Chaos, Solitons and Fractals 175 (2023) 113927

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

journal homepage: www.elsevier.com/locate/chaos

Chaotic behavior learning via information tracking

Jinfu Ren, Yang Liu, Jiming Liu *

Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, PR China

ARTICLE INFO

Keywords:
Chaotic behavior characterization
Machine learning
Information tracking
The Lyapunov exponent
The Hurst exponent

ABSTRACT

Chaotic time series are often encountered in real-world applications, where modeling and understanding such time series present a significant challenge. Existing simulation-based methods for characterizing chaotic behavior may be sensitive to the respective model settings, while data-driven methods cannot adapt well to irregularities and aperiodicity in chaotic time series during the prediction stage, leading to nonlinearly aggregated forecasting errors after a finite number of time steps. In our preliminary work (Ren et al., 2023), we introduced a data-driven method that incorporates the idea of information tracking to capture and adapt to the chaotic changes over time, and tested its effectiveness on a real-world task—decadal temperature prediction. While the practical usefulness of the method has been initially validated on one domain-specific application scenario, a set of open questions is yet to be answered: Is the concept of information tracking generalizable to model and solve a broad spectrum of learning tasks that involve chaotic behavior? If yes, what are the underlying principles and learning behaviors that make the general method work well? Moreover, how can we assess the efficacy of the method in different scenarios? This study constitutes a significant step forward by systematically investigating and addressing the questions mentioned above. Specifically, we first provide a rigorous definition of the formalism and computational mechanism of the general method, with all necessary variables, symbols, and formulations. More importantly, we theoretically characterize and understand the generalizability of the method by mathematically uncovering the fundamental principles and behavior of information tracking-based chaotic behavioral learning, as prescribed by two chaotic behavioral indexes, namely, the Lyapunov and the Hurst exponents. Further, we comprehensively demonstrate the generalizability and robustness of the method by empirically examining a variety of chaotic behavioral learning problems generated by three representative chaotic systems: The logistic map, the double pendulum, and the Lorenz system.

1. Introduction

The chaotic behavior is ubiquitous in various scientific and practical domains, from physics [1,2], chemistry [3,4], biology [5,6] to climatology [7,8], healthcare [3,9], and finance [10,11], among others. By modeling and analyzing such chaotic behavior, one can uncover underlying relationships and dependencies among variables of the chaotic systems and gain an in-depth understanding of the system's dynamics, thus informing decision-making in the corresponding domain. For example, the World Climate Research Programme (WCRP), which consists of scientists and policymakers, launched the well-known Coupled Model Intercomparison Project (CMIP) [12] to model and forecast time series of climate variables (e.g., temperature and precipitation) to assist making plans and taking actions in agricultural production [13], extreme weather response [14], etc. Similarly, analyzing human physiological signals can aid in disease diagnosis and treatment in healthcare [15]; while modeling stock prices can help investors to foresee future financial trends, so as to mitigate the risk or optimize the performance [16].

In pursuit of this goal, simulation-based methods have been widely used to characterize behaviors of chaotic systems. These methods define a set of governing rules and/or a series of parameterized equations and then simulate system behaviors over time under different conditions. Drawing on some well-formulated rules or equations that model the underlying components as well as their interactions in a chaotic dynamical system, simulation-based methods provide not only ways to capture certain behavioral features of the system but also tools to examine the effects of model parameters on the system behavior [17]. While simulation-based methods are generally useful for studying and understanding chaotic dynamics, it is worth noting that, as indicated in previous studies, these methods can be sensitive to the model settings used. The reason is that the chaotic behavior of a system is highly dependent on its initial conditions and parameters [18–20]. Even a slight

* Corresponding author.
E-mail address: jiming@comp.hkbu.edu.hk (J. Liu).

https://doi.org/10.1016/j.chaos.2023.113927
Received 14 June 2023; Received in revised form 25 July 2023; Accepted 7 August 2023
Available online 13 September 2023
0960-0779/© 2023 Elsevier Ltd. All rights reserved.

variation in the parameter setting can lead to vastly different simulation results. In practice, the governing rules or parameters of a chaotic system are generally too complex to be precisely defined [21,22]. This means that a huge amount of effort must be taken when selecting the model settings used in simulation-based methods.

Recently, data-driven methods, particularly machine learning, have been introduced as an alternative to model-based methods. Such data-driven methods do not predefine models but rather exploit the power of learning models from data: they understand the underlying relationships within chaotic systems by analyzing their input–output observations (e.g., historical data of the chaotic time series), enabling the prediction of future values of the system dynamics. With the well-designed learning architectures, such as the gating mechanism in sequence models [23,24], reservoir in echo state networks [25,26], and attention mechanism in transformers [27,28], the complex input–output mapping relationships can be well captured. There are also other ways to utilize machine learning methods to explore and model chaotic data from various perspectives, such as noise removal for chaotic signals [29], detection of systems' dynamical variations [30], and classification of chaotic time series [31]. Despite the effectiveness of current data-driven methods in modeling chaotic time series, a common limitation exists in that input–output relationships and dependencies, once learned, are unable to adapt well to the irregularities and aperiodicity that arise during prediction or forecasting. This limitation can result in nonlinear aggregation of forecasting errors over time, even if they are initially infinitesimal.

To address the aforementioned limitation, in our preliminary work [32], we introduced a data-driven method, which incorporates the concept of information tracking. This method is composed of two distinct components: a dependency learning component (DLC) and an information tracking component (ITC). The DLC is developed to capture the complex input–output mapping relationships in chaotic time series, while the ITC is specifically designed to track and adapt to behavioral changes of chaotic systems using a probabilistic feedback mechanism on the forecast error of the next time step based on the current forecast. The practical effectiveness of this method is initially validated on a domain-specific application scenario—decadal temperature prediction [32]. However, while the method has shown empirical effectiveness in a real-world application, a set of open questions remains unanswered: Can the idea of information tracking be generalized to model and solve various learning tasks that exhibit chaotic behavior? If so, what are the fundamental principles and learning behaviors that guarantee the effectiveness of the general method? Moreover, how can the efficacy of the method be extensively evaluated in various situations?

The goal of this work is to investigate and answer the aforementioned open questions in a systematic way. First, we rigorously define the formalism and computational mechanism of the general method, including the formal statement of the learning problem, the mathematical definition and formulation of information tracking, and the objective function and learning structure of the proposed method. Of greater significance, this work contributes to theoretically characterizing and understanding the generalizability of the method. This is achieved by mathematically revealing the fundamental principles and learning behavior of the information tracking-based learning method in capturing the essence of the chaotic behavior, as indicated by two important chaotic behavioral indexes: the Lyapunov exponent [33] and the Hurst exponent [34]. The Lyapunov exponent is particularly useful to describe chaotic behavior as it quantifies the degree of chaos in a given dynamical system; it measures the rate of divergence of initially close trajectories in a system and reflects how sensitive the system is to initial conditions. The Hurst exponent, on the other hand, measures the degree of self-similarity of a time series and reflects how persistent or anti-persistent the system is, which complements the Lyapunov exponent in measuring the system's complexity. To validate our theoretical findings, we further conduct comprehensive experiments across a variety of chaotic behavioral learning problems (with various dimensions of the chaotic systems, different noise levels, and multiple evaluation metrics) generated by three typical dynamical systems: the logistic map [35], the double pendulum [36], and the Lorenz system [37]. Experimental results demonstrate the generalizability and robustness of our method.

2. Chaotic behavior learning via information tracking

In this section, we define the formalism and computational mechanism of the general method. First, we give a formal statement of the learning problem and an overview of the method, with all the necessary variables and symbols being explained. Then we provide the mathematical definition and formulation of the concept of information tracking. To make the method easily understandable, we elaborate its objective function and specific learning structures of the DLC and ITC components, as well as its procedure in the prediction phase.

2.1. Overview of the method

Mathematically, the problem of predicting chaotic time series can be stated as follows. Given historical observations, i.e., the chaotic time series with N observational variables, we denote the data vector that contains observations at time step i as $\mathbf{x}_i = [x_i^1, x_i^2, \dots, x_i^N] \in \mathbb{R}^{1\times N}$. The observations are composed of J time steps in total, which can be denoted as $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_J$. The target is to learn the underlying input–output relationships from the observations, so that one can make predictions on the future L steps of the time series: $\mathbf{x}_{J+1}, \mathbf{x}_{J+2}, \dots, \mathbf{x}_{J+L}$. To effectively learn chaotic behavior from data, the method must be able to accurately capture the complex input–output dependencies and, more importantly, the irregularities and aperiodicity presented in the observational time series. The method, therefore, comprises two specific components. The first component, namely DLC, is responsible for modeling the input–output relationships; the second component ITC is used to model changes in the system dynamics over time. These two components work together to achieve the goal of characterization and forecast of chaotic time series.

To train the model, we construct training samples from observations. For each training sample, it is composed of the input vector and the output vector. For each output vector $\mathbf{x}_{i+1}$, its corresponding input vector is the concatenation of the previous T data vectors: $\mathbf{X}_i = [\mathbf{x}_{i-T+1}, \mathbf{x}_{i-T+2}, \dots, \mathbf{x}_i] \in \mathbb{R}^{1\times(T\times N)}$, where $i = T, T+1, \dots, J-1$. The structure of DLC is a multilayer perception (MLP) with five layers: one input layer, one output layer, and three hidden layers. The rectified linear unit (ReLU) is utilized as the activation function of the MLP. The output of DLC can be represented as: $\hat{\mathbf{x}}_{i+1} = DLC(\mathbf{X}_i)$, and we use the output vector $\mathbf{x}_{i+1}$ as the ground truth to supervise the learning procedure of DLC. For ITC, its output can be represented as: $\Delta\mathbf{x}_{i+1} = ITC(\mathbf{X}_i)$. As mentioned previously, the purpose of ITC is to track the behavioral changes of chaotic time series. To this end, the first-order difference of the chaotic time series, i.e., $\Delta\mathbf{x}_{i+1} = \mathbf{x}_{i+1} - \mathbf{x}_i$, is used to supervise the learning in ITC. The reason is that the first-order difference, by definition, can be used to calculate the Lyapunov exponent. The summation of the positive Lyapunov exponents is the upper bound of the chaotic system's metric entropy, an index that describes the information change of the system. By using the first-order difference as the supervision, we therefore expect that the changes in chaotic behavior can be captured by ITC. The output of ITC will be added to the output of DLC to become the model's output: $\mathbf{O}_{i+1} = \hat{\mathbf{x}}_{i+1} + \Delta\mathbf{x}_{i+1}$.

2.2. Information tracking

The ITC aims to track the changes in chaotic behavior by providing feedback on the prediction error in a probabilistic way and, at the same time, adjusting the corresponding prediction error. By introducing such a probabilistic setting, we expect that the learning method will not

work in a purely deterministic way, but with the capacity in capturing the stochasticity presented in the chaotic time series. The ITC contains an encoder and a decoder, whose structures are MLPs with ReLU as the activation function. The encoder has the same structure as the DLC while the decoder is a one-layer MLP. To provide the probabilistic feedback, we adopt the reparameterization technique [38]. With this setting, the probabilistic operation will become derivable to ensure the model is trainable. Specifically, the encoder will encode the data vector as the hidden representation, denoted as $\mathbf{r}_i = encoder(\mathbf{X}_i)$. Let $\mathbf{z}$ denote a random variable and its approximate posterior is denoted as $q(\mathbf{z}|\cdot)$, which follows a normal distribution $N(\mu, \sigma)$ and is parameterized by $\mathbf{r}_i$. Here, the mean μ is parameterized by a neural network NN_μ (which is a one-layer MLP), i.e., $\mu = NN_\mu(r_i)$, and the variance σ is parameterized by another neural network NN_σ (which is also a one-layer MLP), i.e., $\sigma = NN_\sigma(\mathbf{r}_i)$. In each training iteration, $\mathbf{z}$ will be sampled from the parameterized distribution and then used as the input of the decoder to generate the feedback value $\Delta\mathbf{x}_{i+1}$.

2.3. Integration of DLC and ITC

By considering both DLC and ITC, the overall objective function is given as follows:

$$\min \sum_{i=T+1}^{J} (\|\hat{\mathbf{x}}_i + \Delta\hat{\mathbf{x}}_i - \mathbf{x}_i\|^2 + \|\Delta\hat{\mathbf{x}}_i - \Delta\mathbf{x}_i\|^2) + D_{KL}(N(\mu,\sigma) \parallel N(0,I)), \quad (1)$$

where $D_{KL(\cdot)}$ denotes the Kullback–Leibler divergence between two distributions. The integration of DLC and ITC is through the supervision of the sum of their outputs by $\mathbf{x}_i$, as shown in the first term in Eq. (1). By supervising the sum of $\hat{\mathbf{x}}_i$ (the output of DLC) and $\Delta\hat{\mathbf{x}}_i$ (the output of ITC) using $\mathbf{x}_i$ (the ground truth of the time series data), i.e., minimizing the prediction error, it is expected that the method can well characterize both underlying input–output dependencies and chaotic behavioral changes and thus deliver accurate prediction results. Moreover, the output of ITC is separately supervised by the first-order difference of the time series, i.e., $\Delta\mathbf{x}_i$, as shown in the second term in Eq. (1), so as to ensure the ITC can track the behavioral changes of chaotic time series.

2.4. Prediction with information tracking

In the prediction phase, none of the ground truth will be available for adjusting or calibrating the learning model; the model itself will generate the output based on its predicted values. For the warm-up, prediction on the first T time steps will involve part of the training data. Specifically, a sliding window with T dimensions will choose the last T dimensions of the N observational variables and then concatenate them into a $(T \times N)$-dimensional vector as the model input for prediction. The predicted values will then be concatenated with the ground truth data, and the sliding window will slide one time step to form the next step's input, i.e., a $(T\times N)$-dimensional vector that contains the previous prediction. The aforementioned process will be repeated, and after T steps, there will be no training data included at all. For each step, the forecast will be generated as $\mathbf{O}_l = \hat{\mathbf{x}}_l + \Delta\mathbf{x}_l$, where l denotes any time step that is not included in the training phase, the model prediction $\mathbf{O}_l$ at the time step l will be used as part of the input for the next step's prediction.

3. Theoretical guarantees of chaotic behavior learning

In this section, our goal is to gain a theoretical understanding of the method's generalizability, which we accomplish by mathematically revealing the fundamental principles and learning behaviors of the method. Firstly, we provide a Lemma, which establishes an upper bound for the forecasting error of the method, establishing the foundation of our analyses. Subsequently, we offer further theoretical guarantees of the method, specifically in relation to two key determinant indexes in chaotic systems: the Lyapunov exponent and the Hurst exponent. The variables used in our theoretical analysis are summarized and explained in Table 1.

Table 1
Variables used in the theoretical analysis and their meaning.

Notions	Descriptions
x_0	Ground truth value at the initial step of the time series to be predicted
L	Lipschitz constant
T	Length of the sliding window
n	Number of predicted steps
ϵ	Initial prediction error of the developed method at the starting step
ϵ_t	Prediction error of the developed method at time step t
λ_f	The Lyapunov exponent of the predicted time series
λ_g	The Lyapunov exponent of the ground truth time series
H_f	The Hurst exponent of the predicted time series
H_g	The Hurst exponent of the ground truth time series
R	Range of cumulative deviate time series of the ground truth
R'	Range of cumulative deviate time series of our prediction
S	Standard deviation of ground truth
S'	Standard deviation of our prediction
ξ, C, K	Positive constants

3.1. Bounding the prediction error on chaotic behavior

Lemma 1. *The prediction error of the developed method at time step t, denoted by ϵ_t, is bounded by:*

$$\epsilon_t \le \frac{\epsilon}{1-TL}, \quad (2)$$

where ϵ is the initial prediction error of the developed method at the starting step, T denotes the length of the sliding window as described in Section 2.1, L is the Lipschitz constant, and TL can be guaranteed to be a positive number satisfying $TL \in (0,1)$.

The proof of this Lemma is provided in the supplementary material. In fact, this Lemma shows that the developed method, by introducing the novel design of information tracking, can effectively alleviate the error accumulation and thus guarantee to achieve a relatively small forecasting error proportional to the initial one. This Lemma will establish the groundwork for the theoretical guarantees on characterizing the Lyapunov and the Hurst exponents, which will be derived in details in the following subsections.

3.2. Theoretical guarantee on the Lyapunov exponent

The Lyapunov exponent is an important index that reflects the level of chaos [33]. To demonstrate that our method is able to capture the essence of chaotic time series, we therefore theoretically bound the characterized Lyapunov exponent by the ground truth in the following theorem.

Theorem 1. *The Lyapunov exponent of the predicted time series, denoted by λ_f, and that of the ground truth, denoted by λ_g, have the following relation:*

$$ln[(x_0e^{n\lambda_g} - \frac{\epsilon}{1-TL})^{\frac{1}{n}}] - \frac{1}{n}ln(x_0+\epsilon) \le \lambda_f \le ln[(x_0e^{n\lambda_g} + \frac{\epsilon}{1-TL})^{\frac{1}{n}}] - \frac{1}{n}ln(x_0+\epsilon), \quad (3)$$

where x_0 represents the ground truth value at the initial step of the time series to be predicted, and n denotes the number of steps that have been predicted.

Proof. According to the definition of the Lyapunov exponent, the ground truth value at the time step n can be denoted as $x_0e^{n\lambda}$, and the prediction error (i.e., the difference between the predicted value and the ground truth) at time step t can be denoted as $\epsilon_t = \left|(x_0+\epsilon)e^{n\lambda_f} - x_0e^{n\lambda_g}\right|$. Then according to Lemma 1, we have the following:

$$\epsilon_t = \left|(x_0+\epsilon)e^{n\lambda_f} - x_0e^{n\lambda_g}\right| \le \frac{\epsilon}{1-TL}. \quad (4)$$

Taking off the absolute value notation, we can derive the following from Eq. (4):

$$\begin{aligned} & (x_0+\epsilon)e^{n\lambda_f} \le x_0 e^{n\lambda_g} + \frac{\epsilon}{1-TL} \\ \Rightarrow \quad & n\lambda_f + ln(x_0+\epsilon) \le ln(x_0 e^{n\lambda_g} + \frac{\epsilon}{1-TL}) \\ \Rightarrow \quad & \lambda_f \le ln[(x_0 e^{n\lambda_g} + \frac{\epsilon}{1-TL})^{\frac{1}{n}}] - \frac{1}{n} ln(x_0+\epsilon). \end{aligned} \tag{5}$$

Similarly, we also can derive the following inequality from Eq. (4) by taking off the absolute value sign:

$$\begin{aligned} & x_0 e^{n\lambda_g} - \frac{\epsilon}{1-TL} \le (x_0+\epsilon)e^{n\lambda_f}, \\ \Rightarrow \quad & \lambda_f \ge ln[(x_0 e^{n\lambda_g} - \frac{\epsilon}{1-TL})^{\frac{1}{n}}] - \frac{1}{n} ln(x_0+\epsilon). \end{aligned} \tag{6}$$

Combining the last inequality in Eq. (5) and that in Eq. (6), we obtain the inequalities stated in Theorem 1. □

Note that as n increases and gradually tends to positive infinity, the influence of $\frac{\epsilon}{1-TL}$ and $\frac{1}{n} ln(x_0+\epsilon)$ in Eqs. (5)–(6) will gradually disappear. As a result, the right-hand side of the last inequality in both Eqs. (5) and (6) will approach λ_g, indicating that the characterized Lyapunov exponent, λ_f, is well bounded by the ground truth λ_g.

3.3. Theoretical guarantee on the Hurst exponent

In addition to the Lyapunov exponent, the Hurst exponent, which describes the property of self-similarity and is related to fractal dimension (i.e., $D = 2 - H$ [39], where D and H denote the fractal dimension and the Hurst exponent, respectively), is another widely used index to describe the level of chaos [40]. In this subsection, we theoretically bound the Hurst exponent of our predicted time series by that of the ground truth in the following theorem.

Theorem 2. *Assume that neither the given time series to be predicted nor our prediction on it is a horizontal line (i.e., with some variations over time), then the Hurst exponent of our prediction, denoted by H_f, and that of the ground truth, denoted by H_g, have the following relation:*

$$\begin{aligned} H_g - \frac{1}{ln(n)} ln(\frac{(S+3p)RK}{max\{R-4np, \xi\}SC}) \quad \le \quad & H_f \quad \le \quad H_g \\ & + \frac{1}{ln(n)} ln(\frac{(R+4np)SC}{max\{S-3p, \xi\}RK}), \end{aligned} \tag{7}$$

where $p = \frac{\epsilon}{1-TL}$, R is the range (i.e., the difference between the maximum value and the minimum value) of the cumulative deviate time series of the ground truth (i.e., for a time series $\mathbf{a} = [a_1, a_2, \dots, a_n]$, the cumulative deviate time series $\mathbf{A} = [A_1, A_2, \dots, A_n]$ is calculated as $A_i = \sum_{j=1}^{i}(a_j - m)$, where m is the mean value of time series $\mathbf{a}$), S is the standard deviation of the ground truth, and ξ, C and K are positive constants. We use R' to denote the range of the cumulative deviate series of our prediction, and S' to denote the standard deviation of our prediction.

Proof. According to the definition of the Hurst exponent, we have:

$$\begin{aligned} & ln(R/S) = lnC + H_g ln(n), \\ & ln(R'/S') = lnK + H_f ln(n), \end{aligned} \tag{8}$$

where C and K are positive constants.

Let a and b denote the ground truth and our prediction of the time series, respectively, A and B denote the cumulative deviate series of the ground truth and our prediction, respectively, and m and m' denote the mean value of the ground truth and our prediction, respectively, we have the following relations:

$$\begin{aligned} & A_i = \sum_{j=1}^{i}(a_j - m), \ B_i = \sum_{j=1}^{i}(b_j - m'), \\ & m = \frac{1}{n}\sum_{i=1}^{n} a_i, \ m' = \frac{1}{n}\sum_{i=1}^{n} b_i, \\ & R = Max(A) - Min(A) > 0, \ R' = Max(B) - Min(B) > 0, \\ & S^2 = \frac{1}{n}\sum_{i=1}^{n}(a_i - m)^2 > 0, \ (S')^2 = \frac{1}{n}\sum_{i=1}^{n}(b_i - m')^2 > 0. \end{aligned} \tag{9}$$

To give the bound on the Hurst exponent, we start by deriving the relationship of the range and standard deviation between our prediction and the ground truth. First, we establish the relationship between the range of our result and the ground truth as follows:

$$\begin{aligned} |R' - R| \quad & = | \ Max(B) - Min(B) - Max(A) + Min(A)| \\ & \le \ |Min(A) - Min(B)| + |Max(A) - Max(B)|. \end{aligned} \tag{10}$$

Then we analyze $|Min(A) - Min(B)|$ and $|Max(A) - Max(B)|$ separately. Let d and q denote the indexes of $Min(A)$ and $Min(B)$, respectively, then for the relation between d and q, there are three different situations:

Situation 1: If $d = q$, then we have the following:

$$\begin{aligned} |Min(A) - Min(B)| &= |\sum_{j=1}^{d}(a_j - m) - \sum_{j=1}^{q}(b_j - m')| \\ &= |\sum_{j=1}^{d}(a_j - b_j) - \sum_{j=1}^{d}(m - m')| \\ &\le 2dp. \end{aligned} \tag{11}$$

Situation 2: If $d < q$, then we have the following:

$$\begin{aligned} |Min(A) - Min(B)| &= |\sum_{j=1}^{d}(a_j - m) - \sum_{j=1}^{d}(b_j - m') - \sum_{j=d+1}^{q}(b_j - m')| \\ &\le |\sum_{j=1}^{d}(a_j - b_j) - \sum_{j=1}^{d}(m - m')| + |\sum_{j=d+1}^{q} b_j - (q-d)m'| \\ &\le 2dp + |\sum_{j=d+1}^{q} b_j - (q-d)m'|. \end{aligned} \tag{12}$$

To bound the term $|\sum_{j=d+1}^{q} b_j - (q-d)m'|$, we construct the following inequality:

$$\begin{aligned} & |(\sum_{j=d+1}^{q} a_j - (q-d)m) - (\sum_{j=d+1}^{q} b_j - (q-d)m')| \\ = \ & |\sum_{j=d+1}^{q}(a_j - b_j) - (q-d)(m - m')| \\ \le \ & 2(q-d)p. \end{aligned} \tag{13}$$

As d and q denote the indexes of $Min(A)$ and $Min(B)$, respectively, it is easy to know that in Eq. (13), the term $(\sum_{j=d+1}^{q} a_j - (q-d)m)$ is positive (otherwise, d will not be the index of $Min(A)$) and the term $(\sum_{j=d+1}^{q} b_j - (q-d)m')$ is negative (otherwise, q will not be the index of $Min(B)$). As a result, we have the following:

$$\begin{aligned} & |(\sum_{j=d+1}^{q} a_j - (q-d)m) - (\sum_{j=d+1}^{q} b_j - (q-d)m')| \ \le \ 2(q-d)p \\ \Rightarrow \ & |\sum_{j=d+1}^{q} a_j - (q-d)m| + |\sum_{j=d+1}^{q} b_j - (q-d)m'| \le 2(q-d)p \\ \Rightarrow \ & |\sum_{j=d+1}^{q} b_j - (q-d)m'| \le 2(q-d)p \end{aligned} \tag{14}$$

Combining Eqs. (12) and (14), we have the following:

$$|Min(A) - Min(B)| \le 2qp \tag{15}$$

Situation 3: If $d > q$, by following the same process in Situation 2, we have the following:

$$|Min(A) - Min(B)| \leq 2dp. \tag{16}$$

Combining the aforementioned three situations, we know that for $|Min(A) - Min(B)|$, it is bounded as follows:

$$|Min(A) - Min(B)| \leq 2Max\{d, q\}p. \tag{17}$$

Similarly, for $|Max(A) - Max(B)|$, let r and k be the indexes of $Max(A)$ and $Max(B)$, respectively, we have the following inequality:

$$|Max(A) - Max(B)| \leq 2Max\{r, k\}p. \tag{18}$$

Combining Eqs. (10), (17), and (18), we have the following:

$$\begin{aligned} |R' - R| &\leq 2(Max\{r, k\} + Max\{d, q\})p \\ &\leq 4np. \end{aligned} \tag{19}$$

Next, we will derive the relation between S and S'. Let $u = (m' - m)$, we have:

$$\begin{aligned} & S^2 = \frac{1}{n}\sum_{i=1}^{n}(a_i - m)^2 \\ \Rightarrow\ & nS^2 = \sum_{i=1}^{n}[(a_i - b_i) + (b_i - m') + u]^2 \\ &= \sum_{i=1}^{n}(a_i - b_i)^2 + 2u\sum_{i=1}^{n}(a_i - b_i) + nu^2 + 2\sum_{i=1}^{n}(a_i - b_i)(b_i - m') \\ &\quad + 2u\sum_{i=1}^{n}(b_i - m') + \sum_{i=1}^{n}(b_i - m')^2 \\ &\leq n(p^2 + 2up + u^2) + 2\sum_{i=1}^{n}(a_i - b_i)(b_i - m') \\ &\quad + 2u\sum_{i=1}^{n}(b_i - m') + n(S')^2. \end{aligned} \tag{20}$$

It is easy to know that

$$\sum_{i=1}^{n}(b_i - m') = \sum_{i=1}^{n} b_i - \sum_{i=1}^{n} m' = nm' - nm' = 0. \tag{21}$$

Moreover, according to the Cauchy–Schwarz inequality, we have

$$\sum_{i=1}^{n}(a_i - b_i)(b_i - m') \leq \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2 \sum_{i=1}^{n}(b_i - m')^2} \leq \sqrt{np^2 n(S')^2} = npS'. \tag{22}$$

Combining Eqs. (20), (21) and (22), we have the following:

$$\begin{aligned} S^2 &\leq p^2 + 2up + u^2 + 2pS' + (S')^2 \\ &= (S' + p)^2 + (u + p)^2 - p^2 \\ &\leq (S' + u + 2p)^2 - p^2 \\ &\leq (S' + 3p)^2 \\ \Rightarrow\ & S \leq S' + 3p. \end{aligned} \tag{23}$$

As the relation between S and S' is symmetric, following the same procedure above, we have the inequality:

$$S' \leq S + 3p. \tag{24}$$

Since $R, R', S, S' > 0$, we have $R, R', S, S' \geq \xi > 0$, where ξ is a positive constant. According to Eq. (8), we have:

$$\begin{aligned} & H_f - H_g = \frac{1}{ln(n)} ln\left(\frac{R'SC}{S'RK}\right) \leq \frac{1}{ln(n)} ln\left(\frac{(R + 4np)SC}{max\{S - 3p, \xi\}RK}\right) \\ \Rightarrow\ & H_f \leq H_g + \frac{1}{ln(n)} ln\left(\frac{(R + 4np)SC}{max\{S - 3p, \xi\}RK}\right). \end{aligned} \tag{25}$$

Similarly, we have:

$$H_f \geq H_g - \frac{1}{ln(n)} ln\left(\frac{(S + 3p)RK}{max\{R - 4np, \xi\}SC}\right). \tag{26}$$

Combining Eqs. (25)–(26), we obtain the inequalities stated in Theorem 2. □

4. Experiments

In this section, we examine the generalizability and robustness of the method in characterizing and predicting chaotic behavior on a variety of chaotic behavioral learning problems, generated by three representative and widely used dynamical systems: the one-dimensional logistic map [35], the two-dimensional double pendulum [36], and the three-dimensional Lorenz system [37], with different levels of noise ratios, and using multiple evaluation metrics. First, we examine the validity of our theoretical analyses by comparing the Lyapunov and the Hurst exponents calculated from the time series predicted by our method and that of the ground truth. To further demonstrate the ability of the developed method in learning chaotic dynamics, we compare its prediction performance with representative data-driven methods on the aforementioned three chaotic systems.

4.1. Experimental settings

We select six representative data-driven methods that have been widely used in chaotic/complex time series prediction as the baseline in our experiments: long short-term memory (LSTM) [24], gated recurrent unit (GRU) [23], echo state network (ESN) [41], dynamics memristor echo state network (DMESN) [42], deep echo state network (Deepesn) [43], and Informer [27]. To evaluate the performance of prediction, We use the dynamic time wrapping (DTW) [44] as the evaluation metric, as it is specifically developed to calculate the similarity of two temporal sequences. The smaller value of DTW indicates better performance. To test the robustness of the developed method, we conduct experiments on both the original systems and the systems with different levels of random noise. We use α to denote the noise level, and we set it to $0, 0.01, 0.02$, and 0.03, respectively. For LSTM and GRU, we use one layer and select the hidden state size from {8, 16, 32, 64} (the parameter with the best performance will be selected, and the corresponding performance will be reported), which is a standard approach for parameter selection [45]. For ESN, we pick the spectral radius from {0.5, 1.0, 1.5} (similarly, the parameter with the best performance will be selected, and the corresponding performance will be reported) [46]. For DMESN, Deepesn, and Informer, we follow the default settings reported in their original papers to select the parameters [27,42,43]. We generate chaotic time series for 10,000 steps with $\Delta t = 0.01$. We use the 80% of the data, i.e., the first 8,000 steps for training and the rest for test.

4.2. Characterization of chaotic behavior by the Lyapunov and the Hurst exponents

Both the Lyapunov and the Hurst exponents are essential indexes to measure the degrees of chaos and complexity of a dynamical system. In this subsection, we thus use these two indexes to evaluate the capability of the developed method in characterizing chaotic behavior. The ground truth of the Lyapunov exponent and that of the Hurst exponent of the three systems with different noise ratios are given in the last column of Tables 2 and 3, respectively. These two exponents of the predicted time series generated by the aforementioned seven methods are also reported in these two tables. We can observe that the exponents calculated from the times series predicted by our method (highlighted in bold) are closer to the ground truth compared with that of other methods, showing the ability of the developed method to capture the chaotic dynamics. Moreover, with the increase of noise ratio in each system, the performance of our method is relatively robust, demonstrating the potential of the developed method in learning chaotic patterns from observations with random noise.

Table 2
The Lyapunov exponent of the ground truth time series and that of the predicted time series generated by the LSTM [24], GRU [23], ESN [41], DMESN [42], Deepesn [43], Informer [27], and our method on the logistic map, double pendulum, and Lorenz systems with different noise ratios. The results of our method are highlighted in bold.

Task	LSTM	GRU	ESN	DMESN	Deepesn	Informer	Ours	Ground Truth
Logistic Map ($\alpha = 0$)	−0.06	0.02	0.08	−0.22	0.01	0.14	**0.27**	*0.26*
Logistic Map ($\alpha = 0.01$)	−0.03	−0.01	0.08	−0.27	0.0	0.73	**0.24**	*0.29*
Logistic Map ($\alpha = 0.02$)	−0.12	0.02	0.08	−0.41	0.01	0.32	**0.24**	*0.27*
Logistic Map ($\alpha = 0.03$)	−0.03	−0.01	0.09	−0.26	0.01	0.32	**0.21**	*0.25*
Double pendulum ($\alpha = 0$)	0.01	0	0.11	−0.51	0.01	0.1	**0.13**	*0.17*
Double pendulum ($\alpha = 0.01$)	0.01	−0.01	0.08	−0.50	0.01	0.13	**0.15**	*0.17*
Double pendulum ($\alpha = 0.02$)	0.01	0.01	0.09	−0.48	0.01	0.1	**0.12**	*0.17*
Double pendulum ($\alpha = 0.03$)	0.01	0.01	0.09	−0.49	0.01	0.1	**0.13**	*0.16*
Lorenz ($\alpha = 0$)	−0.11	−0.02	0.11	−0.41	0.0	0.13	**0.04**	*0.04*
Lorenz ($\alpha = 0.01$)	−0.06	−0.04	0.11	−0.40	0.0	0.15	**0.04**	*0.04*
Lorenz ($\alpha = 0.02$)	−0.04	0	0.08	−0.40	0.0	0.14	**0.04**	*0.04*
Lorenz ($\alpha = 0.03$)	0	0.01	0.08	−0.37	0.01	0.16	**0.05**	*0.04*

Table 3
The Hurst exponent of the ground truth time series and that of the predicted time series generated by the LSTM [24], GRU [23], ESN [41], DMESN [42], Deepesn [43], Informer [27], and our method on the logistic map, double pendulum, and Lorenz systems with different noise ratios. The results of our method are highlighted in bold.

Task	LSTM	GRU	ESN	DMESN	Deepesn	Informer	Ours	Ground Truth
Logistic Map ($\alpha = 0$)	−0.06	0.02	0.16	1.0	0.30	0.84	**0.14**	*0.08*
Logistic Map ($\alpha = 0.01$)	−0.03	−0.01	0.20	0.94	0.66	0.35	**0.12**	*0.05*
Logistic Map ($\alpha = 0.02$)	−0.12	0.02	0.18	0.94	0.30	0.52	**0.09**	*0.06*
Logistic Map ($\alpha = 0.03$)	−0.03	−0.01	0.17	0.94	0.24	0.68	**0.14**	*0.07*
Double pendulum ($\alpha = 0$)	0.32	0.59	0.27	1.0	0.04	0.94	**0.81**	*0.75*
Double pendulum ($\alpha = 0.01$)	0.01	0.33	0.87	0.80	0.04	0.91	**0.81**	*0.75*
Double pendulum ($\alpha = 0.02$)	0.34	0.48	0.63	1.0	0.04	0.90	**0.74**	*0.75*
Double pendulum ($\alpha = 0.03$)	0.33	0.48	0.41	1.0	0.03	0.89	**0.76**	*0.75*
Lorenz ($\alpha = 0$)	0.96	0.94	0.92	0.80	0.42	0.94	**0.91**	*0.88*
Lorenz ($\alpha = 0.01$)	0.86	1	0.95	0.95	0.32	0.90	**0.90**	*0.91*
Lorenz ($\alpha = 0.02$)	1	0.93	0.88	1.0	0.26	0.90	**0.90**	*0.91*
Lorenz ($\alpha = 0.03$)	0.96	0.95	0.83	1.0	0.39	0.93	**0.91**	*0.90*

Table 4
Prediction evaluation results (in terms of the dynamic time wrapping (DTW); the smaller, the better) of LSTM [24], GRU [23], ESN [41], DMESN [42], Deepesn [43], Informer [27], and our method on the logistic map with different noise ratios. The results of our method under each noise ratio setting are highlighted in bold.

Task	LSTM	GRU	ESN	DMESN	Deepesn	Informer	Ours
Logistic Map ($\alpha = 0$)	241.23	239.13	73.26	241.05	240.96	1796.97	**29.94**
Logistic Map ($\alpha = 0.01$)	494.57	250.37	89.97	244.32	244.59	2061.58	**27.31**
Logistic Map ($\alpha = 0.02$)	306.58	231.51	79.98	239.28	239.32	1750.31	**26.83**
Logistic Map ($\alpha = 0.03$)	295.48	231.41	79.43	237.73	237.52	2287.51	**25.49**

4.3. Prediction on one-dimensional system: Logistic map

In this subsection, we test the performance of the developed method using a prediction task on the logistic map [35], which is a one-dimensional (1-D) discrete chaotic system and has been widely used to study the dynamics of population growth. The system equation of the logistic map, with the random noise term, is given as follows:

$$x_{n+1} = (1-\alpha)[rx_n(1-x_n)] + \alpha N(0,1). \tag{27}$$

The predicted time series of the aforementioned seven methods (shown in orange), together with the ground truth (shown in blue), are presented in Fig. 1 . Based on the novel design of information tracking, the developed method (shown in Fig. 1(b)) is the only one that can well predict the chaotic time series over 2,000 time steps (for the clarity and visibility of the figure, we only show the results on the first 100 time steps; the results on the remaining 1,900 time steps exhibit similar patterns). The effectiveness of the developed method is further validated by the evaluation results (in terms of the DTW, the smaller, the better) reported in Table 4.

4.4. Prediction on two-dimensional system: Double pendulum

We further evaluate the effectiveness of our method on the well-known two-dimensional (2-D) chaotic system, double pendulum [36]. The system equations with the random noise term are given as follows:

$$\begin{aligned} x_1 &= (1-\alpha)L_1 sin\theta_1(t) + \alpha N(0,1), \\ y_1 &= -(1-\alpha)L_1 cos\theta_1(t) + \alpha N(0,1), \\ x_2 &= (1-\alpha)(L_2 sin\theta_2(t) + x_1) + \alpha N(0,1), \\ y_2 &= -(1-\alpha)(L_2 cos\theta_2(t) + y_1) + \alpha N(0,1). \end{aligned} \tag{28}$$

The prediction results of the seven methods (shown in orange) and the ground truth (shown in blue) on both the X-axis and Y-axis of the double pendulum system are shown in Fig. 2. Similarly, for the reason of clarity and visibility, we only show the results on the first 500 time steps, while the patterns of the remaining 1,500 steps are similar. Table 5 summarizes the evaluation results on DTW of all the methods. Consistent with the results on the logistic map, our method (as shown in Fig. 2(b) and the last column of Table 5) achieves the best performance among all seven methods in different settings on the double pendulum system, demonstrating the validity of the design of information tracking.

4.5. Prediction on three-dimensional system: Lorenz system

In addition to the experiments on the logistic map and the double pendulum systems, we also validate the developed method on a typical three-dimensional chaotic system, the Lorenz system [37]. The equations of the Lorenz system, with the random noise term, are given as

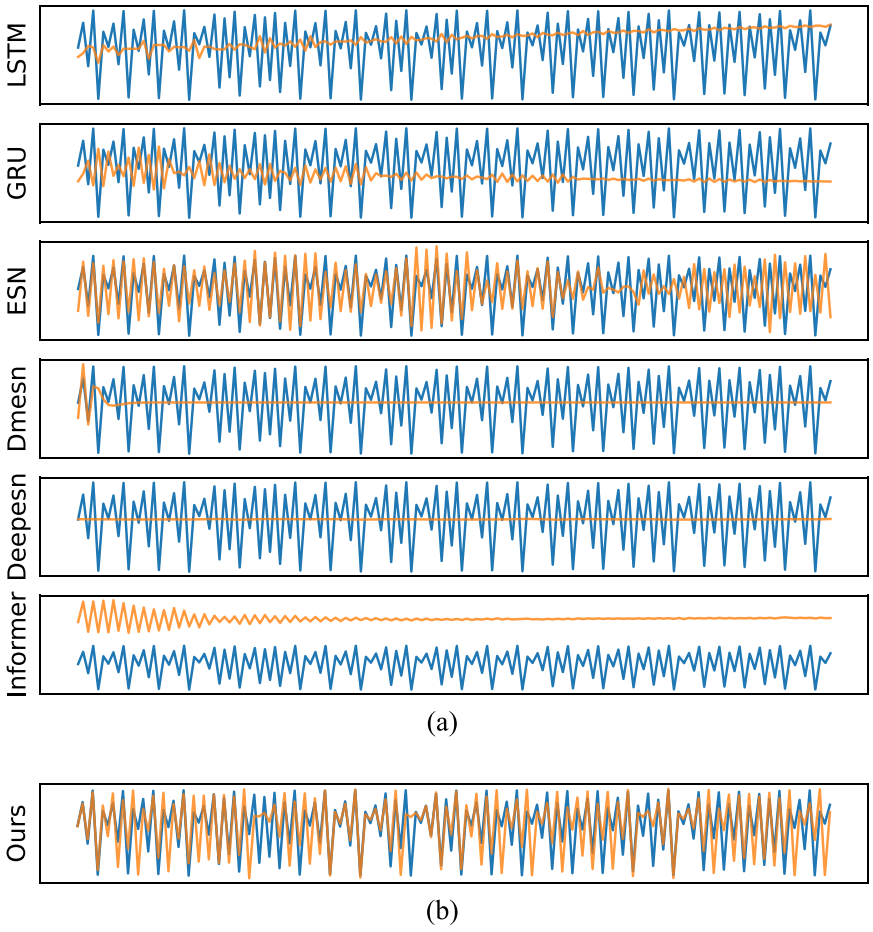


Fig. 1. The chaotic time series predicted by (A) LSTM [24], GRU [23], ESN [41], DMESN [42], Deepesn [43], Informer [27], and (B) our method, together with the ground truth, on the Logistic map without noise. The orange curves are the predictions, and the blue curves are the ground truth. For figure's clarity and visibility, we only show the prediction results of all methods on the first 100 time steps; results on the remaining 1,900 time steps show similar patterns. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5
Prediction evaluation results (in terms of DTW) of LSTM [24], GRU [23], ESN [41], DMESN [42], Deepesn [43], Informer [27], and our method on the double pendulum system with different noise ratios. The results of our method under each noise ratio setting are highlighted in bold.

Task	LSTM	GRU	ESN	DMESN	Deepesn	Informer	Ours
Double pendulum ($\alpha = 0$)	132.13	132.66	27.04	113.74	117.70	2186.16	**23.22**
Double pendulum ($\alpha = 0.01$)	135.14	131.23	22.48	113.91	117.39	2166.25	**16.60**
Double pendulum ($\alpha = 0.02$)	135.31	130.42	31.47	113.48	116.48	2191.18	**20.26**
Double pendulum ($\alpha = 0.03$)	134.39	129.45	32.62	112.90	115.46	2167.98	**22.77**

Table 6
Prediction evaluation results (in terms of DTW) of LSTM [24], GRU [23], ESN [41], DMESN [42], Deepesn [43], Informer [27], and our method on the Lorenz system with different noise ratios. The results of our method under each noise ratio setting are highlighted in bold.

Task	LSTM	GRU	ESN	DMESN	Deepesn	Informer	Ours
Lorenz ($\alpha = 0$)	40.79	37.39	18.73	42.17	60.46	1864.23	**10.18**
Lorenz ($\alpha = 0.01$)	64.16	71.29	10.27	52.74	68.25	1789.48	**5.94**
Lorenz ($\alpha = 0.02$)	100.17	86.19	20.50	68.83	72.45	2203.91	**10.39**
Lorenz ($\alpha = 0.03$)	49.51	52.36	22.22	59.90	67.45	1965.69	**13.31**

follows:

$$\begin{aligned}\frac{dx}{dt} &= (1-\alpha)s(y-x) + \alpha N(0,1),\\ \frac{dy}{dt} &= (1-\alpha)(rx-y-xz) + \alpha N(0,1),\\ \frac{dz}{dt} &= (1-\alpha)(xy-bz) + \alpha N(0,1),\end{aligned} \tag{29}$$

We demonstrate the experimental results on the Lorenz System in Fig. 3 and Table 6. The same as before, our method (as shown in Fig. 3(b) and the last column of Table 6) outperforms the other six methods (as shown in Fig. 3(a) and columns 2–7 of Table 6) in forecasting the chaotic time series. An interesting observation is that on all three dynamical systems, the ESN achieves the second-best

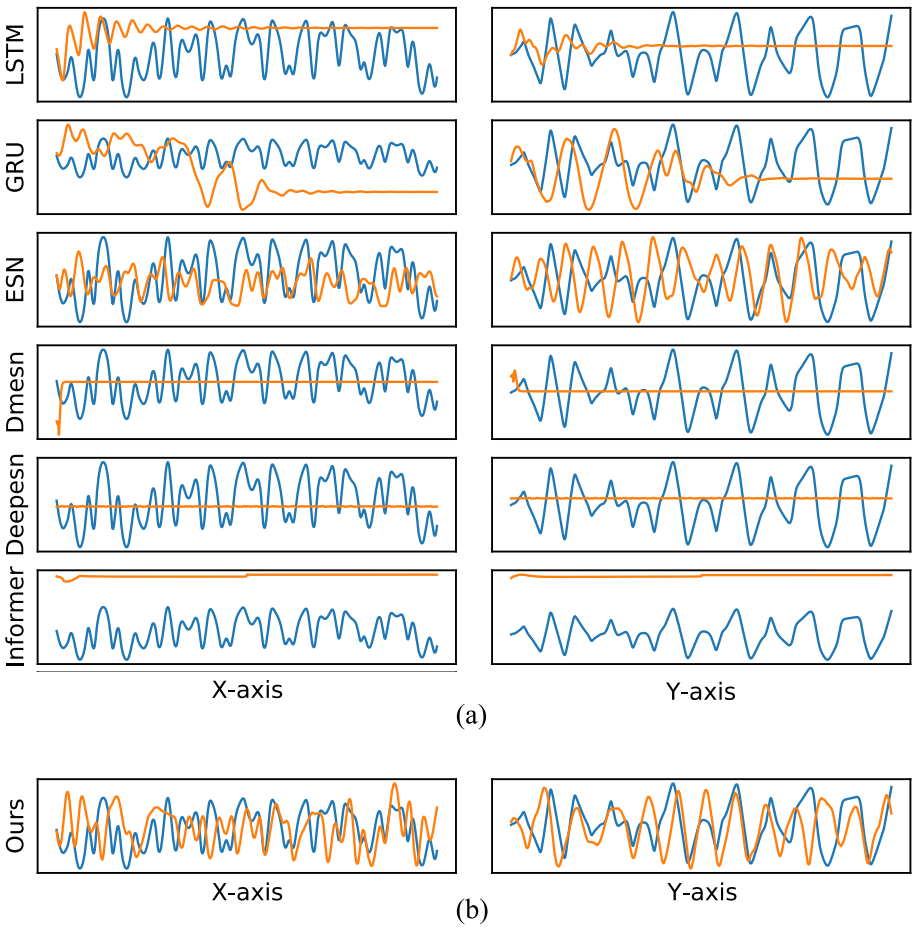


Fig. 2. The chaotic time series predicted by (A) LSTM [24], GRU [23], ESN [41], DMESN [42], Deepesn [43], Informer [27], and (B) our method, together with the ground truth, on the double pendulum system without noise. The orange curves are the predictions, and the blue curves are the ground truth. For figure clarity and visibility, we only show the prediction results of all methods on the first 500 time steps; results on the remaining 1,500 time steps show similar patterns. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

results. A possible reason is that the randomly connected neurons in the maintained reservoir could simulate the chaotic behavior to some extent, thus providing acceptable prediction results on chaotic time series.

5. Conclusion

Building upon our preliminary work [32], this paper constituted a significant step forward by answering the open questions of the generalizability of the concept of information tracking from the perspective of method formalism, theoretical understanding, and systematic validation. Based on a formal mathematical statement of the problem and the general method, our rigorous theoretical analyses revealed that the method is guaranteed to model the essence of the chaotic behavior, as expressed in terms of the Lyapunov and the Hurst exponents. Comprehensive empirical validations on various chaotic behavioral learning problems demonstrated the generalizability and robustness of the method in learning and predicting chaotic time series. In our future work, we will consider applying and tailoring the method as well as our theoretical analyses to more chaotic systems in both well-defined simulations and real-world applications.

CRediT authorship contribution statement

Jinfu Ren: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Yang Liu:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing , Project administration, Funding acquisition. **Jiming Liu:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

The authors thank the Editor-in-Chief, the Section Editor, the Handling Editor, and anonymous reviewers for their constructive comments

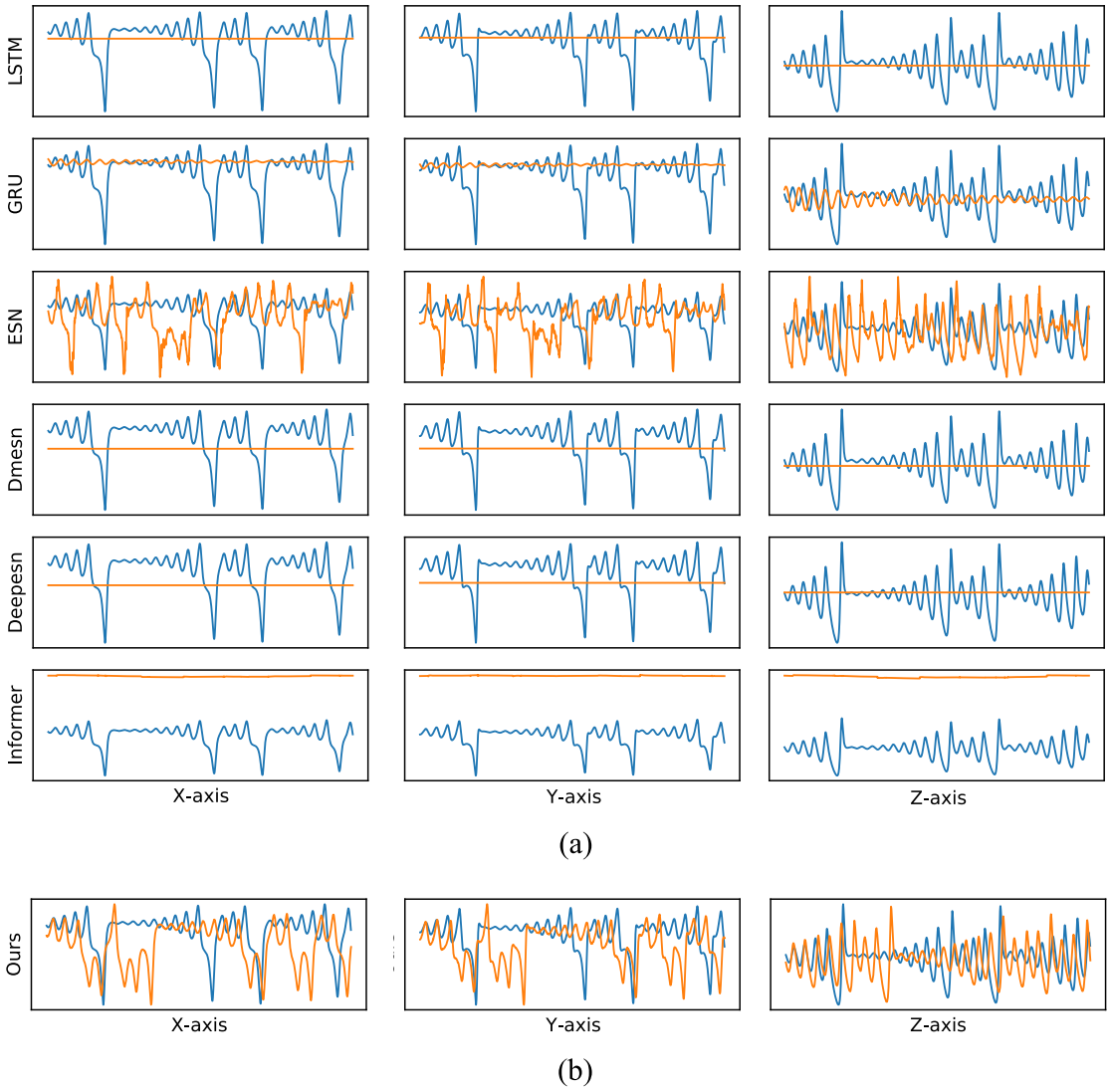


Fig. 3. The chaotic time series predicted by (A) LSTM [24], GRU [23], ESN [41], DMESN [42], Deepesn [43], Informer [27], and (B) our method, together with the ground truth, on the Lorenz system without noise. The orange curves are the predictions, and the blue curves are the ground truth. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and suggestions. This work was supported in part by the National Key Research and Development Program of China from the Ministry of Science and Technology of China under the Projects 2021ZD0112501 and 2021ZD0112502, in part by the National Natural Science Foundation of China/RGC Joint Research Scheme under the Project N_HKBU222/22, and in part by the General Research Fund from the Research Grant Council of Hong Kong SAR under Projects RGC/HKBU12201619 and RGC/HKBU12203122.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.chaos.2023.113927.

References

[1] Pecora LM, Carroll TL. Synchronization in chaotic systems. Phys Rev Lett 1990;64(8):821.
[2] Grassberger P, Procaccia I. Characterization of strange attractors. Phys Rev Lett 1983;50(5):346.
[3] Strogatz SH. Nonlinear dynamics and chaos with student solutions manual: With applications to physics, biology, chemistry, and engineering. CRC Press; 2018.
[4] Scott SK. Chemical chaos, no. 24. Oxford University Press; 1993.
[5] Coffey DS. Self-organization, complexity and chaos: the new biology for medicine. Nat Med 1998;4(8):882–5.
[6] Scharf Y. A chaotic outlook on biological systems. Chaos Solitons Fractals 2017;95:42–7.
[7] Watson PA. Applying machine learning to improve simulations of a chaotic dynamical system using empirical error correction. J Adv Modelling Earth Syst 2019;11(5):1402–17.
[8] Olsen PE, Laskar J, Kent DV, Kinney ST, Reynolds DJ, Sha J, et al. Mapping solar system chaos with the geological orrery. Proc Natl Acad Sci 2019;116(22):10664–73.
[9] Skinner JE, Goldberger AL, Mayer-Kress G, Ideker RE. Chaos in the heart: implications for clinical cardiology. Nature Biotechnol 1990;8(11):1018–24.
[10] Chen W-C. Nonlinear dynamics and chaos in a fractional-order financial system. Chaos Solitons Fractals 2008;36(5):1305–14.
[11] Vogl M. Controversy in financial chaos research and nonlinear dynamics: A short literature review. Chaos Solitons Fractals 2022;162:112444.
[12] Eyring V, Bony S, Meehl GA, Senior CA, Stevens B, Stouffer RJ, et al. Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. Geosci Model Dev 2016;9(5):1937–58.
[13] Solaraju-Murali B, Bojovic D, Gonzalez-Reviriego N, Nicodemou A, Terrado M, Caron L-P, et al. How decadal predictions entered the climate services arena: An example from the agriculture sector. Clim Serv 2022;27:100303.
[14] World Health OrganizationRegional Office for Europe & European Commission. Improving public health responses to extreme weather/Heat-waves: EuroHEAT: Summary for Policy-makers. WHO: World Health Organization; 2009.

[15] Elbert T, Ray WJ, Kowalik ZJ, Skinner JE, Graf KE, Birbaumer N. Chaos and physiology: deterministic chaos in excitable cell assemblies. Physiol Rev 1994;74(1):1–47.
[16] Peters EE. Chaos and order in the capital markets: A new view of cycles, prices, and market volatility. John Wiley & Sons; 1996.
[17] Murphy JM, Sexton DM, Barnett DN, Jones GS, Webb MJ, Collins M, et al. Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature 2004;430(7001):768–72.
[18] Palmer T. Stochastic weather and climate models. Nat Rev Phys 2019;1(7):463–71.
[19] Helbing D. Agent-based modeling. Springer; 2012.
[20] Shinbrot T, Ditto W, Grebogi C, Ott E, Spano M, Yorke JA. Using the sensitive dependence of chaos (the "butterfly effect") to direct trajectories in an experimental chaotic system. Phys Rev Lett 1992;68(19):2863.
[21] Keenlyside N, Latif M, Jungclaus J, Kornblueh L, Roeckner E. Advancing decadal-scale climate prediction in the north atlantic sector. Nature 2008;453(7191):84–8.
[22] on Climate Change (IPCC) IP. Long-term climate change: Projections, commitments and irreversibility pages 1029 to 1076. In: Climate change 2013 – the physical science basis: Working group I contribution to the fifth assessment report of the intergovernmental panel on climate change. Cambridge University Press; 2014, p. 1029–136. http://dx.doi.org/10.1017/CBO9781107415324.024.
[23] Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. 2014, arXiv preprint arXiv:1406.1078.
[24] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9(8):1735–80.
[25] Jaeger H, Haas H. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. Science 2004;304(5667):78–80.
[26] Chen H-C, Wei D-Q. Chaotic time series prediction using echo state network based on selective opposition grey wolf optimizer. Nonlinear Dynam 2021;104(4):3925–35.
[27] Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting. In: Proceedings of the AAAI conference on artificial intelligence, vol. 35, no. 12. 2021, p. 11106–15.
[28] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Advances in neural information processing systems, vol. 30. 2017.
[29] Lou S, Deng J, Lyu S. Chaotic signal denoising based on simplified convolutional denoising auto-encoder. Chaos Solitons Fractals 2022;161:112333.
[30] Kavuran G. When machine learning meets fractional-order chaotic signals: detecting dynamical variations. Chaos Solitons Fractals 2022;157:111908.
[31] Boullé N, Dallas V, Nakatsukasa Y, Samaddar D. Classification of chaotic time series with deep learning. Physica D 2020;403:132261.
[32] Ren J, Liu Y, Liu J. Decadal temperature prediction via chaotic behavior tracking. 2023, arXiv preprint arXiv:2304.09536.
[33] Wolf A, Swift JB, Swinney HL, Vastano JA. Determining Lyapunov exponents from a time series. Physica D 1985;16(3):285–317.
[34] Hurst HE. Long-term storage capacity of reservoirs. Trans Am Soc Civ Eng 1951;116(1):770–99.
[35] Phatak S, Rao SS. Logistic map: A possible random-number generator. Phys Rev E 1995;51(4):3670.
[36] Shinbrot T, Grebogi C, Wisdom J, Yorke JA. Chaos in a double pendulum. Amer J Phys 1992;60(6):491–9.
[37] Lorenz EN. Deterministic nonperiodic flow. J Atmos Sci 1963;20(2):130–41.
[38] Kingma DP, Welling M. Auto-encoding variational bayes. 2013, arXiv preprint arXiv:1312.6114.
[39] Rehman S, Siddiqi A. Wavelet based hurst exponent and fractal dimensional analysis of saudi climatic dynamics. Chaos Solitons Fractals 2009;40(3):1081–90.
[40] Sano M, Sawada Y. Measurement of the Lyapunov spectrum from a chaotic time series. Phys Rev Lett 1985;55(10):1082.
[41] Khovanov IA. Stochastic approach for assessing the predictability of chaotic time series using reservoir computing. Chaos 2021;31(8):083105.
[42] Zhong Y, Tang J, Li X, Gao B, Qian H, Wu H. Dynamic memristor-based reservoir computing for high-efficiency temporal signal processing. Nature Commun 2021;12(1):408.
[43] Gallicchio C, Micheli A, Pedrelli L. Design of deep echo state networks. Neural Netw 2018;108:33–47.
[44] Müller M. Dynamic time warping, information retrieval for music and motion. Springer Berlin Heidelberg; 2007, p. 69–84.
[45] Yang L, Shami A. On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing 2020;415:295–316.
[46] Ozturk MC, Xu D, Principe JC. Analysis and design of echo state networks. Neural Comput 2007;19(1):111–38.