, and the selection results are represented in figure (i).



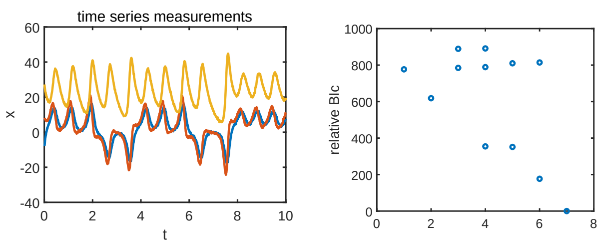
	b_1	b_2	b_3
x	-10.0102	27.9580	0
y	9.99708	-0.98043	0
z	0	0	-2.66852
x^2	0	0	1.00048
xy	0	-0.99955	0
xz	0	0	0

(a) $\eta = 0.1$



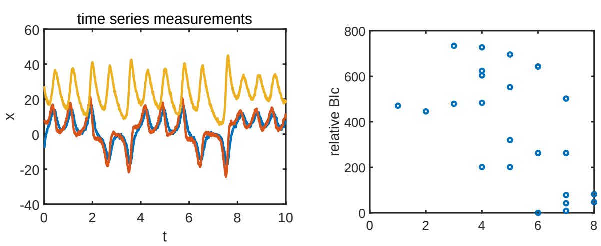
	b_1	b_2	b_3
x	-9.99647	27.8286	0
y	9.97345	-0.9347	0
z	0	0	-2.66891
x^2	0	0	1.0004
xy	0	-0.9971	0
xz	0	0	0

(b) $\eta = 0.2$



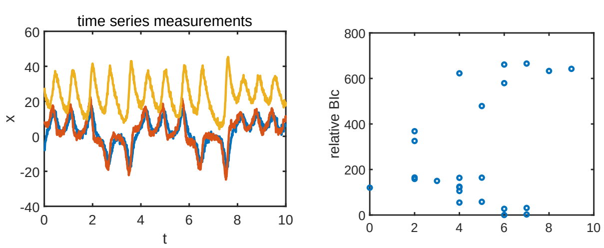
	b_1	b_2	b_3
x	-9.9590	27.6168	0
y	9.9293	-0.8645	0
z	0	0	-2.6678
x^2	0	0	0.99985
xy	0	-0.9927	0
xz	0	0	0

(c) $\eta = 0.3$



	b_1	b_2	b_3
x	-9.8149	25.3871	0
y	9.7815	0	0
z	0	0	-2.6613
x^2	0	0	0.9971
xy	0	-0.9478	0
xz	0	0	0

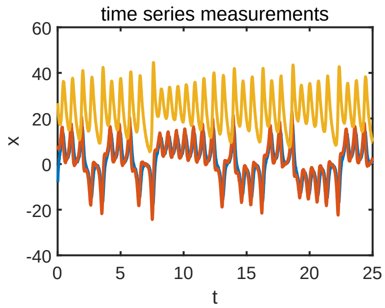
(d) $\eta = 0.5$



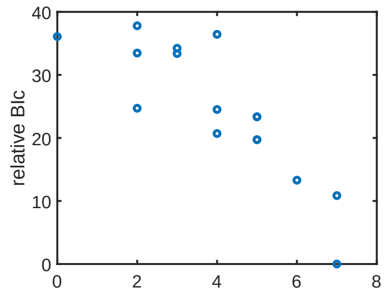
	b_1	b_2	b_3
x	0	24.6909	0
y	4.0911	0	0
z	0	0	-2.6198
x^2	0	0	0.98112
xy	0	0	0
xz	-0.2664	-0.92716	0

(e) $\eta = 1$

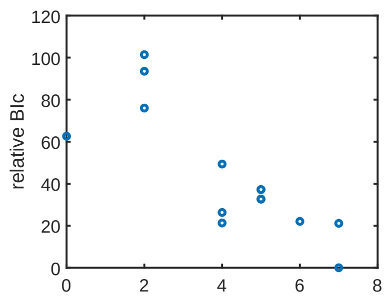
Figure 16. Recover the Lorenz 63 system using the ER-BIC method under different noise amplitudes ($\eta = 0.1, 0.2, 0.3, 0.5, 1$). At each noise amplitude, we show three plots: a time series plot, a BIC-ER selection plot, and the model results for optimal BIC score recovery, respectively.



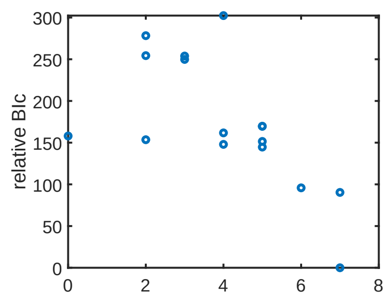
(a) time-series of Lorenz 63



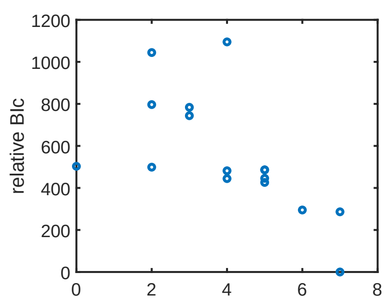
(b) $N = 2$



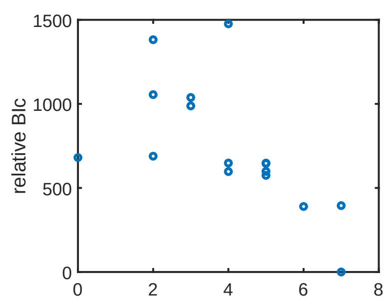
(c) $N = 4$



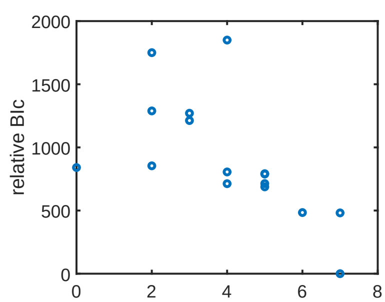
(d) $N = 20$



(e) $N = 60$



(f) $N = 80$



(g) $N = 100$

	b_1	b_2	b_3
x	-9.9512	27.7590	0
y	9.9560	-0.95781	0
z	0	0	-2.6656
x^2	0	0	0.9998
xy	0	-0.9937	0
xz	0	0	0

(h) select result

Figure 17. Select the Lorenz 63 system using the ER-BIC method under different numbers of measurements ($N = 2, 4, 20, 40, 60, 80, 100$). (a) time series of the Lorenz 63 system. (b)-(g) show the BIC score with the models under $N = 2, 4, 20, 40, 60, 80, 100$. (h) displays the selection model via ER-BIC algorithm.

As the number of time series increases, it is clear from the figure that the correctly selected model has the lowest score and the other incorrect models have higher scores. As the number of time series increases, we observe that the BIC scores of the incorrect models also rise, indicating that these scores are deviating further from those of the accurately selected models. Thus, it is clear that not only our ER-BIC algorithm is able to accurately select models for a wide range of time series, but the gap between the BIC score of imprecise and accurate models becomes increasingly pronounced as the number of time series increases.

The success of the ER-BIC algorithm also depends on the duration and sampling of the time series, especially in the case of chaotic systems such as the Lorenz 63 model. Here, the validation experiments are conducted from $t = 0$ to $t_{end} = 1$. Even with the higher noise ($\eta = 0.1, 0.2$, compared to $\eta = 0.01$), the ER-BIC algorithm can still obtain the true model. Again, the sensitivity to quadratic sampling of the data helps distinguish the noisy fitness term from the fundamental term.

5. Conclusion and discussion

In conclusion, we develop the ER-BIC algorithm, based on SINDy-AIC method, for model selection. In our work, first, we obtain the system models via the ER algorithm, whose models are near the Pareto optimal. Second, we compute the BIC score and sort the models based on it. Finally, we select the model of which the BIC score is smallest. Specifically, a large number of sparse regression models are obtained at different tolerance values, and a library of models chosen to be close to the Pareto optimum is composed. After calculating the BIC score for the models, we rank them. According to the definition of BIC, the model with the smallest score value is selected as the optimal model. In this paper, we use BIC scores and display these scores to rank each candidate model and present it in a graph. For time series data with a sufficient data sample, a sufficiently large signal-to-noise ratio, or a sufficiently large set of candidate models, there is only one BIC score closest to zero.

For all examples (such as one-dimensional simple polynomial system and RC circuit system, two-dimensional Duffing system and classical ODE system , three-dimensional Lorenz 63 system and Lorenz 84 system), we observed that ER-BIC is better than SINDy-AIC in recovering dynamical systems from noise data and we extend the scope of the ER algorithm with trigonometric term. Overall, the ER-BIC algorithm produces significantly fewer models than the SINDy algorithm, which allows us to better select the optimal model. The accuracy of the optimal model identified by ER is also elevated, while SINDy is frequently biased in identifying systems containing trigonometric terms. In contrast, we choose the ER algorithm to recover the governing equations more accurately than the SINDy algorithm, and the number of identified models near the Pareto optimum is smaller, which favors us to obtain better results.

There are also subtle differences in model selection methods between ER algorithm and BIC ranking. The ER algorithm, as showed, can describe a dynamical system model as a sparse linear combination of function library via entropy regression, which includes trigonometric and other nonlinear terms. The number of terms in the model obtained by the ER algorithm can be expressed in terms of the number of non-zero terms in the sparse matrix. There exists combinatorial formalism for a large number of physical or biological systems, but the selection of combinatorial models remains difficult, so ER algorithms and BIC ranking provide a systematic and rigorous way to infer and evaluate suitable models in these cases.

In this paper, we obtain the state variables by measurements and obtain the complete state. However, there are many methods that have eliminated the need to measure all state variables, such as combining time-delay coordinates with SINDy [8], and using deep learning autoencoders to reconstruct hidden data [25], thereby obtaining a system model. Therefore, it is of broad interest to further investigate the methodology of the relevant measurement variables used in the model selection process.

In this work, the ER-BIC algorithm successfully selects the exact model needed with sufficient data and low noise magnitude. The ER algorithm uses regression with entropy values, focusing primarily on finding governing equations in short datasets. In cases where we have few measurements, we can use time delay coordinates to increase, and when we have too many measurements, we can use advanced methods from dimensionality reduction and machine learning to extract coherent structures. Furthermore, in this paper, it is recognized that the amplitude of noise can significantly influence the outcome of model selection; excessive noise can yield unsatisfactory results. Therefore, it is crucial to develop future diagnostic methods that can be applied to real data. Combined with BIC ranking, this can constitute a complete set of tools for model identification and validation, and extending diagnostic methods allows ER algorithms and BIC to be useful on complex datasets. Given the effectiveness of ER-BIC in recovering the control equations of nonlinear dynamical systems, we also hope to extend this work to partial differential equations (PDEs) (including high dimensions).

The ER-BIC algorithm has contributed significantly to the development of both model selection and entropy regression algorithms. In particular, the BIC evaluation criteria cannot evaluate a large set of candidate models in the model selection process. In contrast, entropy regression algorithms lack a standard evaluation method for selecting the correct model. The ER-BIC algorithm in this paper addresses both of these problems. The ER algorithm employs entropy regression to accommodate a vast array of candidate models. Subsequently, it leverages BIC evaluation to rank these models, facilitating their assessment and enabling the selection of the precise system model. This approach has important implications for domains where symbolic regression is already used for identification or model selection. We hope that the link between the ER-BIC approach, machine learning and dynamic systems will stimulate the development of automation and improved model selection.

Acknowledgements

The authors express their sincere gratitude to the editors and reviewers for their helpful comments and suggestions, which have significantly improved the quality of the paper.

References

[1] H. Akaike, *A new look at the statistical model identification*, IEEE transactions on automatic control, 1974, 19(6), 716–723.

[2] A. AlMomani, J. Sun and E. Bollt, *How entropic regression beats the outliers*

problem in nonlinear system identification, Chaos: An Interdisciplinary Journal of Nonlinear Science, 2020, 30(1), 013107.

[3] M. Abdul Majid and K. Ibrahim, *On Bayesian approach to composite Pareto models*, Plos one, 2021, 16(9), e0257762.

[4] K. Burnham and D. Anderson, *Model selection and multi-model inference*, Springer, Berlin, 2002.

[5] H. Broer, R. Vitolo, *Dynamical systems modeling of low-frequency variability in low-order atmospheric models*, Discrete and Continuous Dynamical Systems-series B, 2008, 10, 401–419.

[6] S. Brunton, J. Proctor, J. Tu, and J. Kutz, *Compressed sensing and dynamic mode decomposition*, Journal of computational dynamics, 2016, 2(2), 165–191.

[7] S. Brunton, J. Proctor and J. Kutz, *Discovering governing equations from data by sparse identification of nonlinear dynamical systems*, Proceedings of the national academy of sciences, 2016, 113(15), 3932–3937.

[8] S. Brunton, B. Brunton, J. Proctor, et al., *Chaos as an intermittently forced linear system*, Nature communications, 2017, 8(1), 19.

[9] Z. Bai, E. Kaiser, J. Proctor, J. Kutz, et al., *Dynamic mode decomposition for compressive system identification*, AIAA Journal, 2020, 58(2), 561–574.

[10] Z. Babajamali, M. Khabaz and F. Aghadavoudi, *Pareto multi-objective optimization of tandem cold rolling settings for reductions and inter stand tensions using NSGA-II*, ISA transactions, 2022, 130, 399–408.

[11] J. Daron, D. Stainforth, *On quantifying the climate of the nonautonomous Lorenz-63 model*, Chaos: An Interdisciplinary Journal of Nonlinear Science, 2015, 25(4), 043103.

[12] J. Freire, C. Bonatto, C. DaCamara, and J. Gallas, *Multistability, phase diagrams, and intransitivity in the Lorenz-84 low-order atmospheric circulation model*, Chaos: An Interdisciplinary Journal of Nonlinear Science, 2008, 18(3), 033121.

[13] J. Gabelli, J. Berroir and B. Placais, *A coherent RC circuit*, Reports on progress in physics, 2012, 75(12), 126504.

[14] K. Hirose, S. Kawano, S. Konishi and M. Ichikawa, *Bayesian Information Criterion and Selection of the Number of Factors in Factor Analysis Models*, Journal of Data Science, 2011, 9(2), 243–259.

[15] T. Jin, S. Gao, H. Xia and H. Ding, *Reliability analysis for the fractional-order circuit system subject to the uncertain random fractional-order model with Caputo type*, Journal of Advanced Research, 2021, 32, 15–26.

[16] P. Joseph, M. Gerald, G. Sugihara, et al., *Empirical Mode Modeling: A data-driven approach to recover and forecast nonlinear dynamics from noisy data*, Nonlinear Dynamics, 2022, 108(3), 2147–2160.

[17] B. Koopman, *Hamiltonian systems and transformation in Hilbert space*, Proceedings of the National Academy of Sciences, 1931, 17(5), 315–318.

[18] S. Kullback and R. Leibler, *On information and sufficiency*, The annals of mathematical statistics, 1951, 22(1), 79–86.

[19] E. Kaiser, J. Kutz, and S. Brunton, *Data-driven discovery of Koopman eigenfunctions for control*, Machine Learning: Science and Technology, 2021, 2(3), 035023.

[20] S. Kravtsov and A. Tsonis, *Lorenz-63 Model as a Metaphor for Transient Complexity in Climate*, Entropy, 2021, 23(8), 951.

[21] E. Lorenz, *Deterministic nonperiodic flow*, Journal of atmospheric sciences, 1963, 20(2), 130–141.

[22] E. Lorenz, *Irregularity: A Fundamental Property of the Atmosphere*, Tellus A: Dynamic Meteorology and Oceanography, 1984, 36(2), 98–110.

[23] X. Li, Y. Shen, J. Sun and S. Yang, *New periodic-chaotic attractors in slow-fast Duffing system with periodic parametric excitation*, Scientific Reports, 2019, 9(1), 1–11.

[24] K. Liu, Y. Shang, Q. Ouyang and W. Widanage, *A Data-Driven Approach With Uncertainty Quantification for Predicting Future Capacities and Remaining Useful Life of Lithium-ion Battery*, IEEE Transactions on Industrial Electronics, 2020, 68(4), 3170–3180.

[25] P. Lu, J. Ariño Bernad and M. Soljačić, *Discovering sparse interpretable dynamics from partial observations*, Communications Physics, 2022, 5(1), 206.

[26] N. Mangan, J. Kutz, S. Brunton and J. Proctor, *Model selection for dynamical systems via sparse regression and information criteria*, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2017, 473(2204), 20170009.

[27] J. Niu, R. Liu, Y. Shen and S. Yang, *Chaos detection of Duffing system with fractional-order derivative by Melnikov method*, Chaos: An Interdisciplinary Journal of Nonlinear Science, 2019, 29(12), 123106.

[28] A. Ouannas, A. Khennaoui, S. Momani and V. Pham, *The discrete fractional duffing system: Chaos, 0C1 test, c 0 complexity, entropy, and control*, Chaos: An Interdisciplinary Journal of Nonlinear Science, 2020, 30(8), 083131.

[29] N. Papageorgiou, C. Vetro and F. Vetro, *Nonlinear multivalued Duffing systems*, Journal of Mathematical Analysis and Applications, 2018, 468(1), 376–390.

[30] S. Portet, *A primer on model selection using the Akaike Information Criterion*, Infectious Disease Modelling, 2020, 5, 111–128.

[31] G. Schwarz, *Estimating the dimension of a model*, The annals of statistics, 1987, 461–464.

[32] P. Schmid, *Dynamic mode decomposition for numerical and experimental data*, Journal of fluid mechanics, 2010, 656, 5–28.

[33] J. Tu, C. Rowley, D. Luchtenburg, et al., *On dynamic mode decomposition: theory and applications*, Journal of Computational Dynamics, 2014, 1(2), 391–421

[34] L. Veen, *Baroclinic Flow and the Lorenz-84 Model*, International journal of Bifurcation and Chaos, 2003, 13, 2117–2139.

[35] H. Wang, Y. Yu, and G. Wen, *Dynamical Analysis of the Lorenz-84 Atmospheric Circulation Model*, Journal of Applied Mathematics, 2014, 2014, 1–15.

[36] W. Wang, J. Cao, D. Mallick and S. Roy, *Comparison of harmonic balance and multi-scale algorithm in characterizing the response of mono-stable energy harvesters*, Mechanical Systems and Signal Processing, 2018, 108, 252–261.

[37] P. Xia, H. Xu, M. Lei and Z. Ma, *An improved stochastic resonance algorithm with arbitrary stable state matching in underdamped nonlinear systems with a periodic potential for incipient bearing fault diagnosis*, Measurement Science and Technology, 2018, 29(8), 085002.