SCIENCE CHINA
Earth Sciences

•ARTICLE•

July 2025 Vol.68 No.7: 2346–2360
https://doi.org/10.1007/s11430-024-1593-9

Chaotic climate system forecasting using an improved echo state network with sparse observations

Lin DING[1], Yulong BAI[2*], Donghai ZHENG[1], Xiaoduo PAN[1], Manhong FAN[2] & Xin LI[1*]

[1] *National Tibetan Plateau Data Center, State Key Laboratory of Tibetan Plateau Earth System, Environment and Resources, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing 100101, China*
[2] *College of Physics and Electronic Engineering, Northwest Normal University, Lanzhou 730070, China*

Received October 30, 2024; revised May 4, 2025; accepted May 15, 2025; published online June 17, 2025

Abstract Error accumulation in long-term predictions of chaotic climate systems is caused primarily by the model's high sensitivity to initial conditions and the absence of dynamic adjustment mechanisms, leading to gradual forecast divergence. This presents a critical challenge to achieving stable long-term predictions. While current data-driven approaches perform well in short-term forecasting, their accuracy deteriorates significantly over time. To overcome this limitation, we propose an autonomous echo state network with a snow ablation optimizer (AESN-SAO), which significantly improves the adaptability and robustness of data-driven methods under varying initial conditions. This approach not only eliminates the need for manual hyperparameter tuning in traditional AESNs but also effectively mitigates the common issue of initial conditions sensitivity in chaotic climate systems. Furthermore, we introduce a sparse observation insertion mechanism based on the Lyapunov time and valid prediction time (VPT), which enables AESN-SAO to correct errors prior to system divergence, effectively extending the prediction horizon. Numerical experiments conducted on the Lorenz-63 and Climate Lorenz-63 systems demonstrate that integrating sparse observations with AESN-SAO approach extends the VPT to approximately 99 Lyapunov times, markedly reducing error accumulation in long-term forecasts. This study provides a reliable and efficient framework for long-term predictions in climate systems with nonlinear and chaotic dynamics, with promising applications in weather forecasting, climate modeling, and disaster risk assessment.

Keywords Sparse observation, Autonomous echo state network, Snow ablation optimizer, Chaotic climate system

Citation: Ding L, Bai Y, Zheng D, Pan X, Fan M, Li X. 2025. Chaotic climate system forecasting using an improved echo state network with sparse observations. Science China Earth Sciences, 68(7): 2346–2360, https://doi.org/10.1007/s11430-024-1593-9

1. Introduction

The Earth's climate system is composed of multiple chaotic subsystems, each operating on different time scales and exhibiting high sensitivity to initial conditions. Meanwhile, strong nonlinear interactions and complex multiscale coupling within the system pose significant challenges to the development of accurate and reliable physical models (Chen et al., 2020). Compared with traditional physics-based modeling approaches, data-driven methods do not rely on explicit physical equations, but instead learn the dynamic evolution of the system by mining and analyzing observational data, thereby enabling the simulation and prediction of complex climate processes (Wan et al., 2021; Wang et al., 2022; Shen et al., 2023). These approaches not only partially address the limitations of traditional physical models in representing complex systems but also offer new avenues for modeling and predicting chaotic climate systems in Earth system science.

* Corresponding authors: Xin LI (xinli@itpcas.ac.cn), Yulong BAI (baiyulong@nwnu.edu.cn)

1.1 Overview of data-driven approaches

Common data-driven modeling approaches include the Takens embedding theorem (Takens, 2006), sparse identification of nonlinear dynamics (SINDy) (Kaiser et al., 2018; Lin et al., 2021), Koopman theory (Proctor et al., 2018; Parmar et al., 2022), and neural networks (Sangiorgio et al., 2021; Vlachas et al., 2022). Although these approaches differ in modeling formulations and theoretical foundations, they can all be broadly categorized as data-driven approaches, characterized by their reliance on observational data for the identification, reconstruction, and prediction of system dynamics. In recent years, with the breakthrough advances in big data and artificial intelligence (AI) technologies, neural network-based data-driven models have emerged as a promising tool for forecasting the evolution of the Earth's climate system (Li and Guo, 2025). Among them, reservoir computing (RC), a lightweight variant of recurrent neural networks (RNNs) (Lukoševičius and Jaeger, 2009), has gained significant attention in chaotic system modeling and prediction tasks due to its high training efficiency and strong capacity to capture nonlinear temporal features.

Echo state network (ESN) (Jaeger, 2001) and liquid state machines (LSMs) (Maass et al., 2002) are the two principal implementations of the RC framework and have been widely applied to data-driven, model-free prediction tasks in chaotic systems. As a representative method within the RC framework, parallel RC has been shown to maintain low prediction error for approximately 8 Lyapunov times in the time evolution of the Kuramoto-Sivashinsky (K-S) equation (Pathak et al., 2018a), drawing significant attention to the application of AI techniques in chaotic system modeling. Subsequently, extensive research has been conducted on the RC framework focusing on reservoir dynamics and topological optimization (Ren et al., 2024b), adaptive weight adjustment and learning strategies (Ding et al., 2024), as well as architectural extension and integration schemes (Bai et al., 2021; Ding et al., 2023). A variety of improved RC-based frameworks have been proposed, including knowledge-based RC (Pathak et al., 2018b), physics-informed echo state networks (PI-ESNs) (Doan et al., 2020), and Bayesian optimized RC (Platt et al., 2022), aiming to enhance the robustness and generalization ability of AI methods in long-term chaotic system forecasting. Specifically, Ren et al. (2024a) proposed a data-driven method that integrates RC with higher-order dynamic mode decomposition, enabling feature reconstruction and achieving accurate predictions up to 8 Lyapunov times for the K-S equation. Li et al. (2024b) developed a modeling framework that integrates Granger causality inference with RC architectures. This method iteratively identifies higher-order interactions within the system to infer optimal structural relationships, thereby improving the predictive accuracy of complex systems. However, despite their strong performance in short-term prediction tasks, these methods generally face challenges in long-term forecasting due to the persistent accumulation of errors, often leading to prediction divergence.

Currently, improvements in AI methods primarily focus on developing modeling strategies that achieve high predictive performance while maintaining low model complexity. However, another promising pathway involves drawing inspiration from data assimilation concepts (Li et al., 2024a; Li and Liu, 2025), whereby targeted insertion of observational data is employed to combine dynamic modeling with observational information, thereby mitigating error accumulation and enhancing the stability and robustness of long-term predictions. For example, Fan et al. (2020) introduced a coupling coefficient between sparse observations and the output of the conventional RC, allowing the prediction time of the K-S equation to be extended by updating the actual (target) state every one Lyapunov time. In real engineering problems, Jin et al. (2023) utilized sensor observations with certain time delays to correct cumulative errors generated by the invertible Koopman network (IKN). The results show that when nonreal-time measurements with a 256-step time delay are used, the Lorenz-63 system, the Chen system, and the K-S equation achieve longer long-term prediction horizons. Furthermore, Gupta et al. (2023) investigated sparse observations in the K-S equation and found that the prediction accuracy of the reservoir-computing-based recurrent neural networks (RC-RNNs) progressively decreased with increasing sparsity of the observational data. This decline was primarily attributed to the reduction in autocorrelation between observations, further highlighting the high dependence of RC-RNNs on the quality and density of observational data.

1.2 Research motivation and contribution

The ESN leverages a large, fixed and randomly connected reservoir structure to capture the dynamic characteristics of a system. Unlike traditional RNNs, the design of the ESN avoids the problems of gradient vanishing and gradient explosion. Its core feature is that the connection weights within the reservoir remain unchanged after random initialization, and only the connection weights between the reservoir and the output layer are adjusted, thus making the training process more efficient. In the prediction of chaotic systems, the quantity and distribution of observational data are critical to forecast accuracy and reliability. To address the prediction challenges posed by sparse observations, this study proposes a sparse observation insertion strategy aimed at enhancing the stability and accuracy of data-driven methods in long-term predictions. Unlike conventional approaches that improve model performance by increasing data volume, our approach optimizes the timing and placement of sparse ob-

servations within the time series, achieving an effective balance between data availability and prediction performance. The findings provide methodological insights for the effective utilization of sparse data in complex, real-world systems.

This study proposes a parallel data-driven architecture for long-term prediction of chaotic climate systems, which incorporates adaptation weighting in the AESN and sparse observations. A key innovation of this method is the adoption of a parallel optimization strategy, which employs the snow ablation optimizer (SAO) to optimize AESN hyperparameters for chaotic systems across ten different initial conditions. Additionally, we introduce error correction factors based on the Lyapunov time and valid prediction time (VPT) of chaotic systems to guide the sparse sampling of observational data. The proposed framework is validated using two systems with pronounced chaotic characteristics: Lorenz-63 and Climate Lorenz-63 systems. On the basis of the prediction process of the AESN-SAO data-driven approach, we evaluated three scenarios for inserting sparse observational data: (1) inserting observational data at regular intervals of integer Lyapunov time; (2) using VPT to perform real-time assessments when the AESN-SAO prediction is about to diverge and inserting observational data at the point of divergence; and (3) determining the VPT of AESN-SAO and inserting observational data at integer Lyapunov times just before divergence.

2. Methodology

2.1 Autonomous echo state network

The ESN is a data-driven method whose high-dimensional reservoir structure can effectively capture and preserve the dynamic properties of input data. This capability relies on the connection weights between the input layer and the reservoir layer, as well as among neurons within the reservoir layer. These connection weights are randomly initialized and remain fixed throughout the training process (Xu and Han, 2015). Since only the output weights need to be trained, the ESN offers a relatively simple training procedure and high computational efficiency, significantly reducing the demand for computational resources. As a result, the ESN offers significant advantages in simulating and predicting the dynamic behavior of chaotic systems (Vlachas et al., 2020).

The numbers of nodes in the input, reservoir, and output layers of the ESN are denoted as M, K, and N, respectively. The input data $\mathbf{y}_t$ is scaled by the input matrix $\mathbf{W}_{\text{in}}$ to achieve a mapping from the M-dimensional input space to the K-dimensional reservoir space. During the training phase ($1 \leqslant t \leqslant T_a$, where T_a is the length of the data in the training phase), the reservoir mapping form $\mathbf{r}_t$ of $\mathbf{y}_t$ evolves as:

$$\mathbf{r}_t = (1-\gamma)\mathbf{r}_{t-1} + \gamma\tanh\left(\mathbf{W}_{\text{in}}[\sigma_{\text{in}}; \mathbf{y}_t] + \mathbf{W}_{\text{res}}\mathbf{r}_{t-1}\right), \tag{1}$$

where γ is the update rate of the neuron state variables in the reservoir layer, which is the leakage rate; $\tanh(\cdot)$ is the activation function; and σ_{in} is the input bias ($\sigma_{\text{in}} = 1$). The input weight matrix $\mathbf{W}_{\text{in}}$ represents the connection weight between the input layer and the reservoir, with a dimension of $K \times (M+1)$. It is influenced by the input scale IS and follows a uniform distribution within the range of $[-IS, IS]$. $\mathbf{W}_{\text{res}}$ is the connection weight matrix of the $K \times K$-dimensional intrareservoir neurons, expressed as $\mathbf{W}_{\text{res}} = SR \cdot \left[\mathbf{W} / \lambda_{\max}(\mathbf{W})\right]$. The nonzero elements in the sparse matrix $\mathbf{W}$ are uniformly distributed in $[-0.5, 0.5]$, and the percentage of its nonzero elements is denoted by the sparsity degree, which is the percentage of the number of interconnected neurons in a certain size of the reservoir. $\lambda_{\max}(\mathbf{W})$ denotes the maximum eigenvalue of $\mathbf{W}$, and SR is the spectrum radius.

The dimensions of the output matrix $\mathbf{W}_{\text{out}}$ is $N \times (M+K+1)$, which is calculated using the ridge regression method (Lukoševičius and Jaeger, 2009). The formula is shown in eq. (2):

$$\mathbf{W}_{\text{out}} = \mathbf{Y}_{\text{Target}}\mathbf{R}^{\text{T}}\left(\mathbf{R}\mathbf{R}^{\text{T}} + \beta\,\mathbf{I}\right)^{-1}, \tag{2}$$

where β is a regularization factor that prevents overfitting of the model during training ($\beta = 1 \times 10^{-8}$); matrices $\mathbf{R}$ and $\mathbf{Y}_{\text{Target}}$ denote all reservoir state vectors $\mathbf{r}_t$ and target value matrices in the training phase in $(M+K+1) \times T_a$ and $N \times T_a$ dimensions, respectively; and $\mathbf{I}$ is an $(M+K+1)\times(M+K+1)$-dimensional unit matrix.

At the training phase, recursive predictions are generated using the reservoir state vector and output weights, marking the transition of the ESN to an autonomous echo state network (AESN). At this stage, the external input data $\mathbf{y}_t$ at time t are replaced by the prediction data $\mathbf{x}_t$. During the autonomous prediction phase ($T_a + 1 \leqslant t \leqslant T$, where T denotes the length of the input data), the output at time t is represented by the N-dimensional vector $\mathbf{x}_t$. The prediction process of the AESN is defined by eqs. (3) and (4), with $\sigma_{\text{out}} = 1$ serving as the output bias.

$$\mathbf{r}_t = (1-\gamma)\mathbf{r}_{t-1} + \gamma\tanh\left(\mathbf{W}_{\text{in}}[\sigma_{\text{in}}; \mathbf{x}_t] + \mathbf{W}_{\text{res}}\mathbf{r}_{t-1}\right), \tag{3}$$

$$\mathbf{x}_t = \mathbf{W}_{\text{out}}[\sigma_{\text{out}}; \mathbf{x}_{t-1}; \mathbf{r}_{t-1}]. \tag{4}$$

2.2 Snow ablation optimizer

The AESN incorporating SAO goes through the process of population initialization, fitness assessment, selection and updating, and termination conditions during the parallel optimization phase. The SAO algorithm, as a metaheuristic optimization algorithm based on physical phenomena, simulates the process of melting and sublimation of snow

(Deng and Liu, 2023) with the population initialization phase shown below:

$$\mathbf{Z} = \mathbf{L} + r_1 \times (\mathbf{U} - \mathbf{L}), \tag{5}$$

$$\mathbf{Z} = \begin{bmatrix} Z_{1,1} & Z_{1,2} & \cdots & Z_{1,d-1} & Z_{1,d} \\ Z_{2,1} & Z_{2,2} & \cdots & Z_{2,d-1} & Z_{2,d} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Z_{n,1} & Z_{n,2} & \cdots & Z_{n,d-1} & Z_{n,d} \end{bmatrix}, \tag{6}$$

where $\mathbf{L}$ and $\mathbf{U}$ are the lower and upper bounds of the variable search space, respectively; r_1 is a random number in $[0, 1]$; n denotes the number of search agents and the size of the entire population; and d denotes the dimension of the search space, which is related to the number of parameters of the optimization problem to be solved.

During the snow melting process, the degree-day factor (DDF) ranges from 0.35 to 0.6 (Martinec and Rango, 1986). In each iteration, the value of DDF is updated using the formula $DDF = 0.35 + 0.25 \times [(\mathrm{e}^{iter/max\_iter} - 1) / (\mathrm{e} - 1)]$. Accordingly, the snow melt rate MR is calculated as $MR = DDF \times T$, where $T = \mathrm{e}^{-iter/max\_iter}$. Here, $iter$ denotes the current iteration number; max_iter is the maximum number of iterations (or termination conditions), indicating that the algorithm terminates and returns the best solution found when this limit is reached. The snow ablation optimizer updates the position equation via Brownian motion and a two-population mechanism to simulate irregular motion during the snow phase transition. This process allows snow to transform into liquid water or steam through physical processes under specific conditions during the position update process. $\mathbf{Z}_i(iter)$ is the i-th individual in the $iter$-th iteration, and $index_a$ and $index_b$ denote a set of indices of the row numbers of the individuals in the populations P_a and P_b, respectively. When $i \in index_a$ and $i \in index_b$, the position updating equations are shown in eq. (7).

$$\begin{aligned} \mathbf{Z}_i(iter+1) &= \mathbf{Elite}(iter) + \mathbf{BM}_i(iter) \\ &\otimes\big(\theta_1 \times (\mathbf{G}(iter) - \mathbf{Z}_i(iter)) \\ &+(1-\theta_1) \times (\overline{\mathbf{Z}}(iter) - \mathbf{Z}_i(iter))\big), \\ \mathbf{Z}_i(iter+1) &= MR \times \mathbf{G}(iter) + \mathbf{BM}_i(iter) \\ &\otimes\big(\theta_2 \times (\mathbf{G}(iter) - \mathbf{Z}_i(iter)) \\ &+(1-\theta_2) \times (\overline{\mathbf{Z}}(iter) - \mathbf{Z}_i(iter))\big), \end{aligned} \tag{7}$$

where $\mathbf{BM}_i(iter)$ represents the Brownian motion; θ_1 is the value obtained by randomly sampling from a uniform distribution on the interval $[0, 1]$; θ_2 is the value randomly selected from a uniform distribution on the interval $[-1, 1]$; and $\overline{\mathbf{Z}}(iter)$ represents the position of the center of mass of the population, which is computed by the expression $\overline{\mathbf{Z}}(iter) = 1/n \sum_{i=1}^{n} \mathbf{Z}_i(iter)$. $\mathbf{Elite}(iter) \in [\mathbf{G}(iter), \mathbf{Z}_{\mathrm{second}}(iter), \mathbf{Z}_{\mathrm{third}}(iter), \mathbf{Z}_c(iter)]$ is used to denote one randomly selected individual in the population among a set of elite individuals, which maintains the diversity of the population. $\mathbf{G}(iter)$ is the optimal solution at the current iteration; $\mathbf{Z}_{\mathrm{second}}(iter)$ and $\mathbf{Z}_{\mathrm{third}}(iter)$ are the second and third ranked individuals in the current population, respectively, except for the best individual; and $\mathbf{Z}_c(iter) = 1/n_1 \sum_{i=1}^{n_1} \mathbf{Z}_i(iter)$ is the leader, which represents the center-of-mass position of the individual ranked in the top 50% of the fitness. n_1 (n_1=1/2n) and $\mathbf{Z}_i(iter)$ denote the number of leaders and the i-th best leader, respectively.

2.3 The AESN-SAO data-driven approach integrated with sparse observations

The framework of the proposed AESN-SAO data-driven method with sparse observations for chaotic climate system prediction is shown in Figure 1. It comprises three phases: the observable acquisition phase, the parallel optimization phase, and the model-free prediction phase under sparse observations.

The parallel optimization phase can be described as follows.

(1) Ten sets of observations, each corresponding to a different initial state, were obtained from a chaotic climatic system (see Figure 1a), and were subsequently used to construct the training and validation datasets. This approach aims to achieve more robust hyperparameter tuning for the AESN model, with particular attention to the influence of initial condition sensitivity on prediction accuracy.

(2) In the "training-validation" phase, the SAO performs parallel optimization on the 10 sets of input data by evaluating the fitness function. Through iterative optimization of the hyperparameter combinations, the optimal set of hyperparameters for AESN $[IS, SR, \gamma, K, d]$ is obtained. Considering that the mean absolute error (MAE) is less affected by outliers and is more robust, we use MAE as the fitness function to search for the optimal hyperparameter configuration (see eqs. (8) and (9)). During the training process ($1 \leqslant t \leqslant T_a$, where T_a is the length of the data in the training phase), the input data at moment t are composed of s sets of M-dimensional data $\mathbf{y}^s = \left[y_{t,1}^1, y_{t,2}^2, \cdots, y_{t,M}^{10}\right]^{\mathrm{T}}$. In the validation process of parallel optimization ($1 \leqslant t \leqslant T_v$, where T_v is the length of the data in the validation phase), the output prediction result at moment t consists of s sets of N-dimensional data $\mathbf{x}^s = \left[x_{t,1}^1, x_{t,2}^2, \cdots, x_{t,N}^{10}\right]^{\mathrm{T}}$. This is clearly demon-

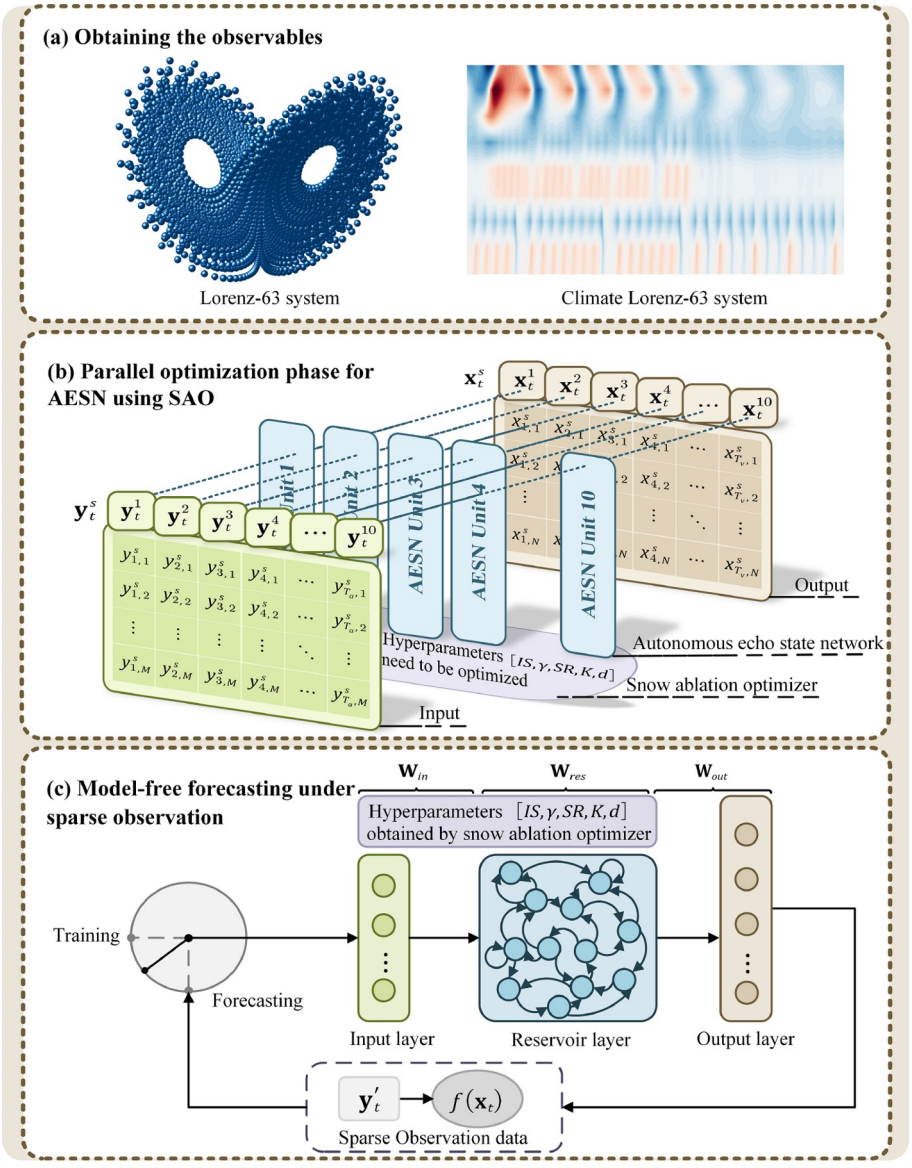

Figure 1 A scheme for integrating sparse observations with the AESN-SAO data-driven approach. (a) Observational data from the Lorenz-63 and Climate Lorenz-63 systems. (b) Parallel optimization of AESN using SAO on 10 sets of observational data $\mathbf{y}_t^{10}$ with different initial conditions, resulting in the optimal combination of five hyperparameters $[IS, SR, \gamma, K, d]$ ($\mathbf{y}_t^s$: the s-th input data at moment t, $s = 1, 2, \cdots, 10$, $t = 1, 2, \cdots, T_a$; $\mathbf{x}_t^s$: the s-th output data at moment t, $s = 1, 2, \cdots, 10$, $t = 1, 2, \cdots, T_v$; IS: input scale; SR: spectrum radius; γ: leakage rate; K: reservoir size; d: the number of interconnected neurons in the reservoir). (c) Model-free forecasting using sparse observation data, coupling the error correction factor $f(\mathbf{x}_t)$ to dynamically update the reservoir state ($\mathbf{W}_{\text{in}}$: input matrix; $\mathbf{W}_{\text{res}}$: reservoir internal connection matrix; $\mathbf{W}_{\text{out}}$: output matrix; $\mathbf{y}'_t$: sparse observation data).

strated in Figure 1b.

$$MAE = \frac{1}{N \times T_v} \sum_{i=1}^{N} \sum_{t=1}^{T_v} \left| y_{t,i} - x_{t,i} \right|, \tag{8}$$

$$fitness(iter) = MAE(iter), \tag{9}$$

where $y_{t,i}$ and $x_{t,i}$ represent the observed data and predicted results, respectively. The optimization is performed during the "training-validation" phase, so the data length corresponds to the length of the validation set T_v, meaning that the validation data are used to evaluate and optimize the performance of the hyperparameters. N is the dimension of the chaotic climate system.

After parallel optimization, model-free forecasting under the sparse observation phase can be described as follows.

(1) Obtain a new set of observation data from the chaotic climate system under a random initial state and construct the training and testing datasets.

(2) The autonomous prediction under sparse observation ($1 \leqslant t \leqslant T_{tr}$, where T_{tr} is the length of the dataset in the testing phase) is shown in Figure 1c. During this phase, the hy-

perparameter $[IS, \gamma, SR, K, d]$ in the AESN are the best hyperparameters obtained by the SAO in Figure 1b.

(3) In the traditional AESN approach, upon entering the forecasting block, the input data $\mathbf{y}_t$ are typically replaced by the prediction result $\mathbf{x}_t$. In this study, an error correction factor $f(\mathbf{x}_t)$ is introduced, integrating sparse observation data $\mathbf{y}'_t$ to dynamically update the reservoir state. Specifically, when predictions reach a designated Lyapunov time of c, sparse observational data $\mathbf{y}'_t$ are employed to adjust the reservoir state; at other time points, the prediction result $\mathbf{x}_t$ continues to guide the evolution of the reservoir state. This approach is mathematically represented as follows in eq. (10). In particular, the Lyapunov time (LT) serves a characteristic timescale in chaotic systems and is used to assess forecasting quality based on the growth rate of prediction errors. The Lyapunov exponent quantifies the separation rate at which neighboring trajectories in a dynamical system diverge or over time.

$$f(\mathbf{x}_t) = \begin{cases} \mathbf{y}'_t, & \text{if } t = c\Lambda_{\max}t, \\ \mathbf{x}_t, & \text{if others,} \end{cases} \tag{10}$$

where $\Lambda_{\max}t$ denotes the Lyapunov time of the chaotic climate system, and c represents the insertion interval of sparse observations (the specific insertion strategy will be described in the subsequent experimental section). The sparse observation $\mathbf{y}'_t$ corresponds to the target observation data at the Lyapunov time points. In this study, potential observation errors are not considered in the method design. This mechanism not only facilitates the correction of predictive errors but also significantly enhances long-term prediction performance of data-driven methods under sparse observation conditions.

3. Experimental design

3.1 Chaotic climate systems

3.1.1 Lorenz-63 system

The Lorenz-63 system, proposed by mathematician and meteorologist Edward N. Lorenz in 1963 (Lorenz, 1963), is a simplified dynamical model designed to simulate atmospheric convection. It characterizes chaotic behavior in deterministic systems and highlights the strong sensitivity of nonlinear systems to initial conditions. The Lorenz-63 system has been widely applied in research on chaos theory and climate dynamics, providing an important theoretical foundation for understanding the uncertainties in Earth's complex climate system. Its governing equations are defined as follows:

$$\begin{aligned} \dot{x} &= \sigma(y - x), \\ \dot{y} &= x(\rho - z) - y, \\ \dot{z} &= xy - bz. \end{aligned} \tag{11}$$

With parameters $\{\sigma, \rho, b\} = [10, 28, 8/3]$, the Lorenz-63 system exhibits typical chaotic properties. This demonstrated that even minor differences in initial conditions could lead to vastly different evolutionary trajectories of the system, a phenomenon famously known as the "butterfly effect", highlighting the sensitivity and unpredictability of the climate system. In this study, the initial values of the Lorenz-63 system are generated from a standard normal distribution, and the dataset is obtained through numerical integration using the fourth-order Runge-Kutta method with a time interval of 0.02.

3.1.2 Climate Lorenz-63 system

To further validate the effectiveness of the proposed approach, numerical experiments are conducted using the Climate Lorenz-63 chaos climate system, which features multiple time scales and varying amplitudes. This system describes the coupling of the extratropical atmosphere and the El Niño-Southern Oscillation (ENSO) (Peña and Kalnay, 2004). In the coupled system, the lowercase variables with the subscript '*e*' represent the extratropical atmosphere. The lowercase variables with the subscript '*t*' and all uppercase variables correspond to the tropical atmosphere and tropical ocean, respectively, both of which collectively represent ENSO. In this system, lowercase and uppercase variables correspond to the fast and slow subsystems, respectively.

Extratropical atmosphere:

$$\begin{aligned} \dot{x}_e &= \sigma(y_e - x_e) - \kappa_e(Sx_t + k_1), \\ \dot{y}_e &= \rho x_e - y_e - x_e z_e + \kappa_e(Sy_t + k_1), \\ \dot{z}_e &= x_e y_e - b z_e. \end{aligned} \tag{12a}$$

Tropical atmosphere:

$$\begin{aligned} \dot{x}_t &= \sigma(y_t - x_t) - \kappa(SX + k_2) - \kappa_e(Sx_e + k_1), \\ \dot{y}_t &= \rho x_t - y_t - x_t z_t + \kappa(SY + k_2) + \kappa_e(Sy_e + k_1), \\ \dot{z}_t &= x_t y_t - b z_t + \kappa_z Z. \end{aligned} \tag{12b}$$

Tropical ocean:

$$\begin{aligned} \dot{X} &= \tau\sigma(Y - X) - \kappa(x_t + k_2), \\ \dot{Y} &= \tau\rho X - \tau Y - \tau SXZ + \kappa(y_t + k_2), \\ \dot{Z} &= \tau SXY - \tau bZ - \kappa_z z_t. \end{aligned} \tag{12c}$$

In eq. (12), σ, ρ, and b are the standard parameters of the Lorenz-63 system, with values of 10, 28, and 8/3, respectively. κ_e, κ_z, and κ represent the coupling strengths among the three subsystems, with values of 0.08, 1, and 1, respectively. This implies weak coupling between the extratropical atmosphere and tropical atmosphere variables and strong coupling between the tropical atmosphere and ocean variables. The parameters τ and S represent the temporal and spatial scale factors, respectively. When $\tau = 0.1$, the slow subsystem evolves at one-tenth the rate of the fast subsystem. $S = 1$ indicates that the fast and subsystems have the same

amplitude scale (it is worth noting that the specific scale may vary depending on the coupling details). k_1 and k_2 denote offsets, with values of 10 and −11, respectively. Eq. (12) is numerically integrated using the fourth-order Runge-Kutta method with a time step of 0.02, and the initial state is randomly generated from a standard normal distribution.

3.2 Data partitioning and parameter setting

In the Lorenz-63 and Climate Lorenz-63 systems, the lengths of the training dataset T_a and the validation dataset T_v are 5300 and 1000, respectively. In the subsequent experiments, different test datasets T_{tr} were used for short-term and long-term forecasting, resulting in different lengths for these two parts. For long-term forecasting, the test data lengths for the Lorenz-63 and Climate Lorenz-63 systems are 5690 and 5680, respectively, corresponding to 100 Lyapunov times. For short-term forecasting, 1250 data points were selected from the dataset for numerical experiments.

The hyperparameters of the AESN (*IS*, γ, *SR*, *SD*) ensure the overall performance of the reservoir by synchronizing the adjustment of multiple matrix elements, which indirectly affects the way the whole network operates. In this paper, we express the sparsity degree in terms of the number of interconnected neurons in the reservoir layer d and the size of the reservoir K. It is denoted as $SD = d/K$. In the Lorenz-63 system, the boundary ranges of *IS*, *SR*, γ, d and K are set as $[0.000001, 1]$, $[0.000001, 2]$, $[0, 1]$, $[7, 18]$, and $[700, 900]$, respectively. In the Climate Lorenz-63 system, the boundary ranges of *IS*, *SR*, γ, d and K are set as $[0.000001, 1]$, $[0.000001, 2]$, $[0, 1]$, $[19, 42]$, and $[1900, 2100]$, respectively. The pseudocode for the AESN-SAO approach is provided in the referenced work (Ding et al., 2024).

4. Experimental results

4.1 Data-driven approach to forecasting without inserting observations

To examine the optimization performance of SAO in MHOAs, we selected the grey wolf optimizer (GWO) (Mirjalili et al., 2014), selective opposition grey wolf optimization (SOGWO) (Dhargupta et al., 2020), the marine predator algorithm (MPA) (Faramarzi et al., 2020), and the chaotic coyote optimization algorithm (CCOA) (Tong et al., 2022) as comparative experiments to validate the potential of the development of different types of metaheuristic optimization algorithms for the fusion of the AESN in two chaotic systems. The hyperparameter values for the AESN approach are referenced in Platt et al. (2022). To verify the generalizability of the optimization algorithm, we record the VPT of the two chaotic systems under 100 sets of different initial conditions in Table 1. VPT is the prediction effectiveness of quantitative prediction methods in chaotic systems, which is correlated with the normalized root mean square error (NRMSE) and the maximum Lyapunov exponent $\Lambda_{\max}$. Specifically, when the normalized root mean square error $\mathrm{NRMSE}(t_f)$ at a time point t_f exceeds a given threshold ϵ for the first time (usually the threshold is set to $\epsilon = 0.5$) (Vlachas et al., 2020; Ren et al., 2024b), the corresponding time point is considered to be the prediction time point that exceeds the threshold. The calculation is shown in eqs. (13) and (14), where σ_i is the long-term standard deviation of the i-dimensional observation data.

$$\mathrm{NRMSE}(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[\frac{y_{t,i}-x_{t,i}}{\sigma_i}\right]^2}, \tag{13}$$

$$\mathrm{VPT} = \frac{1}{\Lambda_{\max}}\underset{t_f}{\operatorname{argmax}}\left\{t_f \,\middle|\, \mathrm{NRMSE}(t) < \epsilon, \forall\, t \leqslant t_f\right\}. \tag{14}$$

As shown in Table 1, without adding the optimization algorithm, the average VPT for 100 sets of data is 4.66 Lyapunov times in the Lorenz-63 system. After introducing various types of the MHOAs, all AESN-based methods combined with optimization exhibit significant improvements, with average VPTs exceeding 9.50 Lyapunov time. Among them, the SOGWO and SAO algorithms demonstrate the best predictive performance, both increasing the average VPT beyond 10 Lyapunov times. In particular, the AESN-SAO method achieves an average VPT of 10.08 Lyapunov times. The optimal hyperparameter configuration obtained by the AESN-SAO approach is $IS = 0.1095$, $SR = 1.1051$, $\gamma = 0.8455$, $K = 709$, and $d = 7.8230$, further confirming its robustness and generalization ability in the prediction task of the Lorenz-63 system.

4.2 Short-term forecasting integrating sparse observations with a data-driven approach

In this paper, to test whether the AESN in sparse observation has prediction performance, this section discusses the experimental results of the AESN-SAO data-driven approach under three observation insertion intervals through the Lorenz-63 and Climate Lorenz-63 systems to illustrate the importance of sparse observation data insertion. In short-term predictions, the inserted sparse observational data remain at the point where the AESN-SAO method first deviates from the target system.

In the Lorenz-63 system, a set of true (target) data generated by the initial values $\{x_0, y_0, z_0\}$=[0.6630, –0.8542, –1.2013] were selected from 100 sets for experiments. The Lyapunov exponent is calculated as follows: $\lambda_1 = 0.8767$, $\lambda_2 = -0.0021$, and $\lambda_3 = -14.5412$. The maximum Lyapunov exponent for this dataset is $\Lambda_{\max} = 0.8767$. Figure 2 presents

Table 1 Valid prediction time for six data-driven approaches in a chaotic climate system based on 100 data samples

Chaotic climate systems	Approaches	VPT_MAX	VPT_MIN	VPT_AVG
Lorenz-63	AESN	11.27	1.73	4.66
	AESN-GWO	15.09	5.08	9.63
	AESN-SOGWO	14.33	5.37	10.01
	AESN-MPA	13.60	5.15	9.88
	AESN-CCOA	14.38	5.38	9.77
	AESN-SAO	15.75	5.12	10.08
Climate Lorenz-63	AESN	6.59	0.46	2.07
	AESN-GWO	8.15	0.96	4.32
	AESN-SOGWO	8.89	1.54	4.77
	AESN-MPA	10.42	0.82	4.31
	AESN-CCOA	9.33	1.03	4.28
	AESN-SAO	11.46	1.80	5.18

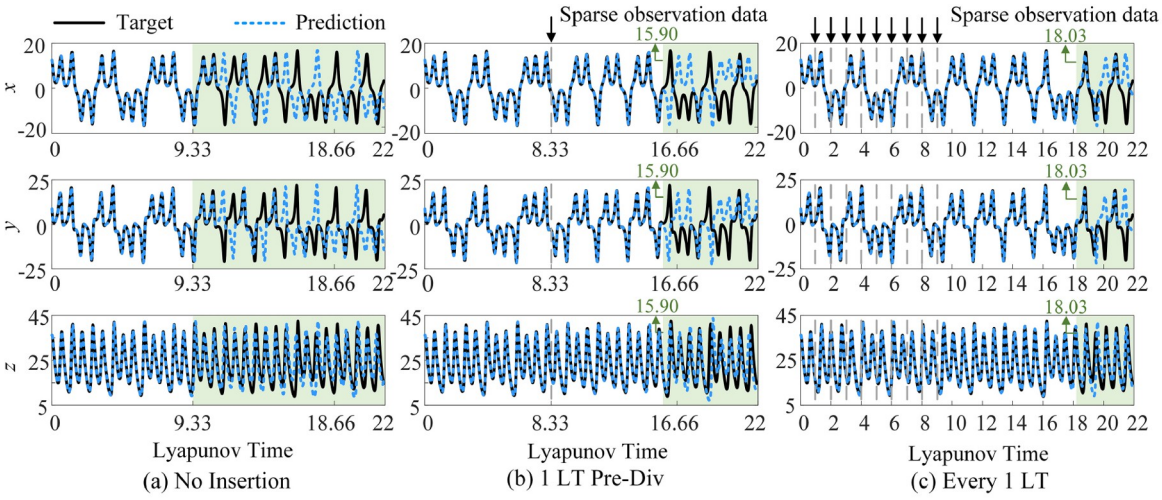

Figure 2 Comparison diagrams of the AESN-SAO approach integrated with sparse observations under different observational data insertion intervals in the Lorenz-63 system. The diagram includes the evolution of both the true (target) approach and the AESN-SAO approach at different observational data insertion intervals. (a) No Insertion: No observational data inserted; (b) 1 LT Pre-Div: Insert observational data at 1 Lyapunov time before divergence; (c) Every 1 LT: Insert observational data every 1 Lyapunov time. The green numbers represent the VPT, and the green shading indicates that the AESN-SAO approach has entered the prediction divergence stage.

the numerical results illustrating the impact of different sparse observation insertion intervals on the prediction performance of the Lorenz-63 system under the above initial conditions. In particular, Figure 2a shows the results of model-free prediction of the Lorenz-63 system via the AESN-SAO approach, the results of which indicate that the data-driven approach starts to diverge after approximately the 9.33 Lyapunov times. Figure 2b shows the prediction results under the same initial conditions, where an observation is inserted one Lyapunov time before the divergence point (at a Lyapunov time of 8.33). The insertion of this observation significantly extends the VPT, delaying the divergence of the data-driven method until approximately 15.90 Lyapunov times. Figure 2c shows the prediction results after inserting observations every 1 Lyapunov time during the prediction process of the AESN-SAO approach. By inserting 9 observations uniformly, the VPT of the AESN-SAO approach can be extended to 18.03 Lyapunov times before divergence occurs. In another set of short-term forecasting experiments, a threshold of $\epsilon = 0.5$ was used to dynamically monitor the NRMSE. When the NRMSE exceeded this threshold, observational data were promptly inserted to correct the predictive state of the AESN-SAO method. Under this insertion strategy, the VPT of the proposed approach reached only 9.92 Lyapunov times (see Table 2), which is significantly lower than those achieved by the previous two insertion strategies. This result indicates that, for the Lorenz-63 system, the timing of sparse ob-

servation insertion plays a crucial role in determining forecasting performance.

In the Climate Lorenz-63 system, a set of true (target) data generated by the initial values x_{e0}, y_{e0}, z_{e0}, x_{t0}, y_{t0}, z_{t0}, X_0, Y_0, and Z_0 of −1.8740, 0.4282, 0.8956, 0.7310, 0.5779, 0.0403, 0.6771, 0.5689, and −0.2557 were selected from 100 sets for experiments. The Lyapunov exponent is calculated as follows: $\lambda_1 = 0.8809$, $\lambda_2 = 0.3484$, $\lambda_3 = 0.0103$, $\lambda_4 = -0.0025$, $\lambda_5 = -0.5008$, $\lambda_6 = -0.8387$, $\lambda_7 = -1.8336$, $\lambda_8 = -12.2213$, and $\lambda_9 = -14.5425$. The maximum Lyapunov exponent for this dataset is $\Lambda_{\max} = 0.8809$. The AESN-SAO approach starts to diverge after 5.51 Lyapunov times when no observations are inserted. We document the results of the numerical experiments of the three insertion methods of sparse observations through Table 2 and Figure 3.

(1) When observational data are inserted every c Lyapunov time, 'Every 1 LT', 'Every 2 LT', 'Every 3 LT', 'Every 4 LT', and 'Every 5 LT' denote observations inserted at intervals of 1, 2, 3, 4, and 5 Lyapunov times. For example, 'Every 1 LT' means that 5 insertions of observations are performed uniformly every 1 Lyapunov time during the prediction process of the AESN-SAO approach, and 'Every 2 LT' means that every 2 Lyapunov times is performed uniformly for 2 insertions of observations. The results show that the data insertion strategies of 'Every 1 LT' and 'Every 2 LT' can extend the VPT of the AESN-SAO approach from 5.51 Lyapunov times to 8.28 Lyapunov times and 8.09 Lyapunov times, respectively.

(2) Taking the threshold $\epsilon = 0.5$ as a judgment condition, we examine the prediction process of the data-driven method AESN-SAO in real time and insert the observation data once the threshold is exceeded. At this time, the VPT can reach 8.35 Lyapunov times, extending 2.84 Lyapunov times.

(3) We discuss the case where observations are inserted c Lyapunov times before the AESN-SAO approach predicts the divergence of results. In Table 2, '1 LT Pre-Div', '2 LT Pre-Div', '3 LT Pre-Div', '4 LT Pre-Div', and '5 LT Pre-Div' denote the observed data inserted at Lyapunov times of 4.51, 3.51, 2.51, 1.51, and 0.51, respectively. The sparse insertions of '1 LT Pre-Div' and '2 LT Pre-Div' can also alleviate the error accumulation of the AESN-SAO data-driven method, and its VPT can be extended to 8.21 and 8.23 Lyapunov times. Therefore, the AESN-SAO approach with sparse observations can significantly extend VPT and correct the error accumulation phenomenon generated by the prediction of data-driven methods.

4.3 Long-term forecasting integrating sparse observations with a data-driven approach

In our next discussion, we focus on the long-term prediction capabilities of the AESN-SAO approach under a sparse number of observations within 100 Lyapunov times via the Lorenz-63 system and the Climate Lorenz-63 system. During the experiment, sparse observations ceased after the 90th Lyapunov time, specifically to assess the critical role that these observations play in our forecasting approach.

In the Lorenz-63 system, the testing set T_{tr} for the 100 Lyapunov time intervals contains 5690 samples. Figure 4 presents the comparison results and NRMSE curves for AESN-SAO during the long-term forecasting process, integrating three sparse observation insertion strategies. To enhance the readability of Figure 4, only the time span from the 70th to the 100th Lyapunov time, rather than the full duration, is presented. This selective focus effectively highlights the dynamic changes within the critical period. We present the outcomes of the numerical experiments for the three sparse observation insertion methods in Table 3.

The following outlines the long-term prediction results for

Table 2 Valid prediction time under different observation insertion intervals for the AESN-SAO approaches[a)]

Insertion interval	Lorenz-63	Climate Lorenz-63	Insertion interval	Lorenz-63	Climate Lorenz-63
No Insertion	9.33	5.51	Div-Point	9.92	8.35
Every 1 LT	18.03	8.28	1 LT Pre-Div	15.90	8.21
Every 2 LT	13.17	8.09	2 LT Pre-Div	11.36	8.23
Every 3 LT	14.62	5.59	3 LT Pre-Div	13.27	5.64
Every 4 LT	10.59	5.59	4 LT Pre-Div	13.33	8.09
Every 5 LT	12.55	5.66	5 LT Pre-Div	13.17	5.55
Every 6 LT	9.28	–	6 LT Pre-Div	11.31	–
Every 7 LT	13.26	–	7 LT Pre-Div	13.24	–
Every 8 LT	11.27	–	8 LT Pre-Div	11.17	–
Every 9 LT	13.33	–	9 LT Pre-Div	9.26	–

a) No Insertion: No observational data inserted; Every c LT: Insert observational data every c Lyapunov time; Div-Point: Insert observational data at the point of divergence; c LT Pre-Div: Insert observational data at c Lyapunov time before divergence.

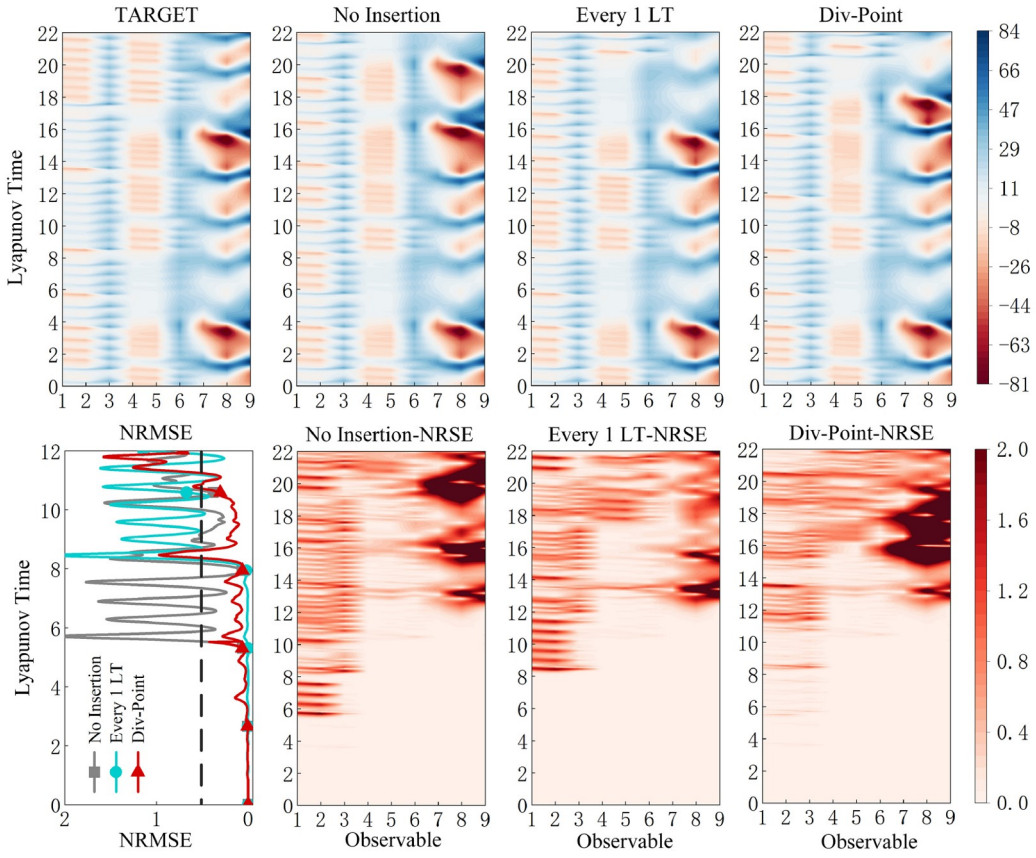

Figure 3 Contour plots of short-term prediction results for the Climate Lorenz-63 system using the AESN-SAO approach integrated with sparse observations under different observational data insertion strategies. The display includes both the evolution of the true (target) and the NRSE contour for the AESN-SAO approach at different observational data insertion intervals (No Insertion, Every 1 LT, Div-Point). The horizontal axis, labeled 'Observable', represents the observable dimension of the chaotic climate system, whereas the vertical axis depicts the system's evolution in Lyapunov time. The curves beside depict the evolution of the average NRSE (NRMSE), with the black dashed line representing the threshold $\epsilon = 0.5$.

the Lorenz-63 system across 100 Lyapunov times.

(1) For the AESN-SAO data-driven approach, under the '1 LT Pre-Div' insertion strategy, a total of 39 observations are inserted, resulting in a VPT of 96.09 Lyapunov times, with the predicted trajectory closely matching the target curve.

(2) When sparse observations are uniformly inserted every 1 Lyapunov time during the prediction process (a total of 90 insertions were inserted), the VPT can be further extended to 98.19 Lyapunov times. The prediction trajectory is aligned with the target, effectively reconstructing the Lorenz-63 attractor. The NRMSE curve indicates that these insertion interval scenarios demonstrate good prediction stability.

(3) The 'Div-Point' insertion scenario implies that when the NRMSE value between the prediction and the target exceeds the threshold ($\epsilon = 0.5$), real-time judgment is made to insert observational data. When the insertion of observational data ceased at the 90th Lyapunov time, the VPT reached 93.60 Lyapunov times (with a total of 84 observations inserted). Although the prediction results of the 'Div-Point' insertion scenario slightly deviate from the target curve, the real-time judgment and insertion of observational data help correct the accumulated errors generated by the data-driven approach, keeping the prediction results roughly consistent with the true trajectory of the target, thereby enabling long-term prediction.

In the Climate Lorenz-63 system, we investigate the long-term prediction capability of the AESN-SAO approach under sparse observation conditions over a time span of 100 Lyapunov times, where the test dataset T_{tr} contains 5680 samples. The following summarizes the long-term prediction results of this method over the specified time horizon.

(1) The prediction of the Climate Lorenz-63 system is less satisfactory when the observations are inserted in the '1 LT Pre-Div', and the VPT reaches only 55.58 Lyapunov times (see Table 3). This suggests that the frequency of observation data insertion is a critical factor when long-term predictions are made for the 9-dimensional Climate Lorenz-63 chaotic system.

(2) When observations are inserted every 1 Lyapunov time, the predictions remain consistent with the trajectory of the target observations for the first 92.94 Lyapunov times (see Figure 5). However, beyond 92.94 Lyapunov times, the

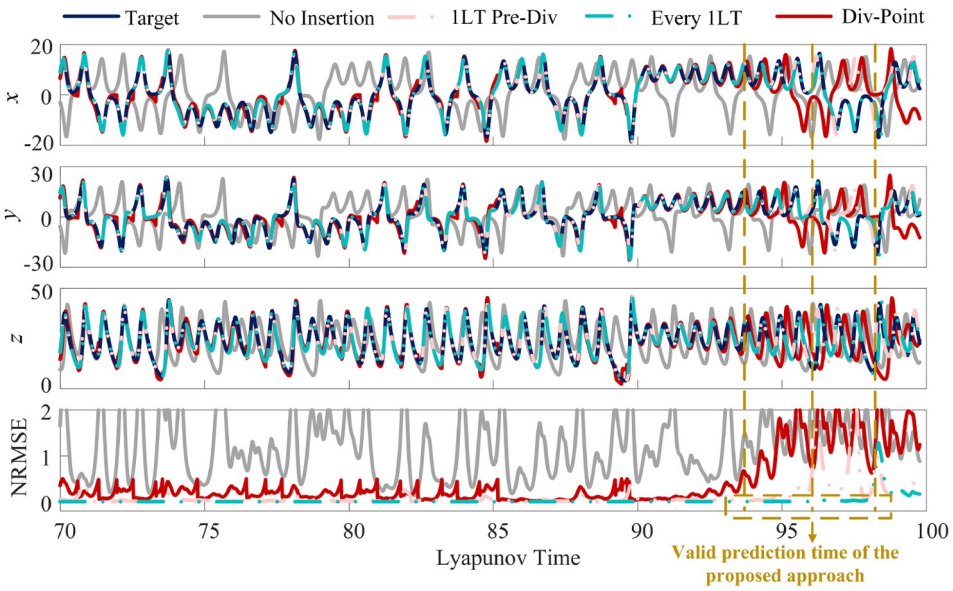

Figure 4 Comparison of diagrams and NRMSE curves from the AESN-SAO approach for long-term forecasts of the Lorenz-63 system using different spare observation insertion intervals. Observations were inserted until the 90th Lyapunov time, after which no further data were added. To improve readability, the time span from the 70th to 100th Lyapunov time was selected rather than displaying the entire duration. The three yellow dashed lines represent the VPT of the AESN-SAO approach under the three data insertion intervals: 1 LT Pre-Div, Every 1 LT, and Div-Point.

Table 3 Valid prediction time of the AESN-SAO approach in long-term forecasting of chaotic climate systems[a)]

Chaotic climate systems	No Insertion	1 LT Pre-Div	Every 1 LT	Div-Point
Lorenz-63	9.33	96.09	98.19	93.60
Climate Lorenz-63	5.51	55.58	92.94	90.23

a) No Insertion: No observational data inserted; 1 LT Pre-Div: Insert observational data at 1 Lyapunov time before divergence; Every 1 LT: Insert observational data every 1 Lyapunov time; Div-Point: Insert observational data at the point of divergence.

predictions start to diverge. This dispersion occurs because the last insertion of sparse observations occurred at the 90th Lyapunov time, after which no further updates were made. In the case of this insertion interval, a total of 90 observations were inserted.

(3) In another insertion strategy, observational data are immediately inserted once the prediction error, measured by the NRMSE, exceeds a predefined threshold ($\epsilon = 0.5$). Under this strategy, the VPT reaches 90.23 Lyapunov times, with a total of 52 observations inserted. By dynamically controlling the insertion frequency through real-time error monitoring, this approach effectively suppresses error accumulation and maintains the NRMSE below 0.5 throughout the prediction process, demonstrating strong long-term forecasting capability and efficient use of sparse data.

5. Discussions

5.1 Advantages of the proposed approach

To assess the effectiveness of the proposed method, we compared it with two state-of-the-art approaches (Fan et al., 2020; Jin et al., 2023), with the corresponding VPT and MAE results provided in Table 4. Forecasting experiments, both short-term (approximately 22 Lyapunov times) and long-term (approximately 100 Lyapunov times), were conducted on the Lorenz-63 and Climate Lorenz-63 systems. In these experiments, sparse observational data were continuously inserted throughout the entire duration, ensuring that the forecasting process remained consistently updated and corrected by incorporating observational data.

A coupling coefficient of $\xi = 0.8$ between the reservoir prediction $\mathbf{x}_t$ and the sparse observational data $\mathbf{y}'_t$ such that the updated input followed the condition $\mathbf{x}_t = \mathbf{x}_t + \xi(\mathbf{y}'_t - \mathbf{x}_t)$, with observational data being inserted at intervals of every 1 Lyapunov time (Fan et al., 2020). When this update scheme was applied to the AESN-SAO approach, the experimental results showed that for the Lorenz-63 system, Fan et al. (2020) achieved low MAEs of 0.0003 for short-term forecasts and 0.0006 for long-term forecasts. These results were second only to the 'Every 1 LT' insertion strategy proposed in this study, which achieved MAEs of 0.0001 (short-term) and 0.0004 (long-term). In contrast, the same insertion strategy performed significantly

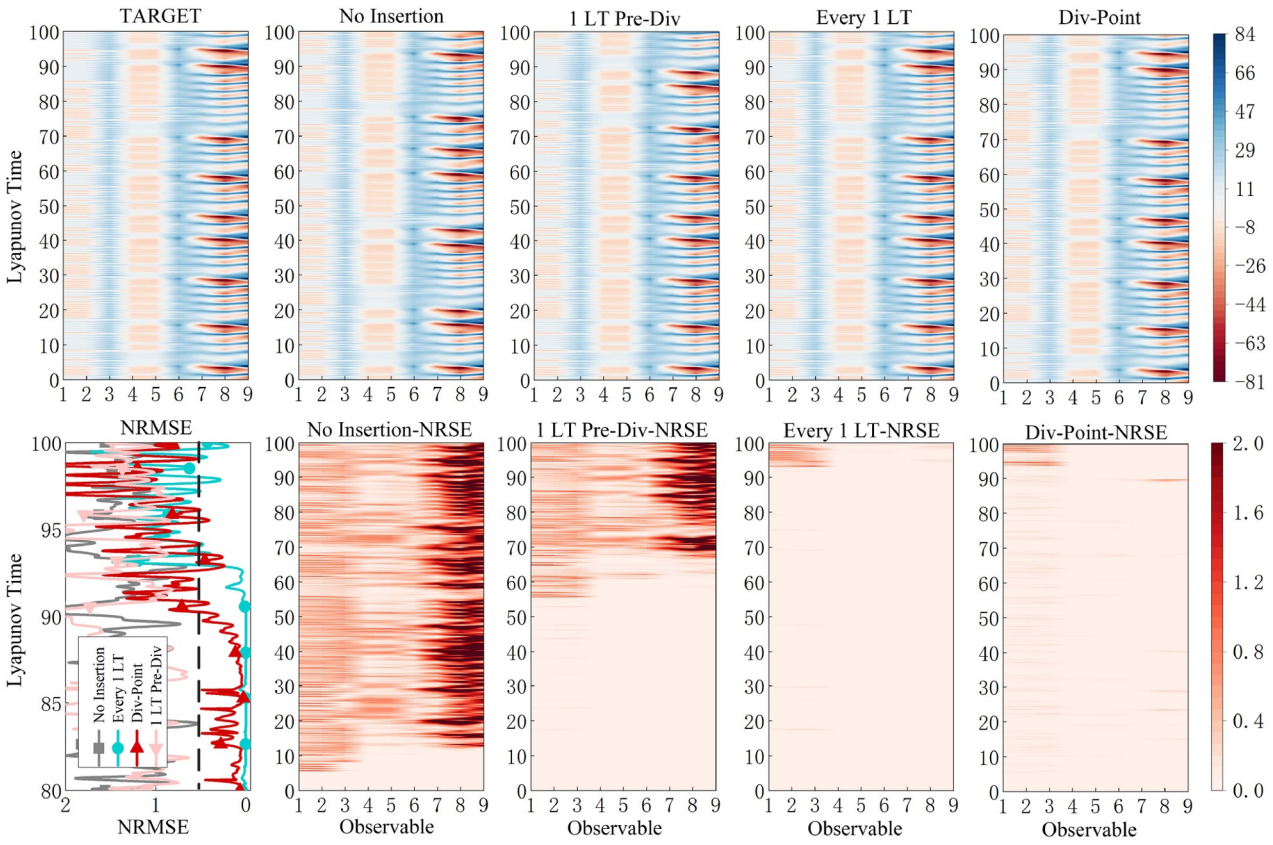

Figure 5 Contour plots of long-term prediction results for the Climate Lorenz-63 system using the AESN-SAO approach integrated with sparse observations under different observational data insertion strategies. The display includes both the evolution of the true (target) and the NRSE contour for the AESN-SAO approach at different observational data insertion intervals (No Insertion, 1 LT Pre-Div, Every 1 LT, and Div-Point). The horizontal axis, labeled 'Observable', represents the observable dimension of the chaotic climate system, whereas the vertical axis depicts the system's evolution in Lyapunov time. Observational data were inserted until the 90th Lyapunov time, with no additional data added thereafter. To improve readability, the NRMSE curves are shown for the time span between the 80th and 100th Lyapunov times. In the accompanying NRMSE curve plot, the black dashed line indicates the threshold $\epsilon = 0.5$. To improve visual clarity, only the critical prediction period from the 80th to the 100th Lyapunov time is shown.

Table 4 Performance indices of the compared approaches in chaotic climate systems for short-term and long-term forecasting

Methods	Lorenz-63				Climate Lorenz-63			
	Short-term		Long-term		Short-term		Long-term	
	VPT	MAE	VPT	MAE	VPT	MAE	VPT	MAE
Fan et al. (2020)	21.92	3.0×10^{-4}	99.77	6.0e-04	8.33	1.34	8.33	1.80
Jin et al. (2023)	15.90	1.66	16.06	5.31	8.23	2.92	8.23	3.98
1 LT Pre-Div	21.92	0.16	99.77	0.24	22.00	0.15	55.58	6.74
Every 1 LT	21.92	1.0×10^{-4}	99.77	4.0e-04	22.00	0.08	100.00	0.10
Div-Point	21.92	0.82	99.77	1.12	21.48	0.83	99.38	0.77

worse in the Climate Lorenz-63 system, with MAE values exceeding those of the 1 LT Pre-Div, Every 1 LT, and Div-Point strategies presented in this study. Its VPT was limited to only 8.33 Lyapunov times. This performance gap may be attributed to the influence of the reservoir prediction $\mathbf{x}_t$ and coupling coefficient ξ settings, which likely compromised the effectiveness of the AESN-SAO method in long-term forecasting.

Additionally, Jin et al. (2023) developed an IKN with physical significance and integrated it with the Transformer model. Although we did not directly compare with this approach, we simulated their experimental setup, which involves inserting observational data every 256 steps. The results showed that this insertion method led to significantly

higher MAE values and lower VPT values in both the short-term and long-term predictions for the Lorenz-63 and Climate Lorenz-63 systems. For example, in the long-term prediction of the Lorenz-63 system, the VPT reached only 16.06 Lyapunov times, which is much lower than the 99.77 Lyapunov times achieved by the other methods. Moreover, the MAE reached 5.31, which was notably higher than those of the other insertion strategies. This suggests that longer insertion intervals fail to provide timely error correction before system divergence, negatively impacting the accuracy of long-term predictions. In the Climate Lorenz-63 system, the performance of this insertion strategy was even less favorable, further validating this conclusion.

Through comparative analysis, the three sparse observation insertion strategies proposed in this paper outperform existing methods in most cases. The specific advantages are as follows.

(1) In the 'Every 1 LT' strategy, the Lyapunov time determined by the largest Lyapunov exponent serves as a critical timescale for quantifying the process of error accumulation in chaotic climate systems. By inserting observational data at fixed intervals of every one Lyapunov time, the AESN-SAO method can update the reservoir state before significant divergence in the prediction occurs, thereby preventing the premature accumulation of forecast errors. This strategy effectively improves the accuracy and stability of long-term predictions in chaotic systems.

(2) The 'Div-Point' strategy continuously monitors the NRMSE of the AESN-SAO approach and inserts observational data immediately once the value exceeds a predefined threshold of $\epsilon = 0.5$, thereby correcting prediction deviations. This strategy offers high flexibility by enabling real-time adjustments based on the dynamic behavior of NRMSE, which significantly extends the forecast horizon. Although its MAE is slightly higher than that of the 'Every 1 LT' and '1 LT Pre-Div' strategies, it maintains good predictive stability and continuity while keeping the NRMSE within the threshold ϵ. Moreover, by inserting observational data only when necessary, this strategy reduces redundant inputs and improves the efficiency of data utilization.

(3) The output weight matrix $\mathbf{W}_{\text{out}}$ of the AESN is trained and optimized using the SAO, eliminating the need for retraining during long-term forecasting and thereby significantly reducing computational costs. This design highlights the reusability of $\mathbf{W}_{\text{out}}$. Furthermore, this study integrates a sparse observation insertion strategy based on the Lyapunov time and VPT, along with the adaptive modeling capability of the reservoir, to develop a novel and efficient data-driven prediction framework. The proposed framework is well-suited for long-term and high-accuracy forecasting in complex dynamical systems, demonstrating strong adaptability and application potential, particularly in climate and physical system modeling.

5.2 Limitations of the proposed approach

The proposed method faces the following three challenges.

(1) Sensitivity of real-time insertion strategies to computational resources. When determining the optimal timing for inserting observational data in real time, the performance of computational resources directly affects the frequency of insertion and the system's response speed. Since each insertion requires a dynamic evaluation of the divergence in prediction results, differences in processing speed, memory capacity, and parallel computing capabilities across platforms significantly influence the efficiency of real-time assessment and data insertion. High-performance computing systems can insert observational data at higher frequencies and with lower latency, enabling timely correction during the early stages of error accumulation. This helps to extend the valid prediction time and improve the forecast horizon. In contrast, systems with lower computational capacity may fail to insert data effectively at critical moments, leading to aggravated error accumulation and degraded predictive performance. Therefore, discrepancies in computational resources not only impact the design of observational data insertion strategies but also constrain the effectiveness of data-driven methods in long-term forecasting tasks.

(2) Generalization limitations of purely data-driven method. Data-driven approaches lack the guidance of physical constraints and guidance, which may lead to an incomplete capture of a system's dynamic evolution. For example, in the Climate Lorenz-63 system, although the '1 LT Pre-Div' strategy can partially characterize the system's local dynamics, the data-driven method proposed in this study exhibits noticeable divergence after 55.58 Lyapunov times during long-term prediction, indicating that the prediction gradually deviates from the true (target) system trajectory. This deviation may stem from the training data's failure to cover the full evolutionary process of the system, thereby limiting the AESN-SAO method's ability to extrapolate when faced with previously unseen dynamical behaviors. This phenomenon underscores a common limitation of purely data-driven methods: their insufficient generalization capability when encountering unfamiliar system states, making it difficult to maintain stability and accuracy in long-term prediction.

(3) Dependence on high-quality observational data. These methods are heavily dependent on high-quality observational data. In Earth system applications, observational data are often noisy or non-Gaussian, introducing uncertainty into model performance. As a result, the sparse observation insertion strategy may become less effective when the input data is inaccurate, leading to unreliable prediction results. To address these challenges, likelihood-based approaches, moment-based approaches, and observational experiment-based estimation techniques can be employed to quantify and mi-

tigate the impact of observational errors and uncertainty (Li et al., 2024a). For instance, a dual-cycle data assimilation method has been proposed, which optimizes the likelihood function using the innovation time series and adjusts both model and observation error parameters through an optimization algorithm (Tian et al., 2022).

6. Conclusions

In recent years, data-driven approaches such as echo state network have received increasing attention for model-free prediction of chaotic systems. However, these methods are usually effective in predicting the dynamics of chaotic systems only in the short term because the predictions tend to diverge after a certain Lyapunov time is reached, and the performance varies across different types of chaotic systems. Therefore, how to effectively extend the prediction horizon of chaotic systems without relying solely on the inherent accuracy of data-driven methods has become a key challenge. To address this issue, this study proposes a data-driven method enhanced by a parallel optimization strategy, incorporating three sparse observational data insertion mechanisms. The approach achieves high-precision long-term forecasting in two representative chaotic climate systems. The following conclusions can be drawn.

(1) The proposed AESN-SAO approach enhances the robustness and general applicability of hyperparameters while improving adaptability of data-driven methods to variations in the initial conditions of chaotic systems. This effectively reduces prediction instability caused by initial perturbations. The method demonstrates strong performance in modeling and forecasting complex chaotic systems, making it particularly suitable for long-term predictions in climate dynamical systems.

(2) By constructing an error correction factor based on Lyapunov time and valid prediction time, a sparse observation insertion strategy is introduced into the AESN-SAO approach. This enables AESN-SAO to fully leverage the reusability of the output weight matrix to continuously update the reservoir state and promptly correct prediction errors before divergence occurs. This strategy effectively mitigates the gradual accumulation of errors during the forecasting process and significantly extends the valid prediction time of purely data-driven methods.

Therefore, the proposed data-driven learning framework, which integrates sparse observations with a parallel optimization strategy, effectively mitigates error accumulation in long-term predictions of chaotic climate system. This approach provides a novel pathway for achieving stable and reliable model-free forecasting. For example, researchers can make moderate modifications to the ESN architecture and incorporate an additional loss function based on the system's governing equations during the training phase, in order to ensure that the predictions comply with fundamental physical laws and enhance generalization capability under limited training data (Doan et al., 2020; Na et al., 2023). Furthermore, the integration of causal inference (Su et al., 2023; Li and Su, 2024) with machine learning holds promise for uncovering causal relationships within the system and extracting key dynamic patterns from large-scale observational datasets. Therefore, future advances in model-free forecasting of chaotic climate systems should prioritize the deep integration of physical models, causal inference, and machine learning to build intelligent, interpretable, and adaptive prediction systems (Li et al., 2023). Such systems will enable more accurate long-term forecasts and offer strong theoretical and methodological support for critical applications such as climate change modeling and disaster risk mitigation.

Data availability

Data sharing is not applicable to this article. The data used in this study are derived from numerical simulations of classical chaotic systems, and the results can be reproduced using publicly available systems and standard numerical methods. Therefore, no additional datasets are required to be shared.

Acknowledgements *This work was supported by the National Natural Science Foundation of China (Grant Nos. 42430112, 42371377).*

References

Bai Y, Liu M D, Ding L, Ma Y J. 2021. Double-layer staged training echo-state networks for wind speed prediction using variational mode decomposition. Appl Energy, 301: 117461

Chen P, Liu R, Aihara K, Chen L. 2020. Autoreservoir computing for multistep ahead prediction based on the spatiotemporal information transformation. Nat Commun, 11: 4568

Deng L, Liu S. 2023. Snow ablation optimizer: A novel metaheuristic technique for numerical optimization and engineering design. Expert Syst Appl, 225: 120069

Dhargupta S, Ghosh M, Mirjalili S, Sarkar R. 2020. Selective opposition based grey wolf optimization. Expert Syst Appl, 151: 113389

Ding L, Bai Y L, Fan M H, Song W, Ren H. 2024. Using a snow ablation optimizer in an autonomous echo state network for the model-free prediction of chaotic systems. Nonlinear Dyn, 112: 11483–11500

Ding L, Bai Y L, Fan M H, Yu Q H, Zhu Y J, Chen X Y. 2023. Serial-parallel dynamic echo state network: A hybrid dynamic model based on a chaotic coyote optimization algorithm for wind speed prediction. Expert Syst Appl, 212: 118789

Doan N A K, Polifke W, Magri L. 2020. Physics-informed echo state networks. J Comput Sci, 47: 101237

Fan H, Jiang J, Zhang C, Wang X, Lai Y C. 2020. Long-term prediction of chaotic systems with machine learning. Phys Rev Res, 2: 012080

Faramarzi A, Heidarinejad M, Mirjalili S, Gandomi A H. 2020. Marine predators algorithm: A nature-inspired metaheuristic. Expert Syst Appl, 152: 113377

Gupta V, Li L K B, Chen S, Wan M. 2023. Model-free forecasting of partially observable spatiotemporally chaotic systems. Neural Netws, 160: 297–305

Jaeger H. 2001. The "echo state" approach to analysing and training recurrent neural networks-with an erratum note. Bonn, Germany: German National Research Center for Information Technology. GMD Technical Report 148. 13
Jin Y H, Hou L, Zhong S, Yi H M, Chen Y S. 2023. Invertible Koopman Network and its application in data-driven modeling for dynamic systems. Mech Syst Signal Process, 200: 110604
Kaiser E, Kutz J N, Brunton S L. 2018. Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. Proc R Soc A, 474: 20180335
Li X, Feng M, Ran Y, Su Y, Liu F, Huang C, Shen H, Xiao Q, Su J, Yuan S, Guo H. 2023. Big Data in Earth system science and progress towards a digital twin. Nat Rev Earth Environ, 4: 319–332
Li X, Guo Y. 2025. Paradigm shifts from data-intensive science to robot scientists. Sci Bull, 70: 14–18
Li X, Liu F. 2025. A mathematical comparison of data assimilation and machine learning in earth system state estimation from a Bayesian inference viewpoint. Inf Geogr, 1: 100001
Li X, Liu F, Ma C, Hou J, Zheng D, Ma H, Bai Y, Han X, Vereecken H, Yang K, Duan Q, Huang C. 2024a. Land data assimilation: Harmonizing theory and data in land surface process studies. Rev Geophys, 62: 45
Li X, Su J. 2024. Towards good governance of data: A case study in geoscience data governance. Chin Sci Bull, 69: 1149–1155
Li X, Zhu Q, Zhao C, Duan X, Zhao B, Zhang X, Ma H, Sun J, Lin W. 2024b. Higher-order Granger reservoir computing: Simultaneously achieving scalable complex structures inference and accurate dynamics prediction. Nat Commun, 15: 2506
Lin M M, Cheng C M, Peng Z K, Dong X J, Qu Y G, Meng G. 2021. Nonlinear dynamical system identification using the sparse regression and separable least squares methods. J Sound Vib, 505: 116141
Lorenz E N. 1963. Deterministic nonperiodic flow. J Atmos Sci, 20: 130–141
Lukoševičius M, Jaeger H. 2009. Reservoir computing approaches to recurrent neural network training. Comput Sci Rev, 3: 127–149
Maass W, Natschläger T, Markram H. 2002. Real-time computing without stable states: A new framework for neural computation based on perturbations. Neural Computat, 14: 2531–2560
Martinec J, Rango A. 1986. Parameter values for snowmelt runoff modelling. J Hydrol, 84: 197–219
Mirjalili S, Mirjalili S M, Lewis A. 2014. Grey wolf optimizer. Adv Eng Software, 69: 46–61
Na X, Li Y, Ren W, Han M. 2023. Physics-informed hierarchical echo state network for predicting the dynamics of chaotic systems. Expert Syst Appl, 228: 120155
Parmar N, Refai H H, Runolfsson T. 2022. A survey on the methods and results of data-driven Koopman analysis in the visualization of dynamical systems. IEEE Trans Big Data, 8: 723–738
Pathak J, Hunt B, Girvan M, Lu Z, Ott E. 2018a. Model-free prediction of large spatiotemporally chaotic systems from data: A reservoir computing approach. Phys Rev Lett, 120: 5
Pathak J, Wikner A, Fussell R, Chandra S, Hunt B R, Girvan M, Ott E. 2018b. Hybrid forecasting of chaotic processes: Using machine learning in conjunction with a knowledge-based model. Chaos-An Interdisciplinary J Nonlinear Sci, 28: 9
Peña M, Kalnay E. 2004. Separating fast and slow modes in coupled chaotic systems. Nonlin Processes Geophys, 11: 319–327
Platt J A, Penny S G, Smith T A, Chen T C, Abarbanel H D I. 2022. A systematic exploration of reservoir computing for forecasting complex spatiotemporal dynamics. Neural Netws, 153: 530–552
Proctor J L, Brunton S L, Kutz J N. 2018. Generalizing Koopman theory to allow for inputs and control. SIAM J Appl Dyn Syst, 17: 909–930
Ren H H, Fan M H, Bai Y L, Ma X Y, Zhao J H. 2024a. Prediction of spatiotemporal dynamic systems by data-driven reconstruction. Chaos Solitons Fractals, 185: 115137
Ren H H, Bai Y L, Fan M H, Ding L, Yue X X, Yu Q H. 2024b. Constructing polynomial libraries for reservoir computing in nonlinear dynamical system forecasting. Phys Rev E, 109: 13
Sangiorgio M, Dercole F, Guariso G. 2021. Forecasting of noisy chaotic systems with deep neural networks. Chaos Solitons Fractals, 153: 111570
Shen C, Appling A P, Gentine P, Bandai T, Gupta H, Tartakovsky A, Baity-Jesi M, Fenicia F, Kifer D, Li L, Liu X, Ren W, Zheng Y, Harman C J, Clark M, Farthing M, Feng D, Kumar P, Aboelyazeed D, Rahmani F, Song Y, Beck H E, Bindas T, Dwivedi D, Fang K, Höge M, Rackauckas C, Mohanty B, Roy T, Xu C, Lawson K. 2023. Differentiable modelling to unify machine learning and physical models for geosciences. Nat Rev Earth Environ, 4: 552–567
Su J, Chen D, Zheng D, Su Y, Li X. 2023. The insight of why: Causal inference in Earth system science. Sci China Earth Sci, 66: 2169–2186
Takens F. 2006. Detecting strange attractors in turbulence. In: Dynamical Systems and Turbulence, Warwick 1980: Proceedings of a Symposium Held at the University of Warwick 1979/80. Springer. 366–381
Tian J, Qin J, Yang K, Zhao L, Chen Y, Lu H, Li X, Shi J. 2022. Improving surface soil moisture retrievals through a novel assimilation algorithm to estimate both model and observation errors. Remote Sens Environ, 269: 112802
Tong H, Zhu Y, Pierezan J, Xu Y, Coelho L S. 2022. Chaotic coyote optimization algorithm. J Ambient Intell Hum Comput, 13: 2807–2827
Vlachas P R, Arampatzis G, Uhler C, Koumoutsakos P. 2022. Multiscale simulations of complex systems by learning their effective dynamics. Nat Mach Intell, 4: 359–366
Vlachas P R, Pathak J, Hunt B R, Sapsis T P, Girvan M, Ott E, Koumoutsakos P. 2020. Backpropagation algorithms and Reservoir Computing in Recurrent Neural Networks for the forecasting of complex spatiotemporal dynamics. Neural Netws, 126: 191–217
Wan Z Y, Dodov B, Lessig C, Dijkstra H, Sapsis T P. 2021. A data-driven framework for the stochastic reconstruction of small-scale features with application to climate data sets. J Comput Phys, 442: 110484
Wang J, Li Y, Gao R X, Zhang F. 2022. Hybrid physics-based and data-driven models for smart manufacturing: Modelling, simulation, and explainability. J Manuf Syst, 63: 381–391
Xu M L, Han M. 2015. Factor echo state network for multivariate chaotic time series prediction (in Chinese). Acta Automat Sin, 41: 1042–1046

(Editorial handling: Huijun WANG)