

PAPER • OPEN ACCESS

Exploring the potential of contemporary deep learning methods in purifying polluted information

To cite this article: Mingyu Wang and Jianping Li 2024 *Mach. Learn.: Sci. Technol.* **5** 045026

View the [article online](#) for updates and enhancements.

You may also like

- [Enhanced deep learning-based water area segmentation for flood detection and monitoring](#)

Thang M Pham, Nam Do, Hanh T Bui et al.

- [Improved deep semantic dictionary learning for multi-label image classification](#)

Shihan Zhang

- [Intensive agriculture, nitrogen legacies, and water quality: intersections and implications](#)

Idhayachandiran Ilampooranan, Kimberly J Van Meter and Nandita B Basu



OPEN ACCESS

RECEIVED
15 July 2024REVISED
13 October 2024ACCEPTED FOR PUBLICATION
21 October 2024PUBLISHED
28 October 2024

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



PAPER

Exploring the potential of contemporary deep learning methods in purifying polluted information

Mingyu Wang¹ and Jianping Li^{1,2,*}

¹ Frontiers Science Center for Deep Ocean Multi-spheres and Earth System (DOMES)/Key Laboratory of Physical Oceanography/Academy of Future Ocean/College of Oceanic and Atmospheric Sciences/ Center for Ocean Carbon Neutrality, Ocean University of China, Qingdao 266100, People's Republic of China

² Laboratory for Ocean Dynamics and Climate, Qingdao Marine Science and Technology Center, Qingdao 266237, People's Republic of China

* Author to whom any correspondence should be addressed.

E-mail: ljp@ouc.edu.cn

Keywords: deep learning methods, purify polluted information, dynamical system deep learning

Abstract

Accurately predicting chaotic dynamical systems is a crucial task in various fields, and recent advancements have leveraged deep learning for this purpose. However, in the era of big data, the inevitable challenge of data contamination caused by invalid information from other interfering systems becomes increasingly prominent and complicates accurate predictions. Although contemporary deep learning methods have shown their potential, very few studies have focused on developing algorithms specifically designed to address the problem posed by such data contamination. Thus, exploring the ability of contemporary deep learning methods to purify polluted information fills an important gap in the current body of research. This study explores the performance and stability of several modern deep learning techniques for predicting chaotic systems using datasets polluted by invalid information. Our findings reveal that while most of the state-of-the-art deep learning methods exhibit reduced and unstable predictive performance owing to such contamination, the dynamical system deep learning (DSDL) method stands out, remaining unaffected by any interference. This breakthrough illustrates DSDL's unique ability to purify invalid data and uncover the inherent rules of chaotic systems. As we move forward, DSDL paves the way for a more reliable and interpretable model, ensuring that we can confidently predict chaotic systems with precision even in the most challenging environments.

1. Introduction

Decades of research have explored that chaotic systems have a wide range of applications across various fields. For instance, chaos systems help explain the sensitivity of weather systems (the 'butterfly effect') and deal with the nonlinear dynamics of the atmosphere and ocean [1]; chaotic systems are used to analyze market fluctuations and predict asset price movements in the field of finance [2]; chaotic systems is also applied in biology and ecology to explain the uncertainty and nonlinearity of complex phenomena such as population dynamics, disease transmission, and gene expression [3]. Although significant advancements across multiple fields, accurately predicting chaotic systems remains a critical and challenging task. Recent improvements in computing power and algorithmic innovation have led to the development and application of many deep learning techniques for predicting chaotic dynamical systems [4] (e.g. artificial neural network [5, 6], long short-term network [6, 7], next generation reservoir computing [8], and reservoir computing-echo state network [6, 9]). However, these deep learning methods often face criticism for being 'black-box' owing to their lack of model interpretability [10, 11]. To address this issue, Wang and Li [12] proposed a novel dynamic-based deep learning method known as dynamical system deep learning (DSDL), which is an innovative approach that combines principles from dynamical systems theory with modern deep learning techniques to enhance the prediction of chaotic systems.

However, in today's era of big data, as the volume and diversity of data sources continue to grow, we will inevitably encounter scenarios where the dataset contains not only information about the target system, but also invalid information from other interfering systems. This problem of data contamination caused by irrelevant information becomes increasingly prominent, which obscures the intrinsic dynamics of the target system, making it significantly harder to build reliable predictive models and complicating accurate predictions. Although recent advancements in deep learning methods have shown their potential, there is still a scarcity of research focused on developing algorithms specifically tailored to address the challenge of data contamination. Therefore, exploring the capacity of contemporary deep learning methods to purify polluted information fills a crucial gap in the current body of research.

This study obtains some of the most novel findings in this field by exploring whether several state-of-the-art deep learning methods have the ability to purify invalid information. Here, we focus on predicting chaotic systems as our target systems and the data contamination can arise from various interfering dynamical systems, irrelevant noise, and even random or periodic systems. By comparing the predictive performance and stability of contemporary deep learning methods under different interference scenarios, we can evaluate their potential in purifying the pollution of invalid information.

2. Methods

2.1. Artificial neural network (ANN)

The ANN framework employed in this study is similar to that described by Dueben and Bauer [5]. This ANN architecture consists of four hidden layers, each of which has 100 neurons with a tanh activation function. The input and output layer to the ANN have 8 neurons. This ANN is stateless, i.e. there is no hidden variable that tracks temporal evolution. Following the procedure used in Dueben and Bauer which involves relating $X(t)$ to its change in $X(t)$ at the next time step, rather than the raw value $X(t + \Delta t)$, this ANN training architecture implicitly contains a reference to the previous time step. Therefore, this ANN's inputs/outputs are chosen to be pairs of $X(t)^{\text{train}}$ and $\Delta X(t)^{\text{train}} = X(t + \Delta t)^{\text{train}} - X(t)^{\text{train}}$. During the prediction phase, by using $X(t)^{\text{test}}$ which is either known from the initial condition or has been previously calculated, $\Delta X(t)^{\text{test}}$ will be predicted. The obtained $\Delta X(t)^{\text{test}}$ is passed through an Adams–Bashforth integration scheme to estimate the next state $X(t + \Delta t)^{\text{test}} = X(t)^{\text{test}} + \frac{1}{2} [3\Delta X(t)^{\text{test}} - \Delta X(t - \Delta t)^{\text{test}}]$. The training process for the ANN was executed using a stochastic gradient descent optimizer, with a learning rate set to 0.001 and a batch size of 100. Mean absolute error was primarily used as the loss function, although using the mean squared error yielded comparable outcomes.

2.2. Long short-term memory network (LSTM)

The LSTM [7] is a deep-learning algorithm suited for predicting sequential data [13]. A major improvement over regular recurrent neural networks (RNNs) is that LSTM has gates that control the information flow into the neural network from previous steps of the time series [14]. The LSTM architecture used in this study closely mirrors the one proposed by Vlachas *et al* [4], which has 50 hidden layers in each cell. The input to the LSTM consists of a time-delay-embedded matrix of $X(t)$, with an embedding dimension q , which is also referred to as the look-back parameter. Through extensive hyper-parameter optimization, mainly through trial and error, the optimal value of q was determined to be 3, as this value provided the longest prediction horizon with the best performance (tested over a range of $q = 1 \sim 22$). The LSTM model predicts $X(t + \Delta t)$ by using the previous q time steps of $X(t)$. The weights for the LSTM layers are determined during the training process. During testing, the prediction for $X(t + \Delta t)$ is generated from the preceding q observations $[X(t - (q - 1)\Delta t), \dots, X(t - \Delta t), X(t)]$ that are either known from the initial condition or have been previously predicted. It was found that the best results came from a stateless LSTM, meaning that the state of the LSTM was reset at the beginning of each training batch. We also experimented with a stateful LSTM, where the states were carried over from one batch to the next, but the stateless configuration performed better during training. There was no evidence of overfitting, as the training and testing accuracies remained consistent throughout.

2.3. Next generation reservoir computing (NG-RC)

Reservoir computing (RC) is a deep-learning algorithm designed to process information generated by dynamical systems using time-series data [15]. NG-RC [8] builds upon traditional RC by replacing the recurrent neural network reservoir with a more efficient nonlinear vector auto-regression (NVAR) approach. Instead of relying on random matrices and optimizing multiple meta-parameters, NG-RC builds a feature

vector from time-delay embeddings of the input data. This feature vector includes both linear and nonlinear transformations of the past k observations:

$$\begin{aligned}\mathbf{O}_{\text{lin}}(t) &= [X(t), X(t + \Delta t), \dots, X(t - (k - 1)\Delta t)] \\ \mathbf{O}_{\text{nonlin}}^{(2)}(t) &= \mathbf{O}_{\text{lin}}(t) \otimes \mathbf{O}_{\text{lin}}(t) \\ \mathbf{O}_{\text{total}}(t) &= \mathbf{c} \oplus \mathbf{O}_{\text{lin}} \oplus \mathbf{O}_{\text{nonlin}}.\end{aligned}$$

The total feature vector is then used to predict the next state via a linear transformation: $\hat{X}(t + \Delta t) = \mathbf{W}_{\text{out}} \mathbf{O}_{\text{total}}(t)$, where \mathbf{W}_{out} is optimized through ridge regression:

$$\mathbf{W}_{\text{out}} = (\mathbf{O}_{\text{total}}^T \mathbf{O}_{\text{total}} + \alpha I)^{-1} \mathbf{O}_{\text{total}}^T \mathbf{Y}_d.$$

Here, α is a regularization parameter to avoid overfitting. By eliminating the need for a neural network reservoir, NG-RC reduces computational costs and requires fewer training data. NG-RC retains the ability to model complex dynamics with fewer data points and a simplified structure, providing robust predictions with much lower computational overhead than traditional RC methods.

2.4. Reservoir computing-echo state network (RC-ESN)

The RC-ESN [9] is a type of recurrent neural network (RNN) characterized by its reservoir of D neurons. These neurons are sparsely connected according to an Erdős–Rényi random graph configuration, with the connectivity described by an adjacency matrix A of size $D \times D$. This matrix is normalized using its largest eigenvalue (Λ_{\max}) and scaled by a factor ρ , typically ranging from 0.05 to 0.7. The RC-ESN is unique because only the output-to-reservoir weights, \mathbf{W}_{out} are trained. The input-to-reservoir and reservoir-to-output weights (\mathbf{W}_{in} and A) remain fixed after being initialized randomly. The state of the reservoir, denoted as $r(t)$, is updated using a hyperbolic tangent function. The reservoir processes the input observables, generating a high-dimensional representation, $r(t + \Delta t)$, which is then used to predict $X(t + \Delta t)$ using a linear transformation. The training process for RC-ESN is notably efficient because it does not involve backpropagation through time. Instead, \mathbf{W}_{out} is computed in a single step using ridge regression, minimizing the difference between the predicted and actual states. This simplicity in training, combined with the avoidance of vanishing and exploding gradients, makes RC-ESN a powerful tool for time series prediction, especially in dynamic systems.

2.5. Dynamical system deep learning (DSDL)

The DSDL [12] is an innovative approach that utilizes the dynamical information embedded in chaotic time series and integrates deep learning's capacity to model complex, nonlinear relationships to achieve the long-term accurate and interpretable predictions of chaotic dynamical systems. The DSDL framework starts with n -dimensional time series data $\{x_1(t), x_2(t), \dots, x_n(t)\}$ from the target system. These time series are treated as primitive system variables, which are used to construct a feature set $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$. In chaotic systems, due to dissipation, the steady-state dynamics are often constrained to a subspace known as the attractor \mathbf{A} , which can be reconstructed using phase-space reconstruction techniques. Based on the delayed embedding theory, a 'delayed attractor' is reconstructed by the time-lagged coordinates of the target variable $x_k(t)$, $k \in [1, n]$ in the form of $\mathcal{D}(x_k(t), x_k(t + \tau), x_k(t + 2\tau), \dots)$ with suitable embedding dimension L and time delay interval τ . Using the generalized embedding theorem, a 'non-delayed attractor' is reconstructed by multiple variables in the form of $\mathcal{N}(x_i(t), x_j(t), x_s(t), \dots)$, $i, j, s, \dots \in [1, n]$. This dual reconstruction approach allows DSDL to capture both the temporal and spatial dimensions of chaotic systems, providing a more comprehensive understanding of the system's underlying dynamics.

Ma *et al* [16] randomly selected index tuple $l = (i, j, s, \dots)$, considering only the linear relationships among system variables. To fully capture the nonlinear interactions among variables, we construct a multi-layer nonlinear network to obtain the candidate variable set \mathbf{X}_N . For each layer i ($i \in [1, m]$), the candidate variable subset is denoted as $\mathbf{X}_{N_{l_i}}^{L_i}$, including all monomials of i th-order based on \mathbf{X} , when $i \geq 2$ the monomials are nonlinear. However, not all variables in the set \mathbf{X}_N have a positive impact on predictions of the target variable $x_k(t)$, thus we select the key variables set $\mathbf{X}_{K_{x_k}}$ that significantly influence the evolution of $x_k(t)$ from \mathbf{X}_N by using the cross-validation-based stepwise regression (CVSR) method. This approach helps identify critical information and eliminate redundancies, thereby reducing the complexity and enhancing interpretability of the DSDL model. As a result, the non-delayed attractor \mathcal{N} is constructed by those key variables we selected.

The DSDL model establishes a diffeomorphic mapping between the two reconstructed attractors, that is, $\Phi : \mathcal{N} \rightarrow \mathcal{D}$, since both attractors are topologically conjugated to the original attractor \mathbf{A} . On this basis, we can obtain the DSDL prediction model for the target variable $x_k(t)$,

$$x_k(t + \tau) = \varphi(x_i(t), x_j(t), x_s(t), \dots).$$

Here, φ represents the mapping learned from the training data. Unlike typical ‘black-box’ deep learning models, DSDL aligns more closely with physical models by learning deterministic rules from training data. This structured approach makes it particularly effective for successive predictions, offering a transparent and explainable path from inputs to outputs.

2.6. Assessments of model predictive performance

To assess the predictive abilities of various deep learning methods, we employ the effective prediction time (EPT) as an indicator of the model’s performance. The EPT is defined as the time it takes for the corresponding prediction error $E(t)$ to first exceed a predefined error threshold ε , calculated as follows:

$$E(t) = |X(t) - \tilde{X}(t)|, \\ \text{EPT} = \text{Time}[E(t) \leq \varepsilon],$$

where $X(t)$ represents the actual time series, and $\tilde{X}(t)$ denotes the predicted series. The error threshold ε is set to one standard deviation of the predicted time series, which is used to determine where the model’s accurate prediction stops, so it cannot be too large or too small to avoid overestimating or underestimating the model’s predictive ability. Through testing, we found that when the error threshold is one standard deviation, the predicted sequence and the original sequence have basically separated, and subsequent predictions will no longer be accurate. A higher EPT indicates a stronger predictive ability of the model. The EPT is expressed in model time units (MTUs), where $1 \text{ MTU} = 100 \Delta t$. To ensure robustness, we compute the mean EPT from 100 distinct training/testing sets to evaluate the predictive performance.

3. Results

3.1. Predicting the Lorenz system with one interference system

To evaluate the predictive performance of several contemporary deep learning methods, we utilize the Lorenz system as our target system. As a paradigmatic chaotic dynamical system, the Lorenz system exhibits deterministic chaos, forming a strange attractor in its phase space trajectory [17]. To challenge the deep learning methods, we introduce five interference systems with distinct properties to contaminate the input dataset: the Rössler system [18], Rucklidge system [19], Hyperchaotic Lorenz system [20], periodic system (trigonometric functions), and stochastic system (random sequences with normal distributions).

When there is no interfering information, the prediction performance of the DSDL method is the best for the Lorenz system, while other deep learning methods also have varying degrees of effective prediction time (figure 1, table 1). When the input dataset is polluted by one interference system, the DSDL method consistently achieves the highest predictive performance across all interference scenarios. Its mean EPT remains stable and basically unaffected by the introduction of invalid information, underscoring its robustness. The DSDL method’s ability to effectively exploit the nonlinear interactions among system variables and removes redundant information by selecting key variables to capture the underlying deterministic rules of the target system enables it to maintain high accuracy.

In contrast, the ANN, LSTM, NG-RC, and RC-ESN methods exhibited a decline in the mean EPT when faced with interference. The LSTM and NG-RC methods, in particular, struggled with stochastic signal pollution, whereas the RC-ESN method demonstrated resilience against periodic and stochastic systems but faltered with chaotic interference. This highlights DSDL’s superiority as a dynamics-based deep learning method in identifying the dynamical attributes of the target system.

We also compared the predictive stability of these five methods using the coefficient of variation (CV), a standardized measure of data dispersion defined as the ratio of the standard deviation of EPTs (achieved by 100 different training/test sets) to the mean value. As shown in figure 2 and table 1, the CV of the DSDL method is the lowest and remains almost unchanged in each scenario, indicating the most stable prediction results despite various invalid information. Notably, the NG-RC method exhibits the highest CV when facing stochastic systems, suggesting significant instability in its predictions due to stochastic signals.

To further illustrate the predictive performance of these methods, figure 3 presents the evolution of mean prediction error over time. The ANN and NG-RC methods often lose the original chaotic properties of the target system, failing to capture its deterministic evolution rules. Conversely, the LSTM, RC-ESN, and DSDL

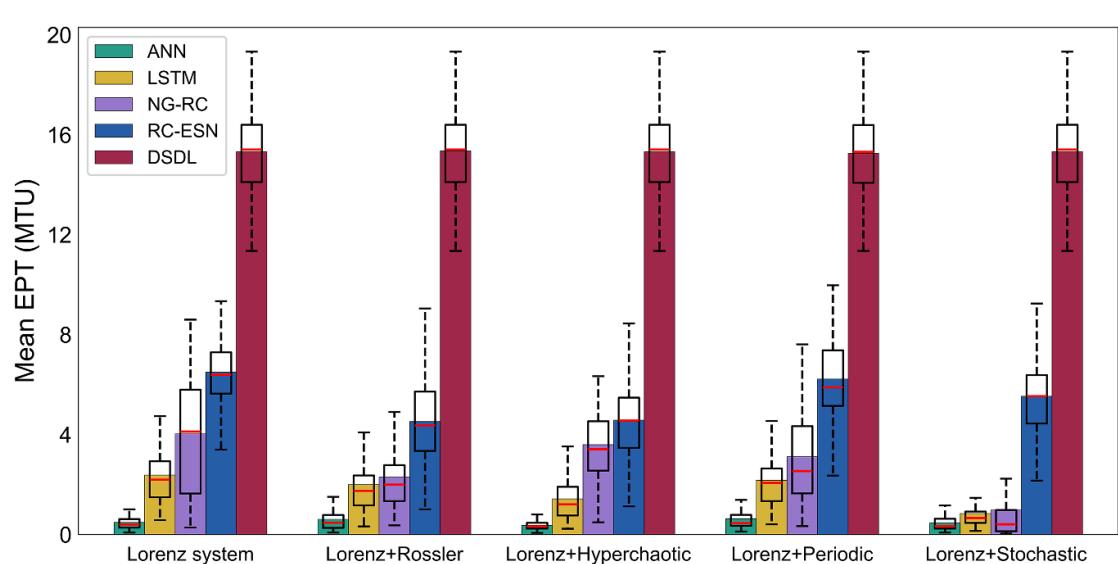


Figure 1. Comparison of predictive performance among different deep learning methods for the Lorenz system without and with various interference systems. The mean EPT (MTU) of each method in every scenario is obtained by 100 different training/test sets. The higher the mean EPT, the better the method performs.

Table 1. Comparison of predictive performance and stability among different deep learning methods for the Lorenz system without and with one interference systems. EPT is measured by dimensionless time units (DTU).

Methods	ANN		LSTM		NG-RC		RC-ESN		DSDL	
Evaluation indicators	EPT	CV	EPT	CV	EPT	CV	EPT	CV	EPT	CV
Lorenz	0.63	0.69	2.38	0.52	4.03	0.58	6.51	0.24	15.35	0.12
Lorenz + Rossler	0.62	0.98	2.00	0.59	2.30	0.63	4.52	0.38	15.32	0.12
Lorenz + Hyperchaotic	0.36	0.84	1.43	0.62	3.61	0.42	4.58	0.38	15.30	0.13
Lorenz + Periodic	0.59	0.72	2.17	0.53	3.12	0.62	6.22	0.29	15.33	0.13
Lorenz + Stochastic	0.50	0.73	0.84	0.84	0.98	1.53	5.53	0.32	15.32	0.12

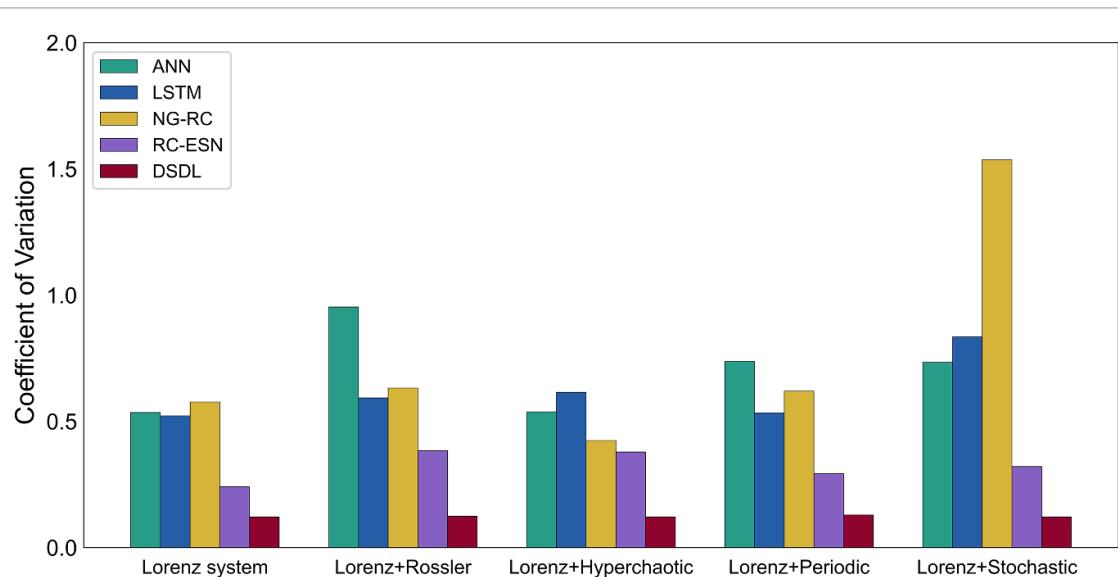


Figure 2. As in figure 1, but for predictive stability measured by the coefficient of variation (CV). The lower the CV, the better the stability of the method.

methods retained these characteristics, with DSDL showing the longest time to reach the error threshold, confirming its optimal performance.

In this study, we focus more on the accurate predictions that different deep learning methods can provide in the presence of other interfering systems. But there is no doubt that capturing the long-term evolution

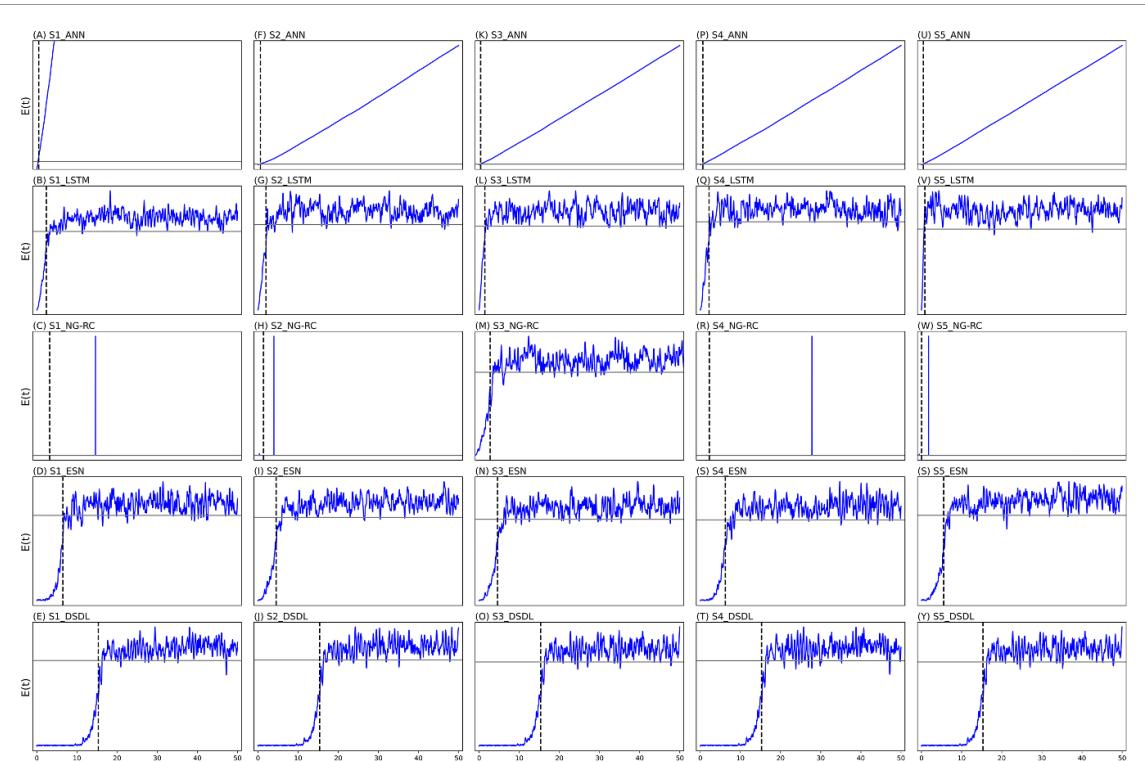


Figure 3. Comparison of mean prediction error among different deep learning methods for the Lorenz system without and with various interference systems. (A)–(E) Scenario 1 (S1): Lorenz system; (F)–(J) Scenario 2 (S2): Lorenz + Rössler systems; (K)–(O) Scenario 3 (S3): Lorenz + Hyperchaotic systems; (P)–(T) Scenario 4 (S4): Lorenz + Periodic systems; (U)–(Y) Scenario 5 (S5): Lorenz + Stochastic systems. The blue line shows the mean prediction error obtained by 100 different training/test sets, the light grey line represents the error threshold for measuring effective prediction while the vertical black dashed line marks the mean EPT.

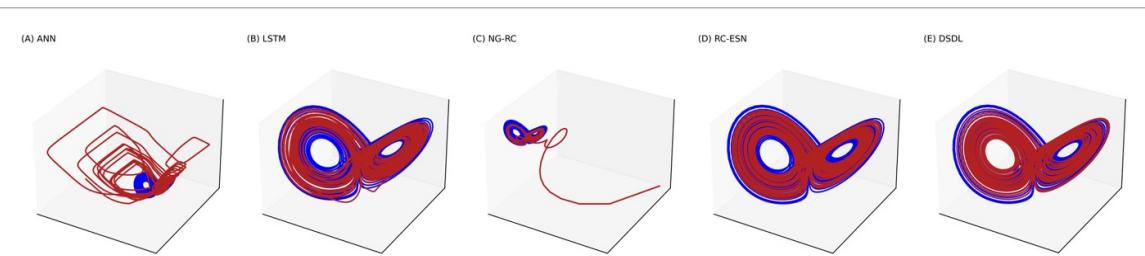


Figure 4. Comparison of long-term evolution behaviour of prediction sequences provided by different deep learning methods when the interfering system is the Rössler system. (A) ANN; (B) LSTM; (C) NG-RC; (D) RC-ESN; (E) DSDL. The blue line represents the real sequence, and the red line represents the predicted sequence.

behavior of chaotic systems is also an important criterion for evaluating the quality of deep learning methods. For example, when the interfering system is the Rössler system, figure 4 illustrates the long-term behavior of the predicted sequences provided by five contemporary deep learning methods. We can clearly see that the prediction results of ANN and NG-RC methods quickly overflow the attractor orbit of the target system, while the prediction sequence of LSTM method, although roughly on the attractor orbit, is actually unstable and has disturbances that deviate from the orbit. The prediction results of RC-ESN and DSDL methods are completely follow the Lorenz attractor orbit, indicating that these two methods can effectively capture the long-term evolution behavior of the target system. However, it can be evidently seen in figures 4(D) and (E) that the long-term behavior of the DSDL predicted sequence coincides better with the Lorenz attractor orbit than the RC-ESN method, especially around the equilibrium points. What's more, as shown in figure 1, the accurate predictions provided by DSDL are far superior to those of the RC-ESN method, which means that the predictive performance of the DSDL method is better than that of the EC-ESN method.

3.2. Predicting the Lorenz system with two interference systems

To further test the capacities of different deep learning methods, we introduce two interference systems simultaneously, increasing the complexity of the dataset. As shown in figure 5 and table 2, the DSDL method continued to exhibit the best predictive performance, remaining largely unaffected by the increased

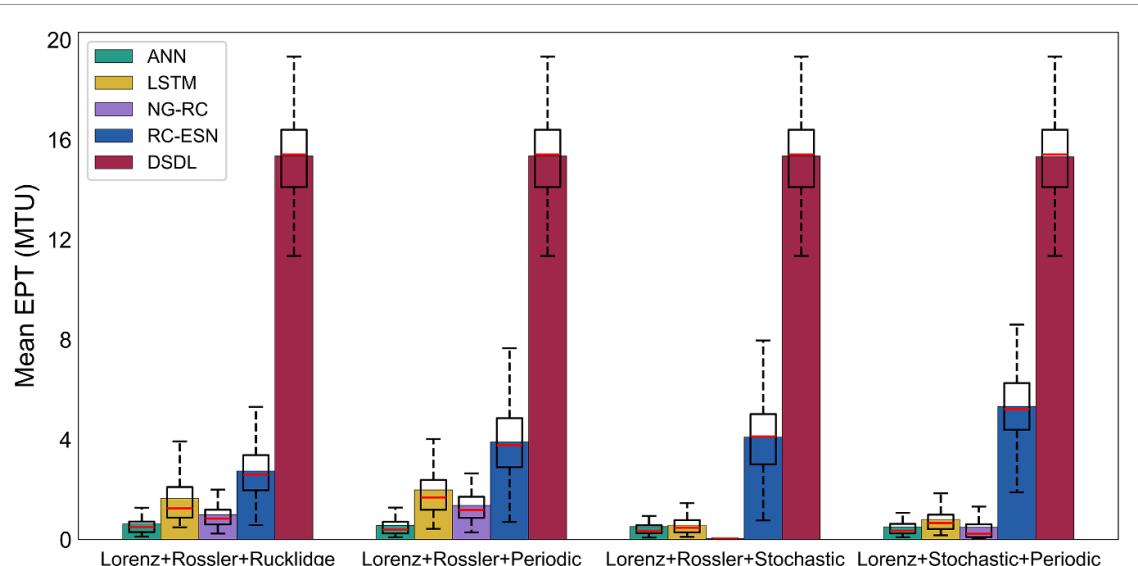


Figure 5. Same as figure 1, but for the Lorenz system with two interference systems for each scenario.

Table 2. Comparison of predictive performance and stability among different deep learning methods for the Lorenz system with two interference systems. EPT is measured by dimensionless time units (DTU).

Methods	ANN		LSTM		NG-RC		RC-ESN		DSDL	
Evaluation indicators	EPT	CV	EPT	CV	EPT	CV	EPT	CV	EPT	CV
Lorenz + Rö + Rucklidge	0.53	1.00	1.64	0.70	0.99	0.63	2.74	0.44	15.33	0.13
Lorenz + Rö + Periodic	0.56	0.98	1.98	0.59	1.36	0.53	3.90	0.41	15.29	0.12
Lorenz + Rö + Stochasitc	0.51	1.08	0.57	0.65	0.04	0.21	4.10	0.40	15.35	0.13
Lorenz + Peri + Stochastic	0.46	0.77	0.80	0.78	0.49	0.43	5.32	0.30	15.32	0.12

complexity. Compared to figure 1 and table 1, its mean EPT did not significantly decline, showcasing its robustness in handling complex, contaminated datasets. The performance of the other methods generally declined with increasing complexity. The ANN and NG-RC methods lost the chaotic properties of the Lorenz system in all scenarios, while the LSTM and RC-ESN methods retained some properties but with reduced mean EPT compared to scenarios with one interference system.

Figure 6 and table 2 show that the CV of the DSDL method remained the lowest and most stable across all scenarios with two interference systems. In contrast, the CV of the other methods increased, indicating higher instability with more complex invalid information. Figure 7 presents the mean prediction error in scenarios with two interference systems. The ANN and NG-RC methods showed significant performance degradation, failing to maintain the chaotic properties of the target system. The LSTM and RC-ESN methods performed better but still showed reduced EPT. The DSDL method's predictive performance remained robust, highlighting its ability to handle complex data contamination.

3.3. Predicting the Chua's circuit system with one interference system

Chaotic systems have a wide range of applications across various fields, including physics, engineering, finance, and so on. For instance, recent studies have explored the use of chaotic systems in secure communications [21] and video encryption [22, 23]. In order to improve the reliability of this study, we further used another chaotic system with different properties from the Lorenz system as the target system to conduct a simple verification of our results, which is the Chua's circuit system. Chua's circuit is currently the most representative one among many chaotic circuits. It has been applied in secure communications, random number generation, pattern recognition and sensing. The equations of Chua's circuit [21] are as follows:

$$\begin{cases} \frac{dx}{dt} = \alpha(y - x - f(x)) \\ \frac{dy}{dt} = x - y + z \\ \frac{dz}{dt} = -\beta y \\ f(x) = m_1 x + 0.5(m_0 - m_1)(|x + 1| - |x - 1|) \end{cases},$$

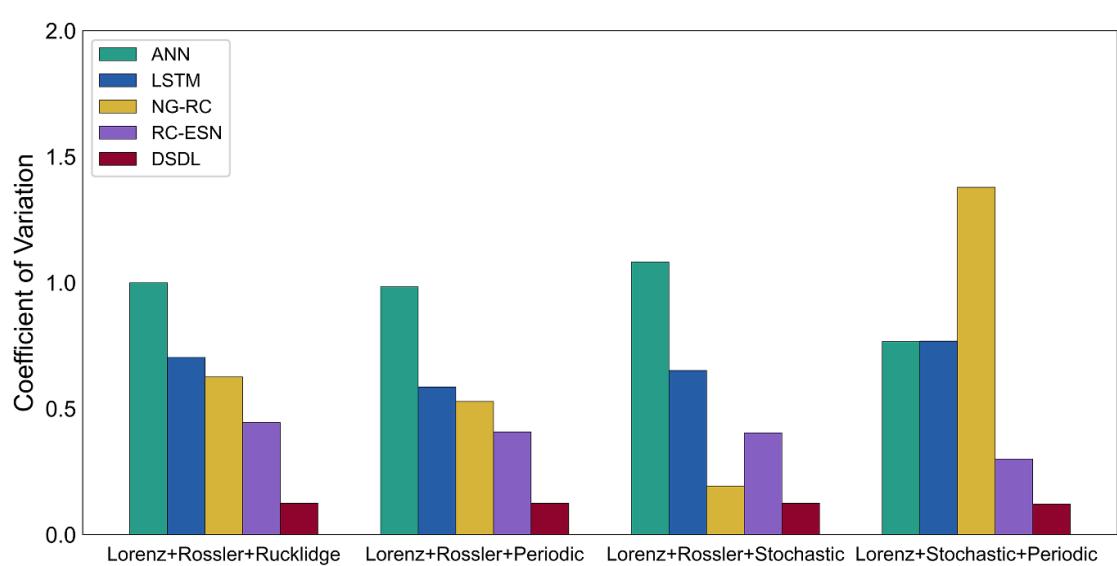


Figure 6. Same as figure 2, but for the Lorenz system with two interference systems for each scenario.

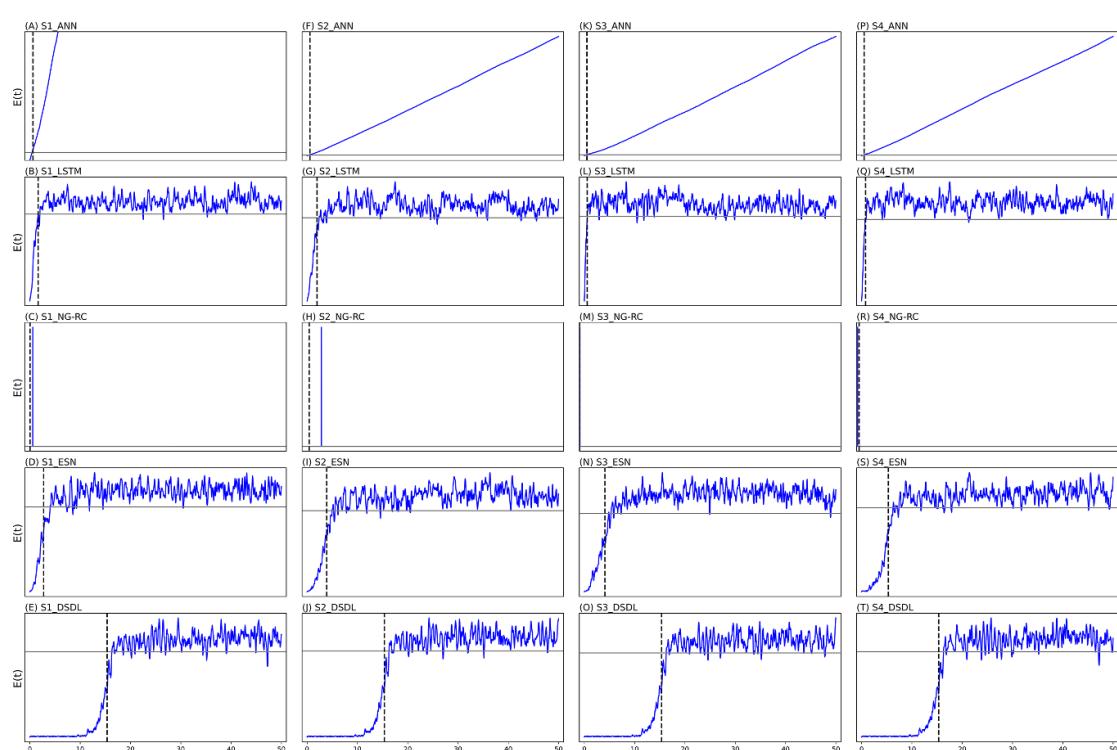


Figure 7. Same as figure 3, but for the target system (the Lorenz system) with two interference systems for each scenario. (A)–(E) Scenario 1 (S1): Lorenz + Rössler + Rucklidge systems; (F)–(J) Scenario 2 (S2): Lorenz + Rössler + Periodic systems; (K)–(O) Scenario 3 (S3): Lorenz + Rössler + Stochastic systems; (P)–(T) Scenario 4 (S4): Lorenz + Stochastic + Periodic systems.

where $\alpha = 15.6$, $\beta = 28$, $m_0 = -\frac{8}{7}$, $m_1 = -\frac{5}{7}$. Using a fourth-order Runge–Kutta scheme with a time step $\Delta t = 0.01$ dimensionless time units (TUs), we integrate forward for 10^7 steps starting from $(x, y, z) = (0.7, 0, 0)$ to generate a large dataset for training and testing, and we discard the first 5000 time points to ensure the chaos of time series.

Focusing on the Chua's circuit, we obtained the results presented in figure 8. To investigate the ability of modern deep learning methods to purify polluted information, we introduced several contamination scenarios into the Chua's circuit system. When no interfering information is present, the DSDL method demonstrates the best predictive performance, outperforming other deep learning methods. However, when the input dataset is contaminated by interference systems with different characteristics, the DSDL method consistently maintains the highest predictive performance across all scenarios, remaining stable and

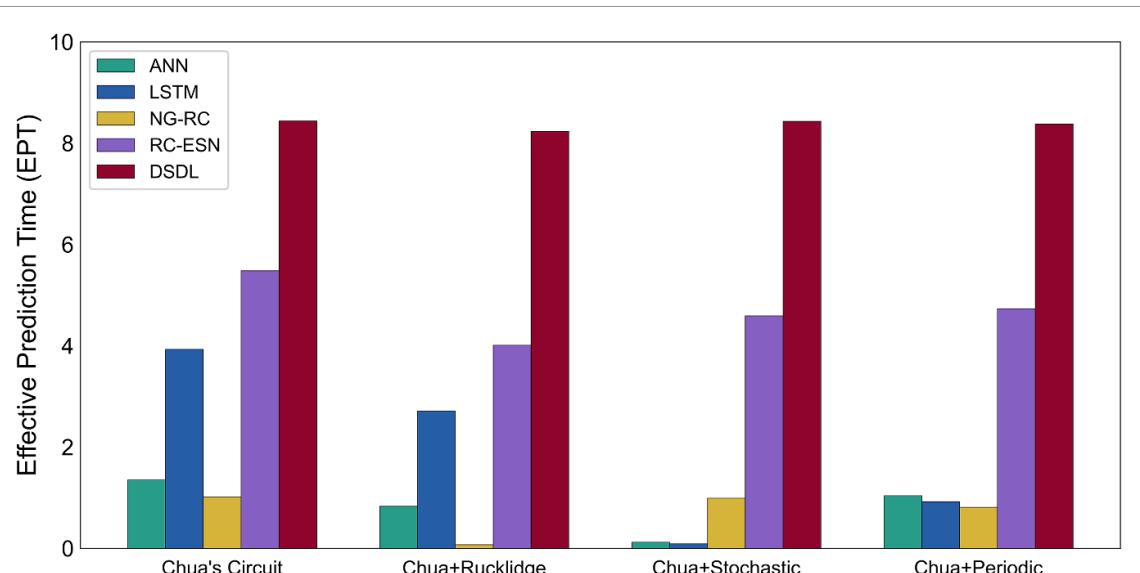


Figure 8. Comparison of predictive performance among different deep learning methods for the Chua's circuit system without and with one interference systems. The higher the EPT, the better the method performs.

unaffected by any form of invalid information. In contrast, methods such as ANN, LSTM, NG-RC, and RC-ESN exhibit a noticeable decline in EPT when facing interference. Specifically, ANN and LSTM methods struggle with stochastic signal pollution, while the NG-RC method shows difficulties in handling other dynamical information for predicting the Chua's circuit. These results underscore DSDL's superiority in capturing the deterministic rules governing the Chua's circuit system, even in the presence of contamination.

4. Conclusion and discussion

As an emerging research field, there is currently almost no similar research to investigate whether contemporary deep learning methods can purify data pollution caused by interfering systems and no such work to develop algorithms specifically tailored to address the challenge of this data contamination. Therefore, this study obtains some of the most novel findings in this field and fills a crucial gap in the current body of this research. By comparing the predictive performance and stability of some state-of-the-art deep learning methods facing interference systems with different properties and complexities, it is clear to see that while most current deep learning methods exhibit reduced and unstable predictive performance owing to such contamination, the DSDL method stands out as being largely unaffected by any interference. This result not only tells us the shortcomings of contemporary deep learning methods, but also makes us realize the DSDL's unique ability to purify invalid data and pave the way for a more reliable and interpretable model, ensuring that we can confidently predict chaotic systems with precision even with the data contamination. The necessity and innovation of this research make it an essential contribution to advancing prediction methods in chaotic systems. The results of this study indicate that DSDL represents a significant step forward in the field of chaotic system prediction and provides a strong foundation for future research in this area.

Despite these significant advancements, some limitations remain. We certainly agree on the necessity of conducting real-world system predictions, which is also a matter of great concern to us. First, the interference systems examined in this study do not encompass all potential real-world complexities. Future research should introduce a broader range of interference systems to simulate more diverse scenarios. Additionally, the primary aim of this study was to explore the performance of various deep learning methods in theoretical systems. Therefore, a key avenue for future work is to apply these findings to real-world systems, which will further highlight the broad utility of the DSDL method and its potential to improve predictions in complex environments.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Acknowledgments

This work was jointly supported by National Key R&D Program of China (2023YFF0805100), National Natural Science Foundation of China (42130607) and Laoshan Laboratory (No. LSKJ202202600). Thanks for the Center for High Performance Computing and System Simulation, Laoshan Laboratory (Qingdao) for providing computing resource.

Author contributions statement

J P L and M Y W contributed equally to this work. J P L conceived the idea. M Y W performed all calculations and wrote the initial manuscript with the help of J P L Both authors contributed to analyses, interpretation and writing of results.

Conflict of interest

The authors declare no competing interests.

ORCID iDs

Mingyu Wang  <https://orcid.org/0000-0002-3246-1464>

Jianping Li  <https://orcid.org/0000-0003-0625-1575>

References

- [1] Palmer T N 1993 Extended-range atmospheric prediction and the Lorenz model *Bull. Am. Meteorol. Soc.* **74** 49–65
- [2] Soramaki K, Bech M L, Arnold J, Glass R J and Beyeler W E 2007 The topology of interbank payment flows *Physica A* **379** 317–33
- [3] Stein R R, Bucci V, Toussaint N C, Buffie C G, Rätsch G, Pamer E G, Sander C and Xavier J B 2013 Ecological modeling from time-series inference: insight into dynamics and stability of intestinal microbiota *PLoS Comput. Biol.* **9** e1003388
- [4] Vlachas P R, Byeon W, Wan Z Y, Sapsis T P and Koumoutsakos P 2018 Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks *Proc. Math. Phys. Eng. Sci.* **474** 20170844
- [5] Dueben P D and Bauer P 2018 Challenges and design choices for global weather and climate models based on machine learning *Geosci. Model Dev.* **11** 3999–4009
- [6] Chattopadhyay A, Hassanzadeh P and Subramanian D 2020 Data-driven predictions of a multiscale Lorenz 96 chaotic system using machine-learning methods: reservoir computing, artificial neural network, and long short-term memory network *Nonlinear Process. Geophys.* **27** 373–89
- [7] Hochreiter S and Schmidhuber J 1997 Long short-term memory *Neural Comput.* **9** 1735–80
- [8] Gauthier D J, Boltt E, Griffith A and Barbosa W A S 2021 Next generation reservoir computing *Nat. Commun.* **12** 5564
- [9] Pathak J, Hunt B, Girvan M, Lu Z and Ott E 2018 Model-free prediction of large spatiotemporally chaotic systems from data: a reservoir computing approach *Phys. Rev. Lett.* **27** 024102
- [10] Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N and Prabhat F 2019 Deep learning and process understanding for data-driven earth system science *Nature* **566** 195–204
- [11] Sun Z H *et al* 2020 A review of earth artificial intelligence *Comput. Geosci.* **159** 105034
- [12] Wang M Y and Li J P 2024 Interpretable predictions of chaotic dynamical systems using dynamical system deep learning *Sci. Rep.* **14** 3143
- [13] Goodfellow I, Bengio Y and Courville A 2016 *Deep Learning* (MIT Press)
- [14] Pascanu R, Mikolov T and Bengio Y 2013 On the difficulty of training recurrent neural networks *Int. Conf. on MachineLearning* pp 1310–8
- [15] Maass W, Natschläger T and Markram H 2002 Real-time computing without stable states: a new framework for neural computation based on perturbations *Neural Comput.* **14** 2531–60
- [16] Ma H F *et al* 2018 Randomly distributed embedding making short-term high-dimensional data predictable *Proc. Natl Acad. Sci. USA* **115** E9994–10002
- [17] Lorenz E N 1963 Deterministic nonperiodic flow *J. Atmos. Sci.* **20** 130–41
- [18] Rössler O E 1976 An equation for continuous chaos *Phys. Lett. A* **57** 397–8
- [19] Rucklidge A M 1992 Chaos in models of double convection *J. Fluid Mech.* **237** 209–29
- [20] Li Y X, Wallace K S and Chen G R 2005 Hyperchaos evolved from the generalized Lorenz equation *Int. J. Circuit Theory Appl.* **33** 235–51
- [21] Kuznetsov N, Kuznetsova O, Leonov G and Vagaitsev V 2011 Hidden attractor in Chua's circuits. *Proc. 8th Int. Conf. on Informatics in Control, Automation and Robotics* vol 1 pp 279–83
- [22] Gao S, Iu H H-C, Wang M, Jiang D, El-Latif A A A, Wu R and Tang X 2024 Design, hardware implementation, and application in video encryption of the 2-D memristive cubic map *IEEE Internet Things* **11** 21807–15
- [23] Gao S *et al* 2024 Development of a video encryption algorithm for critical areas using 2D extended Schaffer function map and neural networks *Appl. Math. Modelling* **134** 520–37